1 **When genomes collide: multiple modes of germline misregulation in a**
2 **dysgenic syndrome of *Drosophila virilis***

3

4 Mauricio A. Galdos*, Alexandra A. Erwin*, Michelle L. Wickersheim, Chris C.
5 Harrison, Kendra D. Marr and Justin P. Blumenstiel

6

7 Department of Ecology and Evolutionary Biology

8 Department of Molecular Biosciences

9 University of Kansas

10 Lawrence, KS

11 66045

12

13 *These authors contributed equally.

14

15 Corresponding Author: Justin Blumenstiel, jblumens@ku.edu

16

1 **Abstract**

2

3 In sexually reproducing species the union of gametes that are not closely related
4 can result in genomic incompatibility. Hybrid dysgenic syndromes represent a
5 form of genomic incompatibility that can arise when transposable element (TE)
6 abundance differs between two parents. When TEs lacking in the female parent
7 are transmitted paternally, a lack of corresponding silencing small RNAs
8 (piRNAs) transmitted through the female germline can lead to TE mobilization in
9 progeny. The epigenetic nature of this phenomenon is demonstrated by the fact
10 that genetically identical females of the reciprocal cross are normal. Here we
11 show that in the hybrid dysgenic syndrome of *Drosophila virilis*, an excess of
12 paternally inherited TE families leads not only to increased expression of these
13 TEs, but also coincides with derepression of TEs in equal abundance within
14 parents. Moreover, TE derepression is stable as flies age and associated with
15 piRNA biogenesis defects for only some TEs. At the same time, TE activation is
16 associated with a genome wide shift in the distribution of endogenous gene
17 expression and an increase in abundance of off-target genic piRNAs. To identify
18 regions of the maternal genome that most protect against dysgenesis, we
19 performed an F3 backcross analysis. We find that pericentric regions play a
20 dominant role in maternal protection. This F3 backcross approach additionally
21 allowed us to clarify the properties of genic paramutation in *D. virilis*. Overall,
22 results support a model in which early germline events in dysgenesis establish a
23 chronic, stable state of mis-expression that is maintained through adulthood.

- 1 Such early events in the germline that are mediated by parent-of-origin effects
- 2 may be important in determining patterns of gene expression in natural
- 3 populations.
- 4

1 **Author Summary**

2

3 Transposable elements (TE) are selfish elements that code for the function of
 4 copying themselves. More than half the human genome is comprised of such
 5 elements. Studies in the fruit flies *Drosophila melanogaster* and *D. virilis* have
 6 been important in demonstrating a role for RNA silencing by piwi-interacting
 7 RNAs (piRNAs) in protecting the genome against these harmful elements. These
 8 small RNAs are capable of recognizing TE mRNAs and mediating their
 9 destruction by Argonaute proteins. They are also transmitted by the female
 10 germline to offspring in order to maintain a stable genome across generations.
 11 When males carrying a particular TE family are crossed with females lacking the
 12 element, the mother is unable to provide genome defense via complementary
 13 piRNAs that target the element. This leads to excess TE activation in the
 14 germline and sterility. This phenomenon is known as hybrid dysgenesis.
 15 In this article we characterize the genomic landscape of TE destabilization that
 16 occurs in hybrid dysgenesis in *D. virilis*. Previous studies had demonstrated that
 17 multiple TEs mobilized during hybrid dysgenesis. We demonstrate that this
 18 mobilization of multiple TEs is associated with activation of additional TEs in the
 19 germline. In addition, we find that TE activation leads to the production of off-
 20 target genic piRNAs that cause reduced expression of highly expressed genes.
 21 Finally, we show that genic off-target effects of piRNA silencing can contribute to
 22 parent-of-origin effects on gene expression. Similar phenomena may influence
 23 patterns of gene expression in the germline of natural populations.

24

1

2 Introduction

3

4 In sexually reproducing species, two unique haploid genomes join together in
5 syngamy to establish each generation. This mixing of genomes introduces
6 potentially advantageous variation under changing environmental conditions, but
7 also provides a condition ripe for exploitation by selfish elements [1] . Because
8 syngamy can introduce selfish elements to new genomes and recombination can
9 separate selfish elements from their harmful consequences, selfish elements
10 such as transposable elements (TEs) can proliferate [2,3]. When the balance of
11 TE burden differs between gametes of mated individuals, a genomic clash known
12 as hybrid dysgenesis may occur [4,5]. Hybrid dysgenesis is associated with
13 sterility and an increased mutation rate that typically occurs when males with a
14 particular TE family are mated with females lacking that family [6,7,8]. Sterility
15 and mutation occurs because paternally inherited TEs become over-activated in
16 the F1 germline [9]. In *Drosophila melanogaster*, three distinct hybrid dysgenic
17 syndromes have been identified. These are driven by the P element, the I
18 element and the hobo element [6,7,8]. In addition, a second form of hybrid
19 dysgenesis has been identified in *D. virilis* [10]. The *D. virilis* syndrome has been
20 proposed to be driven by the *Penelope* element [11,12,13,14], though other
21 elements may contribute to dysgenesis [15,16].

22

23 Hybrid dysgenesis syndromes in *Drosophila* have provided crucial insight into TE
24 dynamics and mechanisms of host genome defense by small RNAs. The

1 classical observation – activation for TEs inherited through the *Drosophila* male
 2 germline – can be explained by the fact that the maternal germline is the primary
 3 agent of transgenerational TE repression via small RNA silencing pathways [17].
 4 Mothers lacking a particular TE family in their genome also lack corresponding
 5 23-30 nt piwi-interacting RNAs (piRNAs) in their eggs. In the absence of the
 6 maternally provisioned piRNAs that target TE mRNAs for Argonaute-mediated
 7 slicing, paternally inherited TEs become activated in the progeny germline. This
 8 has been demonstrated for the P element and the I element systems of hybrid
 9 dysgenesis in *Drosophila melanogaster* [18,19,20,21].

10

11 It remains unclear what precisely follows the initial germline activation of
 12 paternally inherited TEs. Early studies of P element hybrid dysgenesis in *D.*
 13 *melanogaster* indicated downstream activation of additional TEs [22], but this
 14 interpretation was soon called into doubt [23]. Nonetheless, the syndrome of
 15 hybrid dysgenesis in *D. virilis* provided strong support for cascading germline
 16 activation of TEs because multiple TEs were found to mobilize in the germline of
 17 dysgenic progeny [11,15]. While different TE families may contribute to the initial
 18 induction of dysgenesis in *D. virilis*, germline co-mobilization is demonstrated for
 19 two TEs (*Ulysses* and *Telemac*) that are evenly distributed between the two
 20 strains and also mobilize during dysgenesis [16,24,25]. Significantly, a recent
 21 study of the P element system has indicated that the previous conclusion of no
 22 co-mobilization may have been premature [21]. In the face of P element
 23 activation, DNA damage can perturb piRNA biogenesis and this defective piRNA

1 biogenesis in turn leads to mobilization of TEs. Thus, global TE mobilization may
2 be common in syndromes of hybrid dysgenesis that are driven by a single
3 element. Whether the mechanism underlying global TE activation is shared
4 across syndromes is not known.

5

6 Here we use the unique system of hybrid dysgenesis in *D. virilis* to define the
7 landscape of TE derepression and genic mis-expression that is driven by the
8 initial activation of a small number of TE families. We show that germline
9 mobilization of TEs is driven by a multi-layered mechanism. Diverse elements
10 are activated corresponding to TE copy number asymmetry between strains and
11 there is a corresponding activation of multiple TEs that are evenly distributed
12 between strains. This state of chronic TE derepression is maintained as flies age.
13 In turn, highly expressed genes become *de novo* targets of piRNA biogenesis,
14 which leads to a shift in the balance of endogenous gene expression.

15

16

17 **Results**

18

19 Genome Wide Asymmetry in TE abundance in a Dysgenic Cross of *D. virilis*

20

21 Previous studies of hybrid dysgenesis in *D. virilis* have identified several
22 candidate elements that appear to contribute to F1 sterility in crosses between
23 inducer strain 160 fathers and the non-inducer strain 9 mothers. The *Penelope*

1 element is most likely the dominant element. Active copies of *Penelope* are
 2 abundant in strain 160 and only degenerate copies are present in strain 9 [26].
 3 Furthermore, expression of the *Penelope* element is elevated in the ovaries and
 4 testes of F1 dysgenic progeny that have escaped ablation of the gonad [11]. In
 5 addition to the *Penelope* element, three other elements (*Helena*, *Paris* and
 6 *Polyphemus* [16,27]) are also more abundant in the inducer strain, and these
 7 likely contribute to dysgenesis. A complex mode of hybrid dysgenesis, driven
 8 jointly by multiple elements, is supported by the fact that some strains of *D. virilis*
 9 lack *Penelope* piRNAs in the ovary but are "neutral" - capable of preventing
 10 dysgenesis when crossed with strain 160 males but also incapable of induction
 11 [28]. If *Penelope* is the sole cause of paternal induction, it is difficult to explain
 12 how these strains could prevent induction when the female germline lacks
 13 *Penelope* piRNAs.

14

15 To determine whether additional elements to *Penelope* may contribute to
 16 dysgenesis in *D. virilis*, we performed genome sequencing on both strains 9
 17 (non-inducer) and 160 (inducer). To estimate bulk differences in TE abundance
 18 between the two strains, we mapped single end reads against a custom *D. virilis*
 19 TE library (Supplemental 1) that included more TEs than used in the previous
 20 analysis of *Polyphemus* [27]. Relative mapping abundance with BWA-MEM[29],
 21 normalized by read number mapping to the reference, is shown in Figure 1A.
 22 Consistent with previous results, *Penelope* and *Polyphemus* show the greatest
 23 excess in strain 160. Using a 3-fold cutoff as a threshold, we also confirm the

1 previously reported result that copy number for *Helena* and *Paris* is enriched in
2 strain 160[16]. Furthermore, we identify seven additional elements that are
3 greater than 3-fold enriched in strain 160. These elements are all candidates for
4 contributing to the dysgenic syndrome. Interestingly, the telomeric TART
5 elements exhibit higher mapping abundance in strain 160, albeit below the 3-fold
6 enrichment threshold.

