LETTER

doi:10.1038/nature11416

Paramutation in *Drosophila* linked to emergence of a piRNA-producing locus

Augustin de Vanssay[1]†, Anne-Laure Bougé[2]†, Antoine Boivin[1], Catherine Hermant[1], Laure Teysset[1], Valérie Delmarre[1], Christophe Antoniewski[2]† & Stéphane Ronsseray[1]

A paramutation is an epigenetic interaction between two alleles of a locus, through which one allele induces a heritable modification in the other allele without modifying the DNA sequence[1,2]. The paramutated allele itself becomes paramutagenic, that is, capable of epigenetically converting a new paramutable allele. Here we describe a case of paramutation in animals showing long-term transmission over generations. We previously characterized a homology-dependent silencing mechanism referred to as the *trans*-silencing effect (TSE), involved in *P*-transposable-element repression in the germ line[3–5]. We now show that clusters of *P*-element-derived transgenes that induce strong TSE[6,7] can convert other homologous transgene clusters incapable of TSE into strong silencers, which transmit the acquired silencing capacity through 50 generations. The paramutation occurs without any need for chromosome pairing between the paramutagenic and the paramutated loci, and is mediated by maternal inheritance of cytoplasm carrying Piwi-interacting RNAs (piRNAs) homologous to the transgenes. The repression capacity of the paramutated locus is abolished by a loss-of-function mutation of the *aubergine* gene involved in piRNA biogenesis, but not by a loss-of-function mutation of the *Dicer-2* gene involved in siRNA production. The paramutated cluster, previously producing barely detectable levels of piRNAs, is converted into a stable, strong piRNA-producing locus by the paramutation and becomes fully paramutagenic itself. Our work provides a genetic model for the emergence of piRNA loci, as well as for RNA-mediated *trans*-generational repression of transposable elements.

Paramutations have been well described in plants[1,2,8–12]. The best characterized is the *b1* paramutation in maize, which involves a small RNA silencing pathway[13–15], changes in DNA methylation levels and chromatin modifications[16], and shows full penetrance and stability across generations. Paramutation-like phenomena involving microRNAs have been described in mice[17,18]. However, long-term inheritance of a paramutation through generations has not been reported so far in animals.

In *Drosophila melanogaster*, transposition of *P* elements causes hybrid dysgenesis, a syndrome of genetic abnormalities including a high mutation rate, chromosome rearrangements and sterility[19,20]. In natural populations, telomeric *P* elements inserted in heterochromatic telomere-associated sequences (TAS) are master sites for establishing *P*-element repression in the germ line[21–23]. In laboratory lines (for example, *P-1152*), *P-lacZ* transgenes inserted in TAS mimics telomeric *P* elements by repressing germline expression of reporter transgenes inserted at distant euchromatic sites, through a homology-dependent silencing mechanism, TSE[3–5,24]. TSE is strongly sensitive to mutations affecting the piRNA pathway[5,25]. Its establishment involves both genetic and epigenetic components: a chromosomal copy of the telomeric silencer transgene must be either paternally or maternally inherited, and a cytoplasmic component containing small RNAs homologous to the transgene must be maternally inherited[4,5]. In addition to telomeric loci, we found that *T-1*, a tandem repeat cluster of *P-lacZ* transgenes inserted in the middle of chromosome arm 2R (50C), can also trigger a strong TSE[7]. *T-1* and other *P-lacZ* clusters inserted at the same locus (Supplementary Fig. 1) induce ectopic heterochromatin and show variegation of the *white* gene marker in the eye, a phenomenon termed repeat-induced gene silencing[6,26]. However, *T-1* triggers strong silencing of various TSE reporter transgenes in the germ line[7], whereas the other transgene clusters at this locus, including *BX2*, which contains the same number of transgene repeats as *T-1*, did not induce detectable TSE (Supplementary Table 1).

The epigenetic properties of *T-1* were analysed together with those of the *P-1152* telomeric silencer and the *BX2* cluster as controls. *T-1* and *P-1152* showed typical maternal transmission of TSE: strong repression occurred in the germ line of progeny when the silencer was maternally inherited (Fig. 1a), whereas weak or null repression was detected when the silencer was paternally inherited (Fig. 1b). *BX2* showed no repression capacity in these crosses. To analyse the relationship between TSE and piRNAs, we sequenced 19–29-nucleotide RNAs from ovaries of *T-1*, *P-1152* or *BX2* females (Supplementary Table 2). Abundant small RNAs matched the *T-1* sequences in the library from hemizygous females having inherited the *T-1* locus maternally (Fig. 1c), but not paternally (Fig. 1d). Among these species, the 23–28-nucleotide RNAs showed the typical 'ping-pong' signature of piRNA biogenesis[27], including a bias for a 5′ U (1U) and a strong tendency to form sense–antisense pairs with complementarity over their first ten nucleotides (Supplementary Fig. 2). In addition to piRNAs, short interfering RNAs (siRNAs) have been shown to be produced by previously characterized piRNA loci[28]. Similarly, *T-1* produced a significant fraction of 21-nucleotide RNAs (Fig. 1c) that do not show the ping-pong signature of piRNAs and probably correspond to siRNAs (Supplementary Fig. 3a). In agreement with a previous report[29], small RNAs with similar features were produced by *P-1152* in hemizygous females having inherited the *P-1152* locus maternally (Fig. 1f). Homozygous *P-1152* females produced about twice as many piRNAs as these hemizygous females (Supplementary Fig. 4). Finally, only a very low level of small RNAs was produced that matched *BX2* in hemizygous females from the *BX2* line (Fig. 1e). Hence, maternal inheritance of *T-1*, as well as *P-1152*, is associated with both the production of piRNAs derived from these loci and the capacity of these loci to mediate TSE, thereby linking silencing and piRNAs in this system. We next tested epigenetic interactions between the *P-1152* telomeric silencer and *T-1*, and found that chromosomal and maternally transmitted components of *T-1* and *P-1152* can complement each other to induce TSE (Supplementary Fig. 5), consistent with the presence of piRNAs matching *P-lacZ* sequences in ovaries of both *T-1* and *P-1152* females.

