Rapid evolution of promoters from germline-specifically expressed genes including transposon silencing factors

DAVID W. J. MCQUARRIE[1,2], AZAD ALIZADA[3], BENJAMIN CZECH NICHOLSON[3] AND MATTHIAS SOLLER[1,2*]

[1]School of Biosciences, College of Life and Environmental Sciences, University of Birmingham, Edgbaston, Birmingham, B15 2TT, United Kingdom

[2]Birmingham Centre for Genome Biology, University of Birmingham, Edgbaston, Birmingham, B15 2TT, United Kingdom

[3]Cancer Research UK Cambridge Institute, University of Cambridge, Li Ka Shing Centre, Cambridge CB2 0RE, United Kingdom

[*]Corresponding author: m.soller@bham.ac.uk

Running title: Evolution of promoters of germline transposon silencing factors

Key Words: Piwi-interacting RNA (piRNA), germline transposon silencing, RNA transgenerational inheritance, rapid evolution, promoter evolution, speciation, nuclear pore complex, neuronal wiring, transposon silencing

Abstract

Background: The piRNA pathway in animal gonads functions as an 'RNA-based immune system', serving to silence transposable elements and prevent inheritance of novel invaders. In *Drosophila*, this pathway relies on three gonad-specific Argonaute proteins (Argonaute-3, Aubergine and Piwi) that associate with 23-28 nucleotide piRNAs, directing the silencing of transposon-derived transcripts. Transposons constitute a primary driver of genome evolution, yet the evolution of piRNA pathway factors has not received in-depth exploration. Specifically, channel nuclear pore proteins, which impact piRNA processing, exhibit regions of rapid evolution in their promoters. Consequently, the question arises whether such a mode of evolution is a general feature of transposon silencing pathways.

Results: By employing genomic analysis of coding and promoter regions within genes that function in transposon silencing in *Drosophila*, we demonstrate that the promoters of germ cell-specific piRNA factors are undergoing rapid evolution. Our findings indicate that rapid promoter evolution is a common trait among piRNA factors engaged in germline silencing across insect species, potentially contributing to gene expression divergence in closely related taxa. Furthermore, we observe that the promoters of genes exclusively expressed in germ cells generally exhibit rapid evolution, with some divergence in gene expression.

Conclusion: Our results suggest that increased germline promoter evolution, in partnership with other factors, could contribute to transposon silencing and evolution of species through differential expression of genes driven by invading transposons.

45

46 **Background**

47 Eukaryotic organisms are continuously challenged from genomic parasites called
48 transposable elements (TEs) [1-4]. Unchecked transposon activity often results in reduced
49 reproductive fitness [5-8]. To negate the detrimental effects posed by TE mobilisation, vital
50 regulatory pathways evolved that efficiently suppress transposon activity. At the centre of these
51 are 23-28 nucleotide small RNAs, called PIWI-interacting RNAs (piRNAs). piRNAs are
52 generated from transposons and from genomic clusters and bind to Argonaute proteins of the
53 PIWI-clade, including P-element induced wimpy testis (Piwi). Aubergine (Aub) and
54 Argonaute-3 (Ago3). The piRNA pathway is often referred to as an 'RNA-based immune
55 system' thanks to its ability to adopt to new TE insertions and the memory of past transposon
56 activity through piRNA clusters through RNA transgenerational inheritance [1, 2, 9, 10].

57 The piRNA pathway is arguably best understood in the ovary of *Drosophila*
58 *melanogaster*. It was shown that ovaries feature two branches of the silencing machinery [5],
59 one that is active in germ cells and a condensed version that is specific to somatic follicle cells
60 that surround the germline [2, 11]. Somatic follicle cells only feature unistrand piRNA clusters,
61 such as *flamenco* (*flam*), and exclusively produce piRNAs via a Zucchini (Zuc)-dependent
62 biogenesis mechanism that takes place on the mitochondria surface involving Armitage (Armi),
63 Papi, Vreteno (Vret), Gasz, Daedalus (Daed), Minotaur (Mino), Shutdown (Shu), and Sister of
64 Yb (SoYb) (Fig. 1A) [3, 12, 13]. Licensing of *flam* for piRNA production takes place upstream,
65 at so-called Yb-bodies [1, 12]. Somatic cells only express Piwi, which following the association
66 with a mature piRNA shuttles to the nucleus. There, Piwi-piRNA complexes scan for nascent
67 TE transcripts and instruct co-transcriptional gene silencing (cTGS) that requires components
68 of the general chromatin silencing machinery and the piRNA pathway-specific factors Asterix

69 (Arx), Piwi, Panoramix (Panx), Nuclear export factor 2 (Nxf2) and Maelstrom (Mael) (Fig.
70 1A) [1, 2, 12, 14].

71 In germ cells, piRNAs predominantly originate from dual-strand clusters and are
72 produced by a different biogenesis mechanism called the ping-pong amplification cycle (Fig.
73 1B) [1-3, 5, 11, 12, 15]. Dual-strand clusters harbour remnants of transposons and are expressed
74 via a non-canonical transcription process that is centred on a complex containing the HP1a
75 homolog Rhino (Rhi), Deadlock (Del) and Cutoff (Cuff), and further requires Moonshiner
76 (Moon) (Fig. 1B). This Rhi-Del-Cuff complex also recruits a non-canonical export machinery
77 consisting of Bootlegger (Boot) and Nuclear export factor 3 (Nxf3), which transport piRNA
78 precursors for processing to a germ-cell specific perinuclear region termed "nuage" (Fig. 1B).
79 Here, the ping-pong cycle takes place and relies on the two PIWI proteins Aubergine (Aub)
80 and Argonaute 3 (Ago3), as well as several RNA binding proteins and Tudor-domain
81 containing factors including Tejas (Tej), Vasa (Vas), Qin, Squash (Squ), Tapas, Brother of Yb
82 (BoYb), Krimper (Krimp), and Spindle-E (Spn-E) (Fig. 1B) [1-3, 12, 15, 16].

83 Channel nuclear pore proteins (Nups) have been implicated in piRNA regulated
84 transposon silencing in *C. elegans* [17]. Interestingly, in *Drosophila*, Nups were shown to
85 affect transposon silencing in both the germline [16, 18], and in somatic cells of the ovary, here
86 specifically by contributing to the conversion of precursor RNAs from the *flam* locus into
87 piRNAs [19]. Recently, we found that inner and outer ring Nups in the nuclear pore complex
88 (NPC) evolve rapidly through indel accumulation in their promoters [20, 21]. Such promoter
89 indel variability in Nup54 has dominant, pleiotropic effects on sexual differentiation including
90 neuronal wiring important for female post-mating behaviours directed by male-derived sex
91 peptide, that as a result of sexual conflict could drive speciation by such mechanism [17, 20,
92 22]. Intriguingly, nuclear import/export pathways have been linked to speciation through
93 hybrid incompatibility, but are in addition important for many cellular processes beyond

94 piRNA processing [23, 24]. Hence, promoter evolution particularly of piRNA processing genes
95 could have wide impact on transposon silencing and evolution of species through differential
96 expression of genes, but whether regulators of germline transposon silencing generally evolve
97 rapidly through their promoters has not been determined.

98 Here, we analyse the promoters of genes involved in the ovarian piRNA pathway. We
99 compare piRNA factors required in somatic cells (Yb) or in both the somatic and germline
100 compartments (cTGS and piRNA biogenesis, Fig. 1A) with those that are germ cell-specific
101 (dual-strand clusters and the ping-pong cycle, Fig. 1B). We find that the promoters of germline-
102 specific piRNA factors are hot spots for rapid evolution compared to those of somatic and germ
103 cell piRNA pathway genes and the *Drosophila* genome average throughout insect species.
104 Analysis of genes with significant accumulation of promoter changes reveals a mix of indel
105 and base change accumulation. Further, our analysis reveals that rapid promoter evolution is a
106 general feature of genes specifically expressed in germ cells, while soma-expressed genes
107 evolve at a similar rate to the average of all *Drosophila* genes. Through cross-species
108 differential ovary RNA-seq analysis we reveal that germ cell-specific genes minimally diverge
109 in gene expression levels compared to genes expressed in somatic cells. Our findings highlight
110 that promoters of germline genes involved in transposon silencing evolve rapidly and are
111 accompanied by diverging gene expression, suggesting a possible mode of rapid speciation
112 through accumulation of changes in promoters in correlation with additional regulatory factor
113 divergence.

114

115 **Results**

116 **Promoter regions of germline-specific piRNA factors are hot spots for rapid evolution**

117 Since Nups have been shown to function in transposon silencing in the germline (12
118 from 14 tested are above the control) and display fast evolving promoters [16, 20, 21], we

119 analysed the rate of evolution in promoters and coding regions of piRNA pathway genes. We
120 separated all piRNA pathway genes based on their genetic requirement into two categories:
121 piRNA factors essential in somatic cells, including the soma-specific piRNA factor Yb and
122 piRNA factors expressed in both somatic and germline cells (cTGS and piRNA biogenesis,
123 Fig. 1A), and germ cell-specific piRNA proteins (dual-strand clusters and ping-pong cycle, Fig
124 1b). The piRNA factors essential in somatic cells include *armi*, *papi*, *vret*, *Gasz*, *daed*, *mino*,
125 *zuc*, *shu*, *SoYb*, *Yb*, *arx*, *piwi*, *panx*, *nxf2*, and *mael* (Fig. 1A). The group of piRNA factors that
126 are exclusively required in the germline are *aub*, *tej*, *Ago3*, *vas*, *qin*, *squ*, *tapas*, *BoYb*, *krimp*,
127 *spn-E, Boot*, *moon*, *nxf3*, *cuff*, *rhi*, and *del* (Fig. 1B) [3, 14, 25].

128 We defined the promoter containing region as a 350-nucleotide (-30 to -380) window
129 upstream of the predicted TATA box region 30 nucleotides upstream of the transcription start
130 site (TSS) [26]. In a first approach, we performed a genome-wide promoter evolution analysis
131 of germline-specific as well as somatic and germ cell piRNA factors and compared it to all
132 genes in the *Drosophila* genome (Fig 1C-D). We used publicly available PhyloP27way data a
133 conservation score. We observed a significant accumulation of changes in the germ cell-
134 specific piRNA factor promoter regions compared to the genome (Fig. 1C-F). Of note, the
135 somatic and germ cell piRNA factors evolved slower compared to the genome.

136 We next compared indel and base changes in the promoter regions between five closely
137 related *Drosophila* species (*D. melanogaster*, *D. simulans*, *D. sechellia*, *D. yakuba* and *D.*
138 *erecta*). Promoter scores (*d*) were calculated and compared to a control group comprised of the
139 m^6A methylation machinery genes (*Mettl3*, *Mettl14*, *fl(2)d*, *virilizer*, *flacc*, *nito* and *Hakai*,
140 *Ythdc1* and *Ythdf*), chosen due to their high evolutionary conservation and requirement for strict
141 stoichiometry for functionality, making this an ideal control group to monitor promoter
142 evolution [21, 27, 28]. In this analysis, the group of somatic and germ cell piRNA factors
143 showed similar evolution rates compared to the control (Fig. 1E-J), while germ cell-specific

144 piRNA factors as a group evolved at a much higher rate (Fig. 1G and H), accruing a
145 significantly increased number of promoter-located indels (Fig. 1G and H) and base changes
146 (Fig. 1I and J). At the individual gene level, 7 out of 14 of germ cell-specific piRNA factors
147 (*Ago3* was omitted due to its low conservation between closely related *Drosophila* species [29-
148 31] showed significant accumulation of promoter changes (Supplementary Fig. 1A). Among
149 the somatic and germ cell piRNA factors, 3 out of 15 genes had significant accumulation of
150 promoter changes (Supplementary Fig. 1A).

151 To assess general evolution in the coding regions, we used McDonald-Kreitman tests
152 (MKTs) and analysed polymorphisms and divergence within *D. melanogaster* and between *D.*
153 *melanogaster* and *D. simulans* for two ancestral populations (Congo and Zambia) [32].
154 Analysis of grouped germ cell-specific piRNA factors flagged these factors as under positive
155 selection for both populations. Conversely, the somatic and germ cell piRNA factors were not
156 flagged as under positive selection, (Supplementary Fig. 1B).