7
8 Thus, it appears that diverse TE families are in excess in strain 160. This is
9 consistent with the hypothesis that the invasion of the *Penelope* element itself
10 has contributed to genome instability by perhaps mobilizing other TEs within the
11 strain[13]. By contrast, elements enriched in strain 9 do not appear to be
12 transposable elements. However, it is important to note that ascertainment bias
13 may contribute to the observation that no TEs appear in significant excess in
14 strain 9. This is because the reference genome is more closely related to strain
15 160, and TEs that are entirely unique to strain 9 may be missing in the TE
16 libraries generated through computational means from the reference genome.

17
18 TE age analysis identifies different modes for TE asymmetry between strains
19
20 TE abundance between strains can result from different processes. For example,
21 long-resident TEs may be in excess in one strain due to strain or population
22 specific recent re-activation. By contrast, entirely new TEs may have invaded a
23 species and have yet to spread equally throughout the genomes of the

1 individuals within the population. The P element invasion in *D. melanogaster* is
 2 an example of the latter process [30]. It has recently invaded and is only present
 3 in current natural populations, not strains collected seven or more decades ago.
 4
 5 To distinguish among alternative processes contributing to asymmetry in TE
 6 abundance, one can perform an age analysis of given TE families. High
 7 sequence homogeneity within a TE family is an indicator of recent activation or
 8 invasion. A phylogenetic approach using full-length fragments is ideal for this
 9 purpose, but full-length TE assemblies are not available with short read
 10 sequencing technology. Therefore, we estimated TE family "age" by examining
 11 the sequence heterogeneity within mapping reads (Figure 1B). Age was
 12 estimated by considering the average frequency of the most common nucleotide,
 13 across nucleotides within the mapping. A young element that has recently
 14 invaded will show high similarity among copies, nearing 1 for an average
 15 frequency of the major nucleotide variant. Older elements, with patterns of
 16 activation that occurred more distantly in the past, accumulate differences among
 17 insertions. This accumulation of differences among multiple copies is made
 18 evident by lower nucleotide frequencies of the most common variant. Figure 1B
 19 plots these age measures for each TE within the two strains.

20 This analysis reveals two classes of TEs that are enriched in inducer
 21 strain 160. Consistent with previous analysis[26,27,31], *Penelope* and
 22 *Polyphemus* show greater homogeneity among copies in strain 160. This pattern
 23 also applies, albeit to a lesser extent, to *Helena*, *Paris*, *Skippy* and the telomeric

1 TART element. This pattern indicates recent activation from long-term resident
2 status.

3 A second class of elements with excess in strain 160 show a different
4 pattern. *Slicemaster*, *Uvir*, 258 and 734 are all either similarly aged within strains
5 or, in the case of 1069, slightly older in strain 160. In the case of elements like
6 *Slicemaster*, which are very young (>99% similarity), this can be explained by
7 recent invasion of both strains but excess movement in strain 160, rather than
8 "re-activation" from long-time resident status.

9

10 The genomic distribution of paternal induction in hybrid dysgenesis

11

12 Previous studies have indicated that the *Penelope*, *Paris* and *Helena* elements
13 are broadly distributed across multiple chromosomes of the inducer strain 160
14 [16]. Considering the much larger number of TEs that are now known to be in
15 excess in strain 160, we sought to determine the strength of induction across
16 paternally inherited 160 chromosomes. This was achieved by generating F1
17 males heterozygous for strain 160 and strain 9 chromosomes (through a non-
18 dysgenic cross), crossing these males to strain 9 females, scoring for dysgenesis
19 in F2 progeny, and genotyping F2s for chromosomes inherited from the
20 heterozygous father. Overall, we found that induction of dysgenesis is distributed
21 across all tested chromosomes, with the exception of the dot sixth chromosome
22 (Figure 1C). Thus, an excess of diverse TEs in strain 160 is associated with
23 induction distributed across the genome. In light of this, we sought to determine

1 how TE excess in strain 160 corresponded with patterns of increased TE
2 expression in progeny of the dysgenic cross.

3

4 Patterns of TE expression in a dysgenic cross

5

6 To determine how TE excess in strain 160 corresponds with TE expression in
7 dysgenic progeny, we performed mRNA-seq from pooled 0-2 hour old embryos
8 laid by F1 females of both the dysgenic (9 females X 160 males) and non-
9 dysgenic (160 females X 9 males) directions of the cross. One may also measure
10 expression in F1 ovaries, but here this is not preferred since dysgenic ovaries are
11 atrophied and expression analysis from these tissues is confounded by altered
12 ratios of somatic and germline tissue. 0-2 hour old embryos laid by F1 mothers
13 represent a sample of pure germline tissue since zygotic transcription in *D. virilis*,
14 as measured with the early, zygotic *fushi-tarazu* (*ftz*) gene, begins after 2 hours
15 [32]. Confirmation that embryos in this sample were collected prior to the onset of
16 zygotic transcription was obtained by inspecting expression of *ftz*.

17

18 Full penetrance of dysgenesis, evidenced by atrophied gonads, is observed in
19 approximately 50% of male and female progeny from 9 female X 160 male
20 crosses. Therefore, embryos analyzed by mRNA-seq were those laid by mothers
21 that escaped full sterility. For clarity, these tissues will be referred to as dysgenic,
22 even though these tissues escaped complete atrophy. Sexual maturity in *D. virilis*
23 is at about 5 days. To determine the dynamics of TE expression as flies aged, we

1 analyzed mRNA-seq data from embryos laid by F1 mothers 12 to 16 days old, as
2 well as 19 to 21 days old.

3

4 mRNA-seq analysis of dysgenic and non-dysgenic germline indicated different
5 modes of TE mis-expression. TEs with excess abundance in the inducer strain
6 were broadly overexpressed in the dysgenic germline. There were high
7 magnitude differences for some elements such as *Skippy* and *Helena* but for
8 others, such as *Paris*, *Polyphemus* and *TART*, differences were more modest
9 (Figure 2A and B). Coupled with this, there was also a global pattern of
10 overexpression for all TEs in the dysgenic germline, as evidenced by RPKM
11 levels above the diagonal (Figure 2A and B). This pattern of global TE
12 derepression applies largely to TEs that are lowly expressed. It should be noted
13 that some caution must be taken in the interpretation of lowly expressed TEs.
14 RPKM is fundamentally a proportional measure which is sensitive to changes in
15 expression of highly expressed genes (see below). Nonetheless, visual
16 inspection shows that there are at least three elements, not members of the class
17 enriched in strain 160, that show increased expression in the dysgenic germline
18 and are not lowly expressed (>8 RPKM). Importantly, expression level
19 differences do not ameliorate with age of mother (compare Figures 2A and B).

20

21 We next sought to determine whether the degree of expression in dysgenic
22 progeny was largely determined by differences in bulk TE abundance between
23 the two strains (Figure 2C). Ten of eleven elements that are more abundant in

1 the inducer strain were expressed at higher levels in dysgenic progeny. However,
2 many upregulated elements are similar in abundance between strains. In fact,
3 many elements that are *more* abundant in strain 9 are upregulated in dysgenesis
4 (Figure 2C). This includes the three elements that are not lowly expressed.

5

6 Global Genic Mis-expression in a Dysgenic Cross

7

8 To determine how global TE derepression might modulate genic expression, we
9 analyzed endogenous gene expression within the same 0-2 hour old embryos.

10 There was a general increase in expression for lowly expressed genes (<5
11 RPKM) in the F1 progeny of the dysgenic cross. This is most easily
12 demonstrated by observing relative expression level as a function of rank
13 expression level genome wide (Figure 3). This coincided with a modest decrease
14 in expression of the most highly expressed genes (Figure 3, right panel). Since
15 RPKM is a proportional measure, a change in expression in highly expressed
16 genes is expected to increase lowly expressed genes, and vice versa. Thus, it
17 appears that the global distribution of gene expression is tilted in the dysgenic
18 germline. To determine if this result could be explained by differences in the
19 expression of genes primarily residing in heterochromatin, we excluded putative
20 heterochromatic scaffolds smaller than 4 million base pairs and excluded genes
21 residing closer than 1 million base pairs from either end of the remaining
22 scaffolds. To further ensure this result was not driven by TEs that were
23 computationally misidentified as genes, we eliminated all genes lacking orthology

1 to *D. melanogaster*. Even after accounting for these effects, the trend is
2 maintained (Supplemental 2), and is observed in technical replicates within
3 biological samples as well as in flies of different ages.
4
5 Additionally, we performed differential gene expression analysis based on two
6 metrics. We considered both fold-change in expression as well as p-values for
7 differential expression by DESeq2[33] and edgeR[34,35]. Based on these
8 metrics, we performed gene ontology enrichment analysis using GOrilla [36,37],
9 which performs functional enrichment analysis based on ranked lists rather than
10 cutoffs (Supplemental File 3). When sorted by fold expression level, we found
11 that genes involved in positively regulating transcription are significantly enriched
12 among genes more highly expressed in the dysgenic germline, consistent with
13 our observed upregulation of many genes with dysgenesis. Among genes down
14 regulated, genes involved in chitin synthesis were significantly enriched.
15 However, clustering of chitin related genes within the genome causes a lack of
16 independence among these genes. When sorted by p value, there is no
17 enrichment of classes among genes upregulated in dysgenesis, perhaps arising
18 from low statistical power for genes with low expression. For genes
19 downregulated in dysgenesis, we again find that genes in the chitin synthesis
20 pathway are down regulated. In addition, we find that many highly expressed
21 ribosomal proteins as well as members of the SWI/SNF family of chromatin
22 remodelers are enriched for down regulation in dysgenesis. In the former case,

1 this is consistent with results showing a shift toward reduced expression among
2 highly expressed genes.