To investigate possible transfer of epigenetic information between *T-1* and the inactive *BX2* locus, we crossed hemizygous *T-1* females

[1]Laboratoire Biologie du Développement, UMR7622, CNRS-Université Pierre et Marie Curie, 9 quai Saint Bernard, 75005 Paris, France. [2]Drosophila Genetics and Epigenetics, CNRS URA2578 - Institut Pasteur, 25 rue du Dr Roux, 75015 Paris, France. †Present addresses: Institut Jacques Monod, CNRS, UMR 7592, Université Diderot, Sorbonne Paris Cité, F-75205 Paris, France (A.d.V.); Drosophila Normal and Pathological Neurobiology, INSERM U661 - Institut de Génomique Fonctionnelle, 141 rue de la Cardonille, 34094 Montpellier, France (A.-L.B.); Drosophila Genetics and Epigenetics, Laboratoire Biologie du Développement, UMR7622, CNRS-Université Pierre et Marie Curie, 9 quai Saint Bernard, 75005 Paris, France (C.A.).

©2012 Macmillan Publishers Limited. All rights reserved

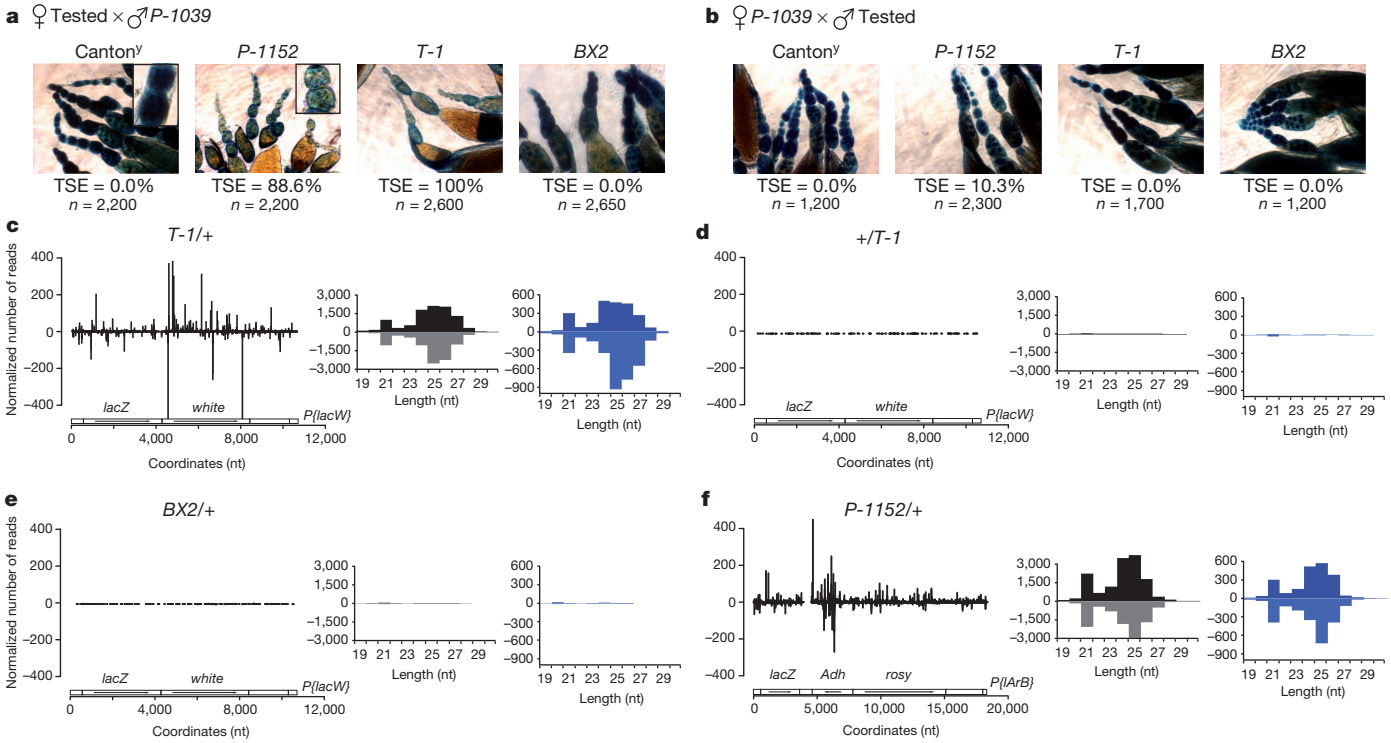

Figure 1 | Maternal inheritance of *P-1152* and *T-1* repression capacities correlates with the presence of *T-1*- or *P-1152*-derived piRNAs in ovaries of female progeny. **a**, **b**, Maternal (**a**) and paternal (**b**) inheritance of TSE mediated by *P-1152*, *T-1* or *BX2* was tested using the *P-1039* TSE reporter transgene. *lacZ* staining of ovaries of G_1 females from the indicated crosses was performed, and TSE was expressed as the percentage of repressed egg chambers among the total number (*n*) of egg chambers analysed. Female and male Canton^y flies are devoid of any transgene and were used as controls. Note that *lacZ* staining of follicle cells surrounding egg chambers (shown at higher magnification in insets) is observed in all ovaries because TSE only occurs in the germ line[4]. Original magnification, ×20. **c**–**f**, Deep sequencing of small RNAs from ovaries of the indicated genotypes in which the maternally inherited allele is always indicated first. Plots show the abundance of 19–30-nucleotide (nt) small RNAs matching *P{lacW}* (**c**–**e**) or *P{lArB}* (**f**). Histograms show the length distributions of small RNAs matching *P{lacW}* or *P{lArB}* (dark bars), or only the *lacZ* sequence in these elements (blue bars). Positive and negative values correspond to sense and antisense reads, respectively.

with hemizygous *BX2* males, and recovered female progeny that had not inherited the *T-1* locus and carrying a paternally inherited *BX2* locus (Fig. 2). These females showed marked silencing of the TSE reporter transgene, indicating that the cytoplasm of *T-1* oocytes can confer new silencing capacities to the inactive allele of the *BX2* locus. This *de novo* silencing allele will be hereafter referred to as *BX2** to differentiate it from the initial *BX2* allele never having been exposed to a *T-1* cytoplasm.