157

158 **Promoters of genes required for germline transposon silencing are fast evolving**

159 Given the rapid evolution of germ cell-specific piRNA factors among closely related
160 *Drosophila* species, we wanted to determine (1) whether this is a general feature of insect
161 evolution, and (2) whether this is a general feature of genes required for transposon silencing
162 with pleiotropic roles that have been identified in a genetic screen (GTS100 genes), since this
163 provides a larger gene group [16]. To get a scale of the persistence of evolutionary changes in
164 promoters of GTS100 genes we used publicly available PhyloP27way data from UCSC
165 genome browser [33, 34]. Genes were ordered based on transposon de-repression scores [16]
166 and d^P was calculated for individual genes, the genome average as a control group, and for the
167 average of GTS100 gene promoter regions (Fig. 2A-C, 3A). Genes without full PhyloP

168 coverage were removed from the analysis (where ≥1 nucleotide of the analysed genomic region
169 contained the *D. melanogaster* sequence only).

170 The promoter score average of GTS100 genes showed a significant increase in
171 promoter nucleotide changes compared to the control (Fig. 2C). Although many of these genes
172 have additional roles, gene ontology (GO) analysis confirmed a primary role in piRNA binding
173 (Fig. 2D) [35-37].

174 To understand whether GTS100 genes were individually fast evolving, individual d^P
175 scores were calculated (Fig. 3A). Compared to the genome average, 49 genes (FEPG, Fast
176 Evolving Promoter Germline genes) were significantly increased in promoter change events,
177 while 12 were significantly decreased (Fig. 3A). Of note *Brother of Yb* (*BoYb*) which is
178 considered the germline replacement of *Yb* [3], also displayed a fast evolving promoter in
179 contrast to its somatic counterpart *Yb* (Fig. 3A). To specifically look at indels at a high
180 confidence level, we analysed FEPG genes for rapid promoter evolution between the
181 previously used five *Drosophila* species (Fig. 3B-D). Here, indels and base changes were
182 measured, showing a significant increase in both indels and base changes compared to the
183 control (Fig. 3C and D).

184

185 **A subset of piRNA factors that are part of distinct protein complexes display hot spots**
186 **for rapid evolution in promoters**

187 In a saturating genetic screen for regulators of transposon silencing in germ cells
188 members of the Nup complex were identified as regulators of germline piRNA silencing, and
189 later shown to rapidly evolve through their promoters [16, 20, 21]. In the same screen,
190 numerous genes whose products are part of distinct protein complexes were identified,
191 including the nuclear cap-binding complex (CBC), THO complex, the exon junction complex
192 (EJC) inner core, outer shell and EJC transient factors, the non-specific lethal (NSL) complex,

193 SUMOylation modifiers and the complex of proteins associated with Set1 (COMPASS)
194 (Supplementary Fig. 2). Given the fast evolution of promoters in germline-specific piRNA
195 pathway genes and the Nup complex, we analysed promoters of the additional protein
196 complexes affecting germline piRNA silencing to see whether their rate of evolution matched
197 that of other piRNA silencing factors. Compared to somatic and germ cell piRNA factors,
198 promoter regions for the nuclear CBC, SUMO modifiers, COMPASS complex, NSL complex
199 and EJC outer shell showed a significant accumulation of sequence changes (Fig. 3E and F).
200 Analysis of individual genes from both significant and non-significant gene groups revealed
201 that 43% of analysed genes have rapidly accumulated sequence changes in their promoters
202 (Supplementary Fig. 2). Specifically looking at genes in fast evolving complexes (Fig. 3E and
203 F) revealed that 56% of genes in these groups rapidly evolve through accumulation of
204 mutations in their promoters (Supplementary Fig. 2).

205 When analysing the base changes and indels of rapid evolving promoters
206 (Supplementary Fig. 3A-D), indel accumulation was observed for the SUMO modifiers,
207 COMPASS complex, NSL complex and EJC outer shell (4 out of 5), while the nuclear CBC
208 showed a significant decrease (Supplementary Fig. 3B). The nuclear CBC, SUMO modifiers
209 and COMPASS complex also showed a significant accumulation of base changes in their
210 promoter regions (3 out of 5) (Supplementary Fig. 3D). Individually, 4 out of 27 genes were
211 significantly enriched in indel events (*Hcf*, *dgt1*, *nsl1*, and *Bin1*), while base change
212 accumulation events indicated 8 out of 27 genes (*Ars2*, *cbp80*, *Ulp1*, *Set1*, *Hcf*, *Cfp1*, *MBD-*
213 *R2*) being significantly increased (Supplementary Fig. 3E).

214 Next, we analysed the rate of evolution in the coding regions of these complexes using
215 MKT tests [32]. This analysis flagged the EJC outer shell, EJC transient facts, NSL complex,
216 and SUMO modifiers as under positive selection in both populations, while the Nuclear CBC
217 was under positive selection in the Zambia population only (Supplementary Fig. 4).

218

219

220 **Analysis of piRNA factor differential gene expression in ovaries of *D. melanogaster* and**

221 ***D. yakuba***

222 Since promoters of FEPG genes are fast evolving, we analysed publicly available RNA-

223 seq data from ovaries between *D. melanogaster* and *D. yakuba* to assess gene expression

224 divergence in the FEPG gene group compared to the somatic and germ cell piRNA factors (Fig.

225 4A) [38-40]. Here, many FEPG genes had a tendency for slight significant divergence in gene

226 expression in both directions (*D. melanogaster* or *D. yakuba*), but this was true for only few

227 somatic and germ cell piRNA factors genes including *arx* (Fig. 4A). Notably, we observed

228 large expression divergence for *CG32152, MEP-1* and *Rbbp5* (Fig. 4A). Aligning the

229 promoters of these three genes revealed an increased number of indels and base changes

230 compared to the 5'UTR or the coding region (Fig. 4B-D), but the cohort was unfortunately too

231 small to detect whether increased changes in promoters correlate with altered expression

232 between the two species. Further limitations stem from the pleiotropic roles of the analysed

233 genes, and therefore expression profiles likely differ across cell types in the analysed tissue.

234

235 **Germ cell-specific genes show rapid evolution of promoters associated with divergent**

236 **gene expression between species**

237 To investigate whether the rapid evolution of promoters in piRNA factors expressed in

238 germ cells is a common feature of germline-expressed genes, we used publicly available RNA-

239 seq data from FACS-sorted vasa-GFP ovaries to identify genes specifically expressed in either

240 germ cells or somatic cells of the ovary. First, using differential RNA-seq analysis of FACS-

241 sorted vasa-GFP+ (germline) and vasa-GFP- (somatic) cells from cultured *D.mel* ovary GSC

242 line (vasa-GFP;hs-bam;bamΔ86) (GSE119862) [41], we defined germline-enriched (vasa-

243 GFP+; log2FC>0, p-val<0.01) and soma-enriched (vasa-GFP-; log2FC<0, p-val<0.01) genes
244 (Fig. 5A and B). Next, using RNA-seq data from ovarian somatic cells (OSCs) (GSE160860)
245 [42], we further filtered the soma-enriched genes that are highly expressed in OSCs.
246 Additionally, we further filtered the germline-enriched genes that show expression in early fly
247 embryos (0-2hr, modENCODE). In this way, we generated a stringent list of 107 genes
248 specifically expressed in germ cells and of 59 soma-expressed genes (Supplementary Dataset
249 1).

250 We then analysed sequence change accumulation in promoters using PhyloP data and
251 calculated d^P for all *D. melanogaster* genes and genes expressed specifically in somatic cells
252 or the germline (Fig. 5C and D) or the average (Fig 5E). Germ cell-specific expressed genes
253 significantly accumulated changes (d^P) in their promoters compared to soma-expressed genes
254 (Fig. 5F). Comparison of the average promoter PhyloP scores and d^P conservation scores
255 between the genome average and somatic or germ cell-specifically expressed genes equally
256 revealed a significant increase in germ cell gene promoter nucleotide changes compared to the
257 genome average (Fig. 5G and H).

258 Since rapid promoter evolution of germline piRNA silencing factors correlated with
259 gene expression divergence in our previous analysis, we hypothesized that this was a general
260 feature of germ cell genes. To test this, we compared gene expression divergence between
261 somatic and germ cell-specific genes using publicly available RNA-seq data of *D.*
262 *melanogaster* and *D. yakuba* ovaries (Fig. 6A) [38-40]. Notably, genes specifically expressed
263 in germ cells showed slightly higher divergence in expression compared to soma-expressed
264 genes (Fig. 6A). When d^P conservation scores were plotted against the gene expression change
265 between *D. melanogaster* and *D. yakuba*, the germ cell-specifically expressed genes showed
266 slightly higher divergence in expression especially in the higher d^P score groups (>50-75, >75-

267 100) compared to soma-expressed genes, which showed relatively equal expression divergence
268 across the different gene groups (Fig. 6C-E).
269
270 **Discussion**
271 Through genomic analysis of germline and somatic transposon silencing genes and
272 their regulators we identify hot spots for rapid evolution in promoter regions (-30 to -380
273 nucleotides upstream of the TSS) of germ cell-specific (dual-strand clusters and the ping-pong
274 cycle), but not soma-expressed piRNA pathway genes (cTGS and piRNA biogenesis). Further,
275 we show that rapid promoter evolution is a general feature of germ cell-specific genes
276 compared to those expressed only in the soma. Our analysis suggests that this mode of
277 evolution in the germline pathway could be a general feature of at least insect evolution.
278 Overall, our results point towards a key role for rapid evolution of gene promoters in the germ
279 cell-specific piRNA pathway which could, coupled with other drivers of expression
280 divergence, impact the expression of germline regulatory genes.
281 Transposon mobility has been attributed to the accumulation of sequence changes in
282 promoter regions because of the presence of open chromatin around the transcription start site
283 of genes [43]. Interestingly, in core NPC genes, rapid evolution mostly led to accumulation of
284 indels [20, 21], while piRNA pathway genes display equal accumulation of indels and base
285 changes. Likely, indels have more profound effects on changes in transcription than single
286 nucleotide alterations. This feature might reflect that compromised transposon silencing causes
287 sterility, hence not allowing for substantial changes in expression in piRNA processing genes.
288 Likewise, stoichiometric changes in protein complexes can drastically alter complex
289 functionality, e.g. the male-specific lethal (MSL) complex binds less to its targets when one
290 component of the complex is missing [44, 45]. Hence, changes in the protein levels of

291 individual piRNA factors likely have dominant effects on their capacity to silence transposon
292 activity.

293 The occurrence of novel transposons can fundamentally impact species reproductive
294 success through a phenomenon call hybrid dysgenesis [46]. Here, if a novel transposable
295 element is transmitted by the male, female fertility is severely compromised and can result in
296 sterility as a result of missing piRNA silencing of the novel invader. Essential to combat these
297 novel active transposable elements is the ping-pong cycle, primed by transcripts from the novel
298 transposon. Accordingly, females can be become resistant to such novel transposons over time
299 through RNA transgenerational inheritance of piRNAs to build up ping-pong cycle
300 amplification in germ cells [9, 10, 47].

301 Our examination of evolution in coding regions flagged germ cell-specific piRNA
302 factors, as well as some pleiotropic piRNA processing associated complexes (EJC outer shell,
303 EJC transient facts, NSL complex, SUMO modifiers, and Nuclear CBC) as under positive
304 selection. This is unsurprising in the case of primary piRNA processing factors in the germline,
305 as there are many examples of these factors being under selection [48-56]. For instance,
306 replacement of *D. melanogaster rhino* and *deadlock* genes with *D. simulans* homologues
307 results in non-functionality [49]. Here, *D. simulans* Rhino binding domains no longer bind to
308 *D. melanogaster* Deadlock, resulting in failed localisation to piRNA clusters [49]. Such effects
309 resulting from few amino acid changes act as powerful driving forces in diverging piRNA
310 processing. In this setting, rapid promoter evolution of piRNA silencing genes could act as an
311 additional factor driving changes in expression levels and factor stoichiometry for piRNA
312 processing divergence, which may be difficult to explore when coupled with functional
313 divergence. Intuitively, one would associate rapid promoter revolution with changes in
314 expression, however, we observed minimal effects on expression. Since transposon silencing
315 is essential, changes in cis regulatory elements of these regulatory factors is likely constrained

316 by compensatory mechanisms and could be accompanied by changes in enhancer regions.
317 Whether this is the case would required a detailed knowledge of relevant enhancers and
318 transcription factor binding.

