



**HAL**  
open science

## **A characterization of piARNs, their biogenesis and their targets in *Spodoptera frugiperda* (Lepidoptera, Noctuidae)**

Imène Seninet, Mélanie Gasser, Sylvie Gimenez, Fabrice Legeai, Karine Durand,  
Nicolas Nègre, Kiwoong Nam, Bernard Duvic, Emmanuelle D'alençon

### ► **To cite this version:**

Imène Seninet, Mélanie Gasser, Sylvie Gimenez, Fabrice Legeai, Karine Durand, et al.. A characterization of piARNs, their biogenesis and their targets in *Spodoptera frugiperda* (Lepidoptera, Noctuidae). 2025. <hal-05374380>

**HAL Id: hal-05374380**

**<https://hal.inrae.fr/hal-05374380v1>**

Preprint submitted on 20 Nov 2025

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

A characterization of piARNs, their biogenesis and their targets in *Spodoptera frugiperda*  
(Lepidoptera, Noctuidae)

Imène Seninet<sup>a§</sup>, Mélanie Gasser<sup>\*a§</sup>, Sylvie Gimenez<sup>a</sup>, Fabrice Legeai<sup>b,c</sup>, Karine Durand<sup>a</sup>, Nicolas Nègre<sup>a</sup>, Kiwoong Nam<sup>a</sup>, Bernard Duvic<sup>a</sup> and Emmanuelle d'Alençon<sup>\*a</sup>

a DGIMI, Univ Montpellier, INRAE, Montpellier, France

b BIPAA, IGEPP, INRAE, Institut Agro, University of Rennes, Rennes, France

c University of Rennes, INRIA, CNRS, IRISA, Rennes

§ These authors contributed equally

\* Corresponding authors :

[emmanuelle.d-alencon@inrae.fr](mailto:emmanuelle.d-alencon@inrae.fr)

[melanie.gasser@gmail.com](mailto:melanie.gasser@gmail.com)

Keywords : piRNAs, PIWI proteins, Transposable Elements, piRNA pathway, gene regulation

## Abstract

PIWI-interacting RNAs (piRNAs) and PIWI proteins were initially described as involved in gametogenesis and preservation of genome integrity through the control of transposable elements (TE). Expressed also in the soma and able to regulate protein coding gene expression, they are involved in multiple biological pathways including host-pathogens interaction, sex determination, reproductive isolation. *Spodoptera frugiperda* is a major invasive insect pest species consisting of two strains with different host-plant ranges. In this paper, we characterized proteins and genomic regions involved in their biogenesis as well as TE and gene transcripts regulated by piRNAs.

By phylogenetic analysis, we identified two new *Piwi* genes conserved in the genus *Spodoptera*, compared to the Lepidopteran model *Bombyx mori*, one more than in *Drosophila*. One of them, more expressed in gonads than soma could be a functional homolog of *Drosophila* PIWI or replace AGO3. A pool of 11 sRNA-Seq libraries was used to annotate piRNA clusters with Shortstack in the genomes of the two strains of *Spodoptera frugiperda*. Identification of TE targeted by piRNAs revealed that active transposons differ between the two strains of *S. frugiperda* despite of a similar TE content, as putative cause or consequence of reproductive isolation. GO analysis of genes targeted by piRNAs shows that some are involved in protein translation initiation. A piRNA cluster in the *Masc* gene suggests that sex determination is regulated by piRNAs in *Spodoptera frugiperda*.

Our analysis supports that piRNAs have additional roles than silencing of transposable elements and contributes to functional annotation of the two strains genomes of *Spodoptera frugiperda*.

## 1. Introduction

Piwi-interacting RNAs (piRNAs) are a class of small non-coding RNAs known to preserve genome integrity by silencing transposable elements (TE) in the germline of animals (<sup>1-4</sup> for review). Piwi protein was initially identified as necessary for germ stem cell self-renewal in *Drosophila* <sup>5,6</sup>. PiRNAs were first described in *Drosophila* involved in the process of Su(Ste) silencing of Stellate <sup>7</sup>. Since that time, in addition to TE, piRNAs have been shown to regulate cellular genes involved in several developmental pathways in animals (<sup>8,9</sup> for review).

Their action is not restricted to the germline, they have been reported to function also in somatic tissues for instance in *Drosophila* in the fat body <sup>10</sup> and the brain <sup>11,12</sup>. In addition to repression of TEs, somatic piRNAs are involved in defense against RNA viruses in the mosquito *Aedes aegypti* <sup>13</sup>. Lewis *et al.* recently found somatic piRNAs targeting TEs and cellular mRNAs in 20 species of Arthropods showing that this somatic pathway is an ancestral trait in their phylogeny <sup>14</sup>.

Like other small non coding RNAs, piRNAs have a specific size range and interact with proteins of the Argonaute family which they guide to their regulatory targets by homology, typically leading to decrease in their expression. The Argonaute family of proteins consists of two clades, the AGO clade and the PIWI clade. AGO clade proteins complex with microRNAs (miRNAs) and small interfering RNAs (siRNAs), which derive from double-stranded RNA (dsRNA) precursors <sup>15,16</sup>. Proteins of the PIWI clade interact with piRNAs. The copy-number of the *Piwi* gene family varies among animals suggesting possible sub-functionalisation or neo-functionalisation, for example *C. elegans* encodes two PIWI proteins from a lineage specific duplication <sup>17</sup>, *Drosophila melanogaster* encodes three PIWI proteins PIWI, AUBERGINE (AUB) and AGO-3 <sup>14</sup>, while annotation of the *A. pisum* genome identified eight copies of the *Piwi* genes and two copies of the *Api-ago3* <sup>18</sup> and some *turbellaria* flatworms show up to eight lineage specific paralogs <sup>19</sup>.

The piRNA biogenesis occurs by two pathways, the primary pathways and the ping-pong amplification loop, the first one provides a pool of piRNAs against all transposable elements (TE) hosted by the animal population, whereas the second allows enrichment of piRNAs against active transposons. piRNAs are produced from long single-stranded RNAs transcribed from genomic loci containing remnants of TE, known as piRNA clusters. These transcripts, usually antisense to TE transcripts, are processed into primary piRNAs by Zucchini endonuclease <sup>20</sup>. They are loaded into PIWI proteins which cleave sense TE transcripts generating secondary piRNAs, which in turn guide AGO-3 to cleave piRNA cluster transcripts through an amplification loop called the Ping-Pong loop <sup>21,22</sup>. PIWI-class members exhibit strand specific interaction with piRNAs, PIWI and AUB are associated with antisense TE-piRNAs; while AGO-3 is associated with sense TE-piRNAs <sup>21,22</sup>. Primary piRNAs have a strong bias for uridine residue at the 5' end <sup>23,24</sup>, while secondary piRNAs have a strong bias for adenine at position ten (10A), and show complementarity over 10 bp to primary piRNAs <sup>22</sup>.

*Spodoptera frugiperda* also called Fall armyworm (FAW) is a Lepidopteran pest of crops, described as two strains with different host-plant range <sup>25,26</sup>, the corn strain mostly found on corn (SfC), the rice strain associated to grasses (SfR). Since publication of their genomes <sup>27</sup>, miRNAs genes have been annotated <sup>28,29</sup> but almost nothing is known about the piRNA pathway except that PIWI protein homologs play a role in antiviral response <sup>30</sup>. Another study of the DNA or non-retroviral RNA viruses embedded in piRNA clusters of 48 Arthropods mentioned that some Rhabdoviridae are enriched in piRNA clusters in the genome of the ovary cell line Sf9 <sup>31</sup> without giving any details on the piRNA clusters themselves. In the silkworm, the lepidopteran model, the ping-pong mechanism has been deciphered *in vitro* in cultured Bm4 germ-cells <sup>32-34</sup>. *Bombyx* lacks orthologs of *Drosophila* PIWI and has

no transcriptional silencing machinery that relies on phased piRNAs. It has only AGO3 and AUB which are cytoplasmic and involved in post-transcriptional silencing of TEs or other gene targets<sup>32</sup>. Among the other gene targets of piRNAs in *Bombyx mori*, the sex determination gene *Masculinizer* is cleaved in females by a piRNA encoded by W chromosome resulting in female developmental process<sup>35</sup>. In this paper, we present a phylogenetic analysis of the Argonaute proteins in insects and the expression pattern of *Spodoptera frugiperda* genes in soma and gonads. We annotated piRNA clusters in the two strains genomes of *Spodoptera frugiperda* thanks to a dataset of sRNA-Seq from the two strains. We then identified TE regulated by piRNAs in the two strain genomes and compared the patterns obtained. Finally, we focused on the genes of the insect regulated by piRNAs and explored their functional annotation and putative involvement in sex-determination.

## 2. Material and Methods

### 2.1 Biological material

The sRNA-Seq data described in <sup>36</sup> and re-used in this study were obtained from two laboratory strains of *S. frugiperda*: the corn-strain, originating from French Guadeloupe and the rice-strain, originating from Florida, USA, both from the native area. Each strain was reared on its principal and alternative host plant (*Zea mays* L. cv B73 or *Oryza sativa* L. japonica cv Arelate) under controlled conditions (temperature: 24°C, photoperiod 16:8 light:dark, relative humidity: 65%). Eggs from *S. frugiperda* were deposited on plant leaves and fed ad libitum. After three generations on plants, larvae of fourth instars (L4) were collected. Therefore, larvae selected for small RNA extraction and sequencing comprised four groups: C strain reared on corn (CC), R strain reared on corn (RC), C strain reared on rice (CR) and R strain reared on rice (RR). Each experiment comprised two independent replicates. To obtain sRNA-Seq from three developmental stages (this study), the corn-strain was reared on the artificial diet “Poitout” <sup>37</sup> under the same controlled conditions.