3

4 Germline transposable element piRNA in a dysgenic cross

5

6 To determine how maternal inheritance of piRNAs might explain persistent TE
7 mis-expression in the dysgenic germline we sequenced piRNAs (defined as
8 small RNAs, 23-30 nt, filtered against known non-piRNA classes) from 0-2 hour
9 old embryos laid by strain 9 and strain 160 mothers. This allows a measure of
10 maternally provisioned piRNA load. In addition, we sequenced piRNAs in
11 embryos laid by F1 females from both reciprocal cross directions between the
12 two strains to determine whether perturbed piRNA biogenesis during dysgenesis
13 might maintain increased TE expression. For F1 germline piRNAs, we collected
14 embryos from the same pool of mothers used for mRNA-seq, but at intermediate
15 maternal age (15-16 days old). This allowed us to determine whether the
16 persistent mis-expression of TEs in the F1 germline of the dysgenic cross could
17 be explained by a persisting defect in piRNA biogenesis.

18

19 Figure 4A shows that there is a large number of TEs, including many of those
20 with greater copy number in Strain 160, with higher levels of 160 maternally
21 provisioned piRNA. Strikingly, despite the asymmetry in maternal provisioning,
22 piRNA levels largely equilibrated in reciprocal F1 germlines (Figure 4B). A
23 significant exception to this is the *Helena* element.

1

2 Figure 4C demonstrates the degree to which asymmetry in maternal provisioning
3 predicts TE expression in reciprocal progeny in the next generation. Many of the
4 elements that are more abundant in strain 160 have greater piRNA abundance in
5 the 160 female germline and nearly all of these show greater expression in the
6 germline of the dysgenic cross. However, at least two elements that are
7 overexpressed in the dysgenic germline (indicated with a *) show equal or higher
8 piRNA abundance in the 9 germline. The second most differentially expressed
9 element in the dysgenic germline shows little difference in piRNA abundance
10 between strains. Therefore, maternal provisioning of piRNA is an imperfect global
11 predictor of TE expression in hybrid dysgenesis.

12

13 To determine whether this chronic level of TE mis-expression might be explained
14 by persistent asymmetry in germline F1 piRNA pools, we investigated the degree
15 to which asymmetry in F1 piRNA levels was predictive of expression in reciprocal
16 progeny. Overall, there is a poor relationship between relative piRNA abundance
17 for a given TE and its relative expression difference in dysgenic compared to
18 non-dysgenic germlines. Furthermore, there is a contrasting pattern within the
19 class of TEs that are higher in copy number in 160 and also more highly
20 expressed in the dysgenic germline. In particular, two elements that become
21 overexpressed during dysgenesis (*Skipper* and *Slicemaster*) in fact have higher
22 piRNA abundances in the dysgenic germline. Thus, it is clear that piRNA
23 biogenesis is maintained, but is not sufficient to maintain TEs at a low level of

1 expression. *Helena* is the only element more abundant in strain 160 (at a 3X
2 level) whose increased expression level in dysgenesis is explained by reduced
3 levels of F1 piRNA.

4

5 Raw abundance measures of piRNAs ignore critical aspects of their biogenesis
6 and two recent studies have demonstrated that globally reduced signatures of
7 robust piRNA biogenesis likely contribute to the mobilization of diverse TEs
8 [21,38]. Overall, we see no evidence that piRNA biogenesis is skewed based on
9 the size distribution of small RNA reads (Figure 5A). For each TE we estimated
10 the percent ping-pong[20] as well as the density of ping-pong pairs in dysgenic
11 and non-dysgenic germline. When comparing metrics directly, there is little
12 evidence that piRNA biogenesis is grossly perturbed in the dysgenic cross
13 though a more sensitive comparison using normalized Z-scores indicates a
14 modest shift in piRNA abundance beyond what would be expected based on
15 patterns of maternal provisioning (Figure 5B). This is observed in the z-score
16 heat maps for abundance and ping-pong pair density, which show an excess of
17 negative z-scores for dysgenesis ($p < 0.0001$ for both differences, Wilcoxon
18 Signed-Rank Test). Importantly, these are both normalized, proportional
19 measures of abundance that are likely influenced by abundance of genic piRNAs
20 in the same library (see below).

21

22 In contrast to these abundance measures, there is no evidence for global ping-
23 pong biogenesis disruption as measured by percent ping-pong. Note, for

1 example, that many row-z scores for percent ping-pong show weaker z-scores
 2 for non-dysgenic compared to dysgenic (compare upper and lower portions of
 3 the percent ping-pong z-score heatmap). A Wilcoxon Signed-Rank test also
 4 found no significant difference ($p=0.22$) in percent ping-pong between
 5 dysgenesis and non-dysgenesis. Thus, in contrast to previous studies, we find no
 6 evidence for a global disruption of piRNA biogenesis.

7 Despite this, one can see that elements that are overexpressed in
 8 dysgenesis (red bars) are in excess among elements with a reduced ping-pong
 9 signature (bottom portion of percent ping-pong z-score heatmap). Furthermore,
 10 there is a slightly significant correlation between difference in percent ping-pong
 11 z-score and fold expression ($p=0.048$, Figure 5C). This trend is driven by the top
 12 eight elements that show higher expression in dysgenesis and all have lower z-
 13 scores in dysgenesis. Overall, this does not support a model of global disruption
 14 in piRNA biogenesis maintained in adult flies. Rather, the persistence of higher
 15 expression for some TEs is driven by idiosyncratic defects in piRNA biogenesis
 16 that occur on a TE-by-TE basis.

17

18

19 Genic piRNAs in hybrid dysgenesis

20

21 Recent studies have demonstrated that in addition to TEs, genes may also be
 22 the target of piRNA silencing. Genic targeting by piRNAs can arise from
 23 neighboring TE insertions that drag flanking sequences into piRNA biogenesis

1 and gene silencing in *cis* [39,40,41,42,43]. Previous important work in the *D.*
 2 *virilis* system of dysgenesis identified a piRNA cluster overlapping the *center*
 3 *divider* (*cdi*) gene (Dvir\GJ14359) [44]. We thus sought to determine how the
 4 global landscape of genic piRNAs was influenced by the activation of diverse
 5 transposable elements. Global gene expression might be modulated if genic
 6 piRNA silencing were either attenuated or enhanced during dysgenesis. To avoid
 7 gene models that might designate TEs to be genes, we only examined the CDS
 8 regions of *D. melanogaster* orthologs.

9

10 In the dysgenic germline the number of genes being processed into
 11 piRNAs was significantly higher than in non-dysgenesis. Figure 6A indicates all
 12 CDS's that have at least 50 piRNAs per 10 million mapped, row normalized by z-
 13 score. A very large number in dysgenesis show a unique signature of being
 14 targeted by piRNA and these *de novo* targets largely become a source of sense
 15 only piRNAs (Figure 6B). The genes that showed a significant increase in piRNA
 16 were also more highly expressed above the genome wide background (Figure
 17 6C, Wilcoxon rank test, $p < 0.0001$) and tend to become more lowly expressed in
 18 the dysgenic germline compared to the non-dysgenic germline (Figure 6D,
 19 Wilcoxon rank test, $p < 0.0002$). GO enrichment analysis indicated that genes that
 20 specifically become sense piRNA biogenesis targets are enriched for highly
 21 expressed ribosomal protein RNAs (FDR p-value: $2E-08$, Supplemental 3). This
 22 is concordant with mRNA-seq results from independent RNA collections that
 23 show that ribosomal protein genes are also enriched among down regulated

1 genes in dysgenesis. The fact that these piRNAs are predominantly sense
 2 suggests that the observed modest levels of reduction in gene expression for
 3 highly expressed ribosomal genes may be explained by being processed into
 4 piRNA biogenesis in an off-target manner without a corresponding anti-sense
 5 pool that amplifies the off-target silencing. One possibility is that the mechanism
 6 underlying perturbed piRNA biogenesis for only a subset of TEs also drives genic
 7 off-targeting of highly expressed genes. This may occur through cytoplasmic,
 8 non-specific slicing of mRNAs that are highly abundant.

9

10 Genetic Analysis of the Cause of Hybrid Dysgenesis

11

12 We have shown that hybrid dysgenesis in *D. virilis* is associated with a persisting
 13 global shift in both TE and gene expression. We next sought to identify the
 14 genomic regions of strain 160 that most protect against hybrid dysgenesis in a
 15 cross with 160 males. Excess TART abundance, as well as previous reports of
 16 unique genic cluster behavior at chromosome ends [44], indicate a possible role
 17 of telomeres. Since maternal repression of dysgenesis is a dominant trait, we
 18 diluted fragments of the 160 genome onto a strain 9 background through
 19 backcrossing hybrid females to strain 9 males. Two generations of backcrossing
 20 through females allowed multiple generations of recombination in the absence of
 21 dysgenesis. We then tested F3 females for protective ability by crossing to
 22 inducer males, followed by genotyping at markers distributed across the genome.
 23 In particular, each F3 chromosome was genotyped for a SNP marker near the

1 telomere, near the centromere and also along the euchromatin of the
2 chromosomal arm.

3

4 With this backcross scheme, we generated a panel of females with significant
5 variation in their ability to protect against induction when crossed to 160 inducer
6 males (Figure 7A). QTL analysis identified three genomic regions from strain 160
7 that significantly protected against F1 sterility (Figure 7B) The region most
8 significantly protective corresponded to the pericentric region of chromosome 5.

9 The pericentric region of the X chromosome was also significantly protective,
10 followed by a euchromatic region in the proximal arm region of chromosome 4. A
11 previous study indicated that the most protective chromosome was also
12 chromosome 5, followed by the X and then chromosome 4 [32]. Interestingly, the
13 most protective region of the *D. virilis* genome corresponds to the pericentric
14 region of the same Muller element that carries cluster 42AB in *D. melanogaster*.

