piRNAs derived from ancient viral processed pseudogenes as transgenerational sequence-specific immune memory in mammals

Nicholas F. Parrish,1,8,9 Kan Fujino,1,8,11 Yusuke Shiromoto,2,10 Yuka W. Iwasaki,3 Hongseok Ha,4 Jinchuan Xing,4 Akiko Makino,1,5 Tomoyuki Honda,1,6 and Keizo Tomonaga1,6,7

1Department of Viral Oncology, Institute for Virus Research, Kyoto University, Kyoto 606-8507, Japan
2Department of Pathology, Medical School and Graduate School of Frontier Biosciences, Osaka University, Osaka 565-0871, Japan
3Department of Molecular Biology, Keio University School of Medicine, Tokyo 160-8582, Japan
4Department of Genetics, Rutgers, The State University of New Jersey, Piscataway, New Jersey 08854, USA
5Center for Emerging Virus Research, Institute for Virus Research, Kyoto University, Kyoto 606-8507, Japan
6Department of Tumor Viruses, Graduate School of Medicine, Kyoto University, Kyoto 606-8507, Japan
7Department of Mammalian Regulatory Network, Graduate School of Biostudies, Kyoto University, Kyoto 606-8507, Japan

ABSTRACT
Endogenous bornavirus-like nucleoprotein elements (EBLNs) are sequences within vertebrate genomes derived from reverse transcription and integration of ancient bornaviral nucleoprotein mRNA via the host retrotransposon machinery. While species with EBLNs appear relatively resistant to bornaviral disease, the nature of this association is unclear. We hypothesized that EBLNs could give rise to antiviral interfering RNA in the form of PIWI-interacting RNAs (piRNAs), a class of small RNA known to silence transposons but not exogenous viruses. We found that in both rodents and primates, which acquired their EBLNs independently some 25–40 million years ago, EBLNs are present within piRNA-generating regions of the genome far more often than expected by chance alone ($P = 8 \times 10^{-3}$–$6 \times 10^{-8}$). Three of the seven human EBLNs fall within annotated piRNA clusters and two marmoset EBLNs give rise to bona fide piRNAs. In both rats and mice, at least two of the five EBLNs give rise to abundant piRNAs in the male gonad. While no EBLNs are syntenic between rodent and primate, some of the piRNA clusters containing EBLNs are; thus we deduce that EBLNs were integrated into existing piRNA clusters. All true piRNAs derived from EBLNs are antisense relative to the proposed ancient bornaviral nucleoprotein mRNA. These observations are consistent with a role for EBLN-derived piRNA-like RNAs in interfering with ancient bornaviral infection. They raise the hypothesis that retrotransposon-dependent virus-to-host gene flow could engender RNA-mediated, sequence-specific antiviral immune memory in metazoans analogous to the CRISPR/Cas system in prokaryotes.

Keywords: endogenous viral elements; retrotransposon; piRNA; CRISPR/Cas; paleovirology

INTRODUCTION
Retroviruses integrate DNA copies of their RNA genome into that of their host, transferring genetic information in a direction not predicted by the central dogma (Crick 1958). Understanding how these gene transfers influence the immunologic distinction between host “self” and viral “nonself” is of practical importance. Interestingly, several host genes derived from retroviruses limit infection by related viruses (Ikeda and Odaka 1983; Hainaut et al. 1990; Mura et al. 2004; Aswad and Katzourakis 2012; Fujino et al. 2014; Yap et al. 2014), demonstrating that lateral gene flow from viruses to their vertebrate hosts can mediate transgenerational immune memory; all known examples are mediated by production of dominant negative proteins. RNA-mediated transgenerational antiviral immunity, while common via the CRISPR/Cas system in prokaryotes...
(Sorek et al. 2008), has thus far been reported in only one metazoan, Caenorhabditis elegans (Rechavi et al. 2011). Immunity through the CRISPR/Cas system requires lateral transfer of genetic information: short nontself nucleic acid “spacers” are integrated into specialized genomic arrays of repeated sequences. Transcripts from these specialized self loci are processed into ribonucleoprotein complexes capable of degrading nontself targets on the basis of Watson–Crick base-pairing with spacer RNA (Barrangou et al. 2007). A conceptually similar system mediated by PIWI-interacting RNAs (piRNAs) appears broadly conserved in metazoans (Grimson et al. 2008), yet the confirmed targets are largely limited to endogenous transposable elements.

In addition to retroviruses, RNA-only viruses have also given rise to sequences in many metazoan genomes, although the specific mechanisms involved in this horizontal RNA-to-DNA information flow are less clear (Belyi et al. 2010; Horie et al. 2010; Katzourakis and Gifford 2010). We have studied one class of these sequences, called endogenous bornavirus-like nucleoprotein elements (EBLNs), because they are the only riboviral endogenous elements known in humans. Homo sapiens EBLNs (hsEBLNs) contain poly(dA:dT) tracts, recognizable transcription start sites, and are flanked by target-site duplications, strongly suggesting that they represent viral mRNA integrated by a retrotransposon (Enault et al. 2000; Belyi et al. 2010). These EBLNs were integrated ~40 million years ago, coincident with the peak of host processed pseudogene formation by a similar mechanism (Zhang et al. 2003). Because species with EBLNs appear relatively protected against modern day bornaviruses (Belyi et al. 2010), negative-strand RNA viruses that can cause neurological disease (Tomonaga et al. 2002), we questioned if they could influence antiviral immunity like some endogenous retroviral elements.