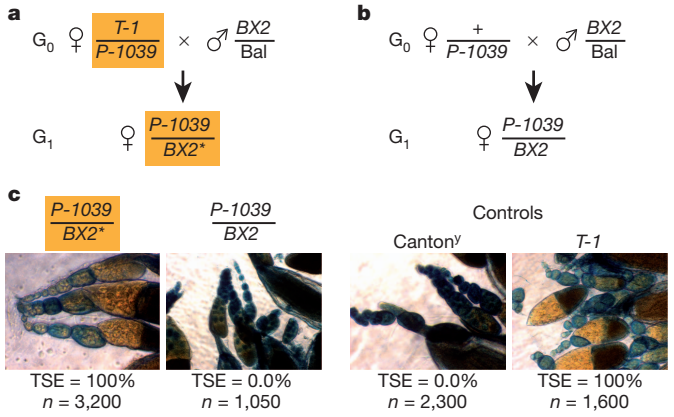

Figure 2 | Epigenetic induction of *BX2* by *T-1*. **a**, *T-1* females carrying the TSE reporter transgene *P-1039* were crossed to *BX2* males carrying a balancer chromosome (Bal). *BX2** female progeny having inherited cytoplasm from *T-1* mothers (orange background) and a *BX2* chromosome from fathers were stained for *lacZ*. **b**, Females carrying only the TSE reporter *P-1039* were crossed to *BX2* males. Female progeny from this cross were stained for *lacZ*. **c**, *P-1039/BX2** female progeny from the cross in **a** showed complete TSE, which was scored as indicated in Fig. 1. *P-1039/BX2* female progeny from the cross in **b** did not show TSE. Controls correspond to crosses between Canton^y (devoid of any transgene) or *T-1* females with *P-1039* males, which resulted in progeny showing null and complete TSE, respectively. Original magnification, ×20.

A *BX2** line was established and analysed in successive generations (Fig. 3a). Notably, second generation (G_2) *BX2** females from test crosses with males carrying a TSE reporter transgene still showed a complete TSE (Fig. 3b). This capacity to mediate TSE was fully maintained over 25 generations of the *BX2** line (TSE = 100%, $n = 4{,}600$). TSE remained very strong between G_{32} and G_{55} (99.4%, $n = 22{,}700$) showing a reversion rate less than 0.5% per generation at 25 °C (Supplementary Discussion). We conclude that maternally inherited factors from the *T-1* strain stably paramutated the *BX2* locus.

In contrast to *BX2* females, ovaries of G_2 *BX2** females contained abundant small RNAs matching the *BX2* sequence (Fig. 3c and Supplementary Table 2) with a profile similar to the one observed in *T-1* females (see Fig. 1c). The size distribution of these small RNAs showed a large peak corresponding to 23–28-nucleotide small RNAs with the piRNA ping-pong signature (Supplementary Fig. 2), as well as a discrete peak corresponding to a 21-nucleotide siRNA-like species of RNAs. Therefore, the acquired capacity of the *BX2** allele to mediate TSE correlates with the *de novo* production of *lacZ*-derived small RNAs from this locus. Finally, *BX2**-derived small RNAs were continuously produced in ovaries over at least 42 generations of a *BX2** line (Fig. 3d and Supplementary Figs 2 and 3). Together, these data indicate that the *BX2** paramutation is associated with stable production of high levels of small RNAs from the *BX2* locus in ovaries.

We next tested whether the paramutated *BX2** allele is paramutagenic. We crossed hemizygous *BX2** females with hemizygous naive *BX2* males and recovered female progeny having inherited the