### 2.2 Identification of *Sf* Argonaute proteins and phylogenetic analysis

To identify Argonaute genes in the genome of *Spodoptera frugiperda* (version 6.0), we used as query the RNA interference genes already annotated in a previous version <sup>38</sup> in a search by TBLASTN in the SfruDB database [http://bipaa.genouest.org/sp/spodoptera\\_frugiperda/](http://bipaa.genouest.org/sp/spodoptera_frugiperda/)

and annotated the corresponding transcripts in Webapollo. We corrected the exon/intron structure based on homology. The Argonaute proteins IDs of all insects used in this study are available in Supplementary Table 1. The evolutionary history was inferred using the Neighbor-Joining method <sup>39</sup>. The evolutionary distances were computed using the Poisson correction method <sup>40</sup>. Evolutionary analyses were conducted in MEGA11 <sup>41</sup>.

### 2.3 Analysis of differential gene expression

To study somatic and germinal expression of Argonaute genes, we re-used RNA-Seq data reported in recent studies <sup>42, 43</sup>. Briefly, for each replicate (three for the fat body, two for the gonads of *S. frugiperda*), total RNA was extracted using the EZNA Total RNA kit I (Omega) from the fat bodies of 25–28 larvae and the testes of five larvae. Library preparation and RNA sequencing (125 bp paired-end reads generated on an Illumina HiSeq 2500 platform) were performed by the Montpellier GenomiX (MGX) sequencing facility (CNRS, Montpellier, France). Raw reads were aligned with the sfC.ver5.fa reference genome sequence for *Spodoptera frugiperda* <sup>44</sup> with STAR v20201 according to <sup>43</sup>. We assessed differences in gene expression between insect tissues (male gonads vs. fat body in fifth instar larvae), using the read count data as input for the R package DESeq2 <sup>45</sup>. DESeq2 uses negative binomial generalized linear models to test for differential expression. An adjusted *p*-value for multiple testing was computed with the Benjamini-Hochberg procedure to control for false discovery rate (FDR). Results with FDR<0.05 were considered statistically significant.

### 2.4 Small RNA extraction, library preparation and sequencing

In a previous paper, we described microRNAs of the corn and rice strain of *Spodoptera frugiperda* reared on plants, corn and rice <sup>28</sup>. In the present paper we re-used the same sRNA-Seq dataset to study piRNAs in SfC and SfR. These datasets had been obtained in the following conditions: Briefly, small RNA was extracted from 12 larvae at the 4<sup>th</sup> instar (two biological replicates) using the mirVana Isolation kit (Ambion). cDNA library preparation was made using the TruSeq <sup>®</sup>Small RNA Sample Prep Kit (Illumina) and sequencing on Illumina HiSeq 2500. A size selection in the range of 15

to 40 nt was performed. Library preparation and sequencing (Single reads, 50 bp) were performed by the platform MGX-Montpellier GenomiX (Montpellier, France). To identify piRNAs candidates involved in sex-determination, additional sRNA-Seq data from three developmental stages (eggs, larvae, pupae) of SfC were obtained, following the same protocol except that the sequencing was performed on the GATC Biotech platform on Illumina HiSeq2000.

### 2.5 Size profiling and mapping of sRNAs to various genomic references

Size profiling of sRNA-Seq reads (Trimmed FastQ files) mapping to the reference genome of SfC (version 5.0) or SfR (version 3.0) was performed using the Galaxy pipeline sRNApipe<sup>46</sup>. Piecharts representing their distribution on various genomic references were obtained with the same pipeline. Briefly sRNApipe aligns sRNA-Seq reads on the reference genome, these reads that align are called “bonafide reads”. The length of these reads are computed and then they are aligned to various genomic references (miRNA precursors, transcripts, TE consensus, rRNA, snoRNA, tRNA) and counted.

### 2.6 Identification of piRNA clusters

Trimmed FastQ files corresponding to smallRNAs extracted from SfC (or from SfR) were combined and filtered to keep only sequences of size range comprised between 24 and 39 nt, expected for piRNAs. These pools were used as input for the software Shortstack version 4.0.3<sup>47, 48, 49</sup> with default settings except -dicermin 24 and -dicermax 39 defining the minimal and maximal lengths of piRNAs. Shortstack is publicly unavailable at

<https://github.com/MikeAxtell/ShortStack/releases> and performs i) alignment of the reads using Bowtie<sup>50</sup> against the reference genome SfC version 5.0 (or SfR, version 3.0, respectively) ii) computation of the depth of small RNA coverage along the reference genomes iii) annotation of piRNAs precursors corresponding to overlapping genomic regions where the depth of piRNA coverage is greater or equal to 1 read per millions (mindepth by default) according to<sup>51</sup>.

<https://github.com/MikeAxtell/ShortStack/blob/master/README.md#usage>

### 2.7 Identification of Long Non-coding RNAs(LNCR)

After trimming, RNA-Seq from hemocytes, fat body<sup>43</sup> and gonads<sup>42</sup> were aligned against the genome reference of *S. frugiperda* version 5.0 using STAR version v2.5.2a. The BAM files were merged using Samtools. The alignment results were assembled using StringTie version 2.1.4 with stringtie\_ignore\_gtf parameter. Identification of candidate lncRNAs from the assembled transcriptome was based on applied filters: exon number  $\geq 2$ , transcript length  $> 200$  bp, using the FEELnc pipeline version 0.2<sup>52</sup>. Prediction of the protein-coding potential was also evaluated using FEELnc using OGS5.0. We could identify a total of 5755 unique sequences of LNCRNAs. The LNCR homolog to TE consensus were identified by Blastn.

### 2.8 Annotation of transposable elements

The annotation of TE in SfC and SfR genome assemblies (v5.0 and v3.0, respectively) was performed using REPET v2.5<sup>53, 54</sup>. Following a Blastn of the genome against itself, the TEdenovo pipeline<sup>53</sup> detects high scoring pairs, which are clustered according to various methods before identification of a consensus sequence representative of each type of repetitive element. For the classification of consensus sequences according to<sup>55</sup>, various features based on homology or structure were taken into account. Consensus sequences were built only if at least 3 similar copies were detected in the genome. In a second step, the TEannot pipeline<sup>54</sup> was used to annotate TE in the genomes. The consensus from TEdenovo were used in a first round of annotation of each genome SfC and SfR independently. Then a

second round of TEannot was performed using a pool of SfC and SfR TE families with full length consensus fragments. This second round of TE annotation allowed identification of TE specific of one strain or the other from their genome coverage.

### *2.9 Mapping of piARNs on TE consensus and gene models of *Spodoptera frugiperda**

Trimmed reads of sRNA-Seq were aligned on different reference sequences using Bowtie1<sup>50</sup> version v0.12.7. For mapping on TE consensus 2 mismatches were allowed –n 2. Mapping with perfect match –n 0 were performed on coding or non coding gene transcripts or genomes. On each reference, mapping was performed with reads mapping uniquely –m 1 or with multimappers. The number of reads mapping to various references was counted using Samtools idxstat version v.0.1.18.

### *2.10 GO term enrichment*

The gene functions and associated Gene Ontology (GO) terms of piRNA sequences were determined by first aligning piRNA sequences to the coding sequences of *Spodoptera frugiperda* genes from the OGS6.1 genome assembly downloaded from the BIPAA platform using Bowtie 1<sup>50</sup>. A strict threshold was applied to identify highly targeted genes ( $\geq 100$  mapped sequences). Functional enrichment analysis was then performed on these target genes using BLAST2GO v6.0.3<sup>56</sup> with default parameters. We identified overrepresented GO terms using Fisher's exact test with a False Discovery Rate threshold  $< 0.1$ .

### *2.11 Identification of ping-pong signatures*

To detect pairs of piRNAs showing a ping pong signature, we used PingPongPro, a script developed at the Institute of Molecular Biology gGmbH in Mainz (Germany). It is freely available at <https://github.com/suhrig/pingpongpro/releases>. This software uses alignment files, either the sequence Alignment Map (SAM) or the Binary-compresses Sequence Alignment/map format (BAM) of small non coding RNAs on the reference genomes to detect stacks on both strands overlapping by 10 nucleotides, starting with a U at their 5' end and with a A at position ten. Then it calculates a score for the validity of the signatures. For the command line tool, we used BAM alignment files.



*Ago1*, *Ago2*, *Ago3* show a 1:1 orthology relationship between *Spodoptera frugiperda* and *Bombyx mori*, the Lepidopteran model, whereas *Spodoptera frugiperda* has three Aubergine genes instead of one in *Bombyx mori*. These three Aubergine genes *Aub1*, *Aub2*, *Aub3* are present in all species of the genus *Spodoptera* studied. In their paper on the role of PIWI in antiviral responses, Xia *et al.*, identified additional copies of Aubergine genes in the genome of an invasive population of *Spodoptera frugiperda*<sup>30,58</sup>. We studied their phylogenetic relationships with the genes identified in the reference genome of native populations (Supplementary Figure 1). They found two proteins clustering with our AUB1 and two others clustering with our AUB3 (only one gene like us for AUB2). We think that these additional copies either evolved during the very recent process of worldwide invasion or that they result from artefactual duplication during the genome assembly process because of high heterozygosity (see Discussion section).

### 3.2 Germinal versus somatic expression of Argonaute proteins

Until 2016, the piRNA pathway was supposed to suppress TE activity in gonads only, when Jones *et al.* showed that it was also functional in somatic tissue *i.e.* the fat body of adult *Drosophila*<sup>10</sup>. The *Piwi* mutant showed depletion of fat body piRNAs and increased TE mobilization in this tissue too. This led us to explore *Ago* gene expression in gonads and fat body in our insect species. For that purpose, we relied on the RNA-Seq analysis described in a recent paper<sup>42</sup>. Briefly, we obtained a mean of 23.9 million reads per replicate (2 for gonads, 3 for fat body), about 18.6 million of which (78%) mapped to unique sites on reference genome version 5.0 of *Spodoptera frugiperda*<sup>42</sup>. The gene expression profiles of the gonads and fat bodies were compared with the DESeq2 software, which can be used to analyze differential gene expression between conditions based on read counts mapping to particular coding sequences, here OGS6.0 from *SfC* genome version 5.0. We performed principal component analysis (PCA; Supplementary Figure 2A) on the various samples (tissues, replicates). The log<sub>2</sub> fold-change for each gene was plotted as a function of mean expression — *i.e.* the mean number of counts normalized by size factors (MA plot; Supplementary Figure 2B). Blue dots correspond to genes displaying significant differential expression between the gonads and the fat body, at a FDR of less than 0.05. The heatmap on Figure 2 shows that genes *Ago1*, 2, 3, and *Aub1*, 2 are expressed in both tissues. *Ago1* and *Ago2* are significantly more expressed in the soma than in gonads (Supplementary Table 2, Tab AGO, DESeq2 results). On the contrary, *Aub1* and *Aub2* are significantly more expressed in gonads than fat body, whereas *Ago3* is expressed to the same level in gonads and fat body. *Aub3* is expressed at a very low level in both tissues.