HsEBLN-1 and -2 contain long open reading frames (ORFs) with the potential to code for proteins of 366 and 225 amino acids, respectively. In cell culture experiments, overexpression of bornaviral nucleoprotein prevents infection, presumably because a specific stoichiometry of replication complex components is critical (Geib et al. 2003). Thus overexpression of EBLN-encoded proteins could potentially have prevented ancient bornaviral replication. Indeed, we recently showed that the Ictidomys tridecemlineatus genome contains an EBLN that was integrated more recently than human EBLNs, shares over 75% of amino acids in common with some extant avian bornaviruses, and can block bornaviral replication when overexpressed in human cells (Fujino et al. 2014). However, most EBLNs in primates and rodents have disrupted ORFs (Horie et al. 2010), and there is no evidence of selection to maintain the ORFs of EBLNs in primates (Kobayashi et al. 2011). Thus most EBLNs either had no function, had a protein-coding function that has been lost, or perhaps had a function not related to encoding a protein.

We recently observed that all seven human EBLNs are expressed as RNAs, some exclusively in the adult testis (K Sofuku, N Parrish, T Honda, and K Tomonaga, in prep.). Because the native promoter sequence is not mobilized during pseudogene formation, the probability of seven out of seven randomly chosen host processed pseudogenes being expressed is low; by the highest estimates, only about one-third of host processed pseudogenes are transcribed (Zheng et al. 2007; Guo et al. 2014). This led us to investigate the possibility that mammalian EBLNs encode antiviral RNAs. The previously noted similarities between the piRNA pathway and the CRISPR/Cas immune system (Karginov and Hannon 2010) made piRNAs an attractive candidate. Analogous to CRISPR guide RNA, piRNA, in complex with a PIWI-clade argonaute protein partner, target transposons for transcriptional and post-transcriptional silencing (Siomi and Kuramochi-Miyagawa 2009; Ishizu et al. 2012). Similar to CRISPR arrays, piRNA precursor molecules are transcribed from discrete loci (“piRNA clusters”) that cover a small percentage of the total genome (Lau et al. 2006; Aravin et al. 2007). It is thought that these loci act as “traps,” in the sense that nucleic acid elements transposing into them will subsequently be silenced (Malone and Hannon 2009). PiRNAs are most abundant in the germline, consistent with a role in genome defense, yet some piRNA pathway components are detectable in somatic cells where their functional relevance is unclear. Here we show that multiple EBLNs, integrated independently in two mammalian lineages at least 20 million years ago, give rise to piRNAs, and present evidence consistent with selection for EBLNs that integrated into piRNA-generating loci.

RESULTS

Rodent EBLNs give rise to small RNA with characteristics of piRNAs

We noted EBLN-derived piRNAs while examining rodent EBLN sequences using a web-based genome browser (Karolchik et al. 2014). Several piRNAs identified in reports initially describing piRNAs (Aravin et al. 2006; Girard et al. 2006; Lau et al. 2006) overlap with Mus musculus and Rattus norvegicus EBLNs (Table 1). Most piRNAs are generated from genomic loci that are unannotated (Girard et al. 2006), as are most EBLNs. Thus to better ascertain the abundance and diversity of EBLN-derived small RNAs, we sequenced small RNAs from the testis of a 6-wk-old mouse. Small RNAs with sequence characteristics of primary piRNAs, namely a length over 26 nucleotides and enrichment of 5′ uridine, mapped to three of the five mouse EBLNs (mmeEBLN-3 through -5) (Fig. 1A). The abundance of small RNAs mapped to mmeEBLN-5 was highest, with thousands of reads mapped to some regions of this sequence. The abundance of small RNAs mapped to mmeEBLN-3 and -4 was lower, yet above
the level of most genomic loci. In contrast, no small RNAs mapped to mmEBLN-1 and -2. Examining the small RNAs mapped to the 100 kb flanking these EBLNs revealed that mmEBLN-3 and -5 were within loci with small RNAs abundantly mapped to a single genomic strand, while mmEBLN-4 was near the end of such a locus (Fig. 1B). Notably, all piRNA-like small RNAs mapped to the antisense strand relative to the hypothetical ancient bornaviral N mRNA whose integration was responsible for EBLN formation. This experiment suggested that small RNAs with characteristics of piRNAs were generated from three of five mmEBLN loci, and that these EBLNs were in or near piRNA clusters. Thus we examined publically available sequence data sets and piRNA cluster annotations to further evaluate this possibility.

