A satellite repeat-derived piRNA controls embryonic development of Aedes

Satellite repeats comprise a substantial portion of eukaryotic genomes, but little is known about their functions. Many satellite repeats are actively transcribed, and some produce small interfering RNAs (siRNAs) that are required for heterochromatin formation and maintenance. We analysed small RNAs in germline (ovaries) and somatic (carcasses) tissues of *Ae. aegypti* and in the somatic cell line Aag2. Although satellite repeats constitute less than 10% of the genome, there was high coverage of these regions by both siRNAs and piRNAs (Extended Data Fig. 1a). PIWI-bound piRNAs8 (Fig. 1b, Supplementary Fig. 1) are 2′-O-methylated at their 3′ end, a common feature of mature PIWI-interacting RNAs (piRNAs). Whereas satellite repeats and piRNA sequences generally evolve extremely quickly8,7, this locus was conserved for approximately 200 million years, suggesting that it has a central function in mosquito biology. piRNA production commenced shortly after egg laying, and inactivation of the more abundant piRNA resulted in failure to degrade maternally deposited transcripts in the zygote and developmental arrest. Our results reveal a mechanism by which satellite repeats regulate global gene expression in trans via piRNA-mediated gene silencing that is essential for embryonic development.

Tandem repeat elements such as the diverse class of satellite repeats occupy large parts of euakaryotic chromosomes, mostly at centromeric, pericentromeric, telomeric and subtelomeric regions. However, some elements are located in euchromatic regions throughout the genome and have been hypothesized to regulate gene expression in cis by modulating local chromatin structure, or in trans via transcripts derived from the repeats3–4. Here we show that a satellite repeat in the mosquito *Aedes aegypti* promotes sequence-specific gene silencing via the expression of two PIWI-interacting RNAs (piRNAs). Whereas satellite repeats and piRNA sequences generally evolve extremely quickly8,7, this locus was conserved for approximately 200 million years, suggesting that it has a central function in mosquito biology. piRNA production commenced shortly after egg laying, and inactivation of the more abundant piRNA resulted in failure to degrade maternally deposited transcripts in the zygote and developmental arrest. Our results reveal a mechanism by which satellite repeats regulate global gene expression in trans via piRNA-mediated gene silencing that is essential for embryonic development.
An evolutionarily conserved satellite repeat produces piRNAs that associate with Piwi4. a. Schematic of the current AAEL017385 annotation (AAeg 5.1) with its six annotated transcripts (RA–RF) and the tapiR satellite repeat locus (top), and small RNA read coverage in Aag2 cells (bottom). Filled boxes represent open reading frames. b. Northern blot of tapiR1 and tapiR2 upon β-elimination in Aag2 cells. mir-2940-3p serves as a positive control. c. tapiR1 and tapiR2 enrichment in the indicated Piwi4 immunoprecipitations (IP) (left) and fraction of total Piwi4-enriched reads (≥ twofold, right). d. Sequence logo of all repeat monomers from tapiR1 with its six annotated transcripts (RA–RF) and the tapiR satellite repeat locus (top), and small RNA read coverage in Aag2 cells (bottom). Filled boxes represent open reading frames.

Fig. 2 tapiR1 silences target RNAs in trans through seed-mediated base pairing. a. Schematic of tapiR1 reporter constructs (left) and luciferase assay with the indicated reporters in Aag2 cells (right). Firefly luciferase; Scr, scrambled RNA control; mm4, mismatch at position t4. b–d. Representative luciferase assays of the indicated reporters, with target sites at different positions in the RNA (b), target sites in the 3′ UTR with three consecutive mismatches (c) or increasing numbers of mismatches (d). e. Expression of mRNAs and long noncoding RNAs (lncRNAs) in Aag2 cells treated with tapiR1 or control antisense oligonucleotides (average from three biological replicates). A pseudocount of one was added to plot values of zero. Diagonal lines indicate a fold change of two. f. RT–qPCR of tapiR1 target genes in Aag2 cells treated with tapiR1 or control antisense oligonucleotides. Data are mean ± s.d.; individual measurements are shown as points (n = 18 (a), n = 6 (b–d), and n = 2 (f); Supplementary Table 7). AO, antisense oligonucleotide.

The biting midge Culicoides nubeculosus or Drosophila. These observations suggest that the locus evolved in the late Triassic after divergence of the Anopheleinae and Culicinae subfamilies of mosquitoes (229–192 million years ago (Ma))\(^2\), but before further divergence of the culicine genera (226–172 Ma)\(^2\). This establishes the tapiR tandem repeat locus as one of the few ancient satellite repeats that have been described\(^1\)–\(^1\). Deep evolutionary conservation suggests that the locus and its associated piRNAs have important and conserved functions that are independent of the upstream AAEL017385 gene (Extended Data Fig. 3g–i, Supplementary Note 1).

We tested the potential for target silencing by tapiR1 in trans using a series of luciferase reporters. We validated that tapiR1 and tapiR2 efficiently silence RNAs containing a single, fully complementary target site in the 3′UTR (Fig. 2a). Addition of an antisense oligonucleotide complementary to tapiR1 relieved silencing of the tapiR1 reporter, but not the tapiR2 reporter, in a concentration-dependent manner; similarly, addition of an antisense oligonucleotide complementary to tapiR2 exclusively relieved silencing of the tapiR2 reporter (Extended Data Fig. 4a). These results confirm that tapiR-mediated silencing is sequence-specific and that the antisense oligonucleotides target the mature piRNAs and do not—or only to a minor extent—target a putative precursor of tapiR1 and tapiR2. Additionally, tapiR antisense oligonucleotide treatment did not influence microRNA (miRNA) stability (Extended Data Fig. 4b). Unlike most miRNAs\(^1\), silencing was independent of the position of the target site in the mature mRNA (Fig. 2b, Extended Data Fig. 5a) but was largely abolished for target sites in an intron or an untranslated transcript expressed from a polymerase III promoter (Extended Data Fig. 5b). We noticed that firefly and Renilla
Fig. 3 | tapiR1 promotes turnover of maternal transcripts and is essential for embryonic development. a, Northern blot of tapiR1 and tapiR2 in Ae. aegypti embryos. Ethidium-bromide-stained rRNA serves as loading control. b–e, Outline of the experimental procedure (b), c, Percentage of embryos that reached the indicated developmental stages (Supplementary Fig. 2e) 2.5 days after injection of antisense oligonucleotides ($\chi^2$ test of independence; $\chi^2$ (2, $N = 521$) = 105.05, $P < 2.2 \times 10^{-16}$). d, Percentage of hatched embryos four days after injection of antisense oligonucleotides. In box plots, the centre line represents the mean, box edges show first and third quartiles, and whiskers show maximum and minimum values ($n = 6$ (c), $n = 5$ (d); Supplementary Table 7). e, Gene expression in embryos injected with tapiR1 or control antisense oligonucleotides 20.5 h after injection (mean counts from five biological replicates plus a pseudocount of one). Diagonal lines indicate a twofold change. f, Cumulative distribution of log$_2$-transformed fold change in expression of genes grouped by the minimum free energy (mfe) of predicted piRNA-target RNA duplexes. g, Expression of tapiR1 target genes in embryos. RT–qPCR was performed on samples shown in a. The Abd-A gene is not targeted by tapiR1 and serves as negative control. Data are mean ± s.d. ($n = 2$; see Supplementary Table 7). h, Fraction of genes grouped by their expression pattern in early embryos.

Luciferase genes, which we used as reporter and normalization control, respectively, contain potential tapiR1 target sites. *Renilla* luciferase is indeed potently suppressed by tapiR1 and firefly luciferase is slightly suppressed by tapiR1 (Extended Data Fig. 5c–e), but mutating these target sites did not affect our conclusions (Extended Data Fig. 5f, g).

To assess sequence requirements for tapiR1 targeting, we introduced mismatches in the piRNA target site. Three consecutive mismatches were tolerated unless they were located in the t1 to t9 region of the target (the nucleotides expected to form base pairs with piRNA positions 1 to 9) (Fig. 2c, Extended Data Fig. 6a), and single mismatches only impaired silencing at positions t3 to t7 (Extended Data Fig. 6a, b), reminiscent of a microRNA seed and comparable to the piRNA seed in *Caenorhabditis elegans*. 

