Tracking the Ancestry of a Deeply Conserved Eumetazoan SINE Domain

Oliver Piskurek1 and Daniel J. Jackson*1
1Courant Research Centre Geobiology, Georg-August University of Göttingen, Göttingen, Germany
*Corresponding author: E-mail: djackson@uni-goettingen.de.
Associate editor: Norihiro Okada

Abstract

Transposable elements (TEs), such as short interspersed elements (SINEs), evolve rapidly and are generally restricted to specific lineages. Here, we demonstrate that a central core of the previously described DeuSINEs located within DeuSINEs (Nishihara et al. 2006) is widely distributed throughout the Metazoa. We characterize five new SINEs with this core sequence from the genomes of cnidarians, molluscs, annelids, and arthropods. Because this domain can be traced back to the cnidarian-bilaterian split >600 Ma, we propose naming it the “Nin” domain (the meaning of the Japanese character “Nin” is to endure and hide). Given that conserved noncoding DNA, such as that derived from the activity of SINES, can be functionally relevant for the host genome (Sasaki et al. 2008), our findings highlight the need to understand these functions and the roles they may have played in supporting the evolution of multicellular genomes.

Key words: short interspersed element, long interspersed element, transposable element, Nin-domain, DeuSINE, Eumetazoa.

Because the sequencing of the Caenorhabditis elegans genome (The C. elegans sequencing consortium 1998) and of various model organisms thereafter, it has become clear that vast regions of genomic DNA are noncoding Transposable elements (TEs) which can in fact play critical roles in gene regulation (Brosius 1999; Kidwell and Lisch 2001; Bejerano et al. 2006; Feschotte and Pritham 2007; Sasaki et al. 2008). Their ubiquitous distribution implies that they are likely to have played an important role in supporting eukaryotic genome evolution (Batzer and Deininger 2002; Kidwell 2002). TEs, for example, account for 45% of the human genome (Lander et al. 2001) and more than 50% of the opossum (Mikkelsen et al. 2007) and Hydra (Chapman et al. 2010) genomes.

Short interspersed elements (SINEs) are nonautonomous TEs (Ohshima and Okada 2005) that are usually composed of three distinct regions: a 5’ transfer RNA (tRNA)-like region (containing RNA polymerase III internal promoter regions), a tRNA-unrelated region, and a 3’ tail that are recognized by the reverse transcriptase of partner long interspersed elements (LINEs) during retrotransposition (Kajikawa and Okada 2002). Because SINE sequences usually mutate and diverge in a regular clock-like manner, it is unusual to detect homologous SINES across a wide phylogenetic range. Currently, the most widespread SINES are the CORE-SINES found in bilaterians (Gilbert and Labuda 1999), V-SINES in vertebrates (Ogiwara et al. 2002), DeuSINES in deuterostomes (Nishihara et al. 2006), and CephSINES in cephalopods (Akasaki et al. 2010). Given the power of some of these SINES to functionally affect genomic outputs (Santangelo et al. 2007; Sasaki et al. 2008) and to therefore affect genome evolution, we were interested to ask whether there are any SINES that have a deeper metazoan ancestry.

To do this, we first searched publicly available metazoan genome assemblies using an in house “ab initio” program (to be described elsewhere) designed to detect SINES flanked by direct repeats with a length of more than 15 nt and a minimum of 90% homology. Sequences detected by this strategy were required to occur more than ten times per genome to be considered for further analysis. Putative SINES were compared with the Repbase database (http://www.girinst.org/) using the RepeatMasker program (www.repeatmasker.org) and further characterized by homology-based searches using BlastN against GenBank databases (wgs and dbEST). Intragenomic consensus SINE sequences were reconstructed if SINE regions ≥50 bp sharing >50% similarity across different metazoan species were detected using CodonCode Aligner (Version 3.5.6) and BioEdit (Hall 1999). Using this strategy, we have characterized five new tRNA-related SINES from disparate eumetazoan species: a cnidian (Nematostella vectensis); two molluscs (Lottia gigantea and Aplysia californica); an annelid (Capitella teleta); and an arthropod (bodies scapularis). All of these SINES possess a conserved central domain that we have named the Nin-domain. This domain was first discovered in the coelacanth Latimeria menadoensis and was subsequently found in many Deuterostomia species and was therefore named the Deu-domain (Nishihara et al. 2006). We have found that a central part of the 340 bp Deu-domain is present in protostome and diploblast taxa (fig. 1). We therefore propose naming this highly conserved SINE domain the Nin-domain. The overall metazoan pairwise identity among Nin-domain consensus sequences using MEGA4 (Tamura et al. 2007) is approximately 63%. The Nin-domain cannot be aligned with conserved regions of other widely distributed SINES, namely CORE-SINES (Gilbert and Labuda 1999), V-SINES (Ogiwara et al., 2002), and CephSINES (Akasaki et al. 2010). From our ab initio scans and subsequent homology-based searches, we detected the Nin-domain in representatives of the Deuterostomia, Lophotrochozoa, Ecdysozoa, and Cnidaria making this SINE domain the most...
phylogenetically widespread, vertically transferred SINE sequence currently known.