**Rodent EBLNs are enriched in pachytene piRNA clusters**

A piRNA is defined as a small RNA in complex with a PIWI protein, thus immunoprecipitation of intact ribonucleoproteins is required to unambiguously confirm the identity of any RNA molecule as a piRNA. Thus we first analyzed sequences from a comprehensive analysis of murine piRNA biogenesis (Li et al. 2013a). Consistent with our initial experiment, small RNAs mapped to three of five mmEBLNs (mmEBLN-3, -4, and -5) (Table 2). These RNAs ranged in length from 25 to 31 nucleotides (nt) and >95% contained uridine as the 5′ nucleotide (Fig. 1C). All were antisense relative to the proposed ancient bornaviral mRNA, consistent with the potential to post-transcriptionally silence such mRNA (Reuter et al. 2011). As expected of piRNAs derived from a primary piRNA transcript, and as opposed to secondary piRNAs generated during the so-called “ping-pong” amplification cycle that occurs prenatally in mice (Beyret et al. 2012), there was no enrichment of adenosine at the 10th nucleotide position (Wang et al. 2014). Li et al. (2013b) defined 417 piRNA precursor transcripts from which over 95% of mature murine piRNA are derived. Together these sequences cover only 0.28% of the mouse genome. Pachytene piRNA transcripts, which give rise to the predominant class of piRNAs found in adult testis, cover only 0.08% of the genome. Thus a short genomic range chosen at random has an exceedingly low likelihood of being within such a transcript. However, two of the five mmEBLNs (mmEBLN-3 and -5) are within pachytene piRNA clusters. Moreover, the piRNA precursors containing these two EBLNs give rise to the third and 13th highest density of mapped piRNAs from the adult mouse testis (Lau et al. 2006). mmEBLN-4 is within 40 kb of the 3′ termini of an annotated pachytene piRNA transcript. To estimate the probability of this apparent enrichment of EBLNs within piRNA-generating genomic loci, we calculated

<table>
<thead>
<tr>
<th>Accession</th>
<th>Name</th>
<th>Coordinates</th>
<th>Length</th>
<th>5′ Nucleotide</th>
</tr>
</thead>
<tbody>
<tr>
<td>DQ549815</td>
<td>mmu_piR_003423</td>
<td>chr9:54093758–54093787</td>
<td>30</td>
<td>U</td>
</tr>
<tr>
<td>DQ48337</td>
<td>mmu_piR_003006</td>
<td>chr9:54093761–54093790</td>
<td>30</td>
<td>U</td>
</tr>
<tr>
<td>DQ69528</td>
<td>mmu_piR_012394</td>
<td>chr9:54093806–54093834</td>
<td>29</td>
<td>U</td>
</tr>
<tr>
<td>DQ700150</td>
<td>mmu_piR_019537</td>
<td>chr9:54093814–54093842</td>
<td>29</td>
<td>U</td>
</tr>
<tr>
<td>DQ703437</td>
<td>mmu_piR_021794</td>
<td>chr9:54093845–54093874</td>
<td>30</td>
<td>U</td>
</tr>
<tr>
<td>DQ724667</td>
<td>mmu_piR_036414</td>
<td>chr9:54093868–54093898</td>
<td>31</td>
<td>U</td>
</tr>
<tr>
<td>DQ706487</td>
<td>mmu_piR_023855</td>
<td>chr18:67195283–67195313</td>
<td>31</td>
<td>U</td>
</tr>
<tr>
<td>DQ711296</td>
<td>mmu_piR_027183</td>
<td>chr18:67195514–67195543</td>
<td>30</td>
<td>U</td>
</tr>
<tr>
<td>DQ725925</td>
<td>mmu_piR_037293</td>
<td>chr18:67195670–67195699</td>
<td>30</td>
<td>U</td>
</tr>
<tr>
<td>DQ698401</td>
<td>mmu_piR_018347</td>
<td>chr18:67195905–67195932</td>
<td>28</td>
<td>A</td>
</tr>
<tr>
<td>DQ551691</td>
<td>mmu_piR_004100</td>
<td>chr18:67195934–67195965</td>
<td>32</td>
<td>U</td>
</tr>
<tr>
<td>NA</td>
<td>mmu_piR_038731</td>
<td>chr18:67196113–67196142</td>
<td>30</td>
<td>U</td>
</tr>
</tbody>
</table>

(U) Uridine, (A) adenosine.
the cumulative binomial distribution of an event with a probability 0.0028 occurring twice or more in, as there are five recognized EBLNs, five trials ($P = 8 \times 10^{-8}$).

Considering the possibility that this enrichment was perhaps limited to a single species, we also examined the genome of the laboratory rat (*Rattus norvegicus*). All EBLNs are in syntenic loci in mice and rats (Horie et al. 2013), suggesting that these EBLNs were integrated between 20 and 30 million years ago into a common ancestor shared by these species (Horie et al. 2013). PiRNA clusters are generally well conserved between mice and rats in terms of their genomic context (Assis and Kondrashov 2009), but not at the level of piRNA sequence (Aravin et al. 2006). As primary piRNA precursor transcripts have not been defined in the rat, we examined piRNA cluster annotations to determine if EBLN-derived piRNAs would be expected to be made in this species. Indeed, both orthologous EBLNs within piRNA clusters in mice were annotated in rat piRNA clusters (Girard et al. 2006; Lau et al. 2006), and *rnEBLN-4* was also annotated within a cluster in one study (Girard et al. 2006). As in the mouse, a low percentage of the overall rat genome gives rise to piRNAs and is annotated as piRNA clusters (0.18% in the more inclusive annotation). The probability of this degree of enrichment of EBLNs in piRNA-generating loci occurring by chance, calculated as above, ranges from $6 \times 10^{-8}$ to $1 \times 10^{-6}$ based on the annotation used (Fig. 2A).