Base pairing at the potential Argonaute cleavage site between nucleotides t10 and t11 was dispensable for silencing, as was a mismatch at t1, suggesting that the first nucleotide of tapiR1 is anchored in a binding pocket of Piwi4, similar to other Argonaute proteins. Unexpectedly, a mismatch at position t2 was also tolerated, but only when the remainder of the target site was perfectly complementary (Extended Data Fig. 6c). Additionally, in contrast to *C. elegans*, G–U wobble pairs were not allowed inside the seed and had the same effect as a mismatch at the same position (Extended Data Fig. 6d). Increasing numbers of mismatches at the 3′ end did not interfere with silencing when at least half of the piRNA could base pair with the target site (Fig. 2d), indicating that the 3′ part of the piRNA is not absolutely required, and that the seed region alone is not sufficient for silencing, a pattern that is more similar to target requirements of miRNAs than those of piRNAs studied to date.

Some satellite repeats can influence gene expression by modulation of the local chromatin environment in *cis*, or via siRNAs targeting homologous repeat insertions. By contrast, given the conservation of the tapiR repeat locus, the efficiency of target-reporter silencing and the targeting requirements of tapiR1, we hypothesized that the tapiR locus has the potential to silence a wide range of remote genes in *trans*, and thus to regulate diverse cellular processes. To test this idea, we predicted tapiR1 target sites and verified that about half were sufficient to mediate silencing in luciferase reporter assays (Extended Data Fig. 7). We then blocked tapiR1-mediated silencing with tapiR1 antisense oligonucleotides in Aag2 cells and assessed global gene expression by RNA-sequencing analysis (RNA-seq). Expression of 134 genes was strongly and significantly increased (Fig. 2e, Supplementary Table 1), and some transposons, though not globally affected, were also upregulated (Extended Data Fig. 8a, Supplementary Table 2). We validated these results by quantitative PCR with reverse transcription.
propose that, analogous to abundant miRNAs in other animals\textsuperscript{22}, Culicinacae mosquitoes have evolved a specific zygotic piRNA to destabilize a defined set of maternally deposited transcripts during early embryonic development, the mechanism of which remains unknown (Extended Data Fig. 10, Supplementary Data 2f, g, Supplementary Note 2). To our knowledge, this is the first demonstration of sequence-specific gene silencing by transcriptional products from a satellite repeat in trans, underlining the regulatory potential of tandem repeat DNA.

**Online content**

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-020-2159-2.

1. Garrido-Ramos, M. A. Satellite DNA: an evolving topic. Genes 8, 230 (2017).
2. Feliciello, I., Akrap, I. & Ugarovik, D. Satellite DNA modulates gene expression in the beetle Tribolium castaneum after heat stress. PLoS Genet. 11, e1005466 (2015).
3. Li, X. Y. & Kirby, M. L. Coordinated and conserved expression of aliphoid repeat and aliphoid repeat-tagged coding sequences. Dev. Dyn. 228, 72–81 (2003).
4. Pezer, Z. & Ugarovik, D. Satellite DNA-associated siRNAs as mediators of heat shock response in insects. RNA Biol. 9, 587–595 (2012).
5. Girard, A., Sachidanandam, R., Hamon, G. J. & Carmell, M. A. A germline-specific class of small RNAs binds mammalian Piwi proteins. Nature 442, 199–202 (2006).
6. Lau, N. C. et al. Characterization of the piRNA complex from rat testes. Science 313, 363–366 (2006).
7. Metters, D. P. et al. Comparative analysis of tandem repeats from hundreds of species reveals unique insights into centromere evolution. Genome Biol. 14, R10 (2013).
8. Czech, B. & Hamon, G. J. One loop to rule them all: the ping-pong cycle and piRNA-guided silencing. Trends Biochem. Sci. 41, 324–337 (2016).
9. Miesen, P., Joostert, J. & van Ri, R. P. Piwi’s go viral. arbovirus-derived piRNAs in vector mosquitoes. PLoS Pathog. 12, e1006017 (2016).
10. Reidenbach, K. R. et al. Phylogenetic analysis and temporal diversification of mosquitoid (Diptera: Culicidae) based on nuclear genes and morphology. BMC Evol. Biol. 9, 298 (2009).
11. Plohl, M. et al. Long-term conservation vs high sequence divergence: the case of an extraordinarily old satellite DNA in bivalve mollusks. Hereditas 104, 543–551 (2010).
12. Martinez-Lage, A., Rodriguez-Fariña, F., González-Tozon, A. & Méndez, J. Origin and evolution of Mytilus mussel satellite DNAs. Genome 48, 247–256 (2005).
13. Chaves, R., Ferreira, D., Mendes-da-Silva, A., Meles, S. & Adegà, F.-S. FA-SAT is an old satellite DNA frozen in several Bilateria genomes. Genome Biol. Evol. 9, 3073–3087 (2017).
14. Bartel, D. P. MicroRNAs: target recognition and regulatory functions. Cell 136, 253–253 (2009).
15. Zhang, D. et al. The piRNA targeting rules and the resistance to piRNA silencing in endogenous genes. Science 359, 587–592 (2018).
16. Shen, E. Z. et al. Identification of piRNA binding sites reveals the Argonauta regulatory landscape of the C. elegans genome. Genome Biol. 172, 1032–1044 (2015).
17. Matsumoto, N. et al. Crystal structure of silkworm PIWI-clade Argonaute Siwi bound to piRNA. J. Genom. 167, 484–497 (2016).
18. Mohn, F., Handler, D. & Brennecke, J. Noncoding RNA: piRNA-guided slicing specifies transcripts for Zucchini-dependent, phased piRNA biogenesis. Science 348, 812–817 (2015).
19. Reuter, M. et al. Miwi catalysis is required for piRNA amplification-independent LINE1 transposon silencing. Nature 480, 264–267 (2011).
20. Goh, W. S. et al. piRNA-directed cleavage of meiotic transcripts regulates spermatogenesis. Genes Dev. 29, 1032–1044 (2015).
21. Grimson, A. et al. MicroRNA targeting specificity in mammals: determinants beyond seed pairing. Mol. Cell 27, 91–105 (2007).
22. Ugarovik, D. Functional elements residing within satellite DNAs. EMBO Rep. 6, 1035–1039 (2005).
23. Vastenhouwen, N. L., Cao, W. X. & Lipshitz, H. D. The maternal-to-zygotic transition revisited. Development 146, dev161747 (2019).
24. Rouget, C. et al. Maternal piRNA denaturation and decay by the piRNA pathway in the early Drosophila embryo. Nature 467, 1128–1132 (2010).
25. Barckmann, B. & Auberge, I.CLP reveals piRNA-dependent decay of miRNAs involved in germ cell development in the early embryo. Cell Rep. 12, 1205–1216 (2015).

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature Limited 2020
Methods

No statistical methods were used to predetermine sample size. The investigators were blinded to allocation during embryo scoring in Fig. 3c. For other experiments, the investigators were not blinded to allocation during experiments and outcome assessment.

Cell culture

Ae. aegypti Aag2 cells were cultured in Leibovitz’s L-15 medium (Invitrogen) supplemented with 10% heat-inactivated fetal bovine serum (PAA Laboratories), 2% tryptose phosphate broth solution (Sigma Aldrich), 1× MEM non-essential amino acids (Invitrogen) and 50 U ml⁻¹ penicillin–streptomycin (Invitrogen) at 25°C. Aag2 cells are a widely used non-clonal cell line of probable embryonic origin26 that expresses all somatic PIWI proteins and produces both primary and secondary piRNAs by ping-pong dependent amplification27,28. For all experiments, cells were seeded the day before and used at 70–80% confluency. Cells were regularly confirmed to be negative for mycoplasma contamination. The cell line was a gift from R. Andino.

Mosquito rearing

Injections of embryos for RNA sequencing and northern blotting presented in Fig. 3a, c–e, g were performed using a cell fusion agent virus-free, isofemale Ae. aegypti strain called Jane. This strain was derived from a field population originally sampled in the Muang District of Kamphaeng Phet Province, Thailand29, and reared for 26 generations at 28 ± 1°C, 75 ± 5% relative humidity, 12:12 h light:dark cycle. Embryos were hatched under low pressure for 30–60 min. Larvae were grown in dechlorinated tap water and fed fish food powder (Tetramin) every two days. Adults were maintained in cages with constant access to a 10% sucrose solution. Female mosquitoes were fed on commercial rabbit blood (BCL) through a membrane feeding system (Hemotek) using pig intestine as membrane. For antisense oligonucleotide injections, female mosquitoes were transferred to 25°C and 70% humidity for at least two days before being allowed to lay eggs, and embryos were then placed back to 28°C immediately after the injection. For the time course experiment in Fig. 3a, g, embryos were kept at 25°C during the course of the experiment. All other injection experiments were conducted with Ae. aegypti Liverpool strain, reared at 28 ± 1°C, 70% humidity as described above, but fed on human blood (Sanquin Blood Supply Foundation), and maintained at 28°C throughout experiments.