15 This provides strong genetic evidence for a role of pericentric, cluster-derived
16 piRNAs in the protection against dysgenesis in *D. virilis*. By contrast, no telomeric
17 region provides significant protection. Thus, amplified TART elements seem to
18 be a passenger of TE destabilization in strain 160 rather than a driver.

19

20 In addition to standard QTL analysis, to determine precise regions of the 160
21 genome critical for protection, we performed whole genome sequencing of the
22 top 6 most protective females for which DNA was available (Figure 7C). A
23 synthetic diploid genome sequence was generated from both strain 160 and

1 strain 9 haploid consensus sequences and reads that strictly mapped uniquely to
2 the strain 160 haploid genome were identified. This allowed us to identify regions
3 of heterozygosity for strain 160 over a strain 9 background for the six most
4 protective mothers. This provided maximal resolution for a test to rule out
5 whether any single genomic region is critical for maternal protection. From this
6 analysis, we found that among the most protective mothers there is no single
7 genomic region of strain 160 always present in each most protective mother.
8 Across all six mothers, we identified at least one mother homozygous for strain 9
9 at each position of the genome (excluding unassembled regions). Thus, no single
10 genomic region appears critical for protection against dysgenesis.

11

12 To determine whether piRNA from any particular TE enriched in 160 was
13 dispensable for protection, we sequenced small RNAs from the ovaries of the six
14 most protective mothers. We first reasoned that the only TEs that are candidate
15 inducers of dysgenesis are those for which piRNA abundance is greater in strain
16 160 than in strain 9. Among the 221 repeats within the library, there are 141 that
17 meet this qualification. These are candidates for elements contributing to the
18 dysgenic syndrome. We further reasoned that if any female with full repressive
19 ability had piRNA abundances for a TE that was similar to strain 9, we could also
20 rule out that TE as a driver of dysgenesis. Of the remaining 141 candidates, 88
21 have at least one protective female that has normalized piRNA abundance less
22 than or equal to strain 9. Thus, we are left with 53 candidate contributing repeats.
23 Using these criteria, we were unable to eliminate any of the elements 3-fold

1 enriched from strain 160. However, the data further demonstrate a minimal role
 2 for the TART element piRNAs as mediators of maternal protection. This is
 3 because the vast majority of TART piRNAs in the protective mothers are derived
 4 from the tip of the X chromosome from 160 (Figure 7C) and three of the six
 5 females that are most protective against dysgenesis lack the X 160 telomere
 6 allele. Notably, the telomeric region of the X has special silencing properties in *D.*
 7 *melanogaster* that may explain the excess of TART piRNAs derived from this one
 8 genomic region [45,46,47,48,49], but in this system this region plays no role in
 9 protection against dysgenesis.

10

11 Genic piRNA cluster behavior across generations

12

13 A previous study by Rozhkov *et al.* demonstrated that strain 160 possesses a
 14 number of piRNA clusters that are absent in strain 9 [50]. This may be related to
 15 the fact that strain 160 also carries many TEs in greater copy number compared
 16 to strain 9. Additionally, it was noted that many of these clusters were telomeric
 17 which is consistent with results here that show a signature of increased TART
 18 activation in strain 160. The 160 and 9 strains in our lab have been separated
 19 from these strains for more than 15 years, so this allows a comparison of cluster
 20 properties among strains of flies that have been long separated. The most well-
 21 characterized cluster from Rozhkov *et al.* was a telomeric cluster at the tip of
 22 chromosome 2 encompassing the gene *center divider* (*cdi*). This dual strand
 23 piRNA cluster was only in strain 160 and absent in strain 9 and this pattern was

1 also observed in our divergent laboratory stocks (Figure 8). Interestingly, a
 2 second cluster identified by Rozhkov *et al*, near the telomere of the 6th
 3 chromosome in strain 160, but absent in strain 9, showed an opposite pattern in
 4 our strains. This cluster in our strain 160 had 318 unique mappers per million
 5 mapped whereas our strain 9 had 3,836 unique mappers per million mapped.
 6 Rather than being independently gained twice, the most parsimonious
 7 explanation for this was that this cluster was originally in both lines, but
 8 independently lost in our strain 160 and their strain 9. Thus, some piRNA clusters
 9 may be prone to losing their activity over time.

10

11 Additionally, we identified the *oysgedart* gene (Dvir\GJ17620) in strain 9 that
 12 displayed a unique mode of genic piRNA targeting. A *Ulysses* element insertion
 13 upstream of this gene in strain 9 is specifically associated with genic silencing
 14 and shunting *oysgedart* into piRNA biogenesis (Figure 8). piRNAs derived from
 15 *oysgedart* are dual-strand in strain 9 ovaries (which includes somatic and
 16 germline tissue) but biased towards the sense strand in the germline.

17

18 To determine how different modes of genic targeting by piRNA influences gene
 19 expression across generations, we examined patterns of gene expression in the
 20 germline in strain 9 and 160 parents (from 0-2 hour embryos), as well as
 21 reciprocal offspring. For reciprocal hybrid offspring, we also examined patterns of
 22 somatic gene expression in carcasses with ovaries removed to determine how
 23 maternally provisioned piRNA might influence somatic expression in the next

1 generation. Finally, we examined cluster behavior in the F3 backcrosses (using
2 ovary piRNA data) to determine how genic clusters are maintained when the
3 original cluster allele is segregated away from a non-cluster allele.

4

5 We found that two modes of genic targeting by piRNAs result in contrasting
6 behavior across generations. The *cdi* dual strand cluster is maintained as a
7 source of germline piRNAs when it is inherited from the mother (Figure 8).
8 However, when inherited paternally, the cluster is inactive in the germline. In
9 addition, *cdi* expression is highly reduced in the germline of strain 160 and the
10 germline silencing of both alleles of *cdi* is maintained when transmitted
11 maternally and made heterozygous in combination with the wild-type strain 9
12 allele. In contrast, when the *cdi* cluster allele is transmitted paternally, piRNA
13 cluster activity is reduced and germline expression levels are maintained near
14 the level of strain 9, with similar contributions from each allele. This asymmetry in
15 gene expression of *cdi* (off in non-dysgenic, on in dysgenic), however, is not
16 observed in the soma. RNA collected from carcasses of F1 females, with ovaries
17 removed, of dysgenic and non-dysgenic crosses (3 individual carcasses
18 performed separately from each direction), showed similar levels of expression
19 from both directions of the cross (Dysgenic Carcass: 17.1 RPKM [9 Allele: 7
20 counts; 160 Allele: 9 counts] vs. Non-dysgenic Carcass: 23.4 RPKM [9 Allele: 5
21 counts; 160 Allele: 12 counts]) Therefore, silencing of germline *cdi* occurs in
22 *trans* when the cluster is inherited maternally, but this has no effect on the
23 somatic expression.

1

2 The *oysgedart* allele in strain 9, being silenced by off-target processing into
3 primarily sense piRNA in the germline, provides a significant contrast. In neither
4 direction of the cross is this form of sense piRNA biogenesis maintained in
5 progeny. Instead, the wild-type allele from strain 160 seems to function in *trans* to
6 limit this mode of silencing. Thus, in both directions of the cross, the expression
7 of *oysgedart* is maintained, but only at about 60% wildtype level, since the
8 *Ulysses* insertion allele is more lowly expressed. While sense piRNA biogenesis
9 is turned off in both reciprocal directions, expression of the 9 allele is always
10 reduced in *cis*. Interestingly, in neither direction of the cross does there appear a
11 maternal effect on somatic expression for *oysgedart*. Even though expression of
12 the 9 allele is reduced in the germline, it remains on in the soma of both dysgenic
13 (RPKM: 62.0, [9 Allele: 21 counts; 160 Allele: 21 counts]) and non-dysgenic
14 (RPKM: 80.9, [9 Allele: 29 counts; 160 Allele: 25 counts]) F1 females. This also
15 demonstrates that the allelic *cis* effects on local silencing by the *Ulysses* insertion
16 are germline, not soma, specific (Fisher's exact test for difference in allele effects
17 between soma and germline. Dysgenic: $p < 0.0001$, Nondysgenic: $p < 0.0001$)

18

19 Using the genome and small RNA data from the six females used for genetic
20 analysis, we were also able to observe the behavior of the *cdi* cluster across
21 multiple generations in cases where the original *cdi* dual strain cluster allele from
22 160 had been segregated away and only a homozygous strain 9 allele was
23 present. From this analysis, the *cdi* cluster demonstrates the ability to modestly

1 paramutate piRNA biogenesis onto the strain 9 allele. To varying degrees, cluster
2 behavior was maintained in most flies homozygous for the strain 9 allele of *cdi*. In
3 one case, cluster behavior was lost. Because the cross scheme was maintained
4 over two generations, we are unable to determine the generation in which the
5 cluster activity is lost. However, it is clear that *cdi* can paramutate in one
6 generation but that this paramutation is not robust across multiple generations.
7 This was never observed for *oysgedart*.

8

9 Discussion

10

11 Females that are the product of reciprocal crosses are genetically identical but
12 differ from each other epigenetically due to parental genotypic effects. Hybrid
13 dysgenesis represents an extreme form of phenotypic variation among
14 genetically identical individuals. Here we show that differences between
15 genetically identical dysgenic and non-dysgenic individuals is manifested in
16 multiple ways. When genomes from two strains of *D. virilis* are brought together,
17 TEs that are more abundant in one genome become more highly expressed in
18 the germline of the next generation. Coincident with this, there is a persisting
19 increase in TE expression for several TEs that are evenly distributed between
20 strains. This can be explained by defects in piRNA biogenesis that occur on a
21 TE-by-TE basis. In addition, there is a global shift in the distribution of
22 endogenous gene expression. Our results suggest that changes to endogenous
23 gene expression are, in part, a consequence of *de novo* genic piRNA production

1 in dysgenic compared to non-dysgenic progeny. Highly expressed genes
2 become off-targets for piRNA biogenesis and their expression is reduced. The
3 mechanism for this is unclear, but may be driven by the same mechanism that
4 leads to idiosyncratic defects in piRNA biogenesis for some TEs.