In summary, three out of the five EBLNs in mice and rats give rise to piRNAs. As we are limited to detecting EBLNs that

**FIGURE 1.** Murine EBLNs give rise to piRNA. (A) Testis small RNAs map to murine EBLNs. Borna disease virus (BDV), nucleoprotein (N) gene, and homologous murine EBLNs (1–5 as numbered) are depicted. Numbers atop each marker represent the EBLN length as defined by translated amino acid homolog to BDV determined in Arensburger et al. (2011). Numbers below each marker indicate the genomic coordinates of the murine EBLNs. Uniquely mapped small RNAs are shown for EBLN-3 through -5; reads antisense to the predicted ancient bornaviral mRNA are plotted in green below each gene marker, sense reads in blue above. The y-axis indicates the number of mapped reads. (B) *mmEBLN* -3 through -5 are in or near clusters of mapped small RNAs. Small RNA reads matching the 50 kb upstream and downstream from *mmEBLN* -3 through -5 are shown. Reads mapping to the 5′ to 3′ strand are shown in blue above the line and those mapped to the opposite strand are shown in green below the line. The y-axis indicates the number of uniquely mapped reads and is arbitrarily truncated at 1000 or 5000 reads. (C) *mmEBLN*-mapped reads have characteristics of piRNAs. The nucleotide composition of small RNAs mapped to murine EBLN-3 through -5 is shown, with each base colored as indicated. Bases 1, 10, 26, and 31 are numbered. The y-axis indicates the percentage nucleotides at each position of all mapped reads corresponding to each base.
<table>
<thead>
<tr>
<th>EBLN</th>
<th>Coordinates (strand)</th>
<th>piRNA precursor transcript</th>
<th>piRNA cluster evolution</th>
<th>Complementary to N mRNA</th>
<th>Complementary to viral genomic RNA</th>
</tr>
</thead>
<tbody>
<tr>
<td>mmEBLN-1</td>
<td>chr1:103490037–103490965 (−)</td>
<td>–</td>
<td>–</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>mmEBLN-2</td>
<td>chr1:93851147–93851916 (−)</td>
<td>–</td>
<td>–</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>mmEBLN-3</td>
<td>chr18:67195203–67196180 (−)</td>
<td>18-461:3-36451.3; chr18:67189100–67226114</td>
<td>Ancestral: human ID 5251; chr18:11631026–11717899 (marmoset chr1:3:45297101–45343400)</td>
<td>144.11</td>
<td>0</td>
</tr>
<tr>
<td>mmEBLN-4</td>
<td>chr6:87894228–87895035 (+)</td>
<td>–</td>
<td>–</td>
<td>5.84</td>
<td>0</td>
</tr>
<tr>
<td>mmEBLN-5</td>
<td>chr9:54093749–54093919 (−)</td>
<td>9-奎5.3–24188.1; chr9:54054980–54097630</td>
<td>Ancestral: human ID 4384; chr5:51535076–51605926 (marmoset chr10:28388101–28440800)</td>
<td>603.48</td>
<td>0</td>
</tr>
<tr>
<td>mnEBLN-1</td>
<td>chr1:1092784712–92785688 (−)</td>
<td>–</td>
<td>–</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>mnEBLN-2</td>
<td>chr1:82624298–82625334 (−)</td>
<td>–</td>
<td>–</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>mnEBLN-3</td>
<td>chr18:63553420–63554192 (−)</td>
<td>cluster 14; chr18:63622000–63658000</td>
<td>Ancestral: human ID 5251; chr18:11631026–11717899 (marmoset chr1:3:45297101–45343400)</td>
<td>ND</td>
<td>0</td>
</tr>
<tr>
<td>mnEBLN-4</td>
<td>chr4:122161250–122161697 (+)</td>
<td>cluster 85; chr4:122158319–122199619</td>
<td>Ancestral: human ID 1264; chr3:128540550–128603591 (marmoset chr1:5:68183501–68217900)</td>
<td>ND</td>
<td>0</td>
</tr>
<tr>
<td>mnEBLN-5</td>
<td>chr8:57747079–57747381 (−)</td>
<td>cluster 6; chr8:57725000–5777000</td>
<td>Ancestral: human ID 4384; chr5:51535076–51605926 (marmoset chr10:28388101–28440800)</td>
<td>ND</td>
<td>0</td>
</tr>
<tr>
<td>hsEBLN-1</td>
<td>chr10:22208838–22209999 (−)</td>
<td>–</td>
<td>–</td>
<td>0.12</td>
<td>0</td>
</tr>
<tr>
<td>hsEBLN-2</td>
<td>chr3:73062206–73063317 (+)</td>
<td>ID 1201; chr3:73061025–73070092</td>
<td>–</td>
<td>0</td>
<td>0.68</td>
</tr>
<tr>
<td>hsEBLN-3</td>
<td>chr9:37086679–37087837 (+)</td>
<td>ID 2841; chr9:37086668–37092983</td>
<td>–</td>
<td>0</td>
<td>1.71</td>
</tr>
<tr>
<td>hsEBLN-4</td>
<td>chr1:74570963–4572180 (+)</td>
<td>–</td>
<td>–</td>
<td>0.07</td>
<td>0</td>
</tr>
<tr>
<td>hsEBLN-5</td>
<td>chr1:23517490–23518542 (−)</td>
<td>–</td>
<td>–</td>
<td>0</td>
<td>0.07</td>
</tr>
<tr>
<td>hsEBLN-6</td>
<td>chr1:75011855–75012971 (−)</td>
<td>ID 3562; chr1:75001149–75023707</td>
<td>–</td>
<td>3.35</td>
<td>0</td>
</tr>
<tr>
<td>hsEBLN-7</td>
<td>chr10:94534165–94535277 (−)</td>
<td>–</td>
<td>–</td>
<td>5.28</td>
<td>0</td>
</tr>
<tr>
<td>cjEBLN-1</td>
<td>chr7:19830726–19831493 (−)</td>
<td>–</td>
<td>–</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>EBLN</td>
<td>Coordinates (strand)</td>
<td>piRNA precursor transcript</td>
<td>pRNA cluster evolution</td>
<td>RPKM&lt;sup&gt;c&lt;/sup&gt;</td>
<td>Complementary to N mRNA&lt;sup&gt;d&lt;/sup&gt;</td>
</tr>
<tr>
<td>---------</td>
<td>----------------------</td>
<td>----------------------------</td>
<td>------------------------</td>
<td>------------------</td>
<td>---------------------------------</td>
</tr>
<tr>
<td>cjEBLN-2</td>
<td>chr1:549481896–49483299 (+)</td>
<td>–</td>
<td>–</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>cjEBLN-2</td>
<td>chr1:138998195–138998975 (+)</td>
<td>–</td>
<td>–</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>cjEBLN-4</td>
<td>GL25135.1:14259–15765 (+)</td>
<td>–</td>
<td>–</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>cjEBLN-5</td>
<td>chr1:945134664–45135692 (−)</td>
<td>–</td>
<td>–</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>cjEBLN-6</td>
<td>chr1:63682272–63683372 (−)</td>
<td>chr11:63677801–63689900</td>
<td></td>
<td>15.81</td>
<td>0</td>
</tr>
<tr>
<td>cjEBLN-7</td>
<td>chr1:281451040–81452148 (−)</td>
<td>–</td>
<td>–</td>
<td>1.23</td>
<td>0</td>
</tr>
</tbody>
</table>