319 The pattern of changes observed in fast evolving promoters like germline specific
320 piRNA factors and Nups resembles the outcome of a *P*-element mutagenesis experiment in the
321 *egghead* (*egh*) locus coding for a glycosphingolipid biosynthesis enzyme [57]. Here, multiple
322 base changes were observed in the region of the first promoter after *P*-element mutagenesis,
323 but not an actual *P*-element insert, resulting in a sex peptide insensitive allele egh^{cm}. Mutations
324 in the *egh* gene result in pleiotropic phenotypes and the egh^{cm} allele disrupts neuronal wiring
325 required for the female post-mating response and development of the optic lobe [22, 57, 58].

326 Since establishing transposon resistance involves forced selection, genes whose
327 expression changes result in pleiotropic effect, like channel Nup54, could combine adaptations
328 in piRNA processing with changes in neuronal wiring resulting in altered behavioural
329 preferences [20]. Of note, in *D. simulans*, a close relative of *D. melanogaster*, projections of
330 *fruitless* P1 sensory neurons that control courtship have changed and alter mate preference [59].

331 The severe impact of hybrid dysgenesis on fertility likely limits fast evolution to species
332 with high numbers of progeny. Perhaps this can explain the differences seen when comparing
333 insect and mammalian promoters that are generally more conserved [26]. However, rapid
334 promoter evolution has also been observed in primate promoters, suggesting a common mode
335 of evolution that has yet to be explored [20, 21, 26, 60].

336

337 **Materials and methods**

338 **Sequence/data retrieval and alignment**

339 Gene and promotor sequences were retrieved from UCSC Genome Browser using the
340 UCSC Table Browser sequence retrieval tool [33, 34]. A standardised region of 2000 bases

341 upstream of the annotated gene TSS was exported for each gene to ensure inclusion of promoter
342 regions. Sequences were imported and aligned with clustalW in MEGA11 [61]. PhyloP27way
343 data were sourced from UCSC genome browser (genome.ucsc.edu) through the Table Browser
344 tool [33, 34]. Data points were collected for a region of 1000 nucleotides upstream and 300
345 nucleotides downstream for each of the analysed genes. Genes without full species coverage
346 when analysing evolution between the 5 *Drosophila* species, where ≥1 species had no
347 conserved genomic sequence were removed from the analysis. Genes without full PhyloP
348 coverage were removed from the analysis (where ≥1 nucleotide of the analysed genomic region
349 contained the *D. melanogaster* sequence only).
350

351 **Molecular evolution of open reading frames analysis**

352 Analysis of adaptive protein evolution was performed via MKTs [32] using the PopFly
353 online database (imkt.uab.cat) which uses Drosophila Genome Nexus project sequence data
354 [62-64]. Polymorphism and divergence data were collected from the ancestral Congo
355 population, chosen since it is a sub-Saharan population with comparably high ancestral stability
356 compared to other populations. The 'Standard MKT' test was used to calculate results for
357 individual genes. Fisher's Exact Test with significance defined as Bonferroni corrected $p \leq 0.05$
358 was used to calculate significance between nonsynonymous to synonymous polymorphisms
359 within *D. melanogaster* or between *D. melanogaster* and *D. simulans*.
360

361 **Identification of promotor hot spots and comparison of substitution rates**

362 All gene models were manually inspected in FlyBase using the JBrowse browser and
363 dominant transcripts were chosen based on comparative modENCODE expression data [65].
364 Gene and promoter sequence alignments were trimmed to -1000/+300 nucleotides from the
365 TSS of each gene. To calculate the frequency of sequence change hotspots in promoter

366 sequences, we calculated the hotspot accumulation score *d* using alignments generated as
367 described, or PhyloP data. Alignments were translated into 'events' for each species compared
368 to *D. melanogaster*, where for each nucleotide in the sequence, 0 signified a conserved
369 sequence and 1 signified a sequence change (base change or indel event) [66]. Events were
370 calculated for all changes, as well as base changes and indels individually. The sum of events
371 was calculated for concatenated gene groups and individual genes at each nucleotide position.
372 A sliding event (*Se*) score was calculated from this using a sliding window of five bases along
373 the sequence, from which heatmaps were generated. To calculate the percentage of events
374 greater than the average control promoter *Se* score (*d*), a 350-nucleotide region upstream of the
375 estimated TATA box region was analysed where the total number of *Se* scores exceeding the
376 control group average sliding event score (Se^C) was divided by the total number of events in
377 that region (*N*).

378 $$d = \frac{\text{Total number of } Se \text{ events where } Se > Se^C}{N} \text{x}100$$

379 To calculate the promoter region *d* scores from PhyloP data, the total number of PhyloP
380 (*p*) scores in the 350 nucleotide regions less than the control group average promoter region
381 (p^C) took the place of the total number of *Se* events where *Se* is greater than Se^C (see equation
382 below).

383 $$d^P = \frac{\text{Total number of } p \text{ scores} < p^C}{N} \text{x}100$$

384 Significance was calculated using non-parametric chi-squared tests compared to the
385 control group *d* score. Significance values where $p \leq 0.05$ with Bonferroni correction were
386 considered statistically significant.

387

388 **Identification of somatic and germ cell-specifically expressed genes in *D. melanogaster***

389 We defined somatic and germline genes using several steps. First, using processed
390 RNA-seq data from vasa-GFP+/- cells sorted from cultured ovary GSC line (vasa-GFP;hs-
391 bam;bamΔ86) (GSE119862) [41], we defined germline-enriched (vasa-GFP+) using logFC>0,
392 p-value<0.01 parameters and soma-enriched (vasa-GFP-) genes using logFC<0, p-value<0.01
393 parameters (Figures 4a and 4b). Next, using processed RNA-seq data from fly ovaries
394 (modENCODE), we filtered genes that are expressed in adult fly ovaries using RPKM>10
395 parameter for both soma and germline genes. To stringently define the germline genes, we
396 additionally used processed RNA-seq data from early fly embryos (0-2hr, modENCODE;
397 mRNA-seq fly embryo 0-2hr; RPKM; mE_mRNA_em0-2hr; FBlc0000086) to further filter
398 the germline-enriched genes using >9 RPKM parameter. To stringently define the somatic-
399 specific genes, we additionally used processed RNA-seq data from ovarian somatic cells
400 (OSCs) (n=4, GSE160860) [42] and filtered the soma-enriched genes using >20 RPKM to
401 select genes that are expressed in OSCs as well as parameter <2 RPKM for early fly embryo
402 (0-2hr, modENCODE) to filter out genes that are expressed in early embryos. This analysis
403 generated a list of 107 germ cell-specifically expressed genes and of 59 somatic cell-
404 specifically expressed genes (Fig. 4A and B, Supplementary Dataset 1).
405
406 **Comparative gene expression analysis between *D. melanogaster* and *D. yakuba***
407 Using publicly available raw RNA-seq data from *D. melanogaster* (n=3) and *D. yakuba*
408 (n=4) ovaries, we trimmed adapter sequences from the raw reads using Cutadapt tool (v1.18,
409 default parameters) and aligned the trimmed reads to the genome assemblies (dm6 and
410 droYak2) using the RNA-seq aligner STAR (v2.7.3a). Gene counts were quantified using the
411 featureCounts tool (Subread package v1.5.3). Orthologues genes were identified using
412 blastDm2FB in UCSC Table Browser. Gene count normalization and differential gene

413 expression analysis between the species were performed using DESeq2 (v1.30.1) with an
414 additional species-specific gene lengths normalization step.
415
416 **Supplementary information**
417 Supplementary materials include Fig. S1-4 and Supplementary Dataset 1.
418
419 **Authors' contributions**
420 DWJM and MS conceived and directed the project. DWJM performed genomic and
421 evolutionary analysis. AA and DWJM analysed differential gene expression. DWJM wrote the
422 initial draft and all authors contributed to the final manuscript. All authors read and approved
423 the final manuscript.
424
425 **Acknowledgements**
426 We thank the Soller lab for helpful discussion.
427
428 **Funding**
429 This work was supported by the Medical Research Council to DWJM [MR/N013913/1] and
430 the Biotechnology and Biological Science Research Council to MS. Work by BCN and AA
431 was funded by Cancer Research UK [A21143] and the Wellcome Trust [110161/Z/15/Z].
432
433 **Availability of data and materials**
434 All data generated or analysed during this study are included in the supplementary information
435 files.
436
437 **Ethics approval and consent to participate**

438 Not applicable.

439

440 **Consent for publication**

441 Not applicable.

442

443 **Competing interests**

444 The authors declare no competing interests.

445

446 **References**

447 1. Iwakawa H-o, Tomari Y: **Life of RISC: Formation, action, and degradation of**
448 **RNA-induced silencing complex**. *Molecular Cell* 2022, **82**(1):30-43.

449 2. Czech B, Munafò M, Ciabrelli F, Eastwood EL, Fabry MH, Kneuss E, Hannon GJ:
450 **piRNA-Guided Genome Defense: From Biogenesis to Silencing**. *Annual Review of*
451 *Genetics* 2018, **52**(1):131-157.

452 3. Yamashiro H, Siomi MC: **PIWI-Interacting RNA in Drosophila: Biogenesis,**
453 **Transposon Regulation, and Beyond**. *Chemical Reviews* 2018, **118**(8):4404-4421.

454 4. Slotkin RK, Martienssen R: **Transposable elements and the epigenetic regulation**
455 **of the genome**. *Nature Reviews Genetics* 2007, **8**(4):272-285.

456 5. Malone CD, Brennecke J, Dus M, Stark A, McCombie WR, Sachidanandam R,
457 Hannon GJ: **Specialized piRNA Pathways Act in Germline and Somatic Tissues of**
458 **the Drosophila Ovary**. *Cell* 2009, **137**(3):522-535.

459 6. Khurana JS, Theurkauf W: **piRNAs, transposon silencing, and Drosophila**
460 **germline development**. *Journal of Cell Biology* 2010, **191**(5):905-913.

461 7. Senti K-A, Brennecke J: **The piRNA pathway: a fly's perspective on the guardian**
462 **of the genome**. *Trends in Genetics* 2010, **26**(12):499-509.

463 8. Sakakibara K, Siomi MC: **The PIWI-Interacting RNA Molecular Pathway:**
464 **Insights From Cultured Silkworm Germline Cells**. *BioEssays* 2018,
465 **40**(1):1700068.
466 9. Moelling K: **Epigenetics and transgenerational inheritance**. *J Physiol* 2023.
467 10. Grentzinger T, Armenise C, Brun C, Mugat B, Serrano V, Pelisson A, Chambeyron S:
468 **piRNA-mediated transgenerational inheritance of an acquired trait**. *Genome Res*
469 2012, **22**(10):1877-1888.
470 11. Wang X, Ramat A, Simonelig M, Liu M-F: **Emerging roles and functional**
471 **mechanisms of PIWI-interacting RNAs**. *Nature Reviews Molecular Cell Biology*
472 2023, **24**(2):123-141.
473 12. Sato K, Siomi MC: **The piRNA pathway in Drosophila ovarian germ and somatic**
474 **cells**. *Proceedings of the Japan Academy, Series B* 2020, **96**(1):32-42.
475 13. Signor S, Vedanayagam J, Kim BY, Wierzbicki F, Kofler R, Lai EC: **Rapid**
476 **evolutionary diversification of the flamenco locus across simulans clade**
477 **Drosophila species**. *PLOS Genetics* 2023, **19**(8):e1010914.
478 14. Czech B, Hannon GJ: **One Loop to Rule Them All: The Ping-Pong Cycle and**
479 **piRNA-Guided Silencing**. *Trends in Biochemical Sciences* 2016, **41**(4):324-337.
480 15. Haase AD: **An introduction to PIWI-interacting RNAs (piRNAs) in the context of**
481 **metazoan small RNA silencing pathways**. *RNA Biology* 2022, **19**(1):1094-1102.
482 16. Czech B, Preall J, B. , McGinn J, Hannon G, J. : **A Transcriptome-wide RNAi**
483 **Screen in the Drosophila Ovary Reveals Factors of the Germline piRNA**
484 **Pathway**. *Molecular Cell* 2013, **50**(5):749-761.
485 17. Brown JS, Zhang D, Chen W, Lee H-C: **Sensitized piRNA reporter identifies**
486 **multiple RNA processing factors involved in piRNA-mediated gene silencing**. In.
487 bioRxiv: Cold Spring Harbor Laboratory; 2023.