5

6 We identify a large number of differences in TE expression, genic expression,
7 TE-derived and genic-derived piRNA biogenesis between dysgenic and non-
8 dysgenic progeny. Therefore, it is difficult to distinguish between causal factors
9 and downstream effects. However, our genetic analyses clearly demonstrate that
10 both induction of and protection against dysgenesis is distributed across the
11 genome. Since multiple TE families are in excess copy number in the inducer
12 strain, the weight of evidence favors a complex mode of hybrid dysgenesis driven
13 jointly by multiple elements. This is supported by the fact that the regions of the
14 genome that are most protective against dysgenesis are located in the pericentric
15 regions, which are known to be critical sources of piRNAs [51].

16

17 In principle, there may be multiple mechanisms underlying the effects of
18 dysgenesis on TE derepression and increased genic targeting. This is because
19 other maternal effects are evident in this system. In particular, germline
20 expression of *cdi* differs between dysgenic and non-dysgenic progeny.
21 Differences in levels of expression of this gene, a tyrosine-protein kinase, could
22 mediate cascading effects in different ways. Alleles of *cdi* share properties with
23 imprinted genes since expression depends on which parent the allele is inherited

1 from. However, *cdi* differs from canonical imprinting in that a silenced *cdi* allele is
 2 capable of silencing the other allele in *trans* when inherited maternally, similar to
 3 paramutations observed in maize, mice and recently in *Drosophila melanogaster*
 4 [52,53,54,55]. Imprinted genes can have a significant downstream effect on
 5 patterns of gene expression [56]. Even in *Drosophila*, where there is no evidence
 6 for DNA methylation, there are significant, albeit poorly understood, parent-of-
 7 origin allelic effects on global gene expression [57,58].

8
 9 Two recent studies have identified globally defective piRNA biogenesis as a
 10 cause for global TE mis-regulation in hybrids between genetically divergent
 11 strains (e.g., the P-element system in *D. melanogaster* and hybrids between *D.*
 12 *melanogaster* and *D. simulans*) [21,38]. In the *D. virilis* system studies here,
 13 differences in maternal piRNA provisioning likely explain the activation of multiple
 14 TEs in dysgenic progeny germline. However, we only find modest evidence that
 15 defects in piRNA biogenesis contribute to global TE mis-expression with
 16 dysgenesis. Even while TE mis-expression persists, piRNA abundance is largely
 17 equilibrated between reciprocal progeny, with the exception of the *Helena*
 18 element. Only modest defects are observed in piRNA biogenesis signatures and
 19 these perturbances only pertain to some TEs.

20
 21 Related to this point, there are several critical differences between this syndrome
 22 of dysgenesis and the *D. melanogaster* P element system that has been most
 23 thoroughly characterized. First, the induction of gonadal atrophy is nearly 100%

1 at the sensitive temperature in the P element system. In the *D. virilis* system,
2 complete sterility is observed in approximately 50% of F1s, and the sterility effect
3 has ameliorated since it was first observed[10]. Second, near complete P
4 element mediated sterility is proposed to be gradually rescued by new TE
5 insertions into piRNA clusters as flies age[21]. These new insertions establish a
6 new pool of protective piRNA that in turn lead to restored TE repression. A
7 significant difference between the P element system and the *D. virilis* system is
8 with respect to the ability of F1 females of the dysgenic cross to again protect in
9 a second generation cross. In *D. virilis*, near complete protection against
10 dysgenesis is present in heterozygous females that have inherited all 160
11 chromosomes paternally [32]. Thus, the ability to transmit protection against
12 dysgenesis is restored in one generation and this is born out by the fact that
13 piRNA pools are largely equilibrated in reciprocal progeny. Functional restoration
14 of maternal protection in the *D. virilis* system argues strongly against a persistent
15 global defect in piRNA biogenesis in the adult F1 germline. Importantly, many of
16 the elements that are in excess in 160 do in fact exist as older variants in strain
17 9. piRNAs targeting these older TE variants are present in strain 9, albeit at a
18 lower level than in strain 160. Previous studies have demonstrated that older
19 divergent piRNA pools play an important role in re-establishing control of re-
20 proliferating TEs [18]. By contrast, multiple generations of female P chromosome
21 transmission are required to re-establish the P cytotype after one generation of
22 paternal transmission [59].

1 It has recently been proposed that parent-of-origin effects made evident from
2 reciprocal F1 crosses may make an important contribution to heritability [56].
3 These effects are mediated through a regulatory network via imprinted genes.
4 Here, using a system of hybrid dysgenesis in *D. virilis*, we show that parent-of-
5 origin effects can influence germline integrity but also influence the expression
6 level of many genes through numerous indirect mechanisms. It will be interesting
7 to determine whether early events in germline establishment mediated by parent-
8 of-origin effects can lead to persistent differences in gene expression among
9 individuals in natural populations.

10

11

1 **Methods**

3 **Custom *D. virilis* TE library**

5 Few annotated TEs are available for *D. virilis*. Therefore, we combined available
6 annotated TE sequences with two computationally predicted libraries (generated
7 with PILER [60] and REaS [61,62]
8 ftp://ftp.flybase.net/genomes/aaa/transposable_elements/) to generate a
9 manually curated library. Several annotations (*Uvir*, *Helena*, *TART*, *Telemac*)
10 were improved by manual curation. Portions of the PILER library were also
11 manually curated. Additional sequence from the *Helena* element was obtained by
12 interrogating a *de novo* assembly of the strain 160 genome. Redundancy was
13 removed from this combined library first by removing repeats with significant
14 blastn hits between and within the PILER and annotated library, with priority to
15 annotated and longer sequences. With this filtered set, further redundancy was
16 removed by blasting this library with and between the REaS library.

17 **Genome Sequencing and TE measurement from strain 9 and 160**

18 Wandering 3rd larvae were collected from strain 9 and 160, rinsed with 50%
19 bleach and DNA was extracted. 100 bp paired end sequencing was performed
20 on an Illumina GAI with 400 to 500 bp fragments. TE abundance estimation was
21 performed with single ends that were trimmed by Sickle
22 (<https://github.com/najoshi/sickle>), mapped to the TE library with BWA-MEM[29]
23 and normalized by read numbers mapping the reference. Homogeneity within
24 mapped reads was measured using piledriver (github.com/arq5x/piledriver) and

1 averaging the frequency of the major allele across all nucleotides within the
2 mapping for each TE.

3 **Estimation of zygotic effects of paternally inherited chromosomes.**

4 From the genome sequences of strain 9 and 160, restriction fragment length
5 polymorphisms were identified that distinguish between the strains at two
6 positions for each chromosome. F1 males were generated from a non-dysgenic
7 cross and these males were crossed to strain 9. 96 F2 progeny were collected,
8 (48 from each class, dysgenic or non-dysgenic) and genotyped for chromosomes
9 inherited paternally with RFLPs. Log-odds ratios for the probability for being
10 dysgenic with a given chromosome were estimated using a generalized linear
11 model for logistic regression (binomial family with a logit link) in R. Some failed
12 genotypes resulted in N=92.

13 **mRNA seq: Dysgenic and Nondysgenic germline, Strains 9 and 160** 14 **germline, dysgenic and non-dysgenic soma.**

15 RNA for sequencing was collected from embryos laid by F1 mothers from the
16 dysgenic and non-dysgenic directions of the cross. Ovaries were not selected
17 because dysgenic ovaries are often atrophied. Therefore, the germline tissue
18 represented in this experiment is derived from mothers that have escaped
19 germline ablation. Paternal effects on embryos that might occur when dysgenic
20 females are mated with sterile dysgenic brothers were minimized by equally
21 mixing males from reciprocal directions of the cross and allocating them in
22 mating cages between reciprocal F1 females. This also ensured improved egg
23 laying from dysgenic females. Females were maintained as continuously laying

1 with a constant supply of yeast and grape plates and eggs were collected after 0-
2 2 hour durations and flash frozen in liquid nitrogen. RNA was pooled from
3 collections of 12 to 16 day old mothers and 19 to 21 day old mothers. From each
4 biological age replicate, two RNA-seq libraries were generated for single-end, 50
5 bp sequencing, for a total of four libraries per condition (dysgenic or non-
6 dysgenic). Reads were quality trimmed at the 3' end (up to 16 bp) and reads with
7 2 bp of quality less than 20 were excluded using the Galaxy server. TE RPKM
8 estimates were obtained by directly mapping with BWA to the annotated TE
9 library and normalizing by mapping to the reference genome. mRNA RPKM
10 estimates were obtained using the RNA-seq tool in CLC. Fold analysis was
11 performed by calculating RPKM(+0.01) ratios. Ranking by p value was performed
12 based on p values estimated with the DEseq2 and edgeR packages from raw
13 counts. p value estimates are not true estimates since among the four libraries
14 used for DEseq2 and edgeR analysis, there are only two biological replicates.
15 However, p values were only used to estimate ranks for ontology enrichment.

16 Additional germline mRNA seq was performed using the same protocol for
17 pure strain 9 and strain 160 (RPKMs averaged for cluster analysis across 2
18 libraries each for 7-15 day old females and 15-25 day old females). Allele counts
19 were determined by direct counting within the RNA-seq mappings (summed
20 across all library mappings) for a SNP known to distinguish the two strains within
21 the transcripts for *cdi* and *oysgedart*. Somatic mRNA seq analysis was likewise
22 performed, but from using RNA from dysgenic and non-dysgenic females with

1 abdomens fully removed (3 libraries per condition, each library from a single
2 female, RPKMs averaged across libraries and allele counts estimated as before).