<sup>a</sup>From Li et al. (2013a), Supplemental Table 2 (0.28% of the genome annotated as a piRNA cluster).

<sup>b</sup>From Assis and Kondrashov (2009).

<sup>c</sup>Mapped from Li et al. (2013b); GEO accession number GSM1096587.

<sup>d</sup>According to hypothetical ancient bornaviral N mRNA.

<sup>e</sup>As a negative-strand RNA virus, the bornavirus genomic strand is the reverse complement of the mRNA strand.

<sup>f</sup>From Lau et al. (2006) (0.106% annotated as a piRNA cluster).

<sup>i</sup>From Hirano et al. (2014) (0.12% annotated as a piRNA cluster).
have maintained recognizable homology with modern bornaviruses, we do not assume that the existing five rodent EBLNs are the only sequences derived from ancient bornaviruses that ever entered the genomes of these species. Nonetheless, given the relative sparsity of piRNA-generating genomic loci, we can reasonably exclude the hypothesis that rodent EBLNs were integrated at random and maintained neutrally thereafter.

**piRNAs are generated from primate EBLNs**

EBLNs are found in many vertebrate genomes (Belyi et al. 2010), and similar to retrotransposon insertions, are predicted to be essentially free of homoplasmy. About 40 million years ago, EBLNs were integrated into a common primate ancestor prior to the divergence of the strepsirrhine and haplorrhine primates (Horie et al. 2010). Accordingly, these primate EBLNs were each integrated into different loci than rodent EBLNs. We hypothesized that if EBLN-derived piRNAs were functional, independent EBLN integrations with piRNA-generating capacity may have evolved convergently in these two lineages. Thus we determined if piRNA-like RNAs or piRNAs were made from the seven EBLNs in human and marmoset genomes, respectively, as deep sequencing of small RNAs from adult testes from both species has recently been performed (Ha et al. 2014; Hirano et al. 2014). Similar to pachytene piRNAs derived from adult mouse testes, the piRNA populations in the testis of adult marmosets are processed from primary piRNA precursor transcripts rather than ping-pong amplification. In humans, three EBLNs (hsEBLN-2, -3, and -6) were found to be within annotated piRNA clusters (Fig. 2B; total 2.4% annotated, \( P = 4.5 \times 10^{-4} \)). hsEBLN-7 was also found to give rise to piRNA-like RNA, yet was shorter than the arbitrary length cutoff used to define clusters (Table 2). In the marmoset, one EBLN (cjEBLN-6) was within an annotated cluster (total 0.12% annotated, \( P = 0.008 \)) and cjEBLN-7 also gave rise to piRNAs associated with the marmoset PIWI-like 1 protein MARWI (Hirano et al. 2014). As in rodents, the piRNAs produced by these EBLNs are antisense to the proposed ancient bornaviral mRNA. Experimental differences could explain why hsEBLN-2 and -3 give rise to piRNA-like small RNAs in the testis while the syntenic marmoset EBLNs in the marmoset do not: the marmoset sequences represent bona fide piRNAs immunoprecipitated with MARWI, while the human testis small RNA was bulk-isolated and enriched for 2'-O-methylated small RNAs (Kirino and Mourelatos 2007). Alternatively, this could reflect loss of piRNA production from cjEBLN-2 and -3 loci in the ~35 million years since the marmoset and human lineages diverged. In any case, these observations confirm that independent EBLN integrations are enriched within piRNA- or piRNA-like RNA-generating loci in two mammalian lineages. The probability of this occurring due to chance, approximated as the union of the most likely probabilities when each lineage is considered independently, is very low (\( P = 6.4 \times 10^{-7} \)).