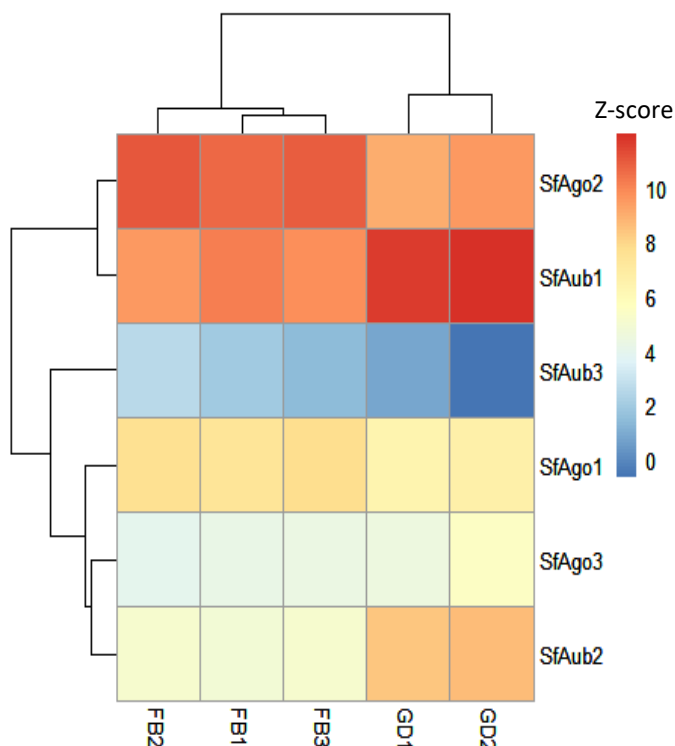


Figure 2. Relative levels of expression in gonadal and somatic tissues of Sf Argonaute genes. The heatmap shows the relative levels of expression of *Ago* and *Piwi* genes in the testis (GD1 & GD2 for gonads) and somatic tissue, the fat body (FB1, FB2, FB3) for each gene, with Z-score normalization.

### 3.3 piRNAs profiling in *Spodoptera frugiperda*

To annotate piRNAs clusters in SfC and SfR genomes, we re-used a dataset of sRNA Seq published in the frame of a previous paper describing microRNAs<sup>28</sup>. Briefly, sRNAs were extracted from 4<sup>th</sup> instar larvae of the lab corn strain fed on corn (CC) or rice plants (CR) as well as larvae of the lab rice strain fed on corn (RC) and rice (RR) - in duplicate. Additionally for the corn strain, sncRNA were extracted from three developmental stages, egg, larvae, pupae fed on artificial diet (CAD), Poitout (this study). The 11 libraries were analyzed by Illumina sequencing as described in<sup>28</sup>. Size profiling of sRNA sequences mapping to the reference genomes (either from SfC or from SfR) and comprised between 18nt to 39nt was performed according to material and methods, the results are shown on Supplementary Figure 3. The piRNAs sequences have a size range between 24 and 39 nucleotides (nt)

and include a major peak between 26-29nt and a minor peak around 33 nt. To identify piRNAs showing a ping-pong signature, we searched for piRNAs mapping to the genome on opposite strands and overlapping on 10 nucleotides using the script Ping-pong pro. We found 12935 loci to which piRNAs with ping-pong signatures mapped, among which 11185 had a score comprised between 0.5 and 1, Supplementary Table 3.

### 3.3 piRNA clusters in SfC and SfR

The pool of SfC piRNAs and of SfR piRNAs of size 24 to 39 nt corresponding to 109,416,742 and 74,370,403 sequences were mapped to the two reference genomes v5.0 and v3.0 for the corn and the rice strain, respectively using the software bowtie1. Sixty-three % of the reads from SfC gave rise to one or more alignment on the reference genome using the software Bowtie. The remaining 36.71% of reads which failed to align may map to the W chromosome which is absent in our genome assemblies built from males (ZZ). piRNAs clusters were detected using the software Shortstack with default parameters (See Material and Methods). As shown on Table 1, 521 and 628 piRNA clusters with more than 1000 reads aligned were detected in SfC and SfR genomes, respectively. The complete list of clusters with their genomic position and the sequence of the the most abundant small RNA expressed are available on Supplementary Table 4. 988 clusters had a size range of 1kb-16kb in the corn strain, 360 clusters had a size range of 1kb-7kb in the rice strain.

Table 1. Distribution of piRNA clusters in the genomes of SfC and SfR

Strain of <i>S. frugiperda</i>	SfC	SfR
Number of clusters with more than 1000 reads aligned	521	628
Number of clusters larger than 1kb	988	360
Clusters on + strand, %	45.8	46.2
Clusters on – strand, %	42.7	50.1
Dual clusters, %	11.5	3.7
Total number of clusters detected	8,889	3,456

### 3.4 Long non-coding RNAs mapping to TE, putative piRNA precursors

Long non-coding RNAs (LNCRNAs) transcribed from piRNA clusters can be precursors of piRNAs. We identified 5755 LNCRNA from contigs assembled from a set of RNA-seq libraries<sup>43, 59</sup> from various tissues of the insect - gonads, hemocytes and fat body- according to Materials and Methods. Among them, 320 show 98% identity over 80% of their length to TE sequences of various orders and can correspond to transcripts of TEs possibly embedded into piRNA clusters (Supplementary Figure 4).

### 3.5 Targets of piRNAs

piRNAs are well known regulators of TE expression, they can also act on gene expression in the germline and in the soma. We first annotated and compared TE content in the two *Sf* strains. Then, we mapped *Sf* piRNAs dataset on TE consensus or gene transcripts, to identify their targets.

#### 3.5.1 Annotation of transposable elements in *SfC* and *SfR* genomes

Using the TE denovo package of the REPET pipeline<sup>53,54</sup>, we identified and classified repeated sequences in the reference genome of of the corn strain *SfC* v5.0<sup>44</sup> and of the rice strain *SfR* v3.0<sup>60</sup> whose assembly quality is equivalent and less fragmented than in our first reference genomes<sup>38</sup> to obtain comparable quality of TE annotation (supplementary Table 5 for statistics of genome assembly). We obtained 2171 repeats consensus in the corn strain and 2032 in the rice strain, with more retrotransposons (959 families in *SfC*, 872 in *SfR*) than DNA Tns (858 families in *SfC*, 821 in *SfR*) in both strains, as shown on Table 2.

Table 2. TE annotation of the two strains genomes *SfC* and *SfR* with the REPET pipeline

Genome assembly :	<i>SfCv5</i> PACBIO Illumina 1000 scaffolds N50=0.9Mb	<i>SfRv3</i> PACBIO Illumina 1054 scaffolds N50= 1.13Mb
Pipeline	REPET v2.5	
	TE DENOVO	
RNA TE	959 (44.17%)	872 (42.91%)
DNA TE	858 (39.52%)	821 (40.40%)
Other repeats	354 (16.31%)	339 (16.69)
Total consensus	2171 (100.00%)	2032 (100.00%)
Pipeline	TE ANNOT	
Cumulative coverage	124313817 bp (113,077,456 bp)	120296323 bp (110,035,603 bp)
Coverage percentage	32.34% (29.42%)	31.67% (28.96%)
Families with full length fragments	1892 (1313)	1830 (1264)

Comparison of TE diversity found in SfC and SfR shows a similar pattern (Figure 3 A and B). The complete list of repeats families and their classification according to <sup>55</sup> are available in Supplementary Table 6. A first round of TE annotation of SfC and SfR genomes using these 2171 and 2032 repeats consensus, respectively, led to a global repeat coverage of 29.42% and 28.96% of the two strains genomes. To optimize TE detection and identify those showing strain specificity, we performed a second round of TE annotation using a pool of the TE families found in the two strains genomes and having full length fragments in the output of the first annotation round (*i.e.* 1313 plus 1264 families). This led to coverages of 32.34% and 31.67% in SfC and SfR respectively (Table 2 , Supplementary Table 7). The coverage of each order of TE in SfC and SfR, displayed on Figure 3C and D, shows similar patterns.

