Recent Expansion of a New Ingi-Related Clade of Vingi non-LTR Retrotransposons in Hedgehogs

Kenji K. Kojima, Vladimir V. Kapitonov, and Jerzy Jurka*

Genetic Information Research Institute

*Corresponding author: E-mail: jurka@girinst.org.

Associate editor: Norihiro Okada

Abstract

Autonomous non-long terminal repeat (non-LTR) retrotransposons and their repetitive remnants are ubiquitous components of mammalian genomes. Recently, we identified non-LTR retrotransposon families, Ingi-1_AAl and Ingi-1_EE, in two hedgehog genomes. Here we rename them to Vingi-1_AAl and Vingi-1_EE and report a new clade “Vingi,” which is a sister clade of Ingi that lacks the ribonuclease H domain. In the European hedgehog genome, there are 11 non-autonomous families of elements derived from Vingi-1_EE by internal deletions. No retrotransposons related to Vingi elements were found in any of the remaining 33 mammalian genomes nearly completely sequenced to date, but we identified several new families of Vingi and Ingi retrotransposons outside mammals. Our data suggest the horizontal transfer of Vingi elements to hedgehog, although the vertical transfer cannot be ruled out. The compact structure and trans-mobilization of nonautonomous derivatives of Vingi can make them useful for in vivo retrotransposition assay system.

Key words: non-LTR retrotransposon, Ingi, Vingi, hedgehog, mammal, nonautonomous, horizontal transfer, reverse transcriptase.
vertebrate, annelid, mollusk, nematode, and insect species.

Based on the reverse transcriptase (RT) phylogeny, *Ingi* and *Vingi* elements were clustered together inside the I group, with a high statistical support (fig. 1 and supplementary fig. S2, Supplementary Material online). Based on the phylogeny and domain structure (fig. 1), *Ingi* and *Vingi* elements can be classified into five clusters (*Ingi*-A, *Ingi*-B, and *Ingi*-C and *Vingi*-A and *Vingi*-B). The monophyly of the *Vingi*-A cluster is also supported by the substitutions

![Figure 1](https://example.com/figure1.png)

**Table 1.** Autonomous and Nonautonomous Families of *Vingi* non-LTR Retrotransposons in the *Erinaceus europaeus* Genome.

<table>
<thead>
<tr>
<th>Name</th>
<th>Length (bp)</th>
<th>Copy number of full-length copies</th>
<th>Corresponding positions in <em>Vingi-1_EE</em></th>
<th>Average identity ± standard deviationa</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Vingi-1_EE</em></td>
<td>3085</td>
<td>242</td>
<td>1-3085</td>
<td>0.9782 ± 0.0058</td>
</tr>
<tr>
<td><em>Vingi-1N1_EE</em></td>
<td>102</td>
<td>92</td>
<td>1-80, 3062-3085</td>
<td>0.9758 ± 0.0154</td>
</tr>
<tr>
<td><em>Vingi-1N2_EE</em></td>
<td>403</td>
<td>20</td>
<td>1-129, 2820-3082</td>
<td>0.9882 ± 0.00715</td>
</tr>
<tr>
<td><em>Vingi-1N3_EE</em></td>
<td>195</td>
<td>6273</td>
<td>1-149, 3036-3081</td>
<td>0.9751 ± 0.0003</td>
</tr>
<tr>
<td><em>Vingi-1N4_EE</em></td>
<td>574</td>
<td>101</td>
<td>1-170, 2679-3085</td>
<td>0.9723 ± 0.0141</td>
</tr>
<tr>
<td><em>Vingi-1N5_EE</em></td>
<td>628</td>
<td>106</td>
<td>1-211, 2669-3085</td>
<td>0.9731 ± 0.0131</td>
</tr>
<tr>
<td><em>Vingi-1N6_EE</em></td>
<td>339</td>
<td>12</td>
<td>1-209, 2953-3082</td>
<td>0.9758 ± 0.1151</td>
</tr>
<tr>
<td><em>Vingi-1N7_EE</em></td>
<td>874</td>
<td>8</td>
<td>1-212, 2324-2405, 2502-3081</td>
<td>0.9777 ± 0.1729</td>
</tr>
<tr>
<td><em>Vingi-1N8_EE</em></td>
<td>680</td>
<td>12</td>
<td>1-222, 2625-3082</td>
<td>0.9743 ± 0.1149</td>
</tr>
<tr>
<td><em>Vingi-1N9_EE</em></td>
<td>619</td>
<td>5</td>
<td>1-243, 2709-3084</td>
<td>0.9814 ± 0.2776</td>
</tr>
<tr>
<td><em>Vingi-1N10_EE</em></td>
<td>730</td>
<td>12</td>
<td>1-271, 2623-3081</td>
<td>0.9838 ± 0.1159</td>
</tr>
<tr>
<td><em>Vingi-1N11_EE</em></td>
<td>675</td>
<td>8</td>
<td>3-321, 2726-3080</td>
<td>0.9849 ± 0.1741</td>
</tr>
</tbody>
</table>

a CpG dinucleotides, prone to quick decay to TpG and CpA, in the consensus sequences were excluded in calculations of identity.