488 18. Handler D, Meixner K, Pizka M, Lauss K, Schmied C, Franz, Brennecke J: **The**
489 **Genetic Makeup of the Drosophila piRNA Pathway**. *Molecular Cell* 2013,
490 **50**(5):762-777.
491 19. Munafò M, Lawless VR, Passera A, Macmillan S, Bornelöv S, Haussmann IU, Soller
492 M, Hannon GJ, Czech B: **Channel nuclear pore complex subunits are required for**
493 **transposon silencing in Drosophila**. *eLife* 2021, **10**.
494 20. Nallasivan MP, Haussmann IU, Civetta A, Soller M: **Channel nuclear pore protein**
495 **54 directs sexual differentiation and neuronal wiring of female reproductive**
496 **behaviors in Drosophila**. *BMC Biology* 2021, **19**(1).
497 21. McQuarrie DWJ, Read AM, Stephens FHS, Civetta A, Soller M: **Indel driven rapid**
498 **evolution of core nuclear pore protein gene promoters**. *Scientific Reports* 2023,
499 **13**(1).
500 22. Haussmann IU, Hemani Y, Wijesekera T, Dauwalder B, Soller M: **Multiple**
501 **pathways mediate the sex-peptide-regulated switch in female Drosophila**
502 **reproductive behaviours**. *Proc Biol Sci* 2013, **280**(1771):20131938.
503 23. Tang S, Presgraves DC: **Evolution of the Drosophila Nuclear Pore Complex**
504 **Results in Multiple Hybrid Incompatibilities**. *Science* 2009, **323**(5915):779-782.
505 24. Tang S, Presgraves DC: **Lineage-Specific Evolution of the Complex Nup160**
506 **Hybrid Incompatibility Between Drosophila melanogaster and Its Sister Species**.
507 *Genetics* 2015, **200**(4):1245-1254.
508 25. Iwasaki YW, Siomi MC, Siomi H: **PIWI-Interacting RNA: Its Biogenesis and**
509 **Functions**. *Annu Rev Biochem* 2015, **84**:405-433.
510 26. Taylor MS, Kai C, Kawai J, Carninci P, Hayashizaki Y, Semple CAM: **Heterotachy**
511 **in Mammalian Promoter Evolution**. *PLoS Genetics* 2006, **2**(4):e30.

512 27. Balacco DL, Soller M: **The m(6)A Writer: Rise of a Machine for Growing Tasks**.
513 *Biochemistry* 2019, **58**(5):363-378.
514 28. Bawankar P, Lence T, Paolantoni C, Haussmann IU, Kazlauskiene M, Jacob D,
515 Heidelberger JB, Richter FM, Nallasivan MP, Morin V *et al*: **Hakai is required for**
516 **stabilization of core components of the m6A mRNA methylation machinery**.
517 *Nature Communications* 2021, **12**(1):3778.
518 29. Van Lopik J, Trapotsi M-A, Hannon GJ, Bornelov S, Czech Nicholson B: **Unistrand**
519 **piRNA clusters are an evolutionary conserved mechanism to suppress**
520 **endogenous retroviruses across the Drosophila genus**. *bioRxiv* 2023.
521 30. Fridrich A, Moran Y: **Some flies do not play ping-pong**. *PLOS Biology* 2023,
522 **21**(6):e3002152.
523 31. Chary S, Hayashi R: **Mechanistic divergence of piRNA biogenesis in Drosophila**.
524 In.: Cold Spring Harbor Laboratory; 2022.
525 32. McDonald JH, Kreitman M: **Adaptive protein evolution at the Adh locus in**
526 **Drosophila**. *Nature* 1991, **351**(6328):652-654.
527 33. Karolchik D: **The UCSC Table Browser data retrieval tool**. *Nucleic Acids Research*
528 2004, **32**(90001):493D-496.
529 34. Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, Haussler,
530 David: **The Human Genome Browser at UCSC**. *Genome Research* 2002,
531 **12**(6):996-1006.
532 35. Consortium GO: **The Gene Ontology resource: enriching a GOld mine**. *Nucleic*
533 *Acids Res* 2021, **49**(D1):D325-d334.
534 36. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP,
535 Dolinski K, Dwight SS, Eppig JT *et al*: **Gene ontology: tool for the unification of**
536 **biology. The Gene Ontology Consortium**. *Nat Genet* 2000, **25**(1):25-29.

537 37. Mi H, Muruganujan A, Ebert D, Huang X, Thomas PD: **PANTHER version 14:**
538 **more genomes, a new PANTHER GO-slim and improvements in enrichment**
539 **analysis tools**. *Nucleic Acids Research* 2018, **47**(D1):D419-D426.
540 38. Sanfilippo P, Wen J, Lai EC: **Landscape and evolution of tissue-specific**
541 **alternative polyadenylation across Drosophila species**. *Genome Biology* 2017,
542 **18**(1).
543 39. Smolko AE, Shapiro-Kulnane L, Salz HK: **The H3K9 methyltransferase SETDB1**
544 **maintains female identity in Drosophila germ cells**. *Nature Communications* 2018,
545 **9**(1).
546 40. Vankuren NW, Vibranovski MD: **A Novel Dataset for Identifying Sex-Biased**
547 **Genes in Drosophila**. *Journal of Genomics* 2014, **2**:64-67.
548 41. Story B, Ma X, Ishihara K, Li H, Hall K, Peak A, Anoja P, Park J, Haug J, Blanchette
549 M *et al*: **Defining the expression of piRNA and transposable elements in**
550 **Drosophila ovarian germline stem cells and somatic support cells**. *Life Science*
551 *Alliance* 2019, **2**(5):e201800211.
552 42. Eastwood EL, Jara KA, Bornelöv S, Munafò M, Frantzis V, Kneuss E, Barbar EJ,
553 Czech B, Hannon GJ: **Dimerisation of the PICTS complex via LC8/Cut-up drives**
554 **co-transcriptional transposon silencing in Drosophila**. *eLife* 2021, **10**:e65557.
555 43. Treiber CD, Waddell S: **Transposon expression in the Drosophila brain is driven**
556 **by neighboring genes and diversifies the neural transcriptome**. *Genome Research*
557 2020, **30**(11):1559-1569.
558 44. Laverty C, Lucci J, Akhtar A: **The MSL complex: X chromosome and beyond**.
559 *Current Opinion in Genetics & Development* 2010, **20**(2):171-178.
560 45. Straub T, Becker PB: **Dosage compensation: the beginning and end of**
561 **generalization**. *Nature Reviews Genetics* 2007, **8**(1):47-57.

562 46. Brennecke J, Malone CD, Aravin AA, Sachidanandam R, Stark A, Hannon GJ: **An**
563 **Epigenetic Role for Maternally Inherited piRNAs in Transposon Silencing**.
564 *Science* 2008, **322**(5906):1387-1392.
565 47. Luo Y, He P, Kanrar N, Fejes Toth K, Aravin AA: **Maternally inherited siRNAs**
566 **initiate piRNA cluster formation**. *Mol Cell* 2023.
567 48. Parhad SS, Yu T, Zhang G, Rice NP, Weng Z, Theurkauf WE: **Adaptive Evolution**
568 **Targets a piRNA Precursor Transcription Network**. *Cell Reports* 2020,
569 **30**(8):2672-2685.e2675.
570 49. Parhad SS, Tu S, Weng Z, Theurkauf WE: **Adaptive Evolution Leads to Cross-**
571 **Species Incompatibility in the piRNA Transposon Silencing Machinery**.
572 *Developmental Cell* 2017, **43**(1):60-70.e65.
573 50. Kelleher ES, Edelman NB, Barbash DA: **Drosophila Interspecific Hybrids**
574 **Phenocopy piRNA-Pathway Mutants**. *PLoS Biology* 2012, **10**(11):e1001428.
575 51. Vermaak D, Henikoff S, Malik HS: **Positive Selection Drives the Evolution of**
576 **rhino, a Member of the Heterochromatin Protein 1 Family in Drosophila**. *PLoS*
577 *Genetics* 2005, **1**(1):e9.
578 52. Obbard DJ, Finnegan DJ: **RNA Interference: Endogenous siRNAs Derived from**
579 **Transposable Elements**. *Current Biology* 2008, **18**(13):R561-R563.
580 53. Blumenstiel JP, Erwin AA, Hemmer LW: **What Drives Positive Selection in the**
581 **Drosophila piRNA Machinery? The Genomic Autoimmunity Hypothesis**. *Yale J*
582 *Biol Med* 2016, **89**(4):499-512.
583 54. Wang L, Barbash DA, Kelleher ES: **Adaptive evolution among cytoplasmic piRNA**
584 **proteins leads to decreased genomic auto-immunity**. *PLOS Genetics* 2020,
585 **16**(6):e1008861.

586 55. Parhad SS, Theurkauf WE: **Rapid evolution and conserved function of the piRNA**
587 **pathway**. *Open Biology* 2019, **9**(1):180181.
588 56. Lawlor MA, Ellison CE: **Evolutionary dynamics between transposable elements**
589 **and their host genomes: mechanisms of suppression and escape**. *Current Opinion*
590 *in Genetics & Development* 2023, **82**:102092.
591 57. Soller M, Haussmann IU, Hollmann M, Choffat Y, White K, Kubli E, Schäfer MA:
592 **Sex-peptide-regulated female sexual behavior requires a subset of ascending**
593 **ventral nerve cord neurons**. *Curr Biol* 2006, **16**(18):1771-1782.
594 58. Fan Y, Soller M, Flister S, Hollmann M, Müller M, Bello B, Egger B, White K,
595 Schäfer MA, Reichert H: **The egghead gene is required for compartmentalization**
596 **in Drosophila optic lobe development**. *Dev Biol* 2005, **287**(1):61-73.
597 59. Seeholzer LF, Seppo M, Stern DL, Ruta V: **Evolution of a central neural circuit**
598 **underlies Drosophila mate preferences**. *Nature* 2018, **559**(7715):564-569.
599 60. Main BJ, Smith AD, Jang H, Nuzhdin SV: **Transcription Start Site Evolution in**
600 **Drosophila**. *Molecular Biology and Evolution* 2013, **30**(8):1966-1974.
601 61. Tamura K, Stecher G, Kumar S: **MEGA11: Molecular Evolutionary Genetics**
602 **Analysis Version 11**. *Molecular Biology and Evolution* 2021, **38**(7):3022-3027.
603 62. Lack JB, Cardeno CM, Crepeau MW, Taylor W, Corbett-Detig RB, Stevens KA,
604 Langley CH, Pool JE: **The Drosophila Genome Nexus: A Population Genomic**
605 **Resource of 623 Drosophila melanogaster Genomes, Including 197 from a Single**
606 **Ancestral Range Population**. *Genetics* 2015, **199**(4):1229-1241.
607 63. Lack JB, Lange JD, Tang AD, Corbett-Detig RB, Pool JE: **A Thousand Fly**
608 **Genomes: An Expanded Drosophila Genome Nexus**. *Mol Biol Evol* 2016,
609 **33**(12):3308-3313.

610 64. Mackay TFC, Richards S, Stone EA, Barbadilla A, Ayroles JF, Zhu D, Casillas S,
611 Han Y, Magwire MM, Cridland JM *et al*: **The Drosophila melanogaster Genetic**
612 **Reference Panel**. *Nature* 2012, **482**(7384):173-178.
613 65. Gramates LS, Agapite J, Attrill H, Calvi BR, Crosby MA, dos Santos G, Goodman
614 JL, Goutte-Gattat D, Jenkins VK, Kaufman T *et al*: **FlyBase: a guided tour of**
615 **highlighted features**. *Genetics* 2022, **220**(4).
616 66. Tang H, Lewontin RC: **Locating Regions of Differential Variability in DNA and**
617 **Protein Sequences**. *Genetics* 1999, **153**(1):485-495.
618

619 **Figure Legends**

620 **Fig. 1: Promoter regions of germ cell-expressed piRNA processing factors are hot spots**
621 **for rapid evolution**

622 A and B) Schematic depiction of somatic and germ cell piRNA factors involved in co-
623 transcriptional gene silencing and piRNA biogenesis (A) and germline-specific piRNA
624 pathway genes involved in dual-strand cluster regulation and the ping-pong cycle (B) in the
625 *Drosophila* ovary.

626 C and D) PhyloP27way conservation score averages (C) and PhyloP27way conservation
627 promoter d^P scores (D) for the 350 nucleotide promoter regions of the somatic and germ cell
628 piRNA factors and the germ cell-specific piRNA factors compared to all genes in the
629 *Drosophila* genome. Statistically significant differences from unpaired student t-tests (C) and
630 non-parametric chi-squared tests (D) are indicated by asterisks (*** $p \leq 0.001$, **** $p \leq 0.0001$
631 following Bonferroni correction).