Experiments presented in Extended Data Figs. 2a, 10a, b were performed with the Ae. aegypti Rockefeller strain, obtained from Bayer. The mosquitoes were maintained at 27 ± 1°C with a 12:12 h light:dark cycle and 70% relative humidity, as described30. For the blood-feeding experiment, mosquitoes were offered human blood (Sanquin Blood Supply Foundation) and five engorged females were selected and killed at each of the indicated time points.

Insects used for tapiR northern blot analyses in Fig. 1f were either laboratory-reared or wild-caught species. Ae. aegypti (Liverpool strain), Culex pipiens p. pipiens, Culex pipiens molestus, Toxorhynchites brevipalpis, Anopheles coluzzii, Anopheles quadriannulatus and Anopheles stephensi mosquitoes, Culicoides nubeculosus biting midges and D. melanogaster fruit flies (genotype w¹¹¹⁸) were laboratory reared, which were stored at −80°C until use. The immunoprecipitation was performed with custom-made antibodies against Piwi4–6 and Ago3 (1:10 dilution) at 4°C for 4 h on

Gene knockdown

Double-stranded RNA was generated by in vitro transcription of T7 promoter-flanked PCR products with T7 RNA polymerase. Primer sequences are given in Supplementary Table 5. The reaction was carried out at 37°C for 3–4 h, then heated to 80°C for 10 min and gradually cooled down to room temperature to facilitate dsRNA formation. The dsRNA was purified with the GeneElute Total RNA Miniprep Kit (Sigma Aldrich).

Aag2 cells were seeded in 24-well plates the day before the experiment and transfected with X-tremeGENE HP transfection reagent (Roche) according to the manufacturer’s instructions, using a ratio of 4 μl reagent per μg of dsRNA. The transfection medium was replaced after 3 h with fully supplemented Leibovitz-15 medium. For Extended Data Figs. 2d, e, f. Sc, cells were collected after 48 h; in other experiments, the knockdown was repeated 48 h after the first transfection and cells were then collected after 24 h. Knockdown was confirmed by RT-qPCR. For the RNA interference (RNAi) screen shown in Extended Data Fig. 10f, a firefly luciferase reporter containing the tapiR1 target site of AAELOJ5SS5 in the 3’UTR, and a Renilla luciferase reporter were co-transfected with the dsRNA during the second transfection.

For injection of embryos with dsRNAs, engorged female Ae. aegypti Liverpool mosquitoes were allowed to lay eggs for 45 min. Embryos were desiccated for 1.5 min, covered with Halocarbon oil (Sigma Aldrich) and injected with 500 ng μl⁻¹ dsRNA with the Pneumatic PicoPump PV820 (World Precision Instruments) with 30 psi injection pressure. Injected embryos were then transferred to a wet Whatman paper and maintained at 28°C and 80% humidity for 21 h. In each experiment, 30–60 embryos were injected per well.

RNA isolation

RNA from cells and mosquitoes was isolated with Isol-RNA lysis buffer (SPRIIM) according to the manufacturer’s instructions. In brief, 200 μl chloroform was added to 1 ml lysis buffer, and centrifuged at 16,060g for 20 min at 4°C. Isopropanol was added to the aqueous phase, followed by incubation on ice for at least one hour, and centrifugation at 16,060g for 20 min at 4°C. The pellet was washed three to five times with 85% ethanol and dissolved in RNase-free water. RNA was quantified on a Nanodrop photospectrometer.

Periodate treatment and β-elimination

Total RNA was treated with 25 mM NaIO₄ in a final concentration of 60 mM borax and 60 mM boric acid (pH 8.6) for 30 min at room temperature. In the control, NaIO₄ was replaced by an equal volume of water. The reaction was quenched with glycerol and β-elimination was induced with a final concentration of 40 mM NaCl for 90 min at 45°C. RNA was ethanol precipitated and analysed by northern blot.

Generation of antibodies

The custom-made antibodies (Eurogentec) against endogenous PIWI proteins have been described previously31,32. In brief, rabbits were immunized with a mixture of two unique peptides (Ago3: (C+)TSGADSSESDDKQSS, (C+)JIYFKKQRMSHEINQF; Piwi4: (C+)HEGGR5PS5RPPAYSS, (C+)HHRESSAGGRERSGN; Piwi5: (C+)DVRISPLSDKYVKQ, CANQGGNWRD-KYRRAI; Piwi6: MANDNPQEGSSSGGRICR+C, (C+)RGRDHQKPYDPRFQEQ). Sera were collected and purified against each peptide separately (antibodies purified against the underlined peptides were used). Specificities of the Piwi5 and Ago3 antibodies were validated previously31,32. Specificities of the Piwi4 and Piwi6 antibodies were confirmed by western blotting of Aag2 cells stably expressing protein A–TEV cleavage site–His, (PTH)-tagged Piwi upon knockdown of the respective PIWI gene or a control knockdown (dsRLuc) (Supplementary Fig. 2a).

Immunoprecipitation and western blotting

Aag2 cells were lysed with RIPA buffer (10 mM Tris-HCl, 150 mM NaCl, 0.5 mM EDTA, 0.1% SDS, 1% Triton-X-100, 10% DOC, 1× protease inhibitor cocktail), supplemented with 1% glycerol and stored at −80°C until use. The immunoprecipitation was performed with custom-made antibodies against Piwi4–6 and Ago3 (1:10 dilution) at 4°C for 4 h on
rotation. Protein A/G Plus beads (Santa Cruz) were added at a dilution of 1:10 and then incubated overnight at 4 °C on rotation. Beads were washed three times with RIPA buffer, and half was used for RNA isolation and protein analysis each. For RNA extraction, beads were treated with proteinase K for 2 h at 55 °C and isolated with phenol–chloroform extraction. Equal amounts of RNA for input and immunoprecipitation were then analysed by northern blotting. For western blotting, the immunoprecipitation samples were boiled in 2× Laemmli buffer for 10 min at 95 °C, separated on 7.5% SDS–polyacrylamide gels, and blotted onto 0.2-μm nitrocellulose membranes (Bio-Rad) in a wet blot chamber on ice. Membranes were blocked for 1 h with 5% milk in PBS-T (137 mM NaCl, 12 mM phosphate, 2.7 mM KCl, pH 7.4, 0.1% (v/v) Tween-20) and incubated with P1W1 (dilution 1:1,000) and tubulin primary antibodies (rat anti-α tubulin, MCA78G, 1:1,000, Sandino) overnight at 4 °C.

The next day, membranes were washed three times with PBS-T and incubated with secondary antibodies conjugated to a fluorescence dye (IRDye 800CW conjugated anti-rabbit, 1:10,000, Li-Cor, and IRDye 680LT conjugated anti-goat-rat, 1:10,000, Li-Cor) for 1 h at room temperature in the dark. After washing three times in PBS-T, signal was detected with the Odyssey-CLX Imaging system (Li-Cor). Alternatively, Ae. aegypti Liverpool embryos were bead-beaten in RIPA buffer supplemented with 1 mM PMSF (Sigma-Aldrich), and boiled in 1× Laemmli buffer. Western blotting was performed as described above, however, the membrane was blocked in 5% BSA in PBS-T, and incubated with primary antibodies for tubulin and pSer2 PolII (rabbit anti-RNA polymerase II C-terminal domain repeat YSPTSPS (phosphorylated serine at position 2), ab5095, 1:1,000, Abcam). Uncropped western blots can be found in Supplementary Fig. 1.