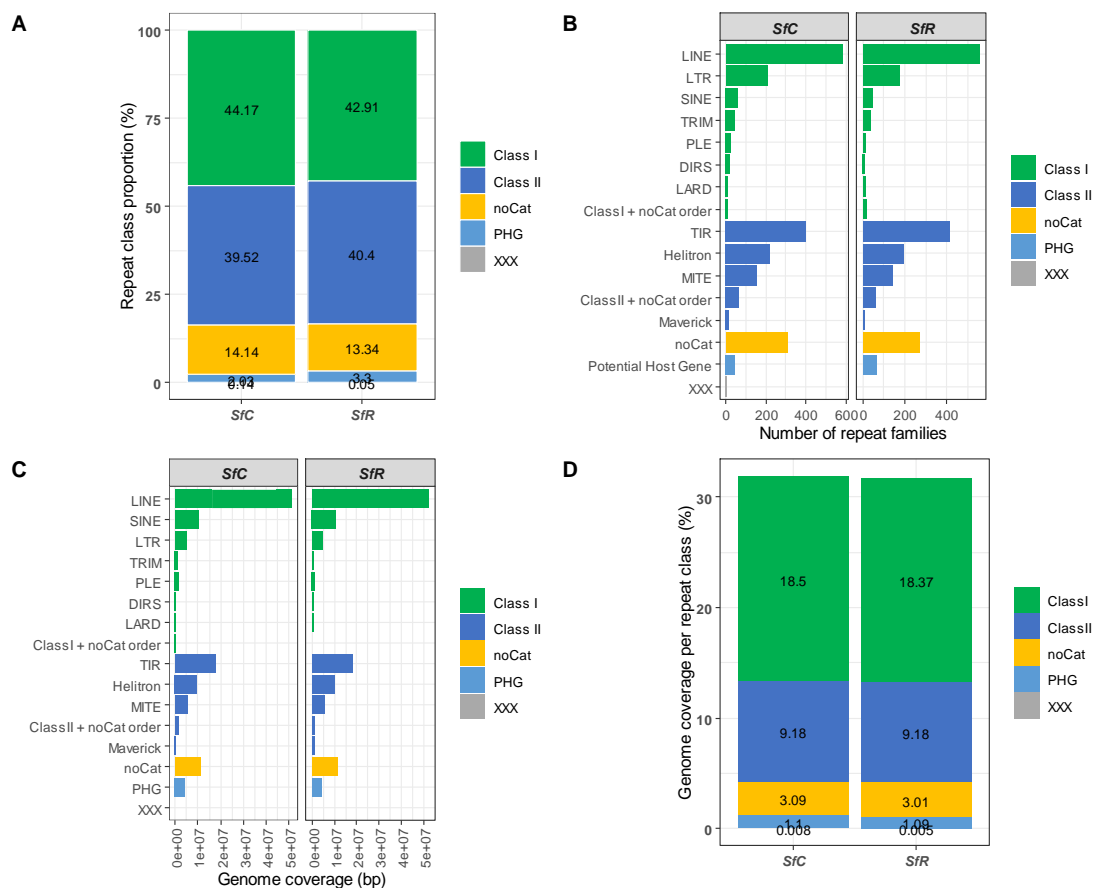


Figure 3. Diversity and coverage of transposable elements in the two strains genomes.

A. Proportion of TE class in SfC and SfR. B. The histograms show the number of transposable elements families in SfC ( on the left, n=2171) and in SfR (on the right, n=2032) belonging to different orders according to <sup>55</sup>. PHG: Potential Host Gene. noCat (repeated sequence not classified at class and order levels). XXX (sequence not classified at class level and with potential several orders) C. The histograms show the percentage of genome covered by different TE orders according to <sup>55</sup>. D. proportion of the genome covered by the different repeats classes.

### 3.5.2 piRNAs regulating TEs

To identify the TE regulated by piRNAs, we searched for piRNAs showing homology to TEs. For that purpose, sequences of size range 24nt to 39nt from SfC (109,416,742 sequences) were pooled as well as those from SfR (74,370,403 sequences) and were aligned using bowtie 1 on TE consensus sequences identified in SfC or SfR, respectively. The percentages of reads uniquely mapped on DNA TE or RT TE were calculated and shown on Figure 4A, the same analysis was performed with piRNA sequences mapping to several locations on consensus sequences (Figure 4B). The detailed counts of piRNAs mapping to the different TE consensus can be found on Supplementary Table 8. In SfC, one of the RT consensus was the most abundantly targeted transposon family, whereas in SfR reads were more evenly distributed between DNA and RT TEs. This pattern was observed both with unique and multiple mappers piRNAs sequences. The differences between histograms highlight variation in the regulation of TEs between the two strains, either because some TE can be specific of one strain or because some may be more active in one strain compared to the other.

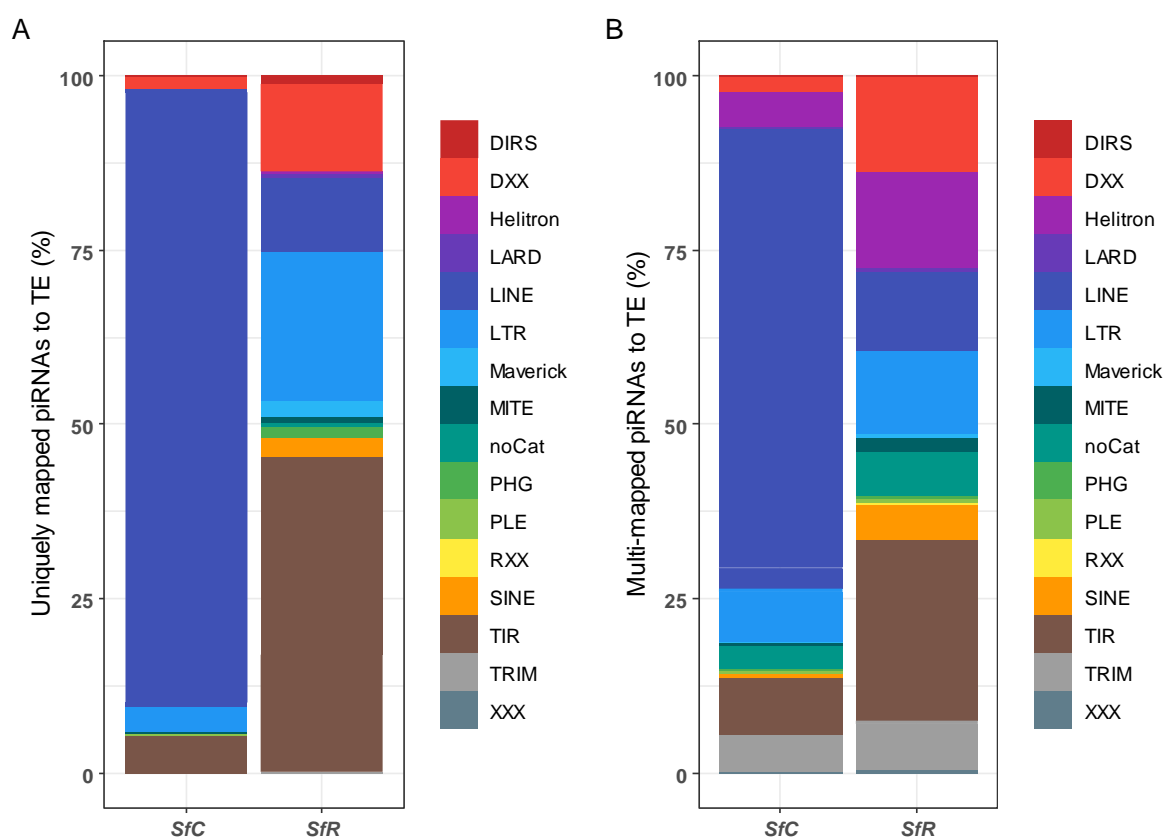


Figure 4. Transposable elements targets of piRNAs.

piRNA were mapped to TE consensus and reads mapping uniquely (A) or at multiple location (B) were counted. The histograms show the percentage of reads mapping to different TE orders in SfC and SfR.

### 3.5.3 piRNAs regulating cellular genes

### 3.5.3.1 Mapping on OGS6.1

PiRNAs were mapped to the coding sequences of *S. frugiperda* genes (OGS6.1). 14577 genes out of 21830 were targeted by one or more piRNAs. 2589 genes out of 21830 were targeted by 100 piRNAs sequences or more (14577 by 1 piRNA or more). We tested whether, among them, the 2059 genes showing gene ontology (GO) annotation were enriched in some specific GO terms compared to the reference set containing 9628 genes. Ninety-five terms showed specific and statistically significant enrichment, among which the top hits were terms related to the biological process translational elongation, formation of cytoplasmic translation initiation complex, fatty acid beta-oxidation and the molecular functions structural constituent of ribosome, ATP binding, proton-transporting ATPase activity, fatty acid binding plus transposase activity (Supplementary table 9, Supplementary Figure 5).

### 3.5.3.2 Attempt to identify a piRNA regulating sex-determination in *Spodoptera frugiperda*.

In *Bombyx mori*, a piRNA is known to regulate sex determination by preventing the protein MASC to induce the splicing pattern of double-sex mRNA specific to the male developmental cascade<sup>35</sup>. It acts by inducing a cleavage of *Masc* mRNA. Regulation of sex determination by a piRNA is not the only mechanism triggering sex determination in Lepidoptera<sup>61</sup>, however involvement of the gene *Masc* seems a general feature in this insect order. In the case of *S. frugiperda*, annotation of piRNA clusters (see previous section) revealed a cluster encompassing the second exon of a gene showing homology to *masculinizer 1* from *Helicoverpa armigera* (Figure 5 A). This piRNA cluster shows a ping-pong signature (Figure 5 B). From the sRNA-Seq data obtained from developmental stages, the candidate for pi-masc is expressed to 28 copies in eggs, 2 in larvae, 0 in pupae whereas the candidate for pi-fem is expressed to 588 copies in eggs, 10 in larvae and 0 in pupae. By mapping of RNA-Seq from testis, we observed two isoforms for *Masc*, one with all 10 exons, one with the two last exons.