**Rodent EBLNs were integrated into existing piRNA clusters**

Precisely what defines a piRNA precursor transcript as such is currently unclear (Vourekas et al. 2015), as are the determinants of innate immune recognition of bornaviral nucleic acids (Martin et al. 2011). In invertebrates, piRNA-like RNAs can be generated from viral sequences (Wu et al. 2010; Morazzani et al. 2012; Léger et al. 2013), and knockdown of piRNA pathway components has a proviral effect in mosquito cells (Schnettler et al. 2013) and flies (Zambon et al. 2006). Viral piRNA biogenesis in these invertebrates is assumed to be due to an interaction between viral genomic or transcript RNAs with PIWI proteins and piRNA pathway components in the absence of an integrated DNA intermediate. Moreover, in Aedes aegypti cells distinct PIWI proteins are required for piRNA biogenesis from viral precursors compared with endogenous transposon precursors. Thus presumably some feature of at least some viral nucleic acids is sufficient to determine their recognition for processing into piRNA-like molecules. The RNA-binding proteins involved in piRNA precursor processing (e.g., MOV10L1) may share common specificities with RNA-binding proteins involved in recognizing infectious viruses (e.g., MOV10 [Wang et al. 2010]), thus it is conceivable that integration of an EBLN into an existing transcriptional unit could have been responsible for that transcript’s definition as a piRNA precursor. If this were the case, the convergent evolution described above could have arisen in a direct, mechanistic fashion. We can reject this possibility in several instances, as some piRNA clusters are syntenic between primates and rodents.

---

**FIGURE 2.** EBLNs are enriched within piRNA clusters. (A) The genome of Rattus norvegicus is depicted. Non-piRNA-generating sequences in the genome (purple) and piRNA clusters (red, 0.18%, as annotated in Girard et al. (2006) are drawn to scale. EBLNs are depicted as blue dots and are drawn larger than scale to allow visualization. The probability of the observed enrichment, estimated as described in the text, is \( 6 \times 10^{-8} \). (B) The genome of Homo sapiens is depicted as above, with 2.4% of the genome annotated as piRNA cluster. The probability of the observed enrichment is \( 4.5 \times 10^{-4} \).

---

Hypothesis: CRISPR/Cas-like immunity in mammals
(Hirano et al. 2014). Indeed, rodent EBLN-3 and -5 are within large intergenic piRNA clusters also present in humans, as is the rat piRNA cluster containing EBLN-4 (Table 2; Fig. 3). The marmoset EBLNs generating bona fide piRNAs (cEBLN-6 and -7) are within the 3′ untranslated regions of genes (NEU3 and TBC1D12) that also generate piRNAs in the mouse, albeit much less abundantly (Robine et al. 2009). Thus at least some EBLNs were integrated into preexisting piRNA clusters. Beyond this, we cannot exclude an influence of EBLN integration on the life/death evolutionary dynamics of piRNA clusters (Assis and Kondrashov 2009).

**DISCUSSION**

Two models could account for the convergent evolution of piRNA-generating EBLNs in two mammalian lineages: one in which EBLNs were preferentially integrated within piRNA clusters, and another in which EBLNs integrated randomly and those within piRNA clusters were positively selected. We cannot exclude the first model, but consider it unlikely based on the distribution of other nucleic acids mobilized by retrotransposons: Among the over 10,000 and 5000 GENCODE-annotated (Harrow et al. 2012) processed pseudogenes in the human and mouse genomes, respectively, few are within piRNA clusters (Ha et al. 2014; Hirano et al. 2014; Watanabe et al. 2015). Furthermore, LINE-1-mobilized sequences in cultured cells (Berry et al. 2006), tumors (Cooke et al. 2014; Tubio et al. 2014), and mice bearing an engineered LINE-1 element (An et al. 2006) are not preferentially targeted to piRNA clusters. Finally, recent processed pseudogene (Abyzov et al. 2013) and retrotransposon insertions in human genomes are not enriched within piRNA-generating loci (Y Zhang and P Gerstein, pers. comm.). Nonetheless, experiments to determine if, under certain circumstances, nonself nucleic acids are preferentially trapped within piRNA clusters are warranted (Kawaoka et al. 2013).

Enrichment of EBLNs within piRNA-generating loci in two mammalian lineages is consistent with natural selection (Fig. 4). EBLN-derived piRNAs may have fortuitously silenced a host gene shared by these lineages, or multiple EBLNs could have hitchhiked with linked alleles. Another parsimonious hypothesis is that EBLN-derived piRNAs silenced bornavirus and thus protected against death or reduced fecundity due to bornaviral infection, reminiscent of transposon (Khurana et al. 2011) or errantiviral (Prud’homme et al. 1995) resistance after sequences from these elements are integrated into Drosophila piRNA clusters. Such an effect is feasible and consistent with current models of the mammalian piRNA system if ancient bornaviruses were transmitted vertically via the gamete. Indeed, the ancient bornaviruses giving rise to EBLNs infected germ cells; otherwise, a noncanonical mechanism of gene transfer from soma to germline, in violation of Weismann’s law, was responsible for EBLN formation (Pittoggi et al. 2006). Of note, modern bornaviruses can be transmitted vertically (Okamoto et al. 2003; Kerski et al. 2012). EBLN integration into piRNA clusters could thus have resulted in viral silencing in germ cells, similar to the transgene silencing observed after insertion of identical sequences into piRNA clusters (Yamamoto et al. 2013). As transcriptional silencing via repressive chromatin modification, rather than post-transcriptional silencing, appears the dominant mechanism of piRNA-mediated silencing, it is notable that modern bornaviruses, unlike most RNA viruses, replicate in the nucleus and interact directly with chromatin (Matsumoto et al. 2012). Further, while an antiviral effect of RNA interference has been detected in certain mammalian systems (Li et al. 2013b), it is considered to have been largely superseded in mammalian antiviral innate immunity by type I interferons (Cullen et al. 2013; Cullen 2014). However, overexpression of interferon α prevents germ-cell development in mice and its receptor is not expressed on pachytene spermatocytes (Satie et al. 2011), perhaps censoring this antiviral mechanism in the germline (Pare and Sullivan 2014).