632 E-J) Heatmaps indicating sequence (F, purple), indel (H, blue) or base change (J, green)
633 accumulation, and their quantification (E-I) depicted as *d* scores for each of the analysed gene
634 groups compared to the control among closely related *D. melanogaster*, *D. simulans*, *D.*

635 *sechellia*, *D. yakuba* and *D. erecta*. Regions of 1000 nucleotides upstream and 300 nucleotides
636 downstream of the TSS were analysed. Analysis was performed for the control group (m^6A
637 writer complex and readers), the somatic and germ cell piRNA factors, and the germ cell-
638 specific piRNA factors. The blue line indicates the promoter region used for quantification of
639 the substitution rate based on the gene and intergenic region schematic. Statistically significant
640 differences from non-parametric chi-squared tests are indicated by asterisks (**** $p \leq 0.0001$
641 following Bonferroni correction).
642
643 **Fig. 2: Promoters of genes required for germline transposon silencing evolve fast**
644 A) PhyloP27way nucleotide scores for the top 100 genes affecting transposon silencing in germ
645 cells based on transposon derepression z scores and ordered by promoter conservation d^P
646 scores. A region of 1000 nucleotides upstream and 300 nucleotides downstream are shown.
647 Red represents lower conservation while blue represents higher.
648 B) Average PhyloP27way nucleotide scores for the top 100 genes affecting transposon
649 silencing in germ cells compared to the genome average.
650 C) Comparison of PhyloP27way conservation promoter scores (d^P) for each of the 100 genes
651 affecting transposon silencing in germ cells compared to the genome average. Statistically
652 significant differences from non-parametric chi-squared tests are indicated by asterisks (***
653 $p \leq 0.001$ following Bonferroni correction).
654 D) Representative GO analysis for the top 100 gene candidates affecting transposon silencing.
655
656 **Fig. 3: Promoters of FEPG genes and their associated complexes are hotspots for rapid**
657 **evolution**
658 A) Individual gene comparison of PhyloP27way conservation promoter scores (d^P) for the top
659 100 genes affecting transposon silencing in germ cells compared to the genome average.

660 Statistically significant differences from non-parametric chi-squared tests are indicated by
661 asterisks (* p≤0.05, ** p≤0.01, *** p≤0.001, **** p≤0.0001 following Bonferroni correction).
662 B-D) Sequence analysis for change accumulation among closely related *D. melanogaster*, *D.*
663 *simulans*, *D. sechellia*, *D. yakuba* and *D. erecta*. Comparisons were performed for sequence
664 change (B), indel (C), and base change (D) *d* scores for each of the positive 49 genes affecting
665 transposon silencing in FEPG gene group compared to the somatic and germ cell piRNA
666 factors. Statistically significant differences from non-parametric chi-squared tests are indicated
667 by asterisks (**** p≤0.0001 following Bonferroni correction).
668 E) Comparison of *d* scores for protein complex genes involved in germ cell transposon
669 silencing compared to the somatic and germ cell piRNA factors. Statistically significant
670 differences from non-parametric chi-squared tests are indicated by asterisks (**** p≤0.0001
671 following Bonferroni correction).
672 F) Heatmaps indicating sequence change accumulation for protein complex genes involved in
673 germ cell transposon silencing compared to the piRNA factors shown in purple among closely
674 related *D. melanogaster*, *D. simulans*, *D. sechellia*, *D. yakuba* and *D. erecta*. Regions of 1000
675 nucleotides upstream and downstream of the TSS were analysed. The blue line indicates the
676 promoter region used for quantification of the substitution rate.
677
678 **Fig. 4: Analysis of differential gene expression of FEPG genes in *D. melanogaster* and *D.***
679 ***yakuba* ovaries**
680 A) Volcano plot showing differential gene expression (DESeq2) between *D. melanogaster* and
681 *D. yakuba* ovaries (RNA-seq data; Supplementary Dataset 1) for FEPG (blue) and the somatic
682 and germ cell piRNA factors (red). Horizontal dashed lines indicate significance thresholds
683 (adjusted p≤0.05).

684 B-D) Alignment of the promoter regions of *CG32152, MEP-1* and *Rbbp5* for *D. melanogaster*
685 and *D. yakuba* with indels (red) and base changes (blue) indicated as boxes. The black lines
686 indicate gene 5'UTRs.
687
688 **Fig. 5: Germ cell-specifically expressed, but not somatic cell-specifically expressed gene**
689 **promoters evolve rapidly and are accompanied by divergence in gene expression**
690 A) Ranking of genes using vasa-GFP signal from FACS-sorted vasa-GFP+/- cells of cultured
691 ovary GSC line (vasa-GFP;hs-bam;bamΔ86) (GSE119862) that defines germline-enriched
692 (vasa-GFP+) and soma-enriched (vasa-GFP-) genes.
693 B) Volcano plot showing differential gene expression between the FACS-sorted vasa-GFP+
694 (germline) and vasa-GFP- (somatic) cells from cultured ovary GSC line (vasa-GFP;hs-
695 bam;bamΔ86) (GSE119862).
696 C and D) PhyloP27way nucleotide scores for somatic cell-specific (C) and germ cell-specific
697 (D) expressed genes ordered by promoter conservation d^P scores. A region of 1000 nucleotides
698 upstream and 300 nucleotides downstream are shown. Red represents lower conservation while
699 blue represents higher.
700 E) Average PhyloP27way nucleotide scores for all genes in the *Drosophila* genome, somatic
701 cell and germ cell-specific genes. Purple represents lower conservation while blue represents
702 higher.
703 F) The percentage of somatic cell and germ cell-specific genes with significant promoter
704 nucleotide changes (d^P) versus all genes in the *Drosophila* genome average, divided by
705 significantly increased (low conservation) or decreased (high conservation). Statistically
706 significant differences from non-parametric chi-squared tests are indicated by asterisks (****
707 $p \leq 0.0001$ following Bonferroni correction).

708 G and H) PhyloP27way conservation score averages (G) and PhyloP27way conservation
709 promoter scores (d^P) (H) for the 350 nucleotide promoter regions compared for all genes in the
710 *Drosophila* genome, somatic cell-specifically expressed and germ cell-specifically expressed
711 genes. Statistically significant differences from unpaired student t-tests (G) and from non-
712 parametric chi-squared tests (H) are indicated by asterisks (** $p \leq 0.01$, **** $p \leq 0.0001$
713 following Bonferroni correction).
714
715 **Fig. 6: Analysis of gene expression divergence between germ cell and somatic cell-**
716 **specifically expressed genes**
717 A) Volcano plot of differential gene expression (DESeq2) between *D. melanogaster* and *D.*
718 *yakuba* ovaries (RNA-seq data; Supplementary Dataset 1) for all (grey), germ cell-specific
719 (blue) and somatic cell-specific (red) gene groups. Horizontal dashed lines indicate
720 significance (adjusted $p \leq 0.05$).
721 B) Scatterplot of individual gene d^P scores plotted against fold-change expression differences
722 between *D. melanogaster* and *D. yakuba* ovaries (RNA-seq data; Supplementary Dataset 1) for
723 germ cell-specific (blue) and somatic cell-specific (red) gene groups. The grey dotted lines
724 indicate log2Fold-change threshold values of ≥ 0.5 and ≤ 0.5.
725 C and E) Comparison of the expression distribution and percentage of genes with significant
726 changes in *D. melanogaster* log2FoldChange (≥ 0.5, D) or significant changes in *D. yakuba*
727 log2FoldChange (≤ 0.5, E) expression changes in PhyloP conservation promoter score (d^P)
728 ranges 25-50, >50-75, and >75-100. Expression changes for germ cell-specific and somatic
729 cell-specific gene groups were analysed between *D. melanogaster* and *D. yakuba*. Statistically
730 significant differences from Mann Whitney U tests (C) and non-parametric chi-squared tests
731 (D and E) are indicated by asterisks (** $p \leq 0.01$, *** $p \leq 0.001$, **** $p \leq 0.0001$ following
732 Bonferroni correction).

733

734 **Supplementary Fig. 1: Germ cell-specific piRNA factors are under positive selection and**
735 **accumulate rapid sequence changes in their promoters**

736 A) Individual gene sequence change, indel, and base change *d* scores and statistically
737 significant accumulation of sequence changes for the somatic piRNA factors and germ cell-
738 specific piRNA factors. Statistically significant differences from non-parametric chi-squared
739 tests are indicated by asterisks (* $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$, **** $p \leq 0.0001$ following
740 FDR correction).

741 B) Evolution of coding regions analysed by MKT tests for polymorphisms and divergence
742 between *D. melanogaster* and *D. simulans* for core genes in the somatic and germ cell
743 transposon silencing pathways. R, replacement; S, synonymous. Statistically significant
744 differences from Fisher's Exact Tests are indicated by asterisks (**** $p \leq 0.0001$ following FDR
745 correction).

746

747 **Supplementary Fig. 2: Genes coding for members of a subgroup of protein complexes**
748 **implicated in piRNA silencing accumulate sequence changes in their promoters**

749 Individual gene *d* scores and statistically significant accumulation of sequence changes for each
750 gene separated by protein complex compared to the somatic piRNA factors. Statistically
751 significant differences from non-parametric chi-squared tests are indicated by asterisks (*
752 $p \leq 0.05$, ** $p \leq 0.01$, **** $p \leq 0.0001$ following Bonferroni correction).

753

754 **Supplementary Fig. 3: Promoter evolution of protein complex genes involved in germ cell**
755 **transposon silencing**

756 A and C) Heatmaps indicating indel (A, blue) or base change (C, green) accumulation in protein
757 complex genes involved in germ cell transposon silencing compared to Yb body genes among

758 closely related *D. melanogaster*, *D. simulans*, *D. sechellia*, *D. yakuba* and *D. erecta*. Regions
759 of 1000 nucleotides upstream and 300 nucleotides downstream of the TSS were analysed. The
760 blue line indicates the promoter region used for quantification of the substitution rate.
761 B and D) Comparison of indel (B, blue) or base change (D, green) *d* scores for each of the
762 protein complex gene groups compared to the somatic piRNA factors. Statistically significant
763 differences from non-parametric chi-squared tests are indicated by asterisks (* $p \leq 0.05$, **
764 $p \leq 0.01$, *** $p \leq 0.001$, **** $p \leq 0.0001$ following Bonferroni correction).
765 E) Individual gene *d* scores and statistically significant accumulation of indels or base changes
766 for each gene separated by protein complex compared to the somatic piRNA factors.
767 Comparison of indel (B, blue) or base change (D, green) *d* scores for each of the protein
768 complex gene groups compared to the control group. Statistically significant differences from
769 non-parametric chi-squared tests are indicated by asterisks (* $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$,
770 **** $p \leq 0.0001$ following Bonferroni correction).
771
772 **Supplementary Fig. 4: Coding sequences evolution of protein complex genes involved in**
773 **germ cell transposon silencing.**
774 A) Impact of protein complex genes involved in germ cell transposon silencing on transposon
775 de-repression. Unprocessed data were taken from [1]. Derepression scores (*z*) were
776 concatenated for protein complexes in each transposon type (HeTA, TAHRE, blood, burdock)
777 and an average of all.
778 B) Evolution of coding regions analysed by MKT tests for polymorphisms and divergence
779 between *D. melanogaster* and *D. simulans* for individual genes from the selected germline
780 transposon silencing implicated protein complexes. R, replacement; S, synonymous.
781 Statistically significant differences from Fisher's Exact Tests are indicated by asterisks (*
782 $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$, **** $p \leq 0.0001$ following FDR correction).