©2012 Macmillan Publishers Limited. All rights reserved

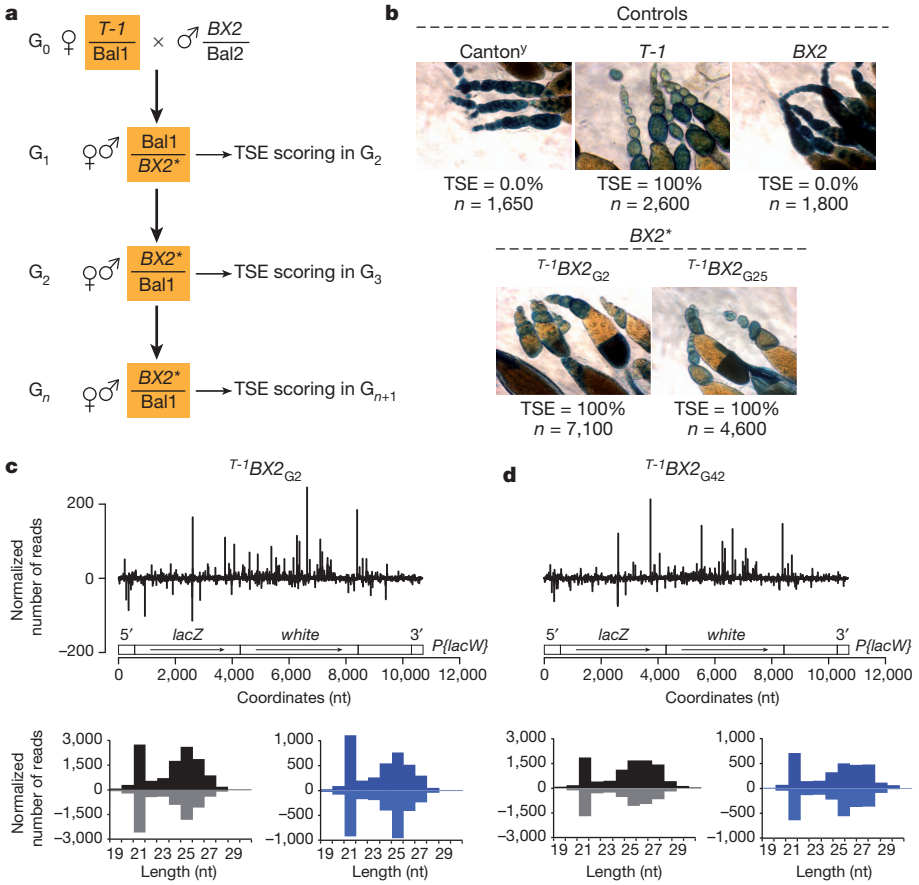

Figure 3 | *BX2 paramutation occurs and is associated to the production of small RNAs by the *BX2* cluster.** **a**, *BX2** lines were established as indicated. Bal1 and Bal2 are balancer chromosomes carrying distinct phenotypic markers. *BX2** siblings were crossed at each generation to perpetuate the *BX2** line. In addition, *BX2** females were crossed at various generations (G_n) to males carrying the *P-1039* reporter, to score the TSE of *BX2** in the G_{n+1} female progeny. **b**, TSE in *BX2** females from generations G_2 and G_{25}, and in progeny of crosses from Canton^y, *T-1* and *BX2* females with *P-1039* males as controls. $^{T\text{-}1}BX2_{G25}$ indicates that *BX2* females inherited cytoplasm from *T-1* females 25 generations before the present cross. TSE was scored as indicated in Fig. 1. Original magnification, ×20. **c**, **d**, Abundance (top) and length distribution (dark histograms) of 19–30-nucleotide small RNAs matching the *P{lacW}* transgene in ovaries from hemizygous *BX2** females from generation G_2 (**c**) and G_{42} (**d**). Length distributions of the subsets of small RNAs only matching *lacZ* are shown as blue histograms. Positive and negative values correspond to sense and antisense reads, respectively.

cytoplasm of *BX2** mothers and the *BX2* locus from fathers (Fig. 4a). This *BX2* allele was then assessed in generation G_2 for its capacity to silence a TSE reporter transgene in the germline. Notably, we observed a complete TSE (Fig. 4a), indicating that the paternally inherited *BX2* allele was paramutated through maternal inheritance of *BX2** cytoplasm. This newly paramutated *BX2* allele, which corresponds to a second-order paramutation, will be hereafter referred to as $BX2^{*2}$. A $BX2^{*2}$ line was established and showed stable TSE over 36 generations (Fig. 4a). Moreover, this line retained the capacity to produce large amounts of $BX2^{*2}$-derived small RNAs after 36 generations (Fig. 4b). Following an identical mating scheme, $BX2^{*2}$ females were able to paramutate a paternally inherited *BX2* locus, generating a third-order $BX2^{*3}$ paramutated allele that showed full TSE capacity over 10 generations. Applying this procedure recurrently, we generated a fifth-order paramutated $BX2^{*5}$ allele that showed full TSE capacity (Supplementary Fig. 6). In conclusion, the conversion of *BX2* to *BX2** by *T-1* maternal cytoplasm has all the properties of a paramutation, because it is stable over generations and the paramutated allele shows secondary paramutagenicity.

Interestingly, *T-1* also fully paramutated *C2*, another seven-copy transgene inserted at the same location (Supplementary Fig. 1), whereas lower-copy-number transgenes at this location were paramutated only transiently (Supplementary Table 3). A similar unstable paramutation interaction was also observed between the non-allelic *P-1152* and *BX2* loci (Supplementary Fig. 7).

As paramutation in this system is correlated with the production of *BX2**-derived piRNAs and siRNAs, we investigated the effect of *aubergine* and *Dicer-2* loss of function on a paramutated *BX2* cluster.

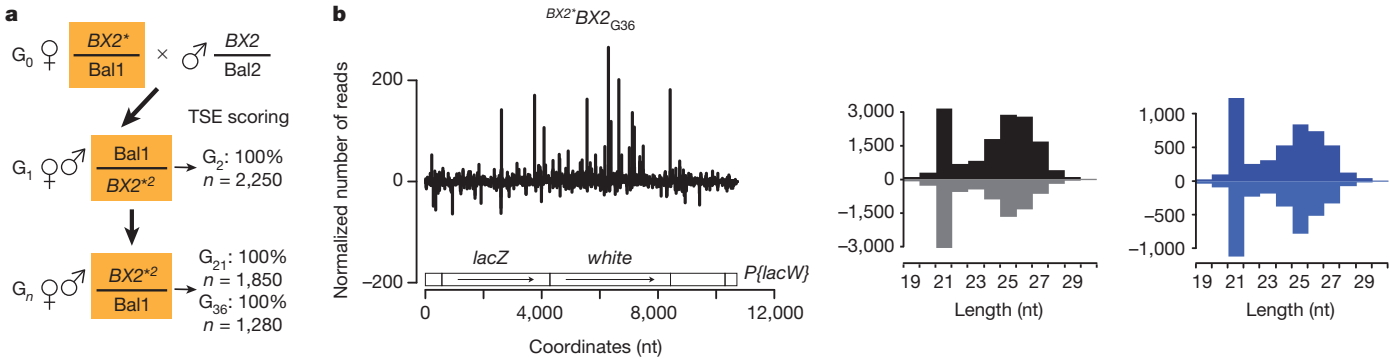

Figure 4 | Paramutated *BX2 is paramutagenic.** **a**, *BX2** females were crossed with *BX2* males and a $BX2^{*2}$ line (second-order paramutation) was established as indicated. Bal1 and Bal2 are balancer chromosomes. $BX2^{*2}$ siblings were crossed at various generations to perpetuate the $BX2^{*2}$ line. In addition, $BX2^{*2}$ females were crossed at each generation (G_n) with males carrying the *P-1039* reporter transgene to score the TSE of $BX2^{*2}$ in the G_{n+1} female progeny. **b**, Abundance (graph on the left) and length distribution (black histogram in the middle) of 19–30-nucleotide small RNAs matching the *P{lacW}* transgene in ovaries from hemizygous $BX2^{*2}$ females from generation G_{36}. Length distribution of the subset of small RNAs only matching *lacZ* is shown in the blue histogram on the right.