Northern blot
piRNAs were detected by northern blot analyses, as published. In brief, RNA was denatured at 80 °C for 2 min in Gel Loading Buffer II (Ambion) and size-separated on 0.5× TBE (45 mM Tris-borate, 1 mM EDTA), 7 M urea, 15% denaturing polyacrylamide gels. RNA was then blotted onto Hybond-NX nylon membranes (GE Healthcare) in a semi-dry blotting chamber for 45 min at 20 V and 4 °C and crosslinked to the membrane with EDC crosslinking solution (127 mM 1 methylimidazole (Sigma-Aldrich), 163 mM N-(3-dimethylaminopropyl)-N'-ethylcarbodiimide hydrochloride (Sigma-Aldrich), pH 8.0) at 60 °C for 2 h. Crosslinked membranes were pre-hybridized in ULTRAHyb-Oligo hybridization buffer (Thermo Scientific) for 1 h at 42 °C and incubated with biotinylated DNA oligonucleotide probes overnight at 42 °C. Membranes were then washed with decreasing concentrations of SSC (300 mM NaCl, 30 mM sodium citrate, pH 7.0; 150 mM NaCl, 15 mM sodium citrate; 15 mM NaCl, 1.5 mM sodium citrate) and 0.1% SDS, and exposed to Carestream BioMax XAR X-Ray films (Kodak). Probe sequences can be found in Supplementary Table 5.

To detect a putative tapiR precursor transcript, RNA was size-separated on a 0.5× TBE, 7 M urea, 6% denaturing polyacrylamide gel, transferred to Hybond-N+ nylon membranes (GE Healthcare) in a semi-dry blotting chamber for 2 h at 200 mA at 4 °C, UV-crosslinked at 150 mJ, pre-hybridized in ULTRAHyb Ultrasensitive hybridization buffer (Thermo Scientific), and probed with 32P-labelled DNA probes overnight at 42 °C, washed as described above, exposed to a Storage Phosphor Screen GP (Kodak), and developed with the Amersham Typhoon S Bioolecular Imager (GE Healthcare). The probe was produced with the Amersham Rediprime DNA Labelling System II (GE Healthcare) from a PCR-amplified sequence of the tapiR locus spanning 1.5 repeat units, phosphorylated with T4 polynucleotide kinase (Roche) at 37 °C for 30 min and ligated into the pMT-GL3 vector. For cloning of 5′UTR reporters, the target site was cloned into the Pmel and SacII restriction sites, while for the 3′UTR reporters the target site and an upstream BamHI site were cloned into NotI and Xhol restriction sites. ORF reporters were constructed by cloning a Kozak sequence followed by the first 45 nt of luciferase and the target site into XhoI and Ncol sites. For designing the intron reporters, the first intron of Rsps7 (AAELO94996) was cloned behind the duplicated first 45 nt of luciferase in the pMT-GL3 vector, the original ATG of firefly luciferase was mutated, 3 stop codons were introduced at the 3′ end of the intron, and a BsaBI restriction site was inserted in the first third of the intron by site directed mutagenesis. Afterwards, oligonucleotides encoding tapiR1 or control target sequences were inserted into BsaBI as described above. IRES-containing reporters were designed by cloning the 5′ UTR of cricket paralysis virus amplified from infected S2 cells into Pmel and SacII restriction sites of the pMT-GL3 vector with a mutated tapiR1 target site (ΔtapiR1) using the HD In Fusion cloning kit (Takara). Afterwards oligonucleotides containing tapiR1 or control target sites were cloned into the SacII site in the 3′ UTR as described above. RNA polymerase III reporters were constructed by cloning the Ae. aegypti U6 promoter and the GL3 3′ UTR including different tapiR1 target sites and a series of six Ts as termination signal into pUC19 with the HD In Fusion cloning kit (Takara). As normalization control, a part of the GL3 ORF was cloned downstream of the U6 promoter. Sequences of the oligonucleotides are provided in Supplementary Table S5. Correct insertion of target sites was confirmed by Sanger sequencing for all clones.

Where indicated, mutated firefly or Renilla luciferase versions were used that contain synonymous mutations destroying the predicted target sites for tapiR1 (firefly luciferase, nucleotide 782: 5′-gagctgcttgtagattgataaggaa-3′ mutated to 5′-gagctgcttgtagattgataaggaa-3′, and Renilla luciferase, nucleotide 462 5′-taagctgcttgtagattgataaggaa-3′ mutated to 5′-taagctgcttgtagattgataaggaa-3′; modified nucleotides underlined). A2g 2 cells were seeded in 96-well plates the day before the experiment and transfected with 100 ng of the indicated plasmids and 100 ng pMT-Renilla16 per well, using 2 μl X-tremeGENE HP DNA transfection reagent per 1 μg plasmid DNA according to the manufacturer’s instructions. Alternatively, 100 ng reporter plasmid and 100 ng pMT-Renilla were co-transfected with the indicated amounts of 5′Cy5-labelled, fully 2′O-methylated antisense RNA oligonucleotide using an additional amount of 4 μl X-tremeGENE HP DNA transfection reagent (Roche) per 1 μg oligonucleotide. Medium was replaced 3 h after reporter plasmid transfection with 0.5 mM CuSO4, in fully supplemented Leibovitz’s L-15 medium to induce the metallothionein promoter. Twenty-four hours later, cells were lysed in 30 μl passive lysis buffer (Promega) and activity of both luciferases was measured in 10 μl of the sample with the Dual Luciferase Reporter Assay system (Promega) on a Modulus Single Tube Reader (Turner Biosystems). Firefly luciferase was normalized to Renilla luciferase activity. For each construct, at least two to three independent clones were measured in triplicate wells to exclude clonal effects.

RT-qPCR
Total RNA (1 μg) was treated with DNase I (Ambion) for 45 min at 37 °C and reverse-transcribed using the Taqman reverse transcription kit (Applied Biosystems) according to the manufacturer’s protocol. Real-time PCR was performed with the GoTag qPCR Master Mix (Promega) and measured on a LightCycler480 instrument (Roche) with 5 min initial denaturation and 45 cycles of 5 s denaturation at 95 °C, 10 s annealing at 60 °C and 20 s amplification at 72 °C. Starting fluorescence values of specific mRNAs were calculated with linear regression method of log
Stem-loop RT–qPCR
Levels of tapiR1 were quantified using the stem-loop RT–qPCR approach previously used for miRNA quantification36. In brief, 100 ng of total RNA was reverse-transcribed with 10 pmol stem-loop reverse transcription primer using 25 U Superscript II reverse transcriptase (Invitrogen) in 1× First Strand buffer, 0.33 mM dNTPs, and 2 U RNase inhibitors. cDNA was then measured by qPCR as described above.

3′ RACE and 5′ RACE
The 3′ rapid amplification of cDNA ends (3′ RACE) was performed using the FirstChoice RLM-RACE Kit (Thermo Fisher Scientific) according to the manufacturer’s instructions. Amplification products were separated on agarose gel, purified and Sanger sequenced. Slicer products were detected from 1 μg total RNA by 5′ RACE following the instructions, but without prior calf intestine alkaline phosphatase and tobacco acid pyrophosphatase treatment. After amplification by PCR, RACE products in the size range of 150–250 nucleotides were purified from agarose gel, cloned into pUC19 with the HD In-Fusion cloning kit (Takara Bio) according to the manufacturer’s instructions, and sequenced by Sanger sequencing. Primer sequences can be found in Supplementary Table 5.

Antisense oligonucleotide treatment and injection
Aeg2 cells were seeded in 24-well plates the day before the experiment. Cells were treated with 500 nM 5′Cy5-labelled, fully 2′O-methylated antisense RNA oligonucleotide in 530 μl medium with 4 μl X-tremeGEME HP DNA transfection reagent (Roche) per 1 μg oligonucleotide. Medium was replaced after 3 h and cells from three independent experiments were collected 48 h after transfection and prepared for RNA sequencing (see below). For 5′ RACE, cells were treated with 200 nM antisense oligonucleotides together with 50 nM siRNA duplexes, and RNA was collected 24 h later.

Ae. aegypti eggs were injected with antisense oligonucleotides as described above. However, the engorged female mosquitoes were kept at 25 °C and 70% humidity and allowed to lay eggs for 45 min. Injection was performed with a Femtotjet 4x (Eppendorf) with 1,200 HaPa pressure. Injected embryos were then transferred to a wet Whatman paper and kept at 27 °C and 80% humidity for the indicated times. Per experiment, 50 to 150 embryos were injected per condition 1–2 h after egg laying.