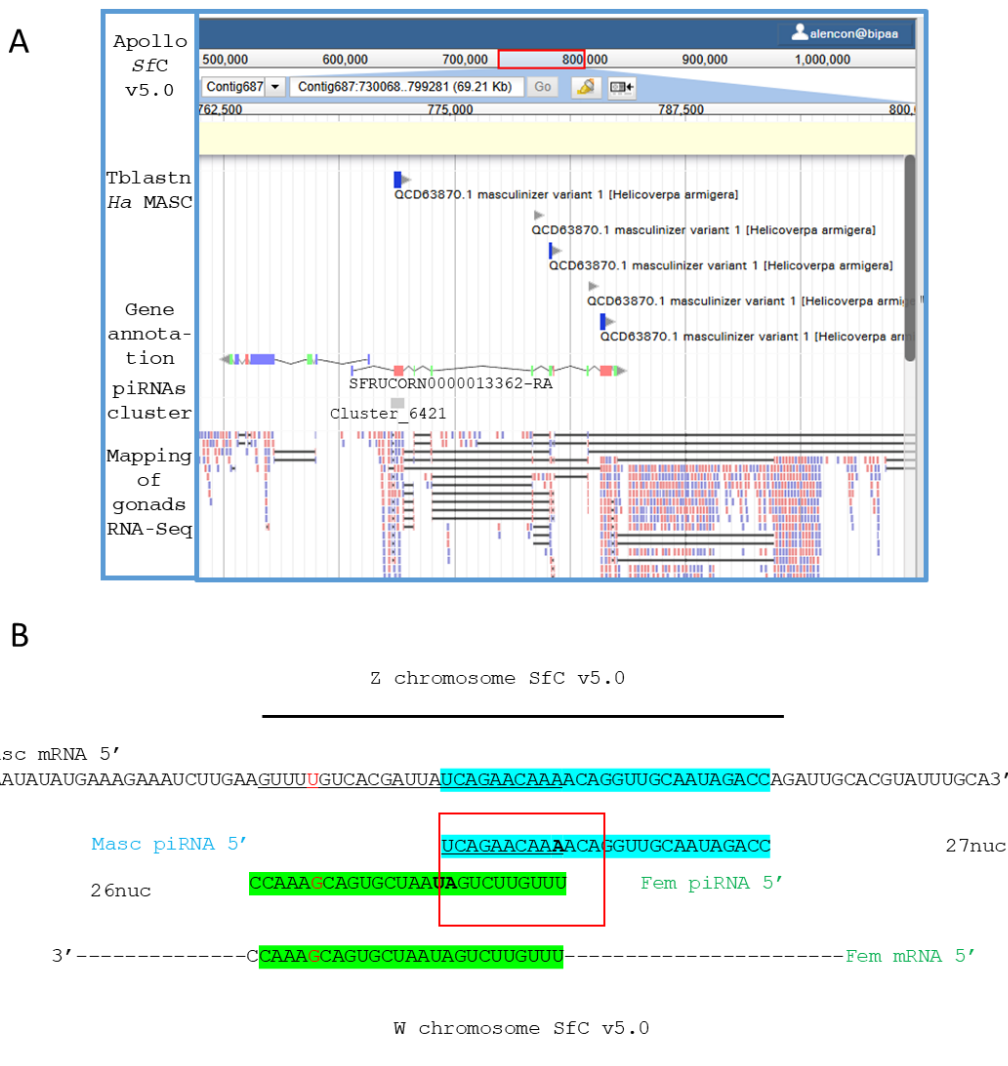


Figure 5. Cluster of piRNAs on *masc* gene

- A. *S. frugiperda* gene models matching to Masculinizer proteins from *Helicoverpa armigera* or *Sf9* cells were identified. A piRNA cluster maps to the second exon of *SfMasc*.
- B. A pair of piRNAs expressed by the cluster shows a ping pong signature.

#### 4. Discussion

In this paper, we described the genes and the genomic loci involved in biosynthesis of piRNAs in *S. frugiperda* as well as the TE or genes regulated by these piRNAs. This led to two main findings i) the discovery of two new *Piwi* genes compared to the Lepidopteran model *Bombyx mori*, these two genes being conserved in the genus *Spodoptera* ii) the fact that the genomes of the two host-plant strains of *S. frugiperda*, despite of a very similar content of transposons diversity and coverage, show different patterns of piRNAs-regulated TE, suggesting a difference in transposons activity.

The phylogenetic analysis of insect Argonaute proteins showed a 1:1 orthology relationship for *Ago1*, *Ago2* and *Ago3* between *Spodoptera* species and *Bombyx mori*, but the existence of three *Aub* genes in the genus *Spodoptera* compared to one in *Bombyx mori* and two in *Drosophila*. In their study focused on the role of PIWI proteins in response to baculovirus infection<sup>30</sup>, Xia *et al.* in 2023 found five paralogs of *Aub* in the genome assembly of an invasive population of *Spodoptera frugiperda*<sup>62</sup>. By phylogenetic analysis, we show here that two of these paralogs cluster with our AUB2 and AUB3 and that they probably result from artefactual duplication during the genome assembly process due to high heterozygosity of invasive population, since the genome size 486Mb is larger than expected<sup>62</sup>. Our genome assemblies were made from lab colonies of SfC and SfR strains originating from native area, they have a size close to expected by flow cytometry (for SfC, 1C = 394.4 +/- 0.2 Mb (N=2), Spencer Johnston, personal communication). We conclude that *Spodoptera frugiperda* has three *Aub* genes rather than five. Gene copy-number variation can occur by tandem duplications, retro-position or unequal crossing over. In our chromosomal genome assembly v7.0<sup>63</sup>, *Aub1* is located on chromosome 11 like *Ago1* while *Aub2* and *Aub3* are located on chromosome 10, like *Ago2*, at a distance between 0.5 and 2Mb from each other, it is thus unlikely that they have arisen by tandem duplication.

In *Bombyx mori*, the two *Piwi* genes, *Ago3* and *Siwi* are highly expressed in gonads and their mRNAs can be detected in other somatic tissues in larvae or pupae<sup>64 65 66 67</sup>. By RNA-seq analysis, we showed that on the contrary *Sf Ago3* is expressed at similar level in gonads during spermatogenesis and in a somatic tissue, the fat body. However, like in *Bombyx*, *Sf Aub1* (ortholog of *Bombyx Siwi*) and *Aub2* mRNAs are more abundant in gonads compared to fat body where they can be readily detected. *Aub3* is expressed at very low level in these two tissues.

This overexpression of the new gene *Aub2* in gonads is in accordance with a function in the control of TEs in the germline - the primary role of piRNA pathway. In *Drosophila*, there are two *Aub* genes, one ortholog to *Siwi* (*SfAub1*) encodes AUB, a cytoplasmic protein involved in post-transcriptional silencing of TE, the other encodes PIWI, a nuclear protein responsible for transcriptional silencing of TE. *Bombyx mori* lacks functional homolog of PIWI<sup>64</sup> and thus could lack transcriptional regulation of TE. However<sup>68</sup> interactions between AGO3/SIWI and the *Bombyx* proteins HP1a and HP1b have been shown suggesting that *Bombyx* PIWI proteins could play the role of transcriptional regulators. *Sf AUB2* could be a functional homolog of *Drosophila* PIWI. Alternatively, AUB2 could replace AGO3 and compensate for its low expression in gonads. We are currently characterizing a KO mutant of this gene to assess its function (Gasser *et al.*, in prep.).

The very low expression of *Aub3* in gonads suggests that it evolved a new function. As reviewed in<sup>69</sup>, diversification of *Piwi* genes in insects led to diversification of the functions of PIWI proteins. For

instance, in the pea aphid<sup>18</sup>, the eight *Piwi* genes have evolved specialization of their expression pattern, two of them, *Piwi2* and 6 are expressed in the germline, two others, *Piwi5* and 3 in the soma, *Piwi5* is more expressed in males, *Piwi6* and *Ago3* are expressed in sexuparae. In mosquitoes, the main role of PIWI proteins and piRNAs is the control of viral infections in different somatic tissues and cell lines ( <sup>69</sup> for review). In Lepidopteran cell cultures from *B. mori* and *Trichoplusia. ni* Santos *et al.*, showed by RNA interference that PIWI proteins are involved in antiviral immunity<sup>70</sup>. In the study that we mentioned in the beginning of the discussion section<sup>30</sup>, Xia *et al.* showed using dsRNA silencing in *Sf9* cells that their Piwi-like-4-5 protein (Our AUB3) significantly promotes the Baculovirus Ac MNPV replication and thus acts as a proviral factor like SIWI and AGO3 in *Bombyx mori*. On the contrary their Piwi-like-1 (Our AUB2) and their Piwi-like-2-3 (Our AUB1) have an antiviral activity against Ac MNPV. In their study these functions related to viral infection seem to be independent on viral piRNAs, they involve only PIWI proteins. This study does not provide any data on the *Piwi* gene expression level in *Sf9* cells.

Biogenesis of piRNAs needs transcription of genomic regions enriched in remnants of transposable elements by RNA-PolIII in *Drosophila* recruited by the HP1 homolog Rhino<sup>71</sup>. The processing of these transcripts to mature piRNAs involves various enzymes PIWI, Zucchini and a 3' to 5' trimming activity (details reviewed in<sup>72</sup>). To annotate these piRNA clusters in the two *S. frugiperda* genomes, we relied on sRNA-Seq data from 11 libraries<sup>28, this study</sup>. The piRNA size profiling highlighted in all libraries, a major peak in the size range of 26-29 nt and a secondary peak of 32-33 nt (Supplementary Figure 3). This secondary peak can correspond to an intermediate of biosynthesis or to mature piRNAs produced by the sur-numerary PIWI proteins found in *S. frugiperda* like in *Drosophila* where each PIWI family member was shown to bind different populations of small RNAs<sup>21</sup>. Since we do not have antibodies against *Sf* PIWI proteins, we could not purify piRNAs further. However from the pie-chart representing the proportion of small RNAs mapping to various genomic references (Supplementary Figure 4), there was a low proportion of sRNAs mapping to rDNA whereas the vast majority of cellular RNA consists of ribosomal RNA (rRNA) - ~ 80 to 90% of total RNA for most cells<sup>73</sup>, showing that contamination of our samples with sRNA sequences resulting from degradation of large RNAs was minimal. We found 500-600 clusters with more than 1000 piRNA reads aligned and between 400-1000 clusters larger than 1 kb scattered on most of the genomic scaffolds in *SfC* and *SfR*, respectively. In *Bombyx mori* genome, depending on the methods and sRNA-Seq libraries used to annotate them, between 560 and 965 clusters were described<sup>74, 75</sup>, which corresponds to a similar magnitude order. This is more than in *Drosophila* where 142 genomic locations were identified as sites of abundant piRNA generation<sup>21</sup>. We found both unidirectional and bi-directional piRNA clusters in *S. frugiperda*. In *Drosophila* piRNA clusters transcribed on both DNA strands are expressed in the germline<sup>21</sup>. Because TEs in piRNA clusters can accumulate mutations that negate their transposition activity, we searched for long non-coding RNAs homolog to TE from a dataset of RNA-Seq<sup>43,59</sup>. We identified 325 such transcripts as putative piRNA precursors. The epigenetic marks associated to piRNA clusters vary : In *Drosophila*, piRNA clusters are embedded in heterochromatin regions<sup>71</sup>. In *Bombyx*, they are mostly found in euchromatin except telomeric ones which are heterochromatic<sup>75</sup>. Accordingly, *Drosophila* Piwi interacts with HP1<sup>76</sup>, whereas *Bombyx* Siwi interacts both with a Histone H3K4 methyl transferase<sup>77</sup> and with HP1<sup>68</sup>.