3 **Small RNA sequencing**

4 All small RNA was size selected from 15% acrylamide, cut between an 18 bp
5 oligo and the 30nt rRNA. Small RNA sequencing for dysgenic and non-dysgenic
6 germline material was performed on embryos laid by the same mothers as for
7 RNA seq, but at 15-16 days old, according to [63]. Small RNAs from strain 9 and
8 strain 160 pooled ovaries were sequenced according to [64] with the oxidation
9 reaction. Small RNAs from ovaries from individual F3 females was performed by
10 Fasteris.

11 **Small RNA Analysis:**

12 Reads were trimmed by removing adapters and filtered by size as piRNA (23-
13 30nt) in CLC Genomics Workbench 7.0. Reads were then filtered by mapping to
14 tRNA, ncRNA, miscellaneous RNA, and miRNA (including pre-miRNA) libraries
15 from the *D. virilis* reference genome. The filtered 23-30 nt small RNA reads were
16 mapped to our curated TE library with BWA.aln[65], using the default
17 parameters. Reads were normalized by non-unique mappers to the *D. virilis*
18 reference genome using BWA.aln defaults. Calculations for ping-pong percent
19 and density of piRNA pairs were done with the R package viRome
20 (<http://www.ark-genomics.org/bioinformatics/virome>), with some modifications.
21 For genic small RNA analysis, reads were mapped uniquely with BWA.aln to the
22 *D. virilis* reference genome, using default parameters. Reads were normalized
23 by non-unique mappers to the genome. BEDTools intersect[66] was utilized to

1 count piRNA hits on genes and CDS sequences. Percent ping-pong was defined
2 as the percent of 23 to 30 nt mapping reads that had a corresponding read on
3 the opposite strand with a 10 bp 5'-5' overlap. We also measured piRNA
4 biogenesis by determining ping-pong pair density. This measure was obtained by
5 counting all non-redundant ping-pong pairs (counting each read only once) per
6 kb.

7

8 **Genetic analysis of genomic regions from strain 160 that maternally protect**
9 **against dysgenesis**

10 F3 females were generated by crossing 160 females to strain 9 males, followed
11 by two rounds of backcrossing to strain 9 males. All but the final cross was
12 performed *en masse*. Dysgenic crosses were performed with >160 single 4 to 5
13 day old tester females mated with three 4 to 5 day old strain 160 males. Adults
14 were transferred to new vials daily and dysgenesis was estimated by counting the
15 number of dysgenic testes in progeny over all testes counted (2 per male) across
16 three broods. Females were then collected, ovaries removed (for small RNA
17 sequencing by Fasteris, top protectors only) and carcasses retained for genomic
18 DNA extraction. Genotyping was performed using the TaqMan Open Array
19 platform on all females, with the exception of the top six females that had the
20 strongest ability to protect against dysgenesis. SNPs distinguishing 160 and 9
21 chromosomes were chosen in pairs for redundancy, one pair at each telomere
22 and pericentric region, as well as one or two euchromatic SNPs. Care was taken
23 to avoid repeat sequences by screening with Blast against the reference and also

1 using RepeatMasker. F3 Females were then genotyped for 160/9 heterozygosity,
2 alongside pure strain 9 and 160 controls, by the National Jewish Health. Single
3 marker regression was carried out with RQTL after dropping individuals with
4 missing genotype data from the analysis. 5000 permutations were done to find
5 the significance threshold at $\alpha = 0.05$. For the top 6 protectors, whole
6 genome sequencing was performed (100 bp, paired-end) using the Nextera
7 library prep protocol.

8 **Genotyping by whole genome sequencing**

9 Reference genome scaffolds from *D. virilis* were concatenated according to their
10 supported positions and orientations on known Muller elements, with a large
11 scaffold arbitrarily generated by concatenating scaffolds from unknown positions.
12 Using this new "assembly" we mapped all strain 9 and strain 160 reads and
13 generated two consensus genomes. A pseudo-diploid heterozygous genome
14 was then assemble by placing these scaffolds into one file.
15 Paired-end reads from the top six protectors were mapped to the hybrid genome,
16 using BWA's default parameters, with the goal of inferring spans of
17 heterozygosity for strain 160 by identifying reads that map uniquely, under high
18 stringency, to the strain 160 haploid reference. The mapping output was piped
19 into SAMtools for filtering by quality score ($-q\ 42$) and post-alignment processing
20 (SAM to BAM conversion and indexing). A relatively high cutoff quality score was
21 used in order to remove reads that could have mapped promiscuously. We were
22 able to remove all reads that mapped to more than one locus/allele leaving us
23 with reads that are specific to either strain 9 or 160. Spans of heterozygosity for

1 strain 160 were visualized with a sliding window for read density along the 160
2 chromosomes within the psuedo-diploid reference genome.

3

4

5 **Acknowledgements**

6 We thank Christine Yoder for early development of small RNA sequencing
7 methods and Lena Hileman for providing comments on the manuscript. Jack
8 Colicchio provided assistance in mRNA-seq analysis. Clark Bloomer from the
9 KUMC Genomics core provided crucial assistance with DNA library prep and
10 mRNA sequencing. We also thank Laura Rogers of the KU CMADP Genome
11 Sequencing Core for additional mRNA sequencing. Additional DNA library prep,
12 small RNA library prep and sequencing was performed by the David H. Murdock
13 Research Institute and Easteris. We also thank Daniel LaFlamme from National
14 Jewish Health for assistance with genotyping. Funding was provided by NSF
15 MCB-1022165, an internal grant through NIH-P20GM103638 and the University
16 of Kansas.

17

References

2

- 3 1. Hickey DA (1982) Selfish DNA: A sexually-transmitted nuclear parasite.
4 Genetics 101: 519-531.
- 5 2. Charlesworth B, Charlesworth D (1983) The population dynamics of
6 transposable elements. Genetical Research 42: 1-27.
- 7 3. Charlesworth B, Langley CH (1989) The population genetics of *Drosophila*
8 transposable elements. Annual Review of Genetics 23: 251-287.
- 9 4. Kidwell MG, Novy JB (1979) Hybrid dysgenesis in *Drosophila melanogaster* -
10 Sterility resulting from gonadal-dysgenesis in the P-M system. Genetics
11 92: 1127-1140.
- 12 5. Kidwell MG, Kidwell JF, Sved JA (1977) Hybrid Dysgenesis in *Drosophila*
13 *melanogaster*: A Syndrome of Aberrant Traits Including Mutation, Sterility
14 and Male Recombination. Genetics 86: 813-833.
- 15 6. Bingham PM, Kidwell MG, Rubin GM (1982) The molecular basis of P-M
16 dysgenesis - The role of the P-element, a P-Strain-Specific Transposon
17 Family. Cell 29: 995-1004.
- 18 7. Bucheton A, Paro R, Sang HM, Pelisson A, Finnegan DJ (1984) The molecular
19 basis of the I-R hybrid dysgenesis syndrome in *Drosophila melanogaster* -
20 Identification, cloning and properties of the I-Factor. Cell 38: 153-163.
- 21 8. Yannopoulos G, Stamatis N, Monastirioti M, Hatzopoulos P, Louis C (1987)
22 *hobo* is responsible for the induction of hybrid dysgenesis by strains of
23 *Drosophila melanogaster* bearing the male recombination factor 23.5MRF.
24 Cell 49: 487-495.
- 25 9. Engels WR (1989) P Elements in *Drosophila melanogaster*. In: Berg DE, Howe
26 MH, editors. Mobile DNA. Washington, D.C.: American Society for
27 Microbiology.
- 28 10. Lozovskaya ER, Scheinker VS, Evgen'ev MB (1990) A hybrid dysgenesis
29 syndrome in *Drosophila virilis*. Genetics 126: 619-623.
- 30 11. Evgen'ev MB, Zelentsova H, Shostak N, Kozitsina M, Barskyi V, et al. (1997)
31 *Penelope*, a new family of transposable elements and its possible role in
32 hybrid dysgenesis in *Drosophila virilis*. Proceedings of the National
33 Academy of Sciences of the United States of America 94: 196-201.
- 34 12. Pyatkov KI, Shostak NG, Zelentsova ES, Lyozin GT, Melekhin MI, et al.
35 (2002) Penelope retroelements from *Drosophila virilis* are active after
36 transformation of *Drosophila melanogaster*. Proceedings of the National
37 Academy of Sciences of the United States of America 99: 16150-16155.
- 38 13. Evgen'ev MB (2013) What happens when Penelope comes?: An unusual
39 retroelement invades a host species genome exploring different
40 strategies. Mob Genet Elements 3: e24542.
- 41 14. Pyatkov KI, Arkhipova IR, Malkova NV, Finnegan DJ, Evgen'ev MB (2004)
42 Reverse transcriptase and endonuclease activities encoded by Penelope-
43 like retroelements. Proceedings of the National Academy of Sciences of
44 the United States of America 101: 14719-14724.