Scoring of embryo development and hatching
Injected embryos were allowed to develop for 2.5 days after injection on a moist Whatman paper and then fixed in 4% paraformaldehyde for 8 h to overnight. Afterwards, the pigment of the endochorion was bleached with Trpss solution37 (0.037 M sodium chloride, 1.45 M acetic acid) for 24 to 48 h. Embryos were washed five times in PBS and images were taken with an EVOS FL imaging system (Thermo Fisher Scientific). Embryos with evident larval segmentation (head, fused thoracic elements and abdomen) were scored as fully developed and embryos without any evident structure of the ooplasm as undeveloped. Individuals that showed first signs of structural rearrangements of the ooplasm but did not complete larval segmentation were scored as intermediate (see Supplementary Fig. 2e). To avoid biases, the scoring was performed blindly. Six independent experiments were performed using the maximum number of embryos that was practically feasible (see Supplementary Table 7).

Hatching rate of injected embryos was analysed four days after injection. Embryos were kept moist for two days and then allowed to slowly dry for the rest of the period. The embryos were transferred to water and then forced to hatch by applying negative pressure for a period of 30 min. The number of hatched L1 larvae was counted immediately afterwards. Five independent experiments were performed with numbers as practically feasible (Supplementary Table 7).

Poly(A) tail length assay
Ligation-mediated poly(A) tail length (PAT) assay38 and RACE PAT39 were performed as published. Superscript IV reverse transcriptase (Invitrogen) was used for cDNA synthesis. RACE-PAT products were amplified by PCR with 3 min initial denaturation at 95 °C, 35 cycles of 30 s denaturation, 30 s annealing at 60 °C and 1 min extension at 72 °C, and final extension for 5 min. PCR for LM-PAT was performed similarly, but with 2 min extension each cycle. PCR products were separated on agarose gel and stained with ethidium bromide. As control, Sindbis virus (SINV) poly(A) tails were compared between in vitro transcribed RNA, produced as previously described40,41, and total RNA isolated from SINV-infected Aag2 cells.

Sequence logo
Repeat monomers from the satellite repeat loci in Ae. aegypti, Ae. albopictus, and Culex quinquefasciatus were extracted manually from the current genome annotations obtained from VectorBase (Ae. aegypti Liverpool AaegL5, Ae. albopictus Foshan Aalof1, Culex quinquefasciatus Johannesburg Cpip2). A repeat unit was defined as the sequence starting from the first tapiR1 nucleotide until one nucleotide upstream of the next tapiR1 sequence. Sequences were aligned using MAFFT (v.7.397)42 (with options –genafpair –leavegappyregion–kimura 1–maxiterate 1000–retree 1) and the sequence logo was constructed with the R package ggseqlogo43.

Small RNA sequencing
Small RNAs from Aag2 cells (input) or PIWI immunoprecipitations were cloned with the TruSeq small RNA sample preparation kit (Illumina) according to the manufacturer’s instructions. For the input sample, size selected 19- to 33-nucleotide small RNAs purified from polyacylamide gel were used to construct the library as previously described44, whereas immunoprecipitation samples were not extracted from gel and isolated as previously described45. Libraries were sequenced on an Illumina HiSeq 4000 instrument by Plateforme GenomEast. Sequenced libraries are available under BioProject number PRJNA594491.

mRNA sequencing
RNA was isolated from Aag2 cells 48 h after antisense oligonucleotide transfection (3 independent experiments), or from embryos at 20.5 h after antisense oligonucleotide injection (about 50 embryos pooled per experiment from 5 independent experiments) with RNAsolv reagent following standard phenol–chloroform extraction. Polyadenylated RNAs were extracted and sequencing libraries were prepared using the TruSeq stranded mRNA Library Prep kit (Illumina) following the manufacturer’s instructions, and sequenced on an Illumina HiSeq 4000 instrument (2× 50 bases). Sequenced libraries are available under BioProject number PRJNA482553.

Analysis of mRNA sequencing
Reads were mapped to the Ae. aegypti genome AaegL5 as provided by VectorBase (https://www.vectorbase.org) with STAR (v.2.5.2b)44 in 2-pass mode: first mapping was done for all samples (options:–readFilesCommand zcat –outSAMtype None –outSAMattrIHasStart 0 –outSAMstrandField intronMotif), identified splice junctions were combined (junctions located on the mitochondrial genome were filtered out, as these are probably false positives), and this list of junctions was used in a second round of mapping with (–sjdbFileChrStartEnd) using default parameters as above. Reads were quantified with the additional option –quantMode GeneCounts to receive reads per gene. Alternatively, reads were quantified on TEfam transposon consensus sequences (https://tefam.biochem.vt.edu/tefam/get fasta.php, downloaded in April 2017) with Salmon (v.0.8.2)45, default settings and libType set to ISR. Statistical and further downstream analyses were performed with DESeq246 from Bioconductor. Significance was tested at a false discovery rate (FDR) of 0.01 and a log2-transformed fold change of 0.5. tapiR1 target

Project number PRJNA482553.
sites were predicted with the online tool from RNAHybrid48 with helix constraints from nucleotide two to seven, and no G:U wobble allowed in the seed. Predictions were made on the AaegL5.1 geneset as provided by VectorBase, and on TEfam transposon consensus sequences. For Fig. 3h, publicly available sequencing datasets49 (accession numbers: SRR923702, SRR923826, SRR923837, SRR923853 and SRR923704) were mapped and quantified as described above. Genes were categorized based on their expression in embryos at 0–2 h versus 12–16 h after egg laying. Genes not detected in the 0–2 h sample, but expressed at 12–16 h were defined as purely zygotic (black bars in Fig. 3h), and genes that did not increase or decrease by more than log(0.5) were defined as maternally provided and stable throughout early development (grey bars). Genes that decreased in expression by more than log(0.5), log(2) and log(5) from 0–2 h to 12–16 h were categorized as maternal unstable fraction (shades of blue). Genes that increased in expression by more than log(0.5), log(2) and log(5) were considered maternally provided and actively transcribed by the zygote (shades of yellow to red). tapIR targets were defined as genes that were significantly upregulated at least twofold in embryos injected with tapIR antisense oligonucleotide and containing a predicted tapIR target site (nfe = 2–4 kcal mol⁻¹).

A list of publicly available datasets that were used in this study is provided in Supplementary Table 6.

Analysis of small RNA sequencing
The 3′ sequencing adapters (TGGAGTTCTCGGTGCCAGG) were trimmed from sequence reads with Cutadapt (v.1.14) and trimmed reads were mapped with Bowtie (v.0.12.7) to the Ae. aegypti LVP AGWG genome sequence AaegL5.1 obtained from VectorBase with at most one mismatch. Reads that mapped to rRNAs or tRNAs were excluded from the analyses. Alternatively, 3′ sequencing adapters ((NNN)TGGAGTTCTCGGTGCCAGG) and three random bases were trimmed from publicly available datasets from Ae. aegypti somatic and germline tissues32 (SRR5961503, SRR5961504, SRR5961505 and SRR5961506) and then processed as described above. Oxidized libraries, immunoprecipitates and input sample were normalized to the total number of mapped reads; all other libraries were normalized to the total number of miRNAs (in millions). piRNAs that were at least twofold enriched in a PIWI immunoprecipitation compared with the corresponding input sample and were present with at least 10 reads per million (rpm) in the immunoprecipitation sample were considered PIWI-bound. Mapping positions were overAPPED with basefeatures and repeatfeatures retrieved from VectorBase and counted with bedtools33. Reads that mapped to two or more features were assigned to only one feature with the following hierarchy: open reading frames > non-coding RNAs (including lncRNAs, pseudogenes, small nucleolar (sno)RNAs, small nuclear (sn)RNAs, miRNAs) > long terminal repeat retrotransposons > non-long terminal repeat retrotransposons (short interspersed nuclear elements, long interspersed nuclear elements, Penelope) > cut-and-paste DNA transposons > other DNA transposons (helitrons, miniature inverted-repeat transposable elements) > satellite and tandem repeat features > low-complexity repeats (DUST) > other or unknown reads. Accordingly, reads that mapped to a repeat feature and an intron or UTR were classified as repeat-derived, whereas all other reads mapping to introns or UTRs were considered as gene-derived. Positions not overlapping with any annotation were summarized as ‘other’. Results were then visualized with ggplot228, or Gviz46 in R.

Reporting summary
Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability
Raw sequence data have been deposited in the NCBI Sequence Read Archive under BioProject numbers PRJNA482553 and PRJNA594491.

Code availability
The source code is available at https://github.com/RebeccaHalbach-Halbach_tapiR_2020.git.