©2012 Macmillan Publishers Limited. All rights reserved

The silencing capacity of the *BX2*²* cluster was completely abolished in homozygous *aubergine* mutants, whereas strong silencing still took place in *Dicer-2* homozygous mutants (Supplementary Fig. 8). Moreover, the *BX2*²* locus still showed full repression capacity after four generations in a *Dicer*-2 homozygous mutant context. Hence, the *BX2** silencing activity requires piRNAs, whereas neither *BX2** activity nor inheritance rely on siRNAs. In maize, paramutation can be induced by a non-allelic transgene producing *b1*-repeat double-stranded RNA (dsRNA) and siRNAs[15] and epigenetic inheritance of the Kit^{tm1Alf} mutant allele in mice seems to result from paternal as well as maternal transmission of small RNAs[17]. These data indicate that paramutations may in some instances involve small RNAs without interactions between alleles at the DNA or chromatin levels. Our findings that, in *Drosophila*, the *BX2* paramutation is triggered by cytoplasmic inheritance strongly support this view.

Finally, we investigated the effect of the paramutation on transcription of the *BX2* locus by quantitative polymerase chain reaction with reverse transcription (RT–qPCR). *BX2* and *BX2** showed similar steady-state levels of both sense and antisense transcripts (Supplementary Fig. 9). This observation suggests that paramutation, rather than increasing the pool of piRNA precursor transcripts, activates their downstream processing into piRNAs. Thus, the maternally transmitted piRNAs could trigger production of primary piRNAs and/or ping-pong amplification of secondary piRNAs in the nuage. As paramutation is accompanied by *de novo* production of high levels of piRNA, it provides an invaluable model to determine the molecular events involved in the genesis of piRNA loci.

METHODS SUMMARY

All crosses were performed at 25 °C. *lacZ* expression assays were carried out using X-gal overnight staining[30]. The *P-lacZ-white* construct (named *P{lacW}*) contains the *P-lacZ* translational fusion and is marked by the mini-*white* gene (Supplementary Fig. 1 and Supplementary Table 4). Small RNA libraries from hand-dissected ovaries were prepared using the Illumina kit and sequenced using an Illumina Genome Analyzer II or an Illumina HiSeq-2000, following the manufacturer's instructions. For library comparisons, read counts were normalized to the total number of small RNAs that matched the *D. melanogaster* genome and did not correspond to abundant cellular RNAs (ribosomal RNA, transfer RNA and small nucleolar RNAs). Overlap signatures were computed for each sequence data set by collecting the appropriate RNA reads matching *P* transgenes and calculating overlap frequencies with RNA reads on the opposite strand.

Full Methods and any associated references are available in the online version of the paper.

Received 6 June 2011; accepted 16 July 2012.
Published online 26 August 2012.