- 1 15. Petrov DA, Schutzman JL, Hartl DL, Lozovskaya ER (1995) Diverse
2 transposable elements are mobilized in hybrid dysgenesis in *Drosophila*
3 *virilis* Proceedings of the National Academy of Sciences of the United
4 States of America 92: 8050-8054.
- 5 16. Vieira J, Vieira CP, Hartl DL, Lozovskaya ER (1998) Factors contributing to
6 the hybrid dysgenesis syndrome in *Drosophila virilis*. Genetical Research
7 71: 109-117.
- 8 17. Aravin AA, Lagos-Quintana M, Yalcin A, Zavolan M, Marks D, et al. (2003)
9 The small RNA profile during *Drosophila melanogaster* development.
10 Developmental Cell 5: 337-350.
- 11 18. Grentzinger T, Armenise C, Brun C, Mugat B, Serrano V, et al. (2012) piRNA-
12 mediated transgenerational inheritance of an acquired trait. Genome Res
13 22: 1877-1888.
- 14 19. Chambeyron S, Popkova A, Payen-Groschene G, Brun C, Laouini D, et al.
15 (2008) piRNA-mediated nuclear accumulation of retrotransposon
16 transcripts in the *Drosophila* female germline. Proc Natl Acad Sci U S A
17 105: 14964-14969.
- 18 20. Brennecke J, Malone CD, Aravin AA, Sachidanandam R, Stark A, et al.
19 (2008) An Epigenetic Role for Maternally Inherited piRNAs in Transposon
20 Silencing. Science 322: 1387-1392.
- 21 21. Khurana JS, Wang J, Xu J, Koppetsch BS, Thomson TC, et al. (2011)
22 Adaptation to P Element Transposon Invasion in *Drosophila*
23 *melanogaster*. Cell 147: 1551-1563.
- 24 22. Gerasimova T, Mizrokhi L, Georgiev G (1984) Transposition bursts in
25 genetically unstable *Drosophila melanogaster*. Nature 309: 714-716.
- 26 23. Eggleston WB, Johnson-Schlitz DM, Engels WR (1988) P-M hybrid
27 dysgenesis does not mobilize other transposable element families in *D.*
28 *melanogaster*. Nature 331: 368-370.
- 29 24. Rozhkov NV, Zelentsova ES, Shostak NG, Evgen'ev MB (2011) Expression
30 of *Drosophila virilis* retroelements and role of small RNAs in their
31 intrastrain transposition. PLoS One 6: e21883.
- 32 25. Scheinker VS, Lozovskaya ER, Bishop JG, Corces VG, Evgen'ev MB (1990)
33 A long terminal repeat-containing retrotransposon is mobilized during
34 hybrid dysgenesis in *Drosophila virilis*. . Proceedings of the National
35 Academy of Sciences of the United States of America 87: 9615-9619.
- 36 26. Lyozin GT, Makarova KS, Velikodvorskaja VV, Zelentsova HS, Khechumian
37 RR, et al. (2001) The structure and evolution of Penelope in the virilis
38 species group of *Drosophila*: an ancient lineage of retroelements. Journal
39 of Molecular Evolution 52: 445-456.
- 40 27. Blumenstiel JP (2014) Whole genome sequencing in *Drosophila virilis*
41 identifies Polyphemus, a recently activated Tc1-like transposon with a
42 possible role in hybrid dysgenesis. Mob DNA 5: 6.
- 43 28. Rozhkov NV, Schostak NG, Zelentsova ES, Yushenova IA, Zatsepina OG, et
44 al. (2013) Evolution and dynamics of small RNA response to a
45 retroelement invasion in *Drosophila*. Mol Biol Evol 30: 397-408.

29. LI H (2013) Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXivorg arXiv:1303.3997v2.
30. Daniels SB, Peterson KR, Strausbaugh LD, Kidwell MG, Chovnick A (1990) Evidence for horizontal transmission of the P transposable element between *Drosophila* species. *Genetics* 124: 339-355.
31. Morales-Hojas R, Vieira CP, Vieira J (2006) The evolutionary history of the transposable element Penelope in the *Drosophila virilis* group of species. *J Mol Evol* 63: 262-273.
32. Blumenstiel JP, Hartl DL (2005) Evidence for maternally transmitted small interfering RNA in the repression of transposition in *Drosophila virilis*. *Proceedings of the National Academy of Sciences of the United States of America* 102: 15965-15970.
33. Love MI, Huber W, Anders S (2014) Moderate estimation of fold change and dispersion for RNA-Seq data with DESeq2. *bioRxiv*.
34. Robinson MD, Oshlack A (2010) A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol* 11: R25.
35. Robinson MD, McCarthy DJ, Smyth GK (2010) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26: 139-140.
36. Eden E, Lipson D, Yogev S, Yakhini Z (2007) Discovering motifs in ranked lists of DNA sequences. *PLoS Comput Biol* 3: e39.
37. Eden E, Navon R, Steinfeld I, Lipson D, Yakhini Z (2009) GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. *BMC Bioinformatics* 10: 48.
38. Kelleher ES, Edelman NB, Barbash DA (2012) *Drosophila* interspecific hybrids phenocopy piRNA-pathway mutants. *PLoS Biol* 10: e1001428.
39. Gu T, Elgin SC (2013) Maternal depletion of Piwi, a component of the RNAi system, impacts heterochromatin formation in *Drosophila*. *PLoS Genet* 9: e1003780.
40. Haynes KA, Caudy AA, Collins L, Elgin SCR (2006) Element 1360 and RNAi components contribute to HP1-dependent silencing of a pericentric reporter. *Current Biology* 16: 2222-2227.
41. Sentmanat MF, Elgin SC (2012) Ectopic assembly of heterochromatin in *Drosophila melanogaster* triggered by transposable elements. *Proc Natl Acad Sci U S A* 109: 14104-14109.
42. Sienski G, Donertas D, Brennecke J (2012) Transcriptional silencing of transposons by Piwi and maelstrom and its impact on chromatin state and gene expression. *Cell* 151: 964-980.
43. Olovnikov I, Ryazansky S, Shpiz S, Lavrov S, Abramov Y, et al. (2013) De novo piRNA cluster formation in the *Drosophila* germ line triggered by transgenes containing a transcribed transposon fragment. *Nucleic Acids Res* 41: 5757-5768.
44. Rozhkov NV, Aravin AA, Zelentsova ES, Schostak NG, Sachidanandam R, et al. (2010) Small RNA-based silencing strategies for transposons in the process of invading *Drosophila* species. *Rna-a Publication of the Rna Society* 16: 1634-1645.

- 1 45. Ronsseray S, Josse T, Boivin A, Anxolabehere D (2003) Telomeric
2 transgenes and trans-silencing in *Drosophila*. *Genetica* 117: 327-335.
- 3 46. Ronsseray S, Marin L, Lehmann M, Anxolabehere D (1998) Repression of
4 hybrid dysgenesis in *Drosophila melanogaster* by combinations of
5 telomeric P-element reporters and naturally occurring P elements.
6 *Genetics* 149: 1857-1866.
- 7 47. Marin L, Lehmann M, Nouaud D, Izaabel H, Anxolabehere D, et al. (2000) P-
8 element repression in *Drosophila melanogaster* by a naturally occurring
9 defective telomeric P copy. *Genetics* 155: 1841-1854.
- 10 48. Simmons MJ, Ragatz LM, Sinclair IR, Thorp MW, Buschette JT, et al. (2012)
11 Maternal enhancement of cytotype regulation in *Drosophila melanogaster*
12 by genetic interactions between telomeric P elements and non-telomeric
13 transgenic P elements. *Genet Res (Camb)* 94: 339-351.
- 14 49. Niemi JB, Raymond JD, Patrek R, Simmons MJ (2004) Establishment and
15 maintenance of the P cytotype associated with telomeric P elements in
16 *Drosophila melanogaster*. *Genetics* 166: 255-264.
- 17 50. Rozhkov NV, Aravin AA, Zelentsova ES, Schostak NG, Sachidanandam R, et
18 al. (2010) Small RNA-based silencing strategies for transposons in the
19 process of invading *Drosophila* species. *RNA* 16: 1634-1645.
- 20 51. Brennecke J, Aravin AA, Stark A, Dus M, Kellis M, et al. (2007) Discrete
21 small RNA-generating loci as master regulators of transposon activity in
22 *Drosophila*. *Cell* 128: 1089-1103.
- 23 52. de Vanssay A, Bouge AL, Boivin A, Hermant C, Teyssset L, et al. (2012)
24 Paramutation in *Drosophila* linked to emergence of a piRNA-producing
25 locus. *Nature* 490: 112-115.
- 26 53. Arteaga-Vazquez M, Sidorenko L, Rabanal FA, Shrivistava R, Nobuta K, et
27 al. (2010) RNA-mediated trans-communication can establish paramutation
28 at the b1 locus in maize. *Proc Natl Acad Sci U S A* 107: 12986-12991.
- 29 54. Alleman M, Sidorenko L, McGinnis K, Seshadri V, Dorweiler JE, et al. (2006)
30 An RNA-dependent RNA polymerase is required for paramutation in
31 maize. *Nature* 442: 295-298.
- 32 55. Chandler VL (2007) Paramutation: from maize to mice. *Cell* 128: 641-645.
- 33 56. Mott R, Yuan W, Kaisaki P, Gan X, Cleak J, et al. (2014) The architecture of
34 parent-of-origin effects in mice. *Cell* 156: 332-342.
- 35 57. Gibson G, Riley-Berger R, Harshman L, Kopp A, Vacha S, et al. (2004)
36 Extensive sex-specific nonadditivity of gene expression in *Drosophila*
37 *melanogaster*. *Genetics* 167: 1791-1799.
- 38 58. Wittkopp PJ, Haerum BK, Clark AG (2006) Parent-of-origin effects on mRNA
39 expression in *Drosophila melanogaster* not caused by genomic imprinting.
40 *Genetics* 173: 1817-1821.
- 41 59. Engels WR (1979) Hybrid dysgenesis in *Drosophila melanogaster* - Rules of
42 inheritance of female sterility. *Genetical Research* 33: 219-236.
- 43 60. Smith CD, Edgar RC, Yandell MD, Smith DR, Celniker SE, et al. (2007)
44 Improved repeat identification and masking in Dipterans. *Gene* 389: 1-9.