We found 32.34% and 31.67% of the genomes of the two strains *SfC* and *SfR* corresponding to repeated sequences, slightly more than in the first genome paper<sup>27</sup> (29.16% - 29.10%, respectively) reflecting the better contiguity of our genome assemblies. The two strains share a similar pattern of TE diversity, with a predominance of Non-LTR retrotransposons and SINES, like in *B. mori*<sup>78</sup>. However, we found variation between the two strains when we compared the patterns of TE targeted by piRNAs showing that active transposons differ between them. Initially described as host-plant strains due to their

association to different plant ranges<sup>79</sup>, the evolutionary status of the two strains has recently been clarified by population genomics analysis: the whole genome sequences of larvae collected on corn or rice cluster in two genetically distinct groups<sup>26</sup> with very few hybrids. They are in the process of incipient speciation and display traits acting as reproductive barriers, reviewed in<sup>80</sup>. In *Drosophila melanogaster*, there are strains called “P strains” with the P transposable element and some which do not contain it called “M strains”<sup>81</sup>. In P strains, P element-derived piRNAs are transmitted from the mother through the oocyte cytoplasm. The transmitted piRNAs then silence the P element. If no piRNA is transmitted from the mother, the P element is not repressed. Consequently, a P male crossed with an M female will have a dysgenic offspring, with abnormally small gonads, increased mutation rates, frequent sterility. This phenomenon, resulting from the fact that the offspring have the P element but no silencing through maternal piRNA, is known as “hybrid dysgenesis”<sup>82</sup>. Similarly in *S. frugiperda* TE families active in one strain and not in the other could impair the fitness of hybrids and reinforce reproductive isolation. We will test this hypothesis in future.

In addition to TE, we identified 2589 genes which were targeted by at least 100 piRNA sequences, likely regulated by piRNAs and PIWI proteins. A GO term enrichment analysis of them showed that among the most enriched biological processes figure translation elongation, macromolecule catabolic process, protein phosphorylation, regulation of mRNA stability, transposition. The fact that among regulated genes, some are related to transposition was expected since our gene annotation was performed without hiding repeated sequences and contains some TE genes. Among the most enriched cellular components terms, the eukaryotic translation initiation factor 3 complex was found as well as eukaryotic 43S and 48S preinitiation complexes. PIWI proteins are known to be involved in translation initiation in germ cells in *Drosophila*<sup>83</sup> as well as in mouse during spermatogenesis<sup>84</sup>, suggesting that this role of PIWI-piRNAs in translation initiation is conserved in *S. frugiperda*. One of the first genes shown to be regulated by a piRNA is the sex determining gene *Masc* in *Bombyx mori*<sup>35</sup>. Since that time, alternative mechanisms to trigger the sex determining cascade have been discovered in Lepidoptera (reviewed in<sup>61</sup>). In the frame of the annotation of piRNA clusters in *S. frugiperda*, we discovered a piRNA cluster in the second exon of *Masc*, coding for a pair of piRNAs showing the ping-pong signature. We are limited at present to study more deeply these piRNAs as candidates for sex-determining gene regulation because the W is absent in our genome assemblies and we could not identify a *fem* precursor transcript in the transcriptomic resources available today. However, because sex determining piRNAs are species specific and small size RNA sequences, they could be of interest in the frame of development of new biological control means for this insect pest populations.

**Acknowledgments:**

We thank Johann Confais at URGI INRAE Versailles for advises with the REPET pipeline. This work was supported by two grants from the Plant Health and Environment Division of INRAE: a “HOLOCEN” grant awarded to Emmanuelle d’Alençon and a “CENTROVIR” grant awarded to Isabelle Darboux. Mélanie Gasser was awarded a post-doctoral fellowship through the ANR-DFG project ORIGINS, ANR-20-CE92-0018-01.

**Data availability:**

SfC v5.0 genome assembly is available at ENA as PRJEB96863 (ERP179489). Its is also available as well as v3.0, the genome assembly of SfR, with gff files corresponding to TE and piRNA clusters annotation on the bioinformatic platform BIPAA at, respectively :  
[https://bipaa.genouest.org/sp/spodoptera\\_frugiperda\\_corn/](https://bipaa.genouest.org/sp/spodoptera_frugiperda_corn/)

[https://bipaa.genouest.org/sp/spodoptera\\_frugiperda\\_rice/](https://bipaa.genouest.org/sp/spodoptera_frugiperda_rice/)

.

## Supplementary materiel

Supplementary Figure 1. Phylogeny and corresponding IDs of the piwi genes identified in native populations of SfC (this study) and those identified in invasive populations in China by Xia et al., 2023

Supplementary Figure 2. PCA and MA plot obtained during the RNA-Seq analysis published in our previous paper (Gimenez et al., 2025) and re-used in this study

Supplementary Figure 3. Size profiling and pie-chart representing the proportion of sRNA sequences mapping to different genomic references.

Supplementary Figure 4. LNCRNAs mapping to transposable elements and DNA repeats in SfC

Supplementary Figure 5. GO term enrichment analysis of the coding genes regulated by piRNAs

Supplementary Table 1. Accession numbers of the proteins used for the phylogenetic analysis of Argonaute proteins in insects.

Supplementary Table 2. Complete DESEQ results from the RNA-Seq analysis published in<sup>59</sup>

Supplementary Table 3. PiRNAs with ping-pong signatures

Supplementary Table 4. Annotation of piRNA clusters in SfC and in SfR

Supplementary Table 5. Statistics of the genome assemblies used in this paper

Supplementary Table 6. Identification of TE consensus using the pipeline REPET TE denovo

Supplementary Table 7. Results of annotation of TE using the pipeline REPET TE annot

Supplementary Table 8. TE regulated by piRNAs

Supplementary Table 9. GO term enrichment analysis of genes regulated by piRNAs

## References

1. Aravin, A.A., Sachidanandam, R., Girard, A., Fejes-Toth, K., and Hannon, G.J. (2007). Developmentally regulated piRNA clusters implicate MILI in transposon control. *Science* *316*, 744–747.
2. Ghildiyal, M., and Zamore, P.D. (2009). Small silencing RNAs: an expanding universe. *Nat Rev Genet* *10*, 94–108. 10.1038/nrg2504.
3. Siomi, M.C., Sato, K., Pezic, D., and Aravin, A.A. (2011). PIWI-interacting small RNAs: the vanguard of genome defence. *Nat Rev Mol Cell Bio* *12*, 246–258. 10.1038/nrm3089.
4. Iwasaki, Y.W., Siomi, M.C., and Siomi, H. (2015). PIWI-Interacting RNA: Its Biogenesis and Functions. *Annu Rev Biochem* *84*, 405–433. 10.1146/annurev-biochem-060614-034258.
5. Lin, H., and Spradling, A.C. (1997). A novel group of pumilio mutations affects the asymmetric division of germline stem cells in the *Drosophila* ovary. *Development* *124*, 2463–2476. 10.1242/dev.124.12.2463.
6. Cox, D.N., Chao, A., Baker, J., Chang, L., Qiao, D., and Lin, H.F. (1998). A novel class of evolutionarily conserved genes defined by piwi  
are essential for stem cell self-renewal. *Genes & Development* *12*, 3715–3727. DOI 10.1101/gad.12.23.3715.
7. Aravin, A.A., Naumova, N.M., Tulin, A.V., Vagin, V.V., Rozovsky, Y.M., and Gvozdev, V.A. (2001). Double-stranded RNA-mediated silencing of genomic tandem repeats and transposable elements in the  
germline. *Current Biology* *11*, 1017–1027. Doi 10.1016/S0960-9822(01)00299-8.
8. Simonelig, M. (2014). piRNAs, master regulators of gene expression. *Cell Research* *24*, 779–780. 10.1038/cr.2014.78.
9. Ramat, A., and Simonelig, M. (2021). Functions of PIWI Proteins in Gene Regulation: New Arrows Added to the piRNA Quiver. *Trends in Genetics* *37*, 188–200. 10.1016/j.tig.2020.08.011.
10. Jones, B.C., Wood, J.G., Chang, C.Y., Tam, A.D., Franklin, M.J., Siegel, E.R., and Helfand, S.L. (2016). A somatic piRNA pathway in the  
fat body ensures metabolic homeostasis and normal lifespan. *Nature Communications* *7*. ARTN 13856  
10.1038/ncomms13856.
11. Ghildiyal, M., Seitz, H., Horwich, M.D., Li, C., Du, T., Lee, S., Xu, J., Kittler, E.L.W., Zapp, M.L., Weng, Z., and Zamore, P.D. (2008). Endogenous siRNAs derived from transposons and mRNAs in *Drosophila* somatic cells. *Science* *320*, 1077–1081. 10.1126/science.1157396.
12. Morazzani, E.M., Wiley, M.R., Murreddu, M.G., Adelman, Z.N., and Myles, K.M. (2012). Production of Virus-Derived Ping-Pong-Dependent piRNA-like Small RNAs in the Mosquito Soma. *Plos Pathogens* *8*. ARTN e1002470  
10.1371/journal.ppat.1002470.
13. Miesen, P., Girardi, E., and van Rij, R.P. (2015). Distinct sets of PIWI proteins produce arbovirus and transposon-derived piRNAs in  
mosquito cells. *Nucleic Acids Research* *43*, 6545–6556. 10.1093/nar/gkv590.
14. Lewis, S.H., Quarles, K.A., Yang, Y.J., Tanguy, M., Frézal, L., Smith, S.A., Sharma, P.P., Cordaux, R., Gilbert, C., Giraud, I., et al. (2018). Pan-arthropod analysis reveals somatic piRNAs as an