Figure 1

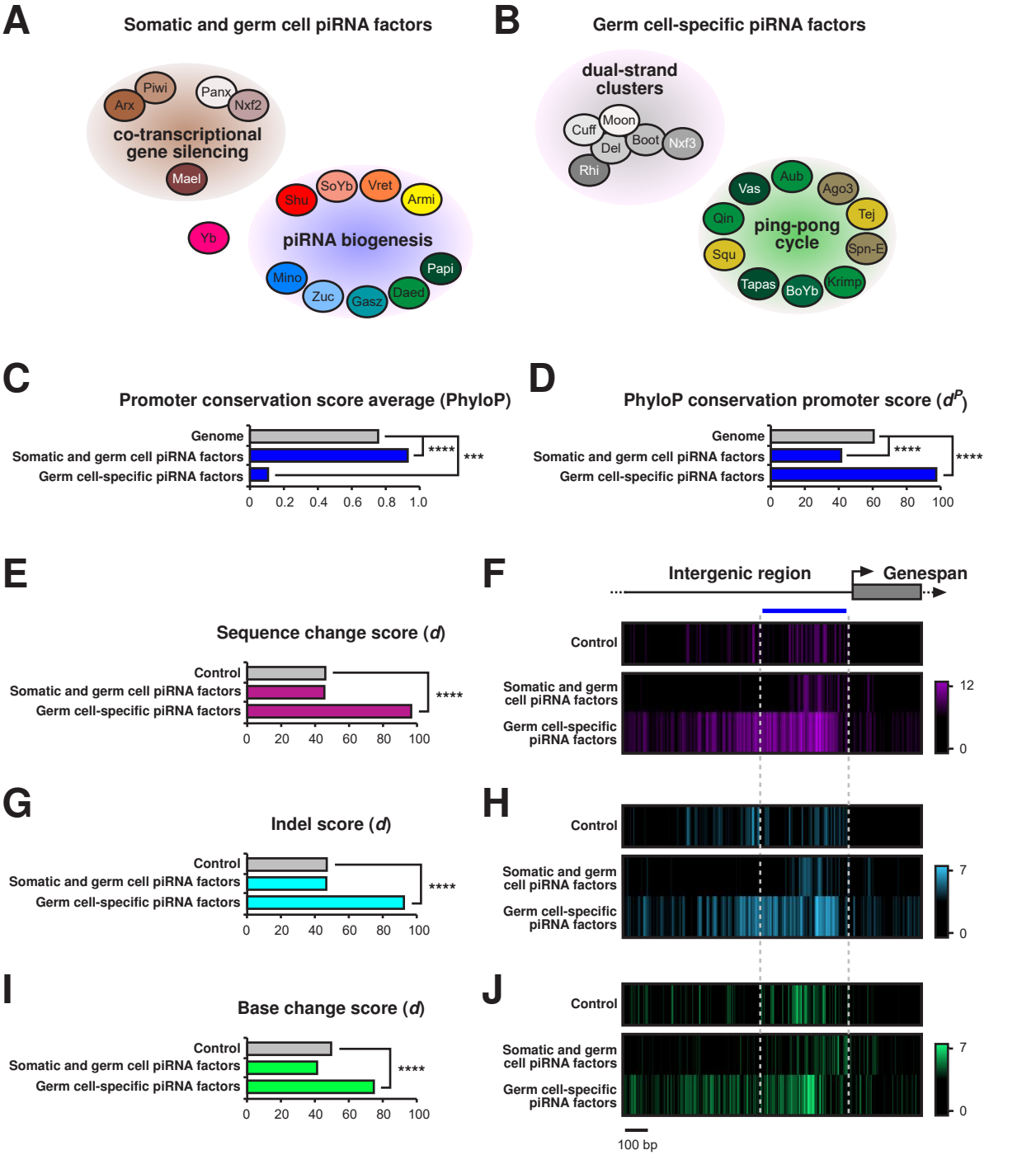

A
Somatic and germ cell piRNA factors
Piwi
Panx
Arx
Nxf2
co-transcriptional gene silencing
Mael
Shu
SoYb
Vret
Armi
Yb
piRNA biogenesis
Papi
Mino
Zuc
Gasz
Daed
B
Germ cell-specific piRNA factors
dual-strand clusters
Moon
Cuff
Boot
Del
Nxf3
Rhi
Vas
Aub
Ago3
Qin
Tej
ping-pong cycle
Spn-E
Squ
Tapas
BoYb
Krimp
C
Promoter conservation score average (PhyloP)
Genome
Somatic and germ cell piRNA factors
Germ cell-specific piRNA factors
****
***
0
0.2
0.4
0.6
0.8
1.0
D
PhyloP conservation promoter score (d^P)
Genome
Somatic and germ cell piRNA factors
Germ cell-specific piRNA factors
****
****
0
20
40
60
80
100
E
Sequence change score (d)
Control
Somatic and germ cell piRNA factors
Germ cell-specific piRNA factors
****
0
20
40
60
80
100
F
Intergenic region
Genespan
Control
Somatic and germ cell piRNA factors
Germ cell-specific piRNA factors
12
0
G
Indel score (d)
Control
Somatic and germ cell piRNA factors
Germ cell-specific piRNA factors
****
0
20
40
60
80
100
H
Control
Somatic and germ cell piRNA factors
Germ cell-specific piRNA factors
7
0
I
Base change score (d)
Control
Somatic and germ cell piRNA factors
Germ cell-specific piRNA factors
****
0
20
40
60
80
100
J
Control
Somatic and germ cell piRNA factors
Germ cell-specific piRNA factors
7
0
100 bp

Figure 2

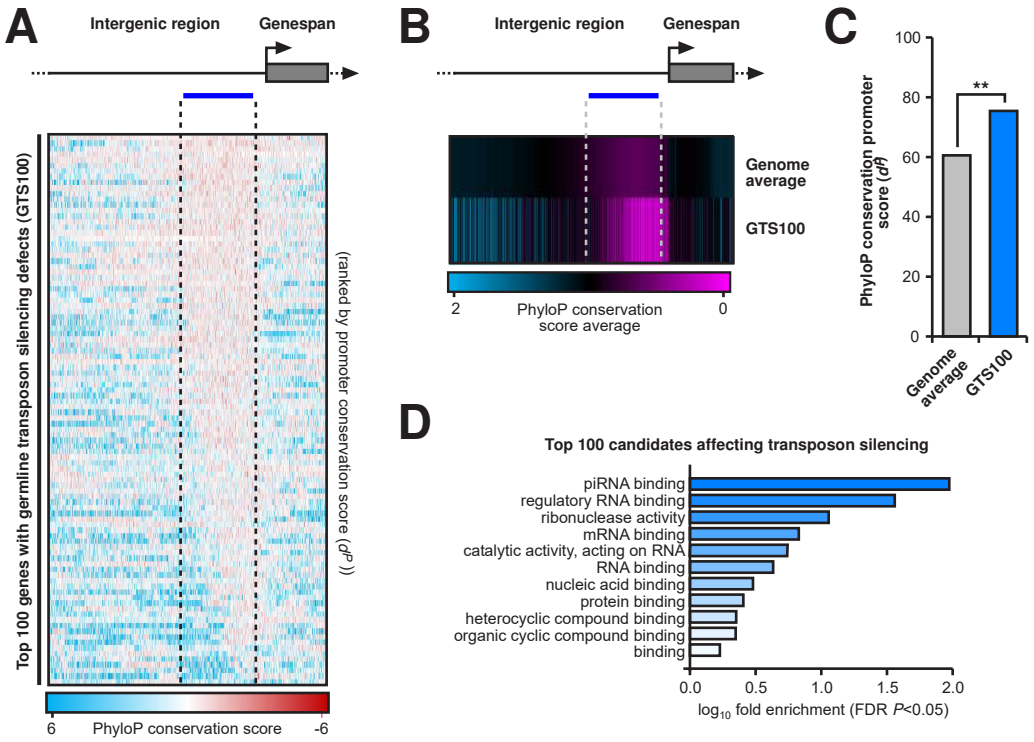

A
Intergenic region
Genespan
Top 100 genes with germline transposon silencing defects (GTS100)
(ranked by promoter conservation score (d^P))
6
PhyloP conservation score
-6
B
Intergenic region
Genespan
Genome average
GTS100
2
PhyloP conservation score average
0
C
PhyloP conservation promoter score (d^P)
100
80
60
40
20
0
**
Genome average
GTS100
D
Top 100 candidates affecting transposon silencing
piRNA binding
regulatory RNA binding
ribonuclease activity
mRNA binding
catalytic activity, acting on RNA
RNA binding
nucleic acid binding
protein binding
heterocyclic compound binding
organic cyclic compound binding
binding
0.0
0.5
1.0
1.5
2.0
$\log_{10}$ fold enrichment (FDR P<0.05)

Figure 3

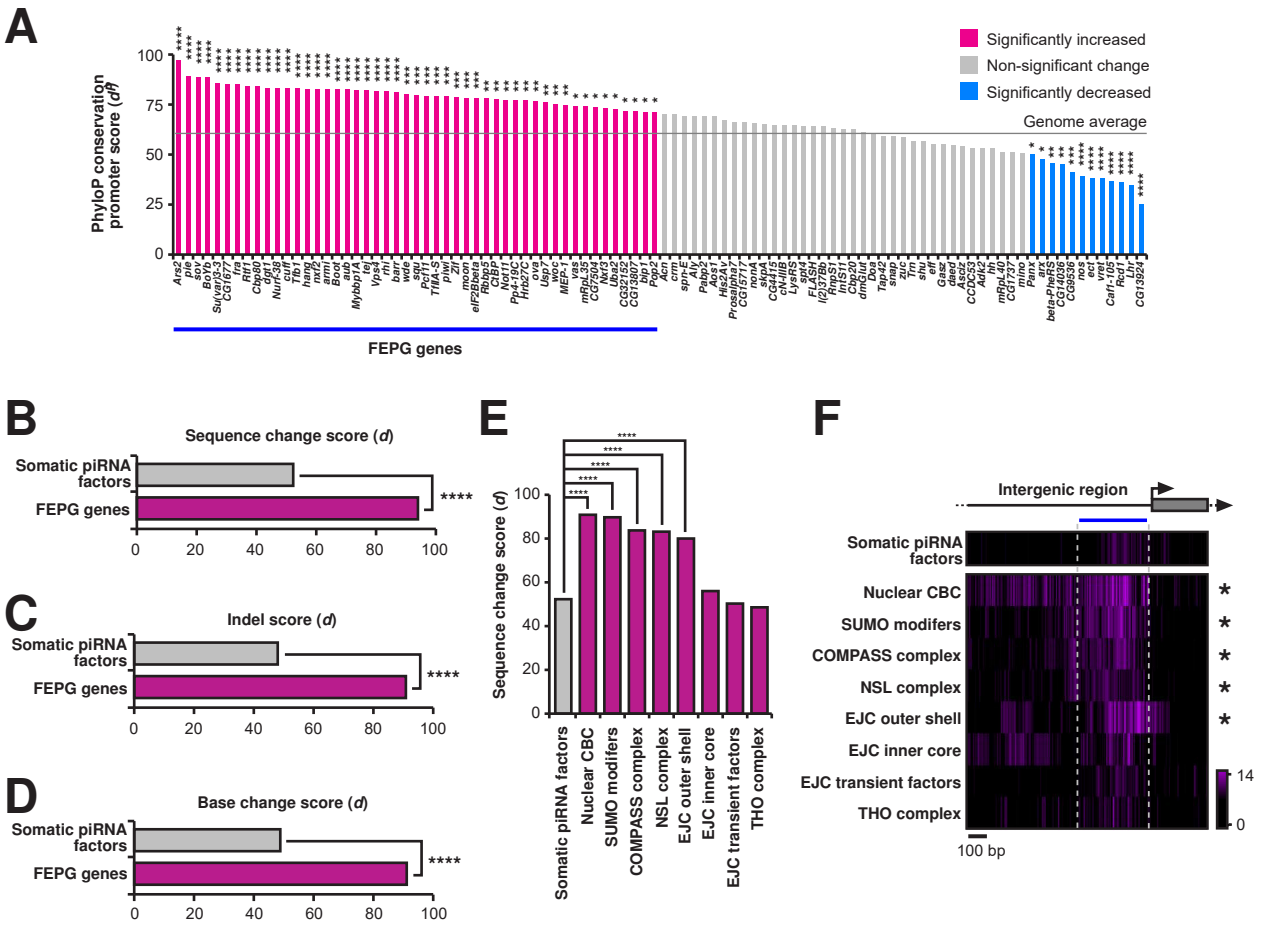

A
PhyloP conservation promoter score (d)
Significantly increased
Non-significant change
Significantly decreased
Genome average
FEPG genes
B
Sequence change score (d)
Somatic piRNA factors
FEPG genes
C
Indel score (d)
D
Base change score (d)
E
Sequence change score (d)
Somatic piRNA factors
Nuclear CBC
SUMO modifers
COMPASS complex
NSL complex
EJC outer shell
EJC inner core
EJC transient factors
THO complex
F
Intergenic region
100 bp