61. Li R, Ye J, Li S, Wang J, Han Y, et al. (2005) ReAS: Recovery of ancestral sequences for transposable elements from the unassembled reads of a whole genome shotgun. PLoS Comput Biol 1: e43.
62. Clark AG, Eisen MB, Smith DR, Bergman CM, Oliver B, et al. (2007) Evolution of genes and genomes on the *Drosophila* phylogeny. Nature 450: 203-218.
63. Vigneault F, Ter-Ovanesyan D, Alon S, Eminaga S, D CC, et al. (2012) High-throughput multiplex sequencing of miRNA. Curr Protoc Hum Genet Chapter 11: Unit 11 12 11-10.
64. Li CJ, Vagin VV, Lee SH, Xu J, Ma SM, et al. (2009) Collapse of Germline piRNAs in the Absence of Argonaute3 Reveals Somatic piRNAs in Flies. Cell 137: 509-521.
65. Li H, Durbin R (2010) Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics 26: 589-595.
66. Quinlan AR, Hall IM BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26: 841-842.

1

2 **Figure Legends**

3

4 Figure 1. Multiple transposable elements are associated with induction of hybrid
 5 dysgenesis. (A) Relative mapping abundance of single-end, 100 bp reads from
 6 strain 9 and strain 160 (normalized by reads mapping to the genome), to a
 7 consolidated repeat library. Eleven elements are in 3-fold excess in strain 160.
 8 TART elements are about 1.7-fold in excess. No apparent TEs were found in
 9 excess in strain 9. (B) Using piledriver (<https://github.com/arq5x/piledriver>) we
 10 assessed homogeneity within reads mapping to the TE library by determining the
 11 average frequency of the major variant in both strains. TEs in excess in strain
 12 160 are either more homogenous in strain 160, or similarly aged between strains.
 13 (C) Induction of sterility by 160 is broadly distributed across the genome, with the
 14 exception of chromosome 6 (the dot chromosome). Log odds ratios for
 15 probability of induction were estimated by crossing F1 males to strain 9,
 16 determining whether F2s had male gonadal atrophy and genotyping F2s to
 17 determine the chromosomes inherited from the father. Estimates were
 18 determined using a generalized linear model for logistic regression (binomial
 19 family with a logit link). Values in red are actual odds ratios. Whiskers are 95%
 20 confidence intervals. Chromosome 5 is significant at 0.1 level only. X
 21 chromosome is not scored because dysgenesis is scored in males and males do
 22 not inherit the X from their fathers (N=92).

23

1 Figure 2. Increased TE expression in the dysgenic germline persists through
 2 adulthood. (A) RPKM+0.01 for TEs (log 10), Dysgenic vs. Non-dysgenic
 3 germline. TEs that are in excess in 160 are more highly expressed, as well as
 4 many TEs that are not in excess. This state of increased global TE expression
 5 persists through adulthood. (B) Fold excess in expression (RPKM+0.01, log 2,
 6 average between both ages) vs. fold excess in abundance in strain 160. Nearly
 7 all TEs that are in excess in 160 show increased expression in the dysgenic
 8 germline (11/12). But many TEs that are equivalent in abundance between
 9 strains are also increased in expression.

10

11 Figure 3. A shifted balance of gene expression in the dysgenic germline: lowly
 12 expressed genes are increased in expression in dysgenesis, highly expressed
 13 genes are decreased. Relative expression of endogenous genes (RPKM+0.01,
 14 Dysgenic vs. Non-dysgenic) ranked from lowly to highly expressed genes for one
 15 mRNA-seq library (12-16 days old females, Library 2). A sliding window for the
 16 average ratio of sets of 250 genes is shown, showing an excess of lowly
 17 expressed genes that have higher relative expression in the dysgenic germline.
 18 Arrow indicates the 5 RPKM boundary of expression. Concordant with this, the
 19 same size sliding window (right panel) shows very highly expressed genes have
 20 reduced expression in the dysgenic germline.

21

22 Figure 4. Despite persisting TE overexpression, piRNA abundances are largely
 23 equilibrated in the dysgenic germline. (A). Normalized (per 10 million mappers)

1 piRNA abundance +1 (log 10) in the strain 9 germline vs. the strain 160
2 germline. A large number of TEs show increased piRNA expression in the strain
3 160 germline, especially TART and others enriched in abundance in strain 160.
4 (B) Normalized (per 10 million mappers) piRNA abundance +1 (log 10) in the
5 dysgenic germline vs. the non-dysgenic germline. piRNAs abundances for many
6 TEs with greater excess in strain 160 become equilibrated in the dysgenic
7 germline. A significant exception to this is the *Helena* element. (C) TE piRNA
8 excess in strain 160 vs. relative expression level in dysgenesis. TE piRNA
9 asymmetry between 160 and 9 is not the sole determinant of increased
10 expression in dysgenesis. Some TEs, such as those denoted with (*), are
11 increased in expression in dysgenesis, despite similar piRNA abundances in 9
12 and 160. (D) TE piRNA excess in the non-dysgenic germline vs. relative
13 expression level in dysgenesis. Many TEs with increased expression in
14 dysgenesis are equilibrated with respect to piRNA abundance, though this is not
15 the rule. *Helena* stands out as a strong exception.

16

17 Figure 5. Signatures of piRNA biogenesis in the dysgenic germline show only
18 modest defects. (A) Size distributions of small RNAs are similar between
19 dysgenic and non-dysgenic germlines. Distribution of all small RNAs (not
20 normalized) from four germline libraries (2 dysgenic, 2 non-dysgenic) filtered for
21 tRNA, rRNA and snoRNA. (B) piRNA biogenesis signature heatmaps. TEs
22 upregulated in the dysgenic germline (a difference of 5 RPKM or higher) are
23 indicated with red bars. TEs upregulated in the non-dysgenic germline (a

1 difference of 5 RPKM or higher) are indicated with purple. On the left are
 2 heatmaps for raw measures of abundance, the density of ping-pong pairs and
 3 percent ping-pong. On the right are heatmaps for the same metrics, but by row z-
 4 score. For raw measures, there are no noticeable effects of dysgenesis on
 5 piRNA biogenesis. Row z-scores in dysgenesis do show lower values for
 6 abundance measures (abundance and ping-pong pair density), but not percent
 7 ping-pong (see text).(C) Fold excess in expression in dysgenesis vs. the
 8 difference in percent ping-pong between dysgenic and non-dysgenic. Of the top
 9 eight that are most differently expresses in dysgenesis, all have lower ping-pong
 10 z-scores in dysgenesis.

11

12 Figure 6. Highly expressed genes are a *de novo* source of sense-piRNAs in the
 13 dysgenic germline. (A) Heat map of *D. melanogaster* ortholog CDS regions that
 14 are piRNA targets (above a threshold of 50 piRNAs per CDS per 10 million
 15 mapped for at least one CDS). A large class of CDS are a *de novo* source of
 16 piRNAs in dysgenesis. (B) Sense vs. Anti-sense abundance for piRNAs in the *de*
 17 *novo* class. piRNAs are highly biased toward the sense strand. (C) Distribution of
 18 expression levels (log 10 RPKM) for *de novo* targets vs. all genes in the genome.
 19 *de novo* targets are more highly expressed. (D) As a source of sense piRNAs in
 20 dysgenesis, the expression of these highly expressed *de novo* targets is
 21 reduced. Shown is the distribution of expression ratios (dysgenic:non-dysgenic)
 22 for all genes and *de novo* targets.

23

1 Figure 7. No single genomic region is necessary for maternal protection against
2 dysgenesis, but pericentric regions are the strongest. (A) Scatterplot showing
3 proportion of dysgenic testes (y axis) observed in the progeny of each F3 female
4 individual (x axis). Red dots indicate F3 females that were selected for whole
5 genome sequencing. (B) Single marker QTL analysis identified 3 putative QTLs:
6 one flanking the centromeres of the 5th and X chromosomes and one of the
7 tested euchromatic regions of the 4th chromosome. (C) Top row: Results from the
8 genotyping assay. Colored rectangles represent the presence of strain 160 SNPs
9 in individuals, ranked from top to bottom (most protective individuals on top).
10 Scatterplots: sequencing results. Each dot represents the average number of
11 base pairs that uniquely mapped to every 10kb of the 160 genome. Valleys
12 indicate regions of 9 homozygosity. Black dots above scatterplots show the
13 location of each SNP used for our genotyping assay. Grey background
14 demonstrates that no region of the genome from 160 is necessary or sufficient to
15 protect against dysgenesis. Right-most column: Number of piRNAs mapped to
16 TART sequences, per million reads, for each F3 female individual. Color intensity
17 is representative of TART piRNA abundance.

18
19

20 Figure 8. Germline and ovary genic cluster behavior across generations for *D.*
21 *virilis* orthologs of *center divider* and *oysgedart* from *D. melanogaster*. piRNA
22 mapping densities are indicated. mRNA-seq RPKM for germline (0-2 H embryo)
23 is also indicated. Allelism was determined by counting mRNA-seq reads based

1 on SNPs that distinguish strain 9 and 160. Strain 160 cluster identity is
2 maintained for *cdi* in non-dysgenic progeny in which strain 160 is the mother.
3 This is correlated with silencing of both alleles in the non-dysgenic germline. In
4 contrast, the cluster is not maintained in the dysgenic germline and both alleles
5 are expressed. Germline cluster identity for *oysgedart* (which in the germline is
6 predominantly sense) is lost in progeny. In this case, expression is even between
7 reciprocal progeny, but germline expression is lower off the 9 allele in both
8 directions of the cross. For cluster behavior in F3 backcrosses, heterozygosity or
9 homozygosity of the respective allele is indicated. Notice how cluster identity is
10 maintained for *cdi* to varying degrees in individuals homozygous for the 9 allele.
11 Cluster identity may be strongest in lines carrying the high TART piRNA
12 abundance allele at the tip of the X chromosome, suggesting cluster activation in
13 *trans*. In contrast, cluster activity is absent in all progeny for *oysgedart*.
14

Figure 1.

A

B

Figure 2.

Figure 3.

Figure 4.

A

B

C

A

B

C

D

pure strain

F1

F3

Germline

Ovary

Germline

Ovary