26. Lan, Q. & Fallon, A. M. Small heat shock proteins distinguish between two mosquito species and confirm identity of their cell lines. Am. J. Trop. Med. Hyg. 42, 669–676 (1990).
27. Vodover, N. et al. Arbovirus-derived piRNAs exhibit a ping-pong signature in mosquito cells. PLoS ONE 7, e30861 (2012).
28. Miesen, P., Girardi, E. & van Rij, R. J. P. Distinct sets of PIWI proteins produce arboravirus and tombusvirus-derived piRNAs in Aedes aegypti mosquito cells. Nucleic Acids Res. 43, 6545–6556 (2015).
29. Fansiri, T. et al. Genetic mapping of specific interactions between Aedes aegypti mosquitoes and dengue viruses. PLoS Genet. 9, e1003621 (2013).
30. Goertz, G. P., Vogele, C. B. F., Grompe, K., Hoenger, A., Kupke, C. & Pijman, G. M. P. Co-infection with Zika and chikungunya virus allows simultaneous transmission without affecting vector competence of Aedes aegypti. PLoS Negl. Trop. Dis. 11, e0005654 (2017).
31. Möhlimann, T. W. R. et al. Community analysis of the abundance and diversity of mosquito species (Diptera: Culicidae) in three European countries at different latitudes. Parasit. Vectors 10, 510 (2017).
32. Joosten, J. et al. The Tudor protein Veneno assembled the ping-pong amplification complex that produces viral piRNAs in Aedes mosquitoes. Nucleic Acids Res. 47, 2546–2559 (2019).
33. Pall, G. S. & Hamilton, A. J. Improved northern blot method for expression analysis of small RNA. Nat. Protoc. 3, 1077–1084 (2008).
34. van Rij, R. P. et al. The RNA silencing endorncus Argonaute 2 mediates specific antiviral immunity in Drosophila melanogaster. Genes Dev. 20, 2985–2995 (2006).
35. Ramakers, C., Ruijter, J. M., Deprez, R. H. & Moorman, A. F. Assumption-free analysis of quantitative real-time polymerase chain reaction (PCR) data. Neurosci. Lett. 339, 62–66 (2003).
36. Chen, C. et al. Real-time quantification of microRNAs by stem-loop RT-PCR. Nucleic Acids Res. 33, e79 (2005).
37. Tripilí, M. A new bleaching and decalcifying method for general use in zoology. Can. J. Zool. 48, 882–893 (1970).
38. Murray, E. L. & Schoenberg, D. R. Assays for determining poly(A) tail length and the polarity of mRNA decay in mammalian cells. Methods Enzymol. 448, 483–504 (2008).
39. Sallés, F. J. & Strickland, S. Analysis of poly(A) tail lengths by PCR: the Pat assay. Methods Mol. Biol. 118, 441–448 (1999).
40. Hahn, C. S., Hahn, Y. S., Brasebona, T. J. & Rice, C. M. Infectious Sindbis virus transmembrane expression vectors for studying antigen processing and presentation. Proc. Natl Acad. Sci. USA 89, 2679–2683 (1992).
41. van Mierlo, J. T. et al. Novel Drosophiiva viruses encode host-specific suppressors of RNAi. PLoS Pathog. 10, e1004256 (2014).
42. Katoh, K., Misawa, K., Kuma, K. & Miyata, T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res. 33, 389–400 (2005).
43. Wagh, G. gp120: a versatile R package for drawing sequence logos. Bioinformatics 33, 3645–3647 (2017).
44. van Cleef, K. W. et al. Mosquito and Drosophila entomobioriviruses suppress dsiRNA- and siRNA-induced RNAi. Nucleic Acids Res. 42, 8732–8744 (2014).
45. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics 29, 15–21 (2013).
46. Patro, R., Duggal, G., Love, M. I., Irizarry, R. A. & Kingsford, C. Salmon provides fast and bias-aware quantification of transcript expression. Nat. Methods 14, 417–419 (2017).
47. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 15, 550 (2014).
48. Rehmsmeier, M., Steffen, P., Hochmann, M. & Giegerich, R. Fast and effective prediction of small RNA. Bioinformatics 25, 2546–2559 (2009).
49. Hitchcock, T. K. & Strauss, J. H. Replication of Sindbis virus. VI. Poly(A) and poly(U) in virus particles. Virology 86, 464–506 (1978).
Organization for Scientific Research (grant number 016.VICI.170.090). A stay of R.H. at Pasteur Institute, Paris, France was supported by ERASMUS+.

**Author contributions** R.H., P.M. and R.P.v.R designed the experiments and analysed the data. R.H. performed the computational analyses and most of the experiments, except for PIWI immunoprecipitations for small RNA-seq (J.J. and E.T.), design and validation of PIWI antibodies (B.P.) and tissue isolations and blood feeding experiment (C.B.F.V. and C.J.K.). C.B.F.V. and C.J.K. provided wild-caught mosquito samples. I.R. assisted with the experiments, and S.H.M. and L.L. helped with optimizing embryo injections. R.H. and R.P.v.R. wrote the paper. All authors read and contributed to the manuscript.

**Competing interests** The authors declare no competing interests.

**Additional information**

**Supplementary information** is available for this paper at https://doi.org/10.1038/s41586-020-2159-2.

**Correspondence and requests for materials** should be addressed to R.P.v.R.

**Peer review information** Nature thanks Francis Jiggins, Susumu Katsuma and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

**Reprints and permissions information** is available at http://www.nature.com/reprints.
Extended Data Fig. 1 | Expression of piRNAs from a satellite repeat locus.