PiRNA-like RNAs and/or PIWI proteins have been described in primate pluripotent cells (Marchetto et al. 2013), human hematopoietic cells (Sharma et al. 2001; Cichocki et al. 2010), and some other somatic cells including neurons (Lee et al. 2011; Yan et al. 2011; Rajasethupathy et al. 2012), a critical target cell of modern bornaviruses. Notably, neurons, like germ cells, are permissive to retrotransposition (Upton et al. 2015) and relatively nonresponsive to type I interferons (Lin et al. 2013; Kreit et al. 2014). Transcripts of the murine PIWI-clade partner of most pachytene piRNAs (MIWI) can be detected in a distribution overlapping highly BDV-susceptible cells in the cerebellum, dentate gyrus, and olfactory bulb (Lein et al. 2007; Ackermann et al. 2010). Thus EBLN-derived piRNAs could potentially have protected stem or somatic cells like neurons from bornavirus-induced pathology, although more work is required to assess the feasibility of this model. A class of piRNA-like RNAs derived from the
same genomic loci that give rise to pachytene piRNAs, yet with different genetic requirements for their biogenesis, have recently been described in a variety of somatic cell types (Ortogero et al. 2014). In addition, dicer-dependent endogenous small interfering RNAs are produced from some piRNA clusters (Watanabe et al. 2008). Thus production of piRNAs from EBLNs in the testis could correlate with production of other, more relevant, RNA molecules from these loci elsewhere.

It has previously been suggested that transposon silencing may not be the sole function of mammalian pachytene piRNAs, most of which are highly complementary only to the locus from which they are transcribed and are derived from unannotated intergenic regions depleted in transposon-derived sequences relative to the genome as a whole (Aravin et al. 2007). While their potential targets are thus unclear, piRNA clusters in rodents (Assis and Kondrashov 2009) and humans (Lukic and Chen 2011) evolve rapidly under positive selection. Recent reports suggest that piRNAs may target miRNAs to which they are only partially complementary, similar to miRNA:target interactions (Zhang et al. 2015). However, the biochemically confirmed functions of PIWI proteins require extensive piRNA:target complementarity (Reuter et al. 2011). Finding that pachytene piRNAs are made from sequences derived from an exogenous virus suggests an alternative explanation for these observations, as well as a potential role for the transcription of piRNA precursors outside the germline: similar to CRISPR spacers (Mojica et al. 2005), they may serve an immunologic function by targeting sequences foreign to the genome from which they are derived (Sagy et al. 2014).

Genetic material laterally transferred from nonretroviral viruses to host genomes has been noted for a wide variety of pathogens and hosts (Zhdanov 1975; Belyi et al. 2010; Gilbert and Feschotte 2010; Kapoor et al. 2010; Katzourakis and Gifford 2010; Liu et al. 2010, 2011). Several of these gene transfer events can be clearly attributed to retrotransposon activity (Ballinger et al. 2012), and in at least one case the gene transfer is associated with viral resistance (Maori et al. 2007). We hypothesize that, lacking RNA-dependent RNA-polymerases used to amplify RNAs for viral silencing (Mojica et al. 2005), they may serve an immunologic function as well as a potential role for the transcription of piRNA precursors outside the germline: similar to CRISPR spacers (Mojica et al. 2005), they may serve an immunologic function by targeting sequences foreign to the genome from which they are derived (Sagy et al. 2014).

Genetic material laterally transferred from nonretroviral viruses to host genomes has been noted for a wide variety of pathogens and hosts (Zhdanov 1975; Belyi et al. 2010; Gilbert and Feschotte 2010; Kapoor et al. 2010; Katzourakis and Gifford 2010; Liu et al. 2010, 2011). Several of these gene transfer events can be clearly attributed to retrotransposon activity (Ballinger et al. 2012), and in at least one case the gene transfer is associated with viral resistance (Maori et al. 2007). We hypothesize that, lacking RNA-dependent RNA-polymerases used to amplify RNAs for viral silencing (Mojica et al. 2005), they may serve an immunologic function as well as a potential role for the transcription of piRNA precursors outside the germline: similar to CRISPR spacers (Mojica et al. 2005), they may serve an immunologic function by targeting sequences foreign to the genome from which they are derived (Sagy et al. 2014).  

Genetic material laterally transferred from nonretroviral viruses to host genomes has been noted for a wide variety of pathogens and hosts (Zhdanov 1975; Belyi et al. 2010; Gilbert and Feschotte 2010; Kapoor et al. 2010; Katzourakis and Gifford 2010; Liu et al. 2010, 2011). Several of these gene transfer events can be clearly attributed to retrotransposon activity (Ballinger et al. 2012), and in at least one case the gene transfer is associated with viral resistance (Maori et al. 2007). We hypothesize that, lacking RNA-dependent RNA-polymerases used to amplify RNAs for viral silencing (Mojica et al. 2005), they may serve an immunologic function as well as a potential role for the transcription of piRNA precursors outside the germline: similar to CRISPR spacers (Mojica et al. 2005), they may serve an immunologic function by targeting sequences foreign to the genome from which they are derived (Sagy et al. 2014).
similar viral cDNA is a source of antiviral interfering RNA in Drosophila (Goic et al. 2013).