---

*Fig. 1* Structure and phylogeny of *Ingi* and *Vingi* elements. (A) Schematic structures of *Ingi* and *Vingi* elements. Groups that elements belong to are shown on the right. EN, endonuclease; ZF, zinc finger motif. (B) Phylogeny of RT domains among *Ingi*, *Vingi*, I, Nimb, and Outcast clades. The root of the tree is placed at the midpoint. Vertebrate *Vingi* lineages are drawn in thick lines. The numbers at each node are the bootstrap values of 100 replicates over 50. The model used is RtREV.
at the highly conserved SDH triplet in the apurinic-like endonuclease domain (supplementary fig. S3, Supplementary Material online).

Vertebrate Vingi elements were clustered into four lineages (fig. 1B, thick lines). Anolis carolinensis was the closest vertebrate related to hedgehogs, but Vingi-1_Acar and Vingi-2_Acar were not closely related to the Vingi elements from hedgehogs (Vingi-1_EE and Vingi-1-AAI). The sister lineage of the hedgehog Vingi elements (Vingi-1_PMa, Vingi-1_Lme, Vingi-1_Lch, Vingi-1_Tcas, and Vingi-1_BM) showed a phylogenetic relationship similar to that of their host species. This relationship is consistent with vertical transmission of Vingi elements in several bilaterian lineages for several hundred million years. In contrast, the hedgehog Vingi lineage does not include any retrotransposons from other species.

The new clade of non-LTR retrotransposons called “Vingi” includes two clusters: Vingi-A and Vingi-B (fig. 1). No Vingi elements encode RNA. The phylogeny of Vingi-1_PMa, Vingi-1_Lme, Vingi-1_Lch, Vingi-1_Tcas, and Vingi-1_BM indicates that the origin of Vingi can be traced back to the last common ancestor of all bilaterians. Although the clustering of Vingi-A and Vingi-B is not well supported statistically in the maximum likelihood phylogeny, these two clusters are generally positioned together independently of different methods of multiple alignment and evolutionary models. The Vingi clade possibly represents an internal branch inside of the Ingi clade, and Vingi-A and Vingi-B are likely to become separate clades after more elements are characterized.

Aside from the two hedgehog species, no Vingi or other elements from the L group could be identified in any of the remaining 33 mammalian genomes sequenced to date. Many fragments of non-LTR retrotransposons that likely predate the eutherian radiation (e.g., for L2 and L3) to date. Many fragments of non-LTR retrotransposons that are likely to be mobilized in trans by the Vingi-1_EE protein. This may be useful for inserting reporter genes in future experimental studies of Vingi.

Materials and Methods

Genomic sequences of various species were obtained mostly from NCBI GenBank, and prototypic sequences of known non-LTR retrotransposons were obtained from Repbase (http://www.girinst.org/repbase). Unpublished genomic sequences of European hedgehog and of other mammals used in this study were deposited in Genbank by Broad Institute (see Acknowledgments).

New Ingi and Vingi non-LTR retrotransposons were identified by repeated Blast (Altschul et al. 1997) and CENSOR (Kohany et al. 2006) searches using genomic sequences of various species with Ingi and Vingi elements as queries. The consensus sequences were derived using the majority rule applied to the corresponding set of multiple aligned copies of retrotransposons. We considered a particular copy of Vingi-1_EE and non-autonomous derivatives to be of full length when it started within the first 10 nucleotides relative to the consensus sequence and ended inside of the terminal AAG repeats.

We aligned the RT domain sequences of non-LTR retrotransposons with the aid of either MAFFT (Katoch et al. 2005) or MUSCLE (Edgar 2004). We constructed maximum likelihood trees by PhyML (Guindon and Gascuel 2003; Guindon et al. 2005) with bootstrap values (100 replicates) or approximate likelihood ratio test nonparametric branch support based on a Shimodaira–Hasegawa–like procedure (Anisimova and Gascuel 2006) in three different amino acid substitution models: RtREV, LG, and WAG. The phylogenetic trees were drawn with the aid of FigTree 1.3.1 (http://tree.bio.ed.ac.uk/software/figtree/).

Supplementary Materials

Supplementary figures S1–S3, tables S1 and S2, and materials S1 and S2 are available at Molecular Biology and Evolution online (http://www.mbe.oxfordjournals.org/).

Acknowledgments

We thank the Broad Institute Genome Sequencing Platform and Genome Sequencing and Analysis Program, Federica Di Palma, and Kerstin Lindblad-Toh for making the data for E. europaeus and other mammalian genomes available. This work was supported by the National Institutes of Health grant 5 P41 LM006252. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Library of Medicine or the National Institutes of Health.

References


Kimmel BE, ole-MoiYoi OK, Young JR. 1987. Ingi, a 5.2-kb dispersed sequence element from Trypanosoma brucei that carries half of a smaller mobile element at either end and has homology with mammalian LINEs. *Mol Cell Biol.* 7:1465–1475.