a, Fraction of siRNAs and piRNAs mapping on genomic features in adult Ae. aegypti female ovaries (germline) or carcasses (soma). Small RNAs that overlapped multiple features were assigned to only one category (Methods). The leftmost bar depicts the fraction of each feature category in the genome. b, Read length distribution of tapiR1 and 2 in Aag2 cells, and adult germline and somatic tissues (β-eliminated, or untreated).
Extended Data Fig. 2  tapiR1 and tapiR2 are expressed in *Ae. aegypti* mosquitoes and associate with Piwi4. **a,** Northern blots of tapiR1 and tapiR2 in different tissues of adult mosquitoes (a), upon dsRNA-mediated knockdown of individual PIWI genes (d) and upon knockdown of miRNA-and siRNA-pathway genes (e), or a control dsRNA treatment (dsFLuc and dsRLuc) in Aag2 cells. U6 snRNA or ethidium-bromide-stained rRNA serve as loading controls. **b,** Western blot analysis of the indicated PIWI proteins before (input) and after immunoprecipitation (IP) used for the small RNA northern blot in c. An immunoprecipitation with empty beads serves as negative control. Tubulin was used to control for nonspecific binding. **c,** Immunoprecipitation of the indicated PIWI proteins from Aag2 cells followed by northern blot analyses for tapiR1 and tapiR2.
Extended Data Fig. 3 | See next page for caption.
Extended Data Fig. 3 | Expression of tapiR1 is conserved in culicine mosquitoes and is independent of AAEL017385. a, Schematic representation of the tapiR satellite repeat locus in Ae. aegypti, Ae. albopictus and Culex quinquefasciatus. Numbers indicate repeat lengths and, for Culex quinquefasciatus, lengths of deviating repeat monomers. b–d, Sequence logos constructed from all individual tapiR repeat units in Ae. aegypti (b), Ae. albopictus (c) or Culex quinquefasciatus (d). Gaps in the sequence logos mainly arise owing to size heterogeneity in few repeat monomers. e, Evolutionary relationships of dipterous genera based on ref. 10. Bar lengths are arbitrary and do not reflect evolutionary distances. f, Northern blot of tapiR1 in Aag2 cells treated with dsRNA targeting different transcripts of AAEL017385 (indicated in g) or, as control, firefly luciferase. Ethidium bromide stained rRNA serves as loading control. g, Top, Schematic of AAEL017385 and the tapiR satellite repeat locus. The primer used for 3′ RACE, and positions targeted by dsRNA in f are indicated with an arrow and wavy lines, respectively. Bottom, 3′ RACE analysis of AAEL017385 transcripts. Indicated are sequences from the current AaegL5 genome annotation and RACE PCR products. The sequences of the 5′ terminal part of tapiR1 and tapiR2 repeats are highlighted with colours. h, Northern blot of a potential tapiR1 or tapiR2 precursor transcript. i, RNA-seq read coverage of the tapiR repeat locus and AAEL017385 (top) and sashimi plot indicating spliced reads (bottom).
Extended Data Fig. 4 | Antisense oligonucleotides relieve tapiR1-mediated silencing. a, Luciferase assay of reporters with a fully complementary target site for tapiR1 (left) or tapiR2 (right) in the 3′ UTR. Aag2 cells were co-transfected with the reporter and increasing amounts of fully 2′O-methylated antisense tapiR RNA oligonucleotides, or a control antisense oligonucleotide. Firefly luciferase activity was normalized to the activity of a co-transfected Renilla luciferase reporter. Indicated are mean, s.d. and individual measurements from a representative experiment measured in triplicate wells. b, Northern blot of tapiR1, tapiR2 and three different miRNAs in Aag2 cells upon treatment with the indicated concentrations of tapiR1, tapiR2 or control antisense oligonucleotides. Ethidium-bromide-stained rRNA serves as loading control.
Extended Data Fig. 5 | Renilla luciferase contains a functional tapiR1 target site. a, Schematic of the different reporter constructs used in this study. pMT, metallothionein promoter; RNAPIII, RNA polymerase III reporter.
b, Representative northern blot (left) and quantification (right) of RNAPIII reporters carrying the indicated tapiR1 target sites. Values are normalized to a non-targeted transfection control. Mean, s.d. and individual measures of three independent experiments (indicated with colours), quantified in triplicate, are shown. The panels are split to reflect that samples were loaded at different locations of the same gel. c, Schematic of predicted tapiR1 target sites and minimum free energy of the indicated structures in the coding sequences of Renilla luciferase (RLuc) or firefly luciferase (FLuc). Numbers indicate the position of the targets relative to the first nucleotide in the ORFs. d, Luciferase assay of Aag2 cells transfected with reporters carrying either a scrambled (scr) site or the predicted target site from firefly luciferase (left) or Renilla luciferase (right) from e in the 3′ UTR of FLuc. e, Luciferase activity of FLuc or RLuc constructs with synonymous mutations introduced into the predicted tapiR1 target site (ΔtapiR1 site), and the parental clones. f, Luciferase assay of reporters carrying tapiR1 target sites or control sequences in the 3′ UTR of either the parental firefly luciferase or the ΔtapiR1 firefly luciferase version. g, Reporter assay with luciferase carrying tapiR1 target sites with single mismatches in the 3′ UTR as used in Extended Data Fig. 6b, using RLuc with a mutated tapiR1 target site (ΔtapiR1 site) for normalization. Left, magnified version of the graph on the right. Shown are mean, s.d. and individual measurements from representative experiments performed with at least two different clones per construct, and each measured in triplicate wells.
Extended Data Fig. 6 | tapiR1 uses a G:U wobble sensitive seed sequence for target recognition. a, Schematic of the reporter constructs used in b and Fig. 2. Numbers indicate the position of the mismatch relative to the 5′ end of the piRNA. b, Luciferase activity of reporters carrying a tapiR1 target site with single mismatches. c, Luciferase activity of reporters with the tapiR1 target site from RLuc and indicated mismatches in the 3′UTR of FLuc (left). Right, predicted tapiR1–RLuc target RNA duplexes. d, Luciferase activity of tapiR1 reporters carrying mismatches or G:U wobble base pairs at the indicated positions. Firefly luciferase activity was normalized to the activity of a co-transfected Renilla luciferase reporter to control for differences in transfection efficiencies. Data represent mean, s.d. and individual measurements of representative experiments with two independent clones per construct and measured in triplicate wells.
Extended Data Fig. 7 | Validation of tapiR1 target genes. a, Predicted structures and minimum free energy of tapiR1–target RNA duplexes analysed in b. Left column, functional target sites resulting in silencing of luciferase reporters; right column, nonfunctional sites. b, Luciferase assay of reporters carrying the predicted target sites from a in the 3′UTR of firefly luciferase. Firefly luciferase activity was normalized to the activity of a co-transfected renilla luciferase reporter to control for differences in transfection efficiencies. Indicated are mean, s.d. and individual measurements from representative experiments performed with one to three independent clones per construct and measured in triplicate wells. c, AAELO01741, AAELO17422 and AAELO00453 were annotated in the previous Aaegl.3 gene set, but not in the current Aaegl.5 gene set. Read coverage in Aag2 cells treated with tapiR1 or control antisense oligonucleotides at these genomic regions suggests that these regions are actively transcribed but repressed by tapiR1. Red boxes indicate the positions of tapiR1 target sites.
Extended Data Fig. 8 | tapiR1 and Piwi4 silence gene expression in Aag2 cells. 
a, mRNA expression of transposable elements in Aag2 cells treated with a tapiR1-specific antisense oligonucleotide or control antisense oligonucleotide. Depicted are the means of three biological replicates. A pseudocount of one was added to all values to plot values of zero. Diagonal lines represent a twofold change. Significance was tested at an FDR of 0.01 and a log2-transformed fold change of 0.5.
b, Luciferase assay of reporters containing different tapiR1 target sites (from Extended Data Fig. 7a) in the 3′UTR of firefly luciferase. Firefly luciferase activity was normalized to the activity of a co-transfected Renilla luciferase reporter to control for differences in transfection efficiencies. Data represent mean, s.d. and individual measurements of representative experiments measured in triplicate wells.
c, d, RT–qPCR of tapiR1 target genes upon dsRNA-mediated knockdown of FLuc (control), Piwi4 or Ago1 in Aag2 cells (c), or after treatment with different concentrations of control, tapiR1 or tapiR2 antisense oligonucleotides (d). Depicted are mean, s.d. and individual measurements of a representative experiment as measured from triplicate wells (c), or from duplicate wells (d). *Even skipped* (eve) does not contain a tapiR1 target site and serves as control.
e, Violin plot of log2-transformed fold changes in mRNA expression of all genes upon treatment with tapiR1 or control antisense oligonucleotides in Aag2 cells (left) and mosquito embryos (right), either with or without predicted tapiR1 target site. f, log2-transformed fold changes in RNA expression of genes upon treatment with tapiR1 or control antisense oligonucleotides in Aag2 cells (left) and mosquito embryos (right) plotted against the minimum free energy of predicted tapiR1-target RNA duplexes. Blue dots indicate target sites that were confirmed to be functional in luciferase reporter assays, and red dots indicate target sites that were not functional (Extended Data Fig. 7b).
Extended Data Fig. 9 | tapiR1 regulates gene expression in mosquito embryos. 

**a**, Western blot analysis of phosphorylated RNA polymerase II (serine 2 of the C-terminal domain repeats, middle) and tubulin (bottom panel) in embryos at the indicated time points after egg laying, and corresponding stem-loop RT–qPCR of tapiR1 measured in technical duplicates (top).

**b, c**, Northern blot of tapiR1 and 2 in developmental stages of Ae. aegypti mosquitoes (b), or at different time points after blood feeding (c). U6 snRNA (b) or ethidium bromide-stained rRNA (c) were analysed to verify equal loading.

**d, e**, Piwi4 RNA-seq read counts in the indicated adult tissues (d) or developmental stages (e). Libraries used for these analyses are listed in Supplementary Table 6.

**f**, log2-transformed mRNA expression of transposable elements in embryos injected with tapiR1 or control antisense oligonucleotides. Mean counts of five biological replicates are shown. Significance was tested at an FDR of 0.01 and a log2-transformed fold change of 0.5. Diagonal lines indicate a twofold change.

**g, h**, RT–qPCR of the indicated tapiR1 target genes 9 h after injection of tapiR1 or control antisense oligonucleotides (g), or dsRNA-mediated knockdown of Fluc (control), Piwi4 or Ago1 (h) 21 h after injection in embryos. Mean, s.d. and individual measurements of a representative experiment are presented. Even skipped (eve) is not a tapiR1 target gene and serves as negative control. 