Figure 4

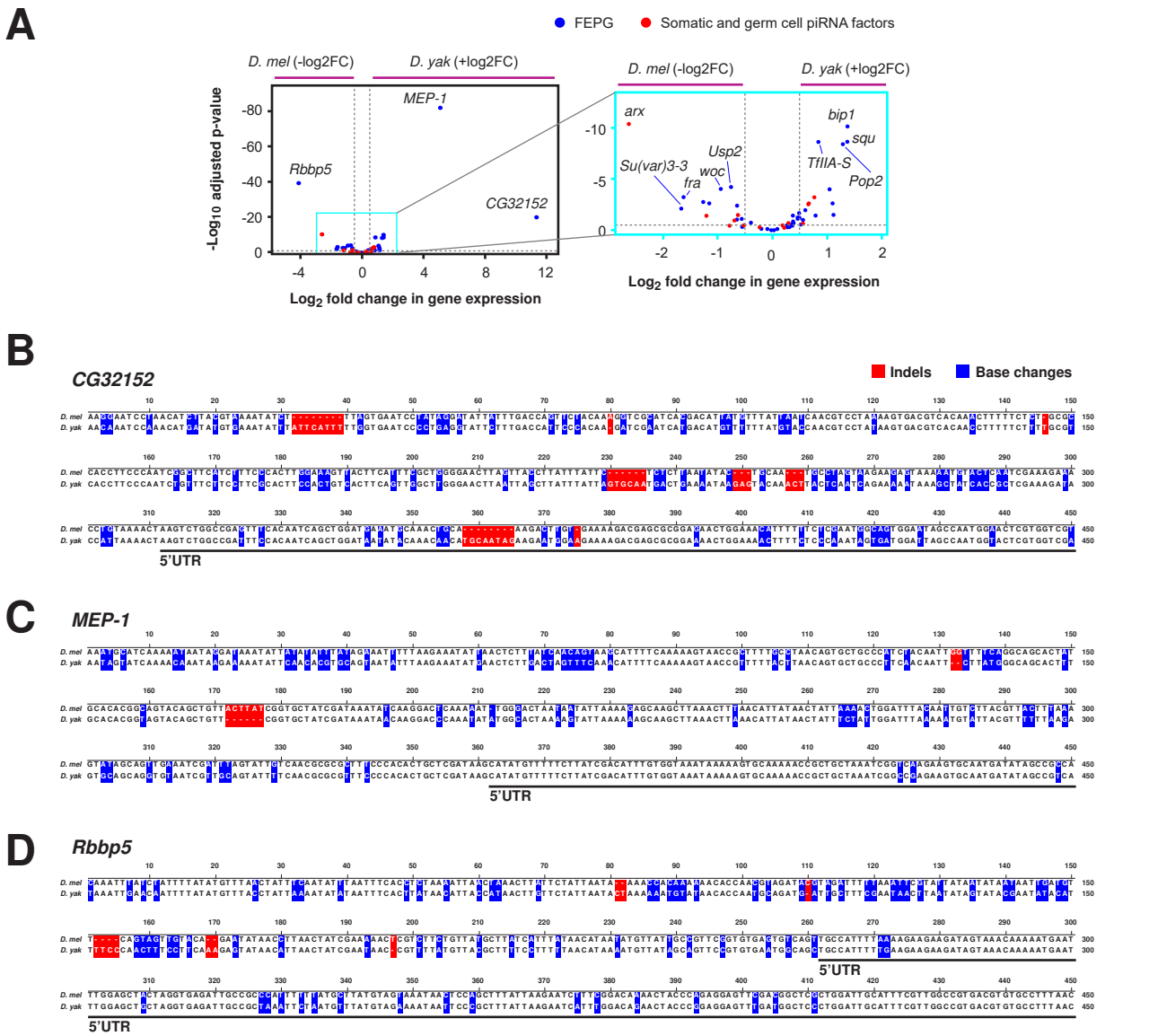

A
FEPG
Somatic and germ cell piRNA factors
D. mel (-log2FC)
D. yak (+log2FC)
-Log10 adjusted p-value
MEP-1
Rbbp5
CG32152
Log2 fold change in gene expression
arx
Su(var)3-3
fra
woc
Usp2
bip1
squ
TfIIA-S
Pop2
B
CG32152
Indels
Base changes
5'UTR
C
MEP-1
5'UTR
D
Rbbp5
5'UTR
5'UTR

Figure 5

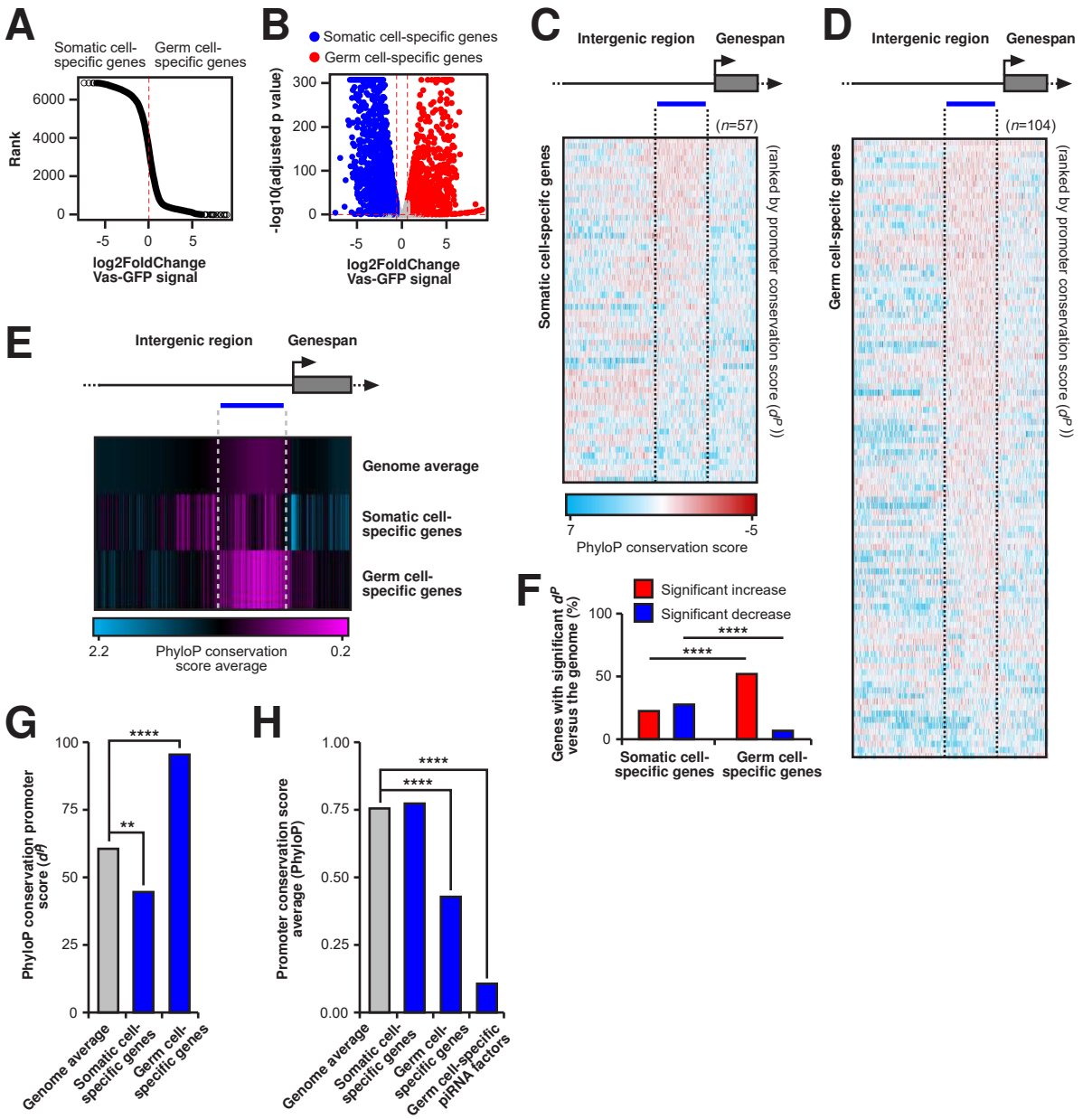
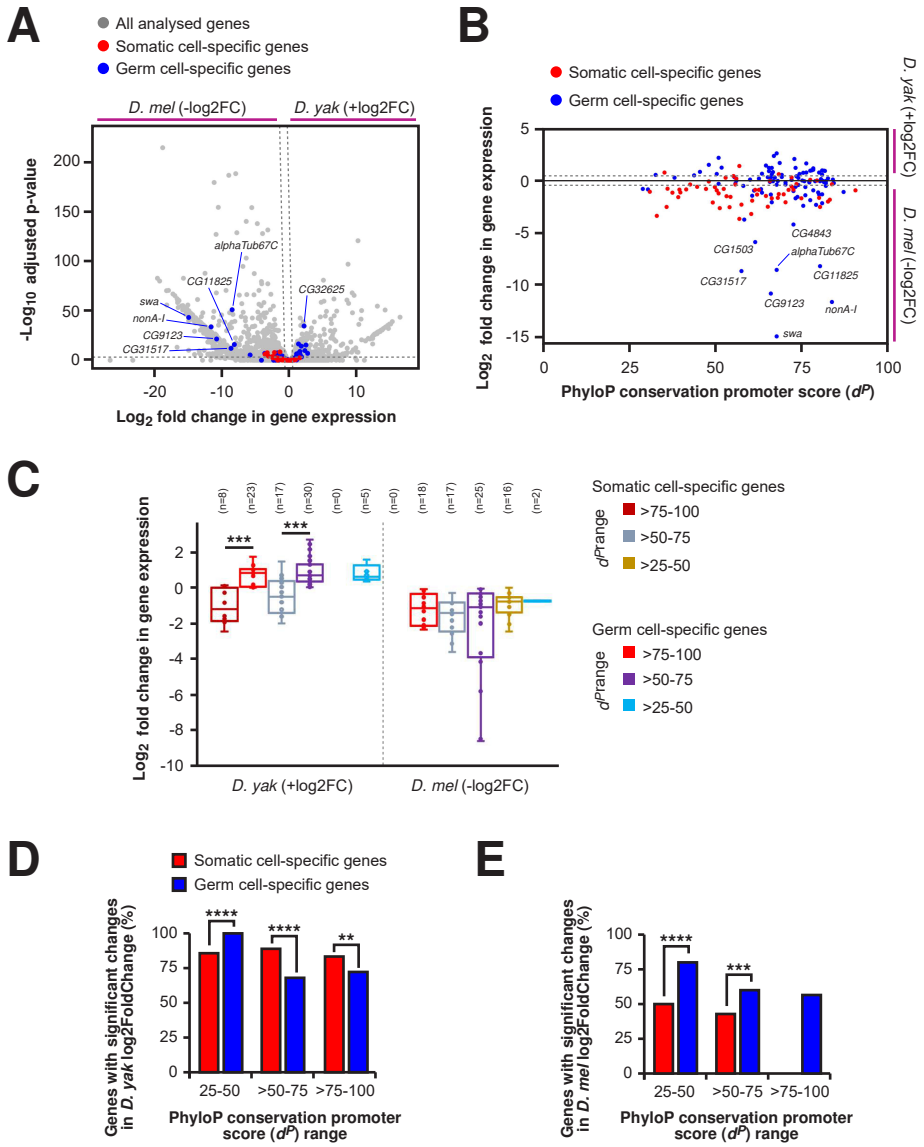
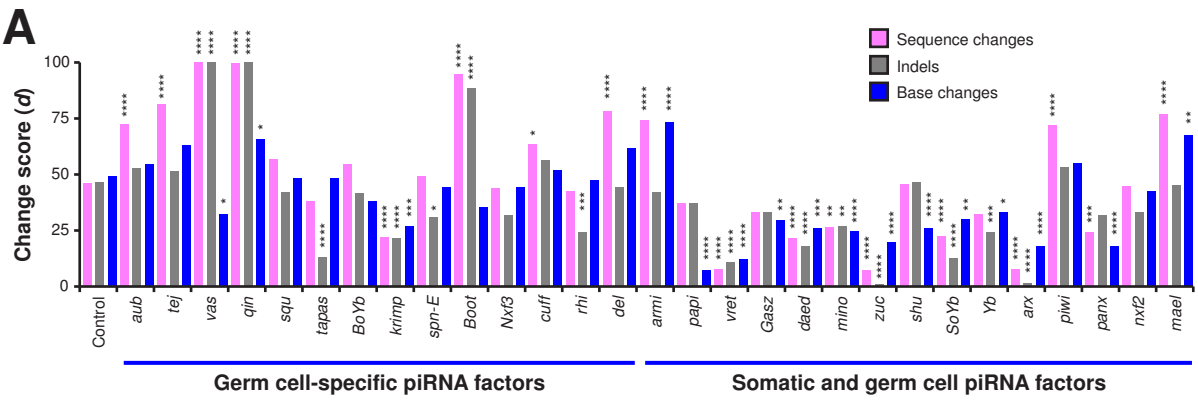