1. Brink, R. A. A genetic change associated with the R locus in maize which is directed and potentially reversible. *Genetics* **41,** 872–889 (1956).
2. Coe, E. H. Jr. A regular and continuing conversion-type phenomenon at the B locus in maize. *Proc. Natl Acad. Sci. USA* **45,** 828–832 (1959).
3. Roche, S. E. & Rio, D. C. *Trans*-silencing by *P* elements inserted in subtelomeric heterochromatin involves the *Drosophila* Polycomb group gene, *Enhancer of zeste*. *Genetics* **149,** 1839–1855 (1998).
4. Josse, T. *et al.* Telomeric trans-silencing in *Drosophila melanogaster*: tissue specificity, development and functional interactions between non-homologous telomeres. *PLoS ONE* **3,** e3249 (2008).
5. Josse, T. *et al.* Telomeric *trans*-silencing: an epigenetic repression combining RNA silencing and heterochromatin formation. *PLoS Genet.* **3,** 1633–1643 (2007).
6. Dorer, D. R. & Henikoff, S. Transgene repeat arrays interact with distant heterochromatin and cause silencing in *cis* and *trans*. *Genetics* **147,** 1181–1190 (1997).
7. Ronsseray, S., Boivin, A. & Anxolabehere, D. *P*-element repression in *Drosophila melanogaster* by variegating clusters of *P-lacZ-white* transgenes. *Genetics* **159,** 1631–1642 (2001).
8. Chandler, V. L. Paramutation: from maize to mice. *Cell* **128,** 641–645 (2007).
9. Hollick, J. B., Patterson, G. I., Coe, E. H., Cone, K. C. & Chandler, V. L. Allelic interactions heritably alter the activity of a metastable maize *pl* allele. *Genetics* **141,** 709–719 (1995).
10. Pilu, R. *et al.* A paramutation phenomenon is involved in the genetics of maize *low phytic acid1-241* (*lpa1-241*) trait. *Heredity* **102,** 236–245 (2009).
11. Sidorenko, L. V. & Peterson, T. Transgene-induced silencing identifies sequences involved in the establishment of paramutation of the maize *p1* gene. *Plant Cell* **13,** 319–335 (2001).
12. Stam, M. Paramutation: a heritable change in gene expression by allelic interactions *in trans*. *Molecular Plant* **2,** 578–588 (2009).
13. Alleman, M. *et al.* An RNA-dependent RNA polymerase is required for paramutation in maize. *Nature* **442,** 295–298 (2006).
14. Dorweiler, J. E. *et al.* mediator of paramutation1 is required for establishment and maintenance of paramutation at multiple maize loci. *Plant Cell* **12,** 2101–2118 (2000).
15. Arteaga-Vazquez, M. *et al.* RNA-mediated *trans*-communication can establish paramutation at the *b1* locus in maize. *Proc. Natl Acad. Sci. USA* **107,** 12986–12991 (2010).
16. Stam, M., Belele, C., Dorweiler, J. E. & Chandler, V. L. Differential chromatin structure within a tandem array 100 kb upstream of the maize *b1* locus is associated with paramutation. *Genes Dev.* **16,** 1906–1918 (2002).
17. Rassoulzadegan, M. *et al.* RNA-mediated non-mendelian inheritance of an epigenetic change in the mouse. *Nature* **441,** 469–474 (2006).
18. Grandjean, V. *et al.* The *miR-124-Sox9* paramutation: RNA-mediated epigenetic control of embryonic and adult growth. *Development* **136,** 3647–3655 (2009).
19. Kidwell, M. G., Kidwell, J. F. & Sved, J. A. Hybrid dysgenesis in *Drosophila melanogaster*: a syndrome of aberrant traits including mutation, sterility, and male recombination. *Genetics* **86,** 813–833 (1977).
20. Engels, W. R. in P *Elements in* Drosophila (eds Berg, D. E. & Howe, M. M.) (American Society for Microbiology, 1989).
21. Ronsseray, S., Lehmann, M., Nouaud, D. & Anxolabehere, D. The regulatory properties of autonomous subtelomeric *P* elements are sensitive to a suppressor of variegation in *Drosophila melanogaster*. *Genetics* **143,** 1663–1674 (1996).
22. Marin, L. *et al.* *P*-element repression in *Drosophila melanogaster* by a naturally occurring defective telomeric P copy. *Genetics* **155,** 1841–1854 (2000).
23. Stuart, J. R. *et al.* Telomeric *P* elements associated with cytotype regulation of the P transposon family in *Drosophila melanogaster*. *Genetics* **162,** 1641–1654 (2002).
24. Poyhonen, M. *et al.* Homology-dependent silencing by an exogenous sequence in the *Drosophila* germline. *G3 (Bethesda)* **2,** 331–338 (2012).
25. Todeschini, A. L., Teysset, L., Delmarre, V. & Ronsseray, S. The epigenetic *trans*-silencing effect in *Drosophila* involves maternally-transmitted small RNAs whose production depends on the piRNA pathway and HP1. *PLoS ONE* **5,** e11032 (2010).
26. Dorer, D. R. & Henikoff, S. Expansions of transgene repeats cause heterochromatin formation and gene silencing in *Drosophila*. *Cell* **77,** 993–1002 (1994).
27. Brennecke, J. *et al.* Discrete small RNA-generating loci as master regulators of transposon activity in *Drosophila*. *Cell* **128,** 1089–1103 (2007).
28. Czech, B. *et al.* An endogenous small interfering RNA pathway in *Drosophila*. *Nature* **453,** 798–802 (2008).
29. Muerdter, F. *et al.* Production of artificial piRNAs in flies and mice. *RNA* **18,** 42–52 (2011).
30. Lemaitre, B., Ronsseray, S. & Coen, D. Maternal repression of the *P* element promoter in the germline of *Drosophila melanogaster*: a model for the P cytotype. *Genetics* **135,** 149–160 (1993).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank O. Sismeiro, J.-Y. Copée, E. Mouchel-Vielh, V. Ribeiro, C. Pappatico and P. Graça for technical assistance, D. Dorer, S. Henikoff and the Bloomington Stock Center for providing stocks, and flybase.org for providing databases. We thank T. Josse for preliminary experiments. We thank J.-R. Huynh, V. Colot, N. Randsholt, A.-M. Pret, C. Carré and F. Peronnet for critical reading of the manuscript. S.R. thanks D. Anxolabéhère and M. Lehmann for previous help. This work was supported by fellowships from the Ministère de l'Enseignement Supérieur et de la Recherche to A.d.V. and C.H., from the Fondation pour la Recherche Médicale to A.d.V., from the Association Nationale de la Recherche (ANR) to A.-L.B., and by grants from the Association pour la Recherche contre le Cancer to S.R. and from the ANR (project ''Nuclear endosiRNAs'') to C.A.

Author Contributions Genetic experiments were conceived by A.d.V., A.B. and S.R., and performed by A.d.V., A.B., C.H., V.D., L.T. and S.R. L.T. conceived and performed molecular mapping of the clusters and Southern blot analysis. Deep-sequencing analysis was conceived by A.d.V., A.-L.B., S.R. and C.A., and performed by A.d.V. and A.-L.B. Bioinformatic analysis was conceived and performed by C.A. RT–qPCR was conceived and performed by A.B. S.R., A.d.V., A.B. and C.A. wrote the paper and all authors discussed the results.

Author Information Small RNA sequences have been deposited at the National Center for Biotechnology Information under accession SRP012172. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to S.R. (stephane.ronsseray@upmc.fr) or C.A. (christophe.antoniewski@upmc.fr).

©2012 Macmillan Publishers Limited. All rights reserved

METHODS

Experimental conditions. All crosses were performed at 25 °C and involved 3–5 couples in most cases. *lacZ* expression assays were carried out using X-gal overnight staining as described previously[30], except that ovaries were fixed for 6 min.

Transgenes and strains. *P-lacZ* fusion enhancer trap transgenes *P-1152*, *BQ16*, *BC69* and *P-1039* all contain an in-frame translational fusion of the *Escherichia coli lacZ* gene to the second exon of the *P* transposase gene and a *rosy* transformation marker[31]. The *P-1152* insertion (Supplementary Table 4) was mapped to the telomere of the X chromosome (cytological site 1A) and consists of two *P-lacZ* insertions in the same TAS unit and in the same orientation[5]. *P-1152* is homozygous, viable and fertile. *BQ16* is located at 64C in euchromatin of the third chromosome[4] (Supplementary Table 4) and is homozygous, viable and fertile. *BC69* is inserted in chromosome 2 (Supplementary Table 4) in the first exon of the *vasa* gene and results in a *vasa* loss-of-function allele; consequently, it is homozygous, female and sterile. *P-1039* is located at 60B on the second chromosome (Supplementary Table 4) and is homozygous lethal. *P-1152* shows no *lacZ* expression in the ovary, *BQ16* and *BC69* are strongly expressed in the nurse cells and in the oocyte and *P-1039* shows strong *lacZ* staining in numerous tissues including the follicle cells, the nurse cells and the oocyte.