The described interactions between infectious viruses and the transposons that comprise the bulk of mammalian host genomes are consistent with the hypothesis that transposons are symbionts integral to genomic stress responses (McClintock 1984), including antiviral immune responses (Zeng et al. 2014; Yu et al. 2015). The piRNA system in mammals is known to silence quasi-nonself transposon nucleic acids; our observations raise the hypothesis that, as for CRISPR/Cas in prokaryotes, truly exogenous nonself nucleic acids from infecting viruses can be targeted by piRNA-like RNAs, but that this requires genetic information flow in an unexpected retrotransposon-dependent manner (Núñez et al. 2015). We refer to this hypothetical mechanism as viral transcript reversion with amanousie piRNA silencing (TRAPS) to indicate the host-directed nature of the reverse transcription involved, its potential role in heritage immune memory, and the specialized genomic loci involved in capturing the nonself information. Testing this hypothesis is of proximal relevance to human health, namely in arbovirus/vector interactions (Arensburger et al. 2011) and resistance to bornaviral disease.

MATERIALS AND METHODS

EBLN annotation

The genomic locations of rodent and primate EBLNs have been described elsewhere (Belyi et al. 2010; Horie et al. 2010, 2013). These studies defined EBLNs on the basis of potential ORFs or regions of amino acid–based homologies with BDV N. In order to evaluate the hypothesis that small RNA generated from these loci could have biological function, EBLN annotation in Table 2 is inclusive of the entire inserted BDV N mRNA-like sequence block (i.e., for primate EBLNs, the sequence flanked by TSDs).

Small RNA sequencing

One 6-wk-old BALB/c mouse was purchased from Charles River Laboratories, Japan. Testis total RNA including small RNA fraction was collected from the mouse using a miRNeasy mini Kit (QIAGEN). RNA quality was confirmed by 2100 Agilent Bioanalyzer (Agilent Technologies). After quality confirmation, cDNA libraries were constructed from the testis RNAs by TruSeq small RNA sample prep kit (Illumina). Small RNA sequencing was performed using an Illumina HiSeq (50SE) machine by Hokkaido Systems Science. The sequence data were mapped onto mm9 using LASTZ alignments of murine piRNA clusters to each of these species to identify the homologous primate clusters in human and marmoset, Kondrashov (2009), who considered rodent clusters in detail. To determine the ancestral piRNA clusters, we used a bioinformatic analysis performed using Galaxy (https://usegalaxy.org).

Cumulative binomial probability

The likelihood of \( x \) EBLNs or more, out of a total of \( n \) EBLNs for a given species, being found within a piRNA cluster if piRNA clusters occupy \( p \) percent of the genome was approximated as the probability \( P \) of \( x \) successes or more in \( n \) Bernoulli trials, each with a probability \( p \):

\[
P(x) = \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x}.
\]

piRNA cluster evolution

Ancestral piRNA clusters were as determined by Assis and Kondrashov (2009), who considered rodent clusters in detail. To identify the homologus primate clusters in human and marmoset, LASTZ alignments of murine piRNA clusters to each of these species were viewed using Ensembl and compared with annotations in Ha et al. (2014) and Hirano et al. (2014).

Murine genes with 3′ UTRs giving rise to piRNA were determined by Robine et al. (2009).

ACKNOWLEDGMENTS

We thank Masayuki Horie, Cedric Feschotte, Craig B. Wilen, Bryan R. Cullen, Alexei A. Aravin, Andrew Z. Fire, Rafi Ahmed, Arnold J. Levine, Sara Cherry, and P. Jeremy Wang for helpful discussions and/or comments on the manuscript. N.F.P. thanks Bruce R. Levin for encouraging a hunt for a mammalian CRISPR-like immune system. N.F.P. was supported by the Japan Society for the Promotion of Science short-term postdoctoral fellowship award #P1507. This study was supported in part by Funding Program for Next Generation World-Leading Researchers (NEXT program), KAKENHI grant numbers 26253027 and 26670225, and the Core-to-Core Program A, Advanced Research Networks from the Japan Society for the Promotion of Science (JSPS) (K.T.); grants from Takeda Science Foundation (K.T.); and KAKENHI grant numbers 25115508 and 25860336 from the Ministry of Education, Culture, Science, Sports and Technology (MEXT) of Japan (T.H.).

Received April 16, 2015; accepted July 8, 2015.

REFERENCES


Parrish et al.
Hypothesis: CRISPR/Cas-like immunity in mammals


Parrish et al.


piRNAs derived from ancient viral processed pseudogenes as transgenerational sequence-specific immune memory in mammals

Nicholas F. Parrish, Kan Fujino, Yusuke Shiromoto, et al.

RNA published online August 17, 2015

This article is distributed exclusively by the RNA Society for the first 12 months after the full-issue publication date (see http://rnajournal.cshlp.org/site/misc/terms.xhtml). After 12 months, it is available under a Creative Commons License (Attribution-NonCommercial 4.0 International), as described at http://creativecommons.org/licenses/by-nc/4.0/.

Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or click here.

To subscribe to RNA go to:
http://rnajournal.cshlp.org/subscriptions

© 2015 Parrish et al.; Published by Cold Spring Harbor Laboratory Press for the RNA Society