- ancestral defence against transposable elements. *Nature ecology & evolution* 2, 174–181. 10.1038/s41559-017-0403-4.
15. Tolia, N.H., and Joshua-Tor, L. (2007). Slicer and the Argonautes. *Nat Chem Biol* 3, 36–43. 10.1038/nchembio848.
  16. Okamura, K., Ishizuka, A., Siomi, H., and Siomi, M.C. (2004). Distinct roles for Argonaute proteins in small RNA-directed RNA cleavage pathways. *Genes Dev* 18, 1655–1666. 10.1101/gad.1210204.
  17. Wynant, N., Santos, D., and Broeck, J.V. (2017). The evolution of animal Argonautes: evidence for the absence of antiviral AGO Argonautes in vertebrates. *Scientific Reports* 7. ARTN 9230  
10.1038/s41598-017-08043-5.
  18. Lu, H.-L., Tanguy, S., Rispe, C., Gauthier, J.-P., Walsh, T., Gordon, K., Edwards, O., Tagu, D., Chang, C.-c., and Jaubert-Possamai, S. (2011). Expansion of genes encoding piRNA-associated argonaute proteins in the pea aphid: diversification of expression profiles in different plastic morphs. *PLoS one* 6, e28051. 10.1371/journal.pone.0028051.
  19. Fontenla, S., Rinaldi, G., and Tort, J.F. (2021). Lost and Found: Piwi and Argonaute Pathways in Flatworms. *Front Cell Infect Mi* 11. ARTN 653695  
10.3389/fcimb.2021.653695.
  20. Han, B.W., Wang, W., Li, C.J., Weng, Z.P., and Zamore, P.D. (2015). piRNA-guided transposon cleavage initiates Zucchini-dependent, phased piRNA production. *Science* 348, 817–821. 10.1126/science.aaa1264.
  21. Brennecke, J., Aravin, A.A., Stark, A., Dus, M., Kellis, M., Sachidanandam, R., and Hannon, G.J. (2007). Discrete small RNA-generating loci as master regulators of transposon activity in *Drosophila*. *Cell* 128, 1089–1103.
  22. Gunawardane, L.S., Saito, K., Nishida, K.M., Miyoshi, K., Kawamura, Y., Nagami, T., Siomi, H., and Siomi, M.C. (2007). A slicer-mediated mechanism for repeat-associated siRNA 5' end formation in *Drosophila*. *Science* 315, 1587–1590.
  23. Vagin, V.V., Sigova, A., Li, C., Seitz, H., Gvozdev, V., and Zamore, P.D. (2006). A distinct small RNA pathway silences selfish genetic elements in the germline. *Science* 313, 320–324.
  24. Saito, K., Nishida, K.M., Mori, T., Kawamura, Y., Miyoshi, K., Nagami, T., Siomi, H., and Siomi, M.C. (2006). Specific association of Piwi with rasiRNAs derived from retrotransposon and heterochromatic regions in the *Drosophila* genome. *Genes Dev* 20, 2214–2222.
  25. Pashley, D.P., and Martin, J.A. (1987). Reproductive incompatibility between host strains of the fall armyworm (Lepidoptera: Noctuidae). *Annals of the Entomological Society of America* 80, 731–733.
  26. Fiteni, E., Durand, K., Gimenez, S., Meagher Jr., R., L., Legeai, F., Kergoat, G., J., Nègre, N., d'Alençon, E., and Nam, K.W. (2022). Host-plant adaptation as a driver of incipient speciation in the fall armyworm ( *Spodoptera frugiperda* ). *Bmc Ecol Evol* 22, 133. 10.1101/2022.09.30.510290.
  27. Gouin, A., Bretaudeau, A., Nam, K., Gimenez, S., Aury, J.-M., Duvic, B., Hilliou, F., Durand, N., Montagné, N., Darboux, I., et al. (2017). Two genomes of highly polyphagous lepidopteran pests (*Spodoptera frugiperda*, Noctuidae) with different host-plant ranges. *Scientific Reports* 7, 1–12. 10.1038/s41598-017-10461-4.
  28. Mone, Y., Nhim, S., Gimenez, S., Legeai, F., Seninet, I., Parrinello, H., Negre, N., and d'Alençon, E. (2018). Characterization and expression profiling of microRNAs in response to plant feeding in two host-plant strains of the lepidopteran pest *Spodoptera frugiperda*. *BMC Genomics* 19, 804. 10.1186/s12864-018-5119-6.
  29. Gimenez, S., Seninet, I., Orsucci, M., Audiot, P., Nègre, N., Nam, K., Streiff, R., and d'Alençon, E. (2021). Integrated miRNA and transcriptome profiling to explore the molecular

- determinism of convergent adaptation to corn in two lepidopteran pests of agriculture. *BMC Genomics* 22, 606. 10.1186/s12864-021-07905-7.
30. Xia, J.M., Fei, S.G., Wu, H.Y., Yang, Y.F., Yu, W.S., Zhang, M.M., Guo, Y.Y., Swevers, L., Sun, J.C., and Feng, M. (2023). The piRNA pathway is required for nucleopolyhedrovirus replication in Lepidoptera. *Insect Science* 30, 1378–1392. 10.1111/1744-7917.13160.
  31. ter Horst, A.M., Nigg, J.C., Dekker, F.M., and Falk, B.W. (2019). Endogenous Viral Elements Are Widespread in Arthropod Genomes and Commonly Give Rise to PIWI-Interacting RNAs. *Journal of Virology* 93. ARTN e02124-18. 10.1128/JVI.02124-18.
  32. Kawaoka, S., Hayashi, N., Suzuki, Y., Abe, H., Sugano, S., Tomari, Y., Shimada, T., and Katsuma, S. (2009). The Bombyx ovary-derived cell line endogenously expresses PIWI/PIWI-interacting RNA complexes. *Rna* 15, 1258–1264.
  33. Xiol, J., Spinelli, P., Laussmann, M.A., Homolka, D., Yang, Z., Cora, E., Couté, Y., Conn, S., Kadlec, J., Sachidanandam, R., et al. (2014). RNA Clamping by Vasa Assembles a piRNA Amplifier Complex on Transposon Transcripts. *Cell* 157, 1698–1711. 10.1016/j.cell.2014.05.018.
  34. Nishida, K.M., Iwasaki, Y.W., Murota, Y., Nagao, A., Mannen, T., Kato, Y., Siomi, H., and Siomi, M.C. (2015). Respective Functions of Two Distinct Siwi Complexes Assembled during PIWI-Interacting RNA Biogenesis in Germ Cells. *Cell reports* 10, 193–203. 10.1016/j.celrep.2014.12.013.
  35. Kiuchi, T., Koga, H., Kawamoto, M., Shoji, K., Sakai, H., Arai, Y., Ishihara, G., Kawaoka, S., Sugano, S., Shimada, T., et al. (2014). A single female-specific piRNA is the primary determiner of sex in the silkworm. *Nature* 509, 633–+. 10.1038/nature13315.
  36. Moné, Y., Nhim, S., Gimenez, S., Legeai, F., Séninet, I., Parrinello, H., NEGRE, N., and D'Alençon, E. (2018). Characterization and expression profiling of microRNAs in response to plant feeding in two host-plant strains of the lepidopteran pest *Spodoptera frugiperda*. *BMC Genomics* 19, 1–15. 10.1186/s12864-018-5119-6.
  37. Poitout, S., and Bues, R. (1974). [Linolenic acid requirements of lepidoptera Noctuidae Quadrifinae Plusiinae: *Chrysodeixis chalcites* Esp., *Autographa gamma* L.' *Macdunnoughia confusa* Sth., *Trichoplusia ni* Hbn. reared on artificial diets]. *Ann Nutr Aliment* 28, 173–187.
  38. Gouin, A., Bretaudeau, A., Nam, K., Gimenez, S., Aury, J.-M., Duvic, B., Hilliou, F., Durand, N., Montagné, N., Darboux, I., et al. (2017). Two genomes of highly polyphagous lepidopteran pests (*Spodoptera frugiperda*, Noctuidae) with different host-plant ranges. *Scientific Reports* 7, 1–12. 10.1038/s41598-017-10461-4.
  39. Saitou, N., and Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* 4, 406–425.
  40. Zuckerkandl, E.a.P., L. (1965). Evolutionary divergence and convergence in proteins. In *Evolving Genes and Proteins* V.B.a.H.J. Vogel, ed. (Academic Press), pp. 97–166.
  41. Tamura, K., Stecher, G., and Kumar, S. (2021). MEGA11 Molecular Evolutionary Genetics Analysis Version 11. *Molecular Biology and Evolution* 38, 3022–3027. 10.1093/molbev/msab120.
  42. Gimenez, S., Eychenne, M., Legeai, F., Gamble, S., and d'Alençon, E. (2025). Towards identification of a holocentromere marker in the lepidopteran model *Spodoptera frugiperda*. *Chromosoma* 134, 2. 10.1007/s00412-025-00828-2.
  43. Robin, S., Legeai, F., Jouan, V., Ogliaastro, M., and Darboux, I. (2023). Genome-wide identification of lncRNAs associated with viral infection in *Spodoptera frugiperda*. *Journal of General Virology* 104. ARTN 001827. 10.1099/jgv.0.001827.
  44. Nam Kiwoong, N.S., Robin Stéphanie, Bretaudeau Anthony, Nègre Nicolas, d'Alençon Emmanuelle (2018). Divergent selection causes whole genome

differentiation without physical linkage among the targets in *Spodoptera*

*frugiperda* (Noctuidae). bioRxiv. <https://doi.org/10.1101/452870>.