***P-lacZ* clusters.** Lines with different numbers of *P-lacZ-white* transgenes[32] located at cytological site 50C on the second chromosome[6,26] were used (Supplementary Table 4). The transgene(s) insertion site is located near the *mRpL53* gene, in an *Ago1* intron. This site is not a piRNA-producing locus, as observed for instance in the deep-sequencing data set from *P-1152* ovaries (data not shown). The *P-lacZ-white* construct contains the *P-lacZ* translational fusion and is marked by the mini-*white* gene (*P{lacW}*, FBtp0000204). *BX2* carries seven *P-lacZ* copies including at least one defective copy inserted in direct orientations. *T-1* derives from *BX2* following X-ray treatments (Supplementary Fig. 1). *T-1* has chromosomal rearrangements including translocations between the second and the third chromosomes. After overnight staining, weak *lacZ* expression is detected in the follicle cells of *BX2* and *T-1* female ovaries, presumably because of a position effect at 50C, but no staining is observed in the germ line (data not shown).

Lines carrying transgenes have M genetic backgrounds (devoid of *P* transposable elements), as do the multi-marked balancer stocks used in genetic experiments. The Canton^{y} and w^{1118} lines were used as controls completely devoid of any *P* element or transgene. Crosses involving *P-1152* were performed with females carrying the telomeric transgenes in the homozygous state (except where indicated), whereas crosses performed with *BX2* or *T-1* were performed with females carrying the cluster in the heterozygous state (referred to as hemizygous in case of insertions) because of the sterility (*BX2*) and lethality (*T-1*) induced by transgene clusters.

Two strong hypomorphic mutant alleles of *aubergine* induced by EMS were used. Both of them are homozygous, female and sterile, and TSE was previously shown to be abolished by a heteroallelic combination of these alleles[5]. aub^{QC42} comes from the Bloomington Stock Center (stock no. 4968) and has not been characterized at the molecular level[33]. aub^{N11} has a 154-bp deletion, resulting in a frameshift which is predicted to add 16 novel amino acids after residue 740 (refs 34, 35). *Dicer-2*^{L811fsX} is a loss-of-function allele induced by EMS that has a sequence variant at residue 811 resulting in a stop codon[36]. It is homozygous, viable and fertile.

Quantification of TSE. When TSE is incomplete, variegation is observed because 'on/off' *lacZ* expression is seen between egg chambers: that is, egg chambers can show strong expression (dark blue) or no expression, but intermediate expression levels are rarely found. TSE was quantified as previously described[5] by determining the percentage of egg chambers with no expression in the germ line.

Deep sequencing analyses. Small RNAs from hand-dissected ovaries were cloned using the DGE-Small RNA Sample Prep Kit and the Small RNA Sample Prep v.1.5 Conversion Kit from Illumina (libraries 1 to 5), following the manufacturer's instructions, or using the TruSeq (TM) SBS v.5 Kit at Fasteris (http://www.fasteris.com/) (libraries 6 to 8). Libraries 1 to 5 were sequenced using an Illumina Genome Analyzer II and libraries 6 to 8 were sequenced using an Illumina Hi-Seq 2000. Sequence reads in fastq format were trimmed from the adaptor sequence 5′-TCGTATGCCGTCTTCTGCTTG-3′ (libraries 1 to 5) or 5′-CTGTAGG CACCATCAAT-3′ (libraries 6 to 8) and matched to the *D. melanogaster* genome release 5.43 using Bowtie[37], as well as to the sequences of the *P*-element constructs *P{lArB}* (FlyBase accession FBtp0000160) and *P{lacW}* (FlyBase accession FBtp0000204). Only 19–30-nucleotide reads matching the reference sequences with 0 or 1 mismatch were retained for subsequent analysis. For global annotation of the libraries (Supplementary Table 2), we used release 5.43 of fasta reference files available in FlyBase, including transposon sequences (dmel-all-transposon_r5.43.fasta) and release 18 of miRNA sequences from miRBase (http://www.mirbase.org).

Sequence length distributions, small RNA mapping and frequency maps were generated using in-house Python scripts and R (http://www.r-project.org/) to analyse Bowtie outputs. Scripts were integrated and run in a Galaxy instance hosted by the laboratory. The corresponding Mississippi suite of analysis workflows and codes is accessible from http://www.drosophile.org upon request. For library comparisons, read counts were normalized (Supplementary Table 2) to the total number of small RNAs that matched the *D. melanogaster* genome and did not correspond to abundant cellular RNAs (rRNA, tRNA and snoRNAs). For small RNA mapping, we matched each individual RNA sequence to *P{lArB}* or *P{lacW}* and gave to each matched position a weight corresponding to the normalized occurrence of the sequence in the small RNA library. When RNA sequences matched *P{lArB}* or *P{lacW}* repeatedly, the weight was divided by the number of hits to these *P*-element constructs.

Distributions of piRNA overlaps (ping-pong signatures) were computed by collecting, for each sequencing data set, all the 23–28-nucleotide RNA reads matching *P{lArB}* or *P{lacW}* whose 5′ ends overlapped with another 23–28-nucleotide RNA read on the opposite strand. Then, for each possible overlap of 1–28 nucleotides, the number of read pairs was counted. Distributions of siRNA overlaps were computed using a similar procedure, except that 20–22-nucleotide RNA reads were collected instead of the 23–28-nucleotide RNA reads. The distributions of piRNA/siRNAs overlaps were computed by collecting separately the 20–22-nucleotide and 23–28-nucleotide RNA reads matching *P{lArB}* or *P{lacW}*, and counting for each possible overlap of 1–22 nucleotdies the number of read pairs across these two distinct read data sets. To plot the overlap signatures, a *z*-score was calculated by computing, for each overlap of 1 to i nucleotides, the number $O(i)$ of read pairs and converting it using the formula $z(i) = (O(i) - \text{mean}(O))/\text{standard deviation}(O)$.