**i**, Overlap of upregulated tapiR1 target genes (log2-transformed fold change ≥ 1, with a predicted target site with mfe ≤ −24 kcal/mol) in Aag2 cells and Ae. aegypti embryos.
Extended Data Fig. 10 | See next page for caption.
**Extended Data Fig. 10 | tapiR1 regulates gene expression at a post-transcriptional level.**

- **a**, Schematic of different modes of silencing of the three small RNA silencing pathways in *Drosophila*. 
- **b**, Top, sequences of tapiR1 target gene AAEL026349, an siRNA targeting the gene at the same position, and Sanger sequencing results of 5′ RACE of Aag2 cells treated with tapiR1 antisense oligonucleotide and siRNAs. The tapiR1 target site is indicated in blue. RACE-sequencing adaptor in yellow and gene sequence in dark grey. The predicted slice site between nucleotides 10 and 11 is marked with a red vertical line. Bottom, summary of the results from 5′ RACE in the indicated conditions. Numbers refer to the number of sequenced clones with the 5′ RACE adaptor ligated to the predicted slice site and the total number of sequenced clones is shown between brackets. 
- **c**, Small RNA coverage in Aag2 cells (not normalized) and individual reads (direction of the arrow indicates the strand) on tapiR1 target genes. Red boxes indicate positions of tapiR1 target sites on the mRNA. 
- **d**, Schematic (top) and luciferase expression (bottom) of IRES-containing reporter constructs. Depicted are mean, s.d. and individual measurements of a representative experiment performed in triplicate wells with two different reporter clones. Bottom right, firefly luciferase (FLuc) activity normalized to *Renilla* luciferase (RLuc); raw luciferase counts from the same experiment are shown in the left (FLuc) and middle (RLuc). 
- **e**, Luciferase activity of a reporter harbouring the tapiR1 target site of AAEL001555 in the 3′ UTR of FLuc upon dsRNA mediated knockdown of the indicated genes. Symbols are colour-coded according to the indicated RNA decay pathways. FLuc expression was normalized to RLuc expression to control for differences in transfection efficiencies and expressed relative to non-targeting control dsRNA (Sindbis virus dsRNA). Depicted are mean and standard deviation of one experiment performed in triplicates. Horizontal lines indicate a fold change of 1 and 1.5. 
- **f**, AAEL00831 target gene expression as measured by RT-qPCR upon knockdown of the indicated genes. Primer sets located 5′ (upstream) and 3′ (downstream) to the tapiR1 target site were used for PCR. Mean, s.d. and individual measurements of one out of three experiments performed in triplicate are shown. The other two experiments are presented in Supplementary Fig. 2f, g. The horizontal line indicates a twofold change. 
- **g**, Schematic illustration of the two PAT assays and expected results of genes with increasing poly(A) tail lengths. LM-PAT, ligation mediated-PAT; RACE-PAT, rapid amplification of cDNA ends-PAT. 
- **h–j**, Electrophoretic analysis with ethidium-bromide-stained agarose gels of a LM-PAT assay (**h**) and RACE-PAT assay (**i**) of different tapiR1 target genes upon treatment with two concentrations of tapiR1 or control antisense oligonucleotides. As positive control, poly(A) tail length was measured from SINV RNA in vitro transcribed from a plasmid (IVT), or from infected Aag2 cells, in which the poly(A) tail is elongated during viral replication (J). White asteriks indicate primer dimers.
Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistical parameters

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. figure legend, table legend, main text, or Methods section).

- n/a
- Confirmed

- The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- An indication of whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided
  - Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistics including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g. $F$, $t$, $r$) with confidence intervals, effect sizes, degrees of freedom and $P$ value noted
  - Give $P$ values as exact values whenever suitable.
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen’s $d$, Pearson’s $r$), indicating how they were calculated
- Clearly defined error bars
  - State explicitly what error bars represent (e.g. SD, SE, CI)

Our web collection on statistics for biologists may be useful.

Software and code

Policy information about availability of computer code

Data collection

No software was used to collect data.

Data analysis

Sequencing data was analyzed with custom code that will be made available on GitHub upon publication. For details see Methods.

Open source programs that were used include:
- STAR (version 2.5.2b)
- Salmon (v.0.8.2)
- Cutadapt (version 1.14)
- Bowtie (version 0.12.7)
- samtools (version 1.9)
- bedtools (v2.27.1)
- ImageJ (v1.49)
- RNAHybrid (web service: https://bibiserv2.cebitec.uni-bielefeld.de/rnahybrid)
- R version 3.4.0 and Bioconductor version 3.5

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.
Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Raw sequencing data have been deposited in the NCBI Sequence Read Archive. Summaries of the statistical analyses of mRNA sequencing experiments are provided as Supplementary Tables.

Field-specific reporting

Please select the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

- Life sciences
- Behavioural & social sciences
- Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/authors/policies/ReportingSummary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size
Sample size was not predetermined with statistical methods beforehand, but chosen to be adequate based on the magnitude and consistency of the effects measured.

Data exclusions
No data were excluded from the analyses.

Replication
All experiments were reproducible at replication.

Randomization
Samples were not randomized.

Blinding
For scoring of developmental stages of embryos, investigators were blinded during sample preparation and analysis.

Reporting for specific materials, systems and methods

Materials & experimental systems

| n/a | Involved in the study |
|-----|-----------------------|
| [ ] | Unique biological materials |
| [ ] | Antibodies |
| [ ] | Eukaryotic cell lines |
| [ ] | Palaeontology |
| [ ] | Animals and other organisms |
| [ ] | Human research participants |

Methods

| n/a | Involved in the study |
|-----|-----------------------|
| [ ] | ChIP-seq |
| [ ] | Flow cytometry |
| [ ] | MRI-based neuroimaging |

Unique biological materials

Policy information about availability of materials

Obtaining unique materials
Wild-caught mosquitoes have been collected from the field in different places in Sweden, Italy and the Netherlands. Material from rare species may be limited or unavaiable.

Antibodies

Antibodies used
Antibodies against Aedes aegypti PIWI proteins were custom-made antibodies generated by Eurogentec by immunization of two rabbits per antibody with a mix of two unique peptides (Ago3: TSGADSSESDDKQSS, IYKRKQIMSENQF; Piwi4: HEGRGSPSPAYGK, HHRESSAGGRERSGN; Piwi5: DIVRSRPLDSKVQK, CANQGGNWRDNYKRAI; Piwi6: MADCNPQEGSSGGRIR,
RGDHRQKPYDRPEQS). After 87 days and a total of four immunizations (t=0, 14, 28, 56 days), sera of both rabbits were collected, pooled, and purified against each peptide separately. The antibodies were used at a dilution 1:1000 for Western blot and 1:10 for IP.

Other antibodies:
- rat anti-Tubulin alpha, MCA78G, 1:1000, Sanbio
- rabbit anti-RNA polymerase II CTD YSPTSPS (phosphor S2), ab5095, 1:1000, abcam

Secondary antibodies:
- IRDye 800CW conjugated goat anti rabbit, 1:10,000, Li-Cor
- IRDye 680LT conjugated goat anti rat, 1:10,000, Li-Cor

Validation
Specificity of the PIWI antibodies was confirmed by Western blotting of Aag2 cells stably expressing PTH (ProteinA, TEV cleavage site, 6x His-tag)-tagged PIWI constructs upon knockdown of the respective PIWI protein or control knockdown (dsRLuc) (Supplementary Fig 2A and Joosten et al., Nucleic Acids Res 2019).

Rat anti-Tubulin and rabbit anti-RNA polymerase II are widely used antibodies and validated by the manufacturers.

Eukaryotic cell lines

| Policy information about | cell lines |
|--------------------------|------------|
| **Cell line source(s)**  | Aag2 are a widely used Aedes aegypti cell line. The respective clone used in this study was a gift from Raul Andino, UCSF. |
| **Authentication**       | Cell lines were not authenticated, but RNA-seq data confirm that they are Ae. aegypti derived. |
| **Mycoplasma contamination** | Cells were tested negative for Mycoplasma. |
| **Commonly misidentified lines** | No cell line listed in the database of commonly misidentified cell lines have been used in this study. |

Animals and other organisms

| Policy information about | studies involving animals; ARRIVE guidelines recommended for reporting animal research |
|--------------------------|------------------------------------------|
| **Laboratory animals**   | Injection and northern blot of embryos was performed using a Cell fusion agent virus-free, isofemale Aedes aegypti strain Jane. This strain was initiated from a field population originally sampled in the Muang District of Kamphaeng Phet Province, Thailand (described in Fansiri T et al., PLoS Genetics 2013), and reared for 26 generations in the lab. In vivo knockdown in embryos was performed with the sequenced reference strain Aedes aegypti, Strain Black Eye Liverpool, provided by the NIH/NIAID Filariasis Research Reagent Resource Center for distribution by BEI Resources, NIAID, NIH (NR-48921). All other in vivo experiments were performed with the Ae. aegypti Rockefeller strain, obtained from Bayer AG, Monheim, Germany. Only females were used for the blood-feeding experiments, while embryos were of both sexes. |
| **Wild animals**         | No wild animals were used in this study. |
| **Field-collected samples** | Wild-caught mosquitoes have been described in Möhlmann TWR, Parasit Vectors 2017. |