45. Anders, S., and Huber, W. (2010). Differential expression analysis for sequence count data. *Genome biology* *11*, R106. 10.1186/gb-2010-11-10-r106.
46. Pogorelcnik, R., Vaury, C., Pouchin, P., Jensen, S., and Brasset, E. (2018). sRNAPipe: a Galaxy-based pipeline for bioinformatic in-depth exploration of small RNAseq data. *Mobile DNA* *9*. ARTN 25  
10.1186/s13100-018-0130-7.
47. Axtell, M.J. (2013). ShortStack: comprehensive annotation and quantification of small RNA genes. *RNA* *19*, 740–751. 10.1261/rna.035279.112.
48. Johnson, N.R., Yeoh, J.M., Coruh, C., and Axtell, M.J. (2016). Improved Placement of Multi-mapping Small RNAs. *G3-Genes Genom Genet* *6*, 2103–2111. 10.1534/g3.116.030452.
49. Shahid, S., and Axtell, M.J. (2014). Identification and annotation of small RNA genes using ShortStack. *Methods* *67*, 20–27. 10.1016/j.ymeth.2013.10.004.
50. Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* *10*, R25. 10.1186/gb-2009-10-3-r25.
51. Johnson, N.R., Yeoh, J.M., Coruh, C., and Axtell, M.J. (2016). Improved Placement of Multi-mapping Small RNAs. *G3 (Bethesda)* *6*, 2103–2111. 10.1534/g3.116.030452.
52. Wucher, V., Legeai, F., Hedan, B., Rizk, G., Lagoutte, L., Leeb, T., Jagannathan, V., Cadieu, E., David, A., Lohi, H., et al. (2017). FEELnc: a tool for long non-coding RNA annotation and its application to the dog transcriptome. *Nucleic Acids Res* *45*, e57. 10.1093/nar/gkw1306.
53. Flutre, T., Duprat, E., Feuillet, C., and Quesneville, H. (2011). Considering transposable element diversification in de novo annotation approaches. *PLoS One* *6*, e16526, 2011. 10.1371/journal.pone.0016526.
54. Quesneville, H., Bergman, C.M., Andrieu, O., Autard, D., Nouaud, D., Ashburner, M., and Anxolabehere, D. (2005). Combined evidence annotation of transposable elements in genome sequences. *PLoS Comput Biol* *1*, 166–175.
55. Wicker, T., Sabot, F., Hua-Van, A., Bennetzen, J.L., Capy, P., Chalhoub, B., Flavell, A., Leroy, P., Morgante, M., Panaud, O., et al. (2007). A unified classification system for eukaryotic transposable elements. *Nat Rev Genet* *8*, 973–982.
56. Conesa, A., and Gotz, S. (2008). Blast2GO: A comprehensive suite for functional analysis in plant genomics. *Int J Plant Genomics* *2008*, 619832. 10.1155/2008/619832.
57. Felsenstein, J. (1985). Confidence limits on phylogenies: An approach using the bootstrap. *Evolution*

39

783–791.

58. Xiao, H., Ye, X., Xu, H., Mei, Y., Yang, Y., Chen, X., Yang, Y., Liu, T., Yu, Y., Yang, W., et al. (2020). The genetic adaptations of fall armyworm *Spodoptera frugiperda* facilitated its rapid global dispersal and invasion. *Mol Ecol Resour* *20*, 1050–1068. 10.1111/1755-0998.13182.
59. Gimenez, S., Eychenne, M., Legeai, F., Gamble, S., and d'Alençon, E. (2025). Towards identification of a holocentromere marker in the lepidopteran model. *Chromosoma* *134*. ARTN 2

10.1007/s00412-025-00828-2.

60. Nam, K., Nhim, S., Robin, S., Bretaudeau, A., Nègre, N., and D'Alençon, E. (2020). Positive selection alone is sufficient for whole genome differentiation at the early stage of speciation process in the fall armyworm. *Bmc Evolutionary Biology* *20*. ARTN 152

10.1186/s12862-020-01715-3.

61. van't Hof, A.E., Yoshido, A., and Marec, F. (2025). Sex determination in moths and butterflies: Masculinizer as key player. *Current Opinion in Insect Science* 70. ARTN 101375  
10.1016/j.cois.2025.101375.
62. Xiao, H.M., Ye, X.H., Xu, H.X., Mei, Y., Yang, Y., Chen, X., Yang, Y.J., Liu, T., Yu, Y.Y., Yang, W.F., et al. (2020). The genetic adaptations of fall armyworm facilitated its rapid global dispersal and invasion. *Molecular Ecology Resources* 20, 1050–1068. 10.1111/1755-0998.13182.
63. Yainna, S., Tay, W.T., Durand, K., Fiteni, E., Hilliou, F., Legeai, F., Clamens, A.-L., Gimenez, S., Asokan, R., Kalleshwaraswamy, C., M., et al. (2022). The evolutionary process of invasion in the fall armyworm (*Spodoptera frugiperda*). *Scientific Reports* 12, 21063. 10.1038/s41598-022-25529-z.
64. Kawaoka, S., Minami, K., Katsuma, S., Mita, K., and Shimada, T. (2008). Developmentally synchronized expression of two Piwi subfamily genes, and in germ-line cells. *Biochem Bioph Res Co* 367, 755–760. 10.1016/j.bbrc.2008.01.013.
65. Tatsuke, T., Sakashita, K., Masaki, Y., Lee, J.M., Kawaguchi, Y., and Kusakabe, T. (2010). The telomere-specific non-LTR retrotransposons SART1 and TRAS1 are suppressed by Piwi subfamily proteins in the silkworm, *Bombyx mori*. *Cellular & Molecular Biology Letters* 15, 118–133. 10.2478/s11658-009-0038-9.
66. Li, Z.Q., You, L., Yan, D., James, A.A., Huang, Y.P., and Tan, A.J. (2018). histone methyltransferase is essential for silkworm piRNA-mediated sex determination. *Plos Genetics* 14. ARTN e1007245  
10.1371/journal.pgen.1007245.
67. Swevers, L., Liu, J.S., Huvenne, H., and Smagghe, G. (2011). Search for Limiting Factors in the RNAi Pathway in Silkworm Tissues and the Bm5 Cell Line: The RNA-Binding Proteins R2D2 and Translin. *Plos One* 6. ARTN e20250  
10.1371/journal.pone.0020250.
68. Tatsuke, T., Zhu, L., Li, Z.Q., Mitsunobu, H., Yoshimura, K., Mon, H., Lee, J.M., and Kusakabe, T. (2014). Roles of Piwi Proteins in Transcriptional Regulation Mediated by HP1s in Cultured Silkworm Cells. *Plos One* 9. ARTN e92313  
10.1371/journal.pone.0092313.
69. Santos, D., Feng, M., Koliopoulou, A., Taning, C.N.T., Sun, J.C., and Swevers, L. (2023). What Are the Functional Roles of Piwi Proteins and piRNAs in Insects? *Insects* 14. ARTN 187  
10.3390/insects14020187.
70. Santos, D., Verdonck, T.W., Mingels, L., Van den Brande, S., Geens, B., Van Nieuwerburgh, F., Koliopoulou, A., Swevers, L., Wynant, N., and Vanden Broeck, J. (2022). PIWI Proteins Play an Antiviral Role in Lepidopteran Cell Lines. *Viruses-Basel* 14. ARTN 1442  
10.3390/v14071442.
71. Andersen, P.R., Tirian, L., Vunjak, M., and Rennecke, J.B. (2017). A heterochromatin-dependent transcription machinery drives piRNA expression. *Nature* 549, 54–+. 10.1038/nature23482.
72. de Araujo, A.C., Huguet, E., Herniou, E.A., Drezen, J.M., and Josse, T. (2022). Transposable element repression using piRNAs, and its relevance to endogenous viral elements (EVEs) and immunity in insects. *Current Opinion in Insect Science* 50. ARTN 100876

10.1016/j.cois.2022.100876.

73. Palazzo, A.F., and Lee, E.S. (2015). Non-coding RNA: what is functional and what is junk? *Frontiers in genetics* 6, 2. 10.3389/fgene.2015.00002.

74. Lee, J., Fujimoto, T., Sahara, K., Toyoda, A., and Shimada, T. (2025). Comprehensive genome annotation of

p50ma strain, a newly developed standard strain. *Sci Data* 12. ARTN 359

10.1038/s41597-025-04679-5.

75. Kawaoka, S., Hara, K., Shoji, K., Kobayashi, M., Shimada, T., Sugano, S., Tomari, Y., Suzuki, Y., and Katsuma, S. (2013). The comprehensive epigenome map of piRNA clusters. *Nucleic Acids Research* 41, 1581–1590. 10.1093/nar/gks1275.

76. Brower-Toland, B., Findley, S.D., Jiang, L., Liu, L., Yin, H., Dus, M., Zhou, P., Elgin, S.C.R., and Lin, H.F. (2007). PIWI associates with chromatin and interacts directly with HP1a. *Genes & Development* 21, 2300–2311. 10.1101/gad.1564307.

77. Katsuma, S., Shoji, K., Suzuki, Y., and Kiuchi, T. (2021). CRISPR/Cas9-mediated mutagenesis of and

in silkworm cultured cells. *Gene* 768. ARTN 145314

10.1016/j.gene.2020.145314.

78. Osanai-Futahashi, M., Suetsugu, Y., Mita, K., and Fujiwara, H. (2008). Genome-wide screening and characterization of transposable elements and their distribution analysis in the silkworm, *Bombyx mori*. *Insect Biochem Mol Biol* 38, 1046–1057.

79. Pashley, D.P. (1986). Host associated genetic differentiation in fall armyworm: a sibling species complex? *Annals of the Entomological Society of America* 79, 898–904.

80. Nam, K., Nègre, N., and Benjumea, C.I.S. (2024). Two host-plant strains in the fall armyworm. *Insect Science* 31, 1675–1683. 10.1111/1744-7917.13346.

81. Kelleher, E.S. (2016). Reexamining the P-Element Invasion of *Drosophila melanogaster* Through the Lens of piRNA Silencing. *Genetics* 203, 1513–1531. 10.1534/genetics.115.184119.

82. Hill, T., Schlötterer, C., and Betancourt, A.J. (2016). Hybrid Dysgenesis in *Drosophila simulans*

Associated with a Rapid Invasion of the

P-Element. *Plos Genetics* 12. ARTN e1005920

10.1371/journal.pgen.1005920.

83. Ramat, A., Garcia-Silva, M.R., Jahan, C., Naït-Saïdi, R., Dufourt, J., Garret, C., Chartier, A., Cremaschi, J., Patel, V., Decourcelle, M., et al. (2020). The PIWI protein Aubergine recruits eIF3 to activate translation in the germ plasm. *Cell Research* 30, 421–435. 10.1038/s41422-020-0294-9.

84. Dai, P., Wang, X., Gou, L.T., Li, Z.T., Wen, Z., Chen, Z.G., Hua, M.M., Zhong, A., Wang, L.B., Su, H.Y., et al. (2019). A Translation-Activating Function of MIWI/piRNA during Mouse Spermiogenesis. *Cell* 179, 1566–+. 10.1016/j.cell.2019.11.022.