RT–qPCR experiments. Total RNA was extracted (Qiagen kit) from ovaries dissected from *1A-6*, *BX2* and *BX2** females and quantified (NanoDrop). Four to six biological replicates were made for each genotype. For each sample, 10 μg of RNA was treated with DNase (Fermentas). 1 μg of DNase-treated RNA was used for reverse transcription (Fermentas) using either no primer (control RT) or two primers simultaneously (specific RT): one specific to the *nanos* transcript used as the sample RNA quantification reference (5′-GGATTCGCCCTCTCTAAACC-3′) and the second specific to a region of the *P{lacW}* transgene. *P{lacW}* RT primers were designed to be specific to the sense (s) or to the antisense (a) transcripts of five regions of the *P{lacW}* transgene: 5′*P*, 5′*lacZ*, 3′*lacZ*, 5′*white* and 3′*P*. Sequences are: a1 (5′-ATTCAAACCCCACGGACAT-3′), a2 (5′-AGTA CGAAATGCGTCGTTTAGAGC-3′), a3 (5′-GGGGAAAACCTTATTTATCAG CCG-3′), a4 (5′-GCTGTTTGCCTCCTTCTCTG-3′), s1 (5′-GTTTTCCCAGT CACGACGTT-3′), s2 (5′-AATGCGCTCAGGTCAAATTC-3′), s3 (5′-TATGG AAACCGTCGATATTCAGCC-3′), s4 (5′-ATTTTTGTGGGTCGCAGTTC-3′), s5 (5′-TTAAGTGTATACTTCGGTAAGCTTCG-3′), s6 (5′-TTTGGGAGTT TTCACCAAGG-3′). One primer was both antisense and sense (as) because it is located in the inverted repeat of the *P* element. It is (5′-TGATGA AATAACATAAGGTGGTCCCGTCG-3′). RT primers are shown on the transgene map (Supplementary Fig. 9). qPCR was then performed on triplicates of each RT with a primer pair specific for the *nanos* gene in order to quantify the *nanos* transcripts. Simultaneously, qPCR was performed on triplicates of the same RT using different primer pairs corresponding to the former five regions of interest of *P{lacW}*. qPCR primer sequences are: 5′ *P* (5′-CTGCAAAGCTGTGACTGGAG-3′ and 5′-TTTGGGAGTTTTCACCAAGG-3′), 5′ *lacZ* (5′-GAGAATCCGACGG GTTGTTA-3′ and 5′-AAATTCAGACGGCAAACGAC-3′), 3′ *lacZ* (5′-ACT ATCCCGACCGCCTTACT-3′ and 5′-GTGGGCCATAATTCAATTCG-3′), 5′ *white* (5′-GTCAATGTCCGCCTTCAGTT-3′ and 5′-GGAGTTTTGGCACAGC ACTT-3′) and 3′ *P* (5′-CCACGGACATGCTAAGGGTTAA-3′ and 5′-GTCGG CAAGAGACATCCACT-3′). The same series of dilutions composed of a mix of different RT preparations was used to normalize the quantity of *nanos* transcripts in all RT preparations leading to standard quantity (Sq) values for *nanos* transcripts in specific RT (using *nanos* primer = Sq(*nanos*)) or in control RT (without primer = Sq(control *nanos*)) preparations. A series of dilutions of a plasmid containing the *P{lacW}* transgene was used to normalize the quantity of transcripts of the clusters leading to Sq values for cluster transcript (Sq(specific) and Sq(control specific)). Variations between technical triplicates seem to be very low when compared to variations between biological replicates. The mean of the three technical replicates was then systematically used ($\overline{\text{Sq}}$). The measure of the quantity of transcripts from a given region for one biological sample was then calculated using the formula: ($\overline{\text{Sq}}$(specific) − $\overline{\text{Sq}}$(control specific))/($\overline{\text{Sq}}$(*nanos*) − $\overline{\text{Sq}}$ (control *nanos*)). This allowed us to eliminate the background noise due to both sense and antisense transcripts (Sq(control transcript)) and to take into account variations in the quantity of RNA between biological samples (Sq(*nanos*)).

31. O'Kane, C. J. & Gehring, W. J. Detection *in situ* of genomic regulatory elements in *Drosophila*. *Proc. Natl Acad. Sci. USA* **84,** 9123–9127 (1987).

©2012 Macmillan Publishers Limited. All rights reserved

32. Bier, E. *et al.* Searching for pattern and mutation in the *Drosophila* genome with a P-lacZ vector. *Genes Dev.* **3,** 1273–1287 (1989).
33. Schupbach, T. & Wieschaus, E. Female sterile mutations on the second chromosome of *Drosophila melanogaster*. II. Mutations blocking oogenesis or altering egg morphology. *Genetics* **129,** 1119–1136 (1991).
34. Wilson, J. E., Connell, J. E., Schlenker, J. D. & Macdonald, P. M. Novel genetic screen for genes involved in posterior body patterning in *Drosophila*. *Dev. Genet.* **19,** 199–209 (1996).
35. Harris, A. N. & Macdonald, P. M. *aubergine* encodes a *Drosophila* polar granule component required for pole cell formation and related to eIF2C. *Development* **128,** 2823–2832 (2001).
36. Lee, Y. S. *et al.* Distinct roles for *Drosophila* Dicer-1 and Dicer-2 in the siRNA/miRNA silencing pathways. *Cell* **117,** 69–81 (2004).
37. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10,** R25 (2009).

©2012 **Macmillan Publishers Limited. All rights reserved**