A
Somatic cell-specific genes
Germ cell-specific genes
Rank
6000
4000
2000
0
-5
0
5
log2FoldChange Vas-GFP signal
B
Somatic cell-specific genes
Germ cell-specific genes
-log10(adjusted p value)
300
200
100
0
-5
0
5
log2FoldChange Vas-GFP signal
C
Intergenic region
Genespan
(n=57)
Somatic cell-specific genes
(ranked by promoter conservation score (dP))
7
-5
PhyloP conservation score
D
Intergenic region
Genespan
(n=104)
Germ cell-specific genes
(ranked by promoter conservation score (dP))
E
Intergenic region
Genespan
Genome average
Somatic cell-specific genes
Germ cell-specific genes
2.2
PhyloP conservation score average
0.2
F
Significant increase
Significant decrease
Genes with significant dP versus the genome (%)
100
50
0
****
****
Somatic cell-specific genes
Germ cell-specific genes
G
PhyloP conservation promoter score (dP)
100
75
50
25
0
****
**
Genome average
Somatic cell-specific genes
Germ cell-specific genes
H
Promoter conservation score average (PhyloP)
1.00
0.75
0.50
0.25
0.00
****
****
Genome average
Somatic cell-specific genes
Germ cell-specific genes
Germ cell-specific piRNA factors

Figure 6


A
All analysed genes
Somatic cell-specific genes
Germ cell-specific genes
D. mel (-log2FC)
D. yak (+log2FC)
-Log10 adjusted p-value
Log2 fold change in gene expression
alphaTub67C
CG11825
CG32625
swa
nonA-I
CG9123
CG31517
B
Somatic cell-specific genes
Germ cell-specific genes
Log2 fold change in gene expression
PhyloP conservation promoter score (dP)
CG4843
CG1503
alphaTub67C
CG11825
CG31517
CG9123
nonA-I
swa
D. yak (+log2FC)
D. mel (-log2FC)
C
Log2 fold change in gene expression
D. yak (+log2FC)
D. mel (-log2FC)
Somatic cell-specific genes
dPrange
>75-100
>50-75
>25-50
Germ cell-specific genes
dPrange
>75-100
>50-75
>25-50
D
Somatic cell-specific genes
Germ cell-specific genes
Genes with significant changes in D. yak log2FoldChange (%)
25-50
>50-75
>75-100
PhyloP conservation promoter score (dP) range
E
Genes with significant changes in D. mel log2FoldChange (%)
25-50
>50-75
>75-100
PhyloP conservation promoter score (dP) range

Supplementary Figure 1

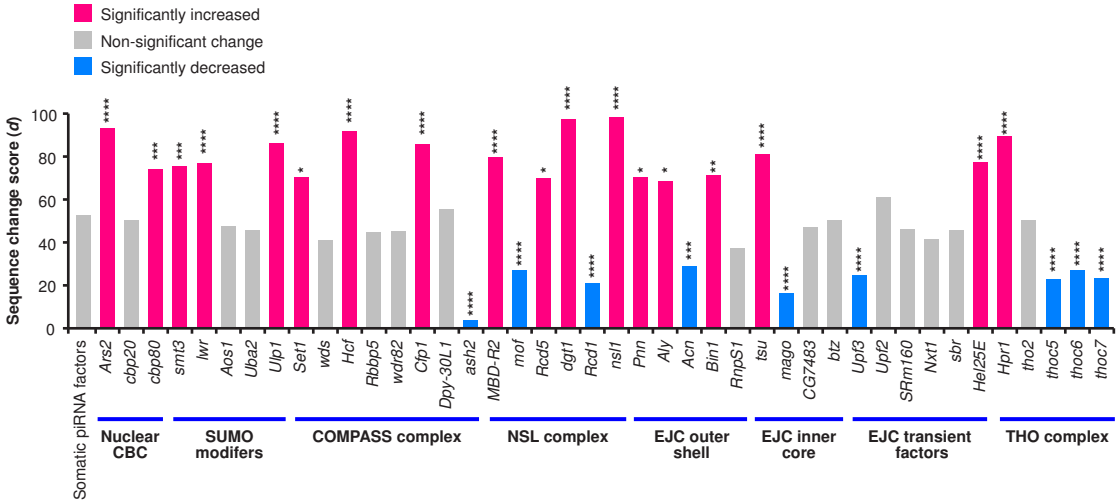
A



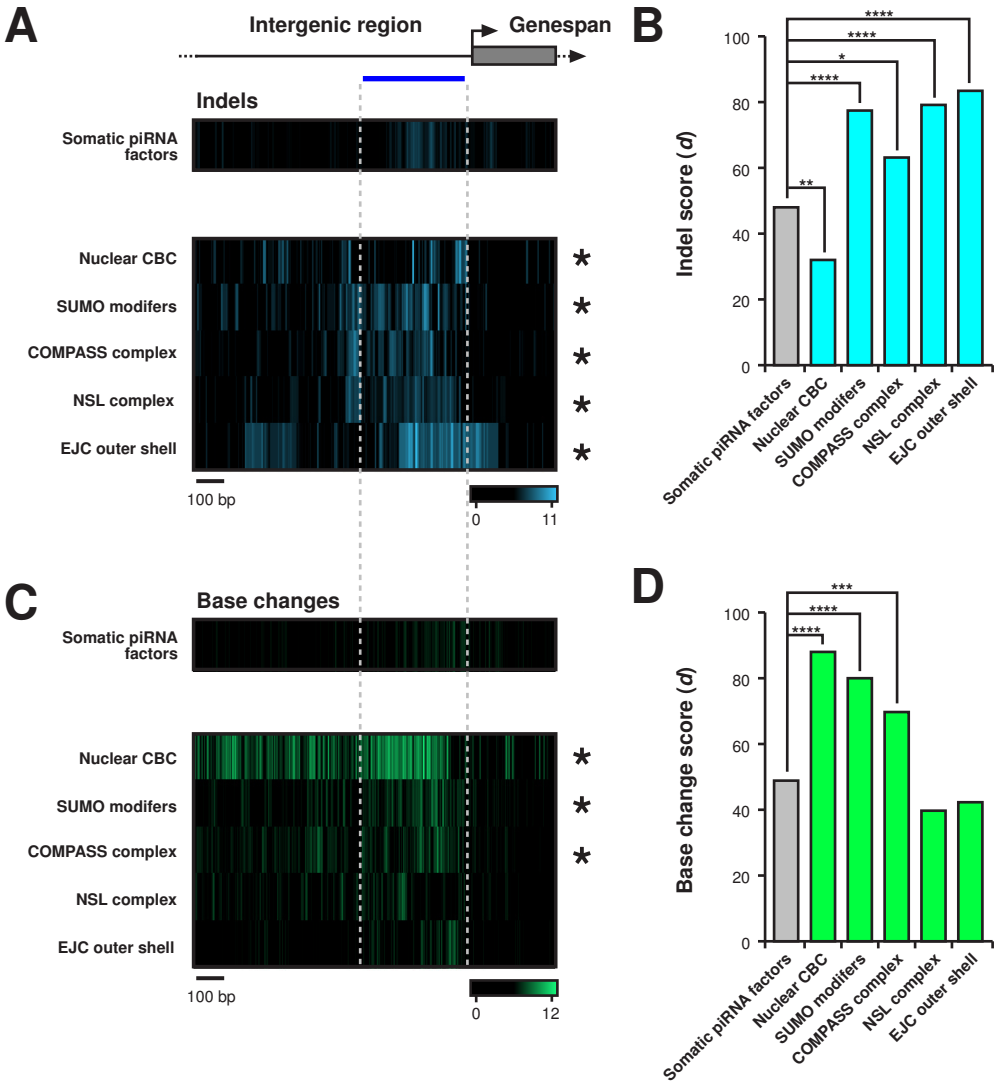
B

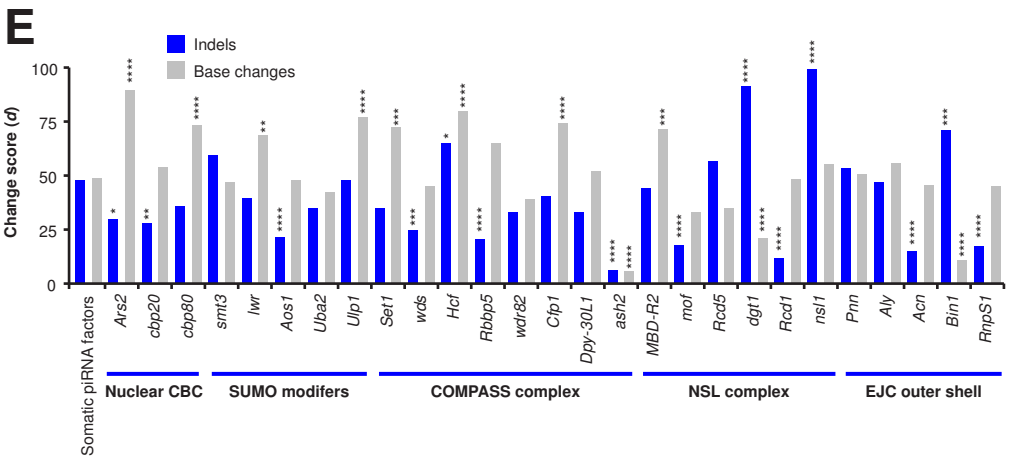
	Population	Polymorphism R	Polymorphism S	Polymorphism R/S	Divergence R	Divergence S	Divergence R/S	Fisher's exact *P*-value (FDR)	
Germ cell-specific piRNA factors	Congo	133	102	1.304	1271	519	2.449	0.000023	****
	Zambia	645	384	1.680	1204	473	2.545	0.000004	****
Somatic and germ cell piRNA factors	Congo	149	97	1.536	571	400	1.428	0.663171	
	Zambia	588	446	1.318	530	338	1.568	0.090960	

Supplementary Figure 2



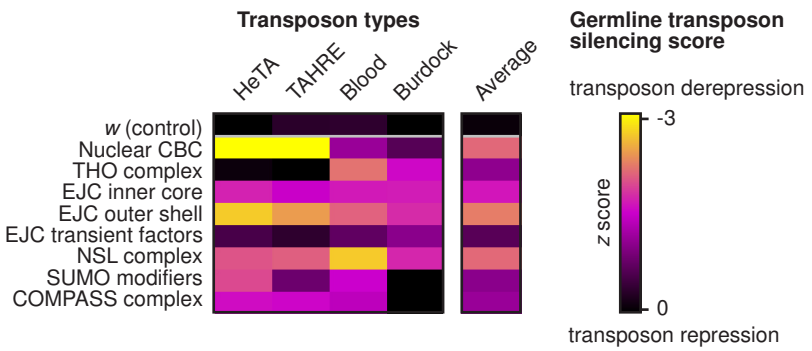
Supplementary Figure 3





Supplementary Figure 4

A



B

	Population	Polymorphism R	Polymorphism S	Polymorphism R/S	Divergence R	Divergence S	Divergence R/S	Fisher's exact *P*-value (FDR)	
Nuclear CBC	Congo	9	20	0.450	86	106	0.811	0.330182	
	Zambia	25	67	0.373	85	98	0.867	0.005210	**
THO complex	Congo	24	29	0.828	77	124	0.621	0.573867	
	Zambia	102	136	0.750	70	108	0.648	0.594249	
EJC inner core	Congo	7	11	0.636	26	52	0.500	0.895886	
	Zambia	32	68	0.471	24	46	0.522	0.926331	
EJC outer shell	Congo	15	26	0.577	274	135	2.030	0.000533	***
	Zambia	64	74	0.865	268	127	2.110	0.000042	****
EJC transient factors	Congo	18	66	0.273	110	124	0.887	0.000157	***
	Zambia	100	259	0.386	100	103	0.971	0.000003	****
NSL complex	Congo	42	62	0.677	366	253	1.447	0.000914	***
	Zambia	206	321	0.642	345	220	1.568	7.44E-12	****
SUMO modifiers	Congo	17	19	0.895	349	147	2.374	0.014222	*
	Zambia	94	97	0.969	337	133	2.534	0.000001	****
COMPASS complex	Congo	6	41	0.146	32	90	0.356	0.106720	
	Zambia	47	124	0.379	29	73	0.397	0.889564	