All Nin domain–containing SINEs (Nin-DC-SINEs) possess a 5′ tRNA-related Pol III promoter region, a tRNA-unrelated region (within which lies the Nin-domain), and a variable 3′ tail sequence (often with a repeat motif) which is unique to each metazoan lineage (figs. 1 and 2, supplementary figs. S1–S3, Supplementary Material online). We distinguished 17 new subfamilies of Nin-DC-SINEs in five eumetazoan species (six in N. vectensis, three in L. gigantea, two in A. californica, three in C. teleta, and three in I. scapularis) based on diagnostic nucleotides, including different head and tail regions (see supplementary fig. S2a–e, Supplementary Material online). Nin-DC-SINEs exist as high copy genomic elements (table 1) and are flanked by clearly recognizable target site duplications (TSD) that are diagnostic of amplification via RNA intermediates. Thus, Nin-DC-SINES in eumetazoan genomes are retrotranspositionally active. Interestingly, the Nin-DC-SINEs previously described in Amniote genomes (AmnSINE1) lack these TSDs and have therefore lost their retrotranspositional activity (Nishihara et al. 2006). Furthermore, the recently characterized tRNA-derived “Ruka” SINE identified in the genomes of several tick species (Sunter et al. 2008),

![Fig. 1. Schematic illustration of Nin-DC-SINE architecture and distribution across the Metazoa. The previously described Deu-SINEs found in deuterostomes (Deu-domain highlighted green), and all newly characterized Nin-DC-SINEs from this study share the Nin-domain (shaded red) which is located within the central part of their tRNA-unrelated regions. The 3′ tail regions of all new Nin-DC-SINEs are unique and of unknown function and origin except in Nematostella vectensis. Dots indicate that there are known LINE partners with identical 3′ tail regions. TSDs, target site duplications.](image)

![Fig. 2. Twenty-seven Nin-domain consensus sequences in eumetazoan genomes of human (Homo sapiens = Hs), chicken (Gallus gallus = Gg), coelacanth (Latimeria menadoensis = Lm), salmon (Oncorhynchus mykiss = Om), zebrafish (Danio rerio = Dr), catfish (Ictalurus punctatus = Ip), dogfish shark (Squalus acanthias = Sa), hagfish (Eptatretus burgeri = Eb), amphioxus (Branchiostoma floridae = Bf), sea urchin (Strongylocentrotus purpuratus = Sp), owl limpet (Lottia gigantea = Lg), sea hare (Aplysia californica = Ac), annelid worm (Capitella teleta = Ct), tick (Ixodes scapularis = Is), and sea anemone (Nematostella vectensis = Nv). A similarity threshold of >50% was used for shading. See supplementary figure S2a–e, Supplementary Material online for complete sequence information.](image)
is in fact a Nin-DC-SINE. In addition to these shorter Ruka-like Nin-DC-SINEs (240 bp), we identified an additional Nin-DC-SINE of 290 bp in the genome of the blacklegged tick *I. scapularis* (fig. 1). Interestingly, the subfamily *Isc*-Nin-DC-SINE3 is part of an uncharacterized long terminal repeat-Gypsy element (supplementary fig. S2, Supplementary Material online).

All *N. vectensis* Nin-DC-SINE copies possess a highly variable 3′ tail (supplementary figs. S2a and S3, Supplementary Material online). Because this 3′ region is involved in the mobilization of SINE elements (Kajikawa and Okada 2002), it can be expected that many different LINEs are responsible for the retrotransposition of *N. vectensis* Nin-DC-SINES. To this end, we have identified three putative Nin-SINE-LINE partners in the *N. vectensis* genome (supplementary fig. S3, Supplementary Material online). These LINEs all belong to the CR1/LINE 2 clade (Repbase Update database: http://www.girinst.org/, Jurka et al. 1995; Smit 1996).

We were also interested in whether Nin-DC-SINES could resolve any of the recently proposed and conflicting hypotheses regarding early metazoan evolution. According to Dunn et al. (2008) sponges form a sister group to the Cnidaria with ctenophores representing the earliest branching metazoan lineage. In contrast, Pick et al. (2010) placed sponges as the earliest diverging monophyletic metazoan lineage, whereas according to Schierwater et al. (2009), placozoans are the earliest branching member of a “Diploblasta” clade that is sister to the Bilateria. Exhaustive ab initio and homology-based searches of publicly available draft genomes did not reveal any Nin-DC-SINES from the choanoflagellate *Monosiga brevicollis* (King et al. 2008), the demosponge *Amphimedon queenslandica* (Srivastava et al. 2010), or the placozoan *Trichoplax adhaerens* (Srivastava et al. 2008). Unfortunately, the lack of a publicly available ctenophore genome prevents us from comprehensively distinguishing between all of the above phylogenetic scenarios. However, based on the simple presence/absence of Nin-DC-SINES in the available genomes (regardless of genomic location), the Diploblasta hypothesis (Schierwater et al. 2009) can be rejected on the basis of parsimony and the assumption that Nin-DC-SINES were not horizontally transferred between cnidarian and bilaterian ancestors (supplementary fig. S4, Supplementary Material online). This result supports a recent reanalysis of the Schierwater data (Philippe et al. 2011). However, it should be mentioned that certain ecdysozoan lineages such as nematodes and terrestrial insects (*Drosophila, Apis, Anoplophora, Bombyx*, and *Tribolium*) lack clear Nin-DC-SINES in their genomes (in contrast to *I. scapularis*), indicating that Nin-DC-SINES can be lost.

The evolutionary persistence of Nin-DC-SINES is presumably related to the function of the Nin-domain. Gilbert and Labuda (1999) characterized the CORE-SINES which all share a conserved 65 bp central region that gave rise to a number of different SINE families including mammalian-wide interspersed repeats (Jurka et al. 1995). They hypothesize that CORE-SINES recruited the 5′ internal promoter regions of highly transcribed host tRNAs as well as the 3′ sequence of actively amplifying LINEs. It has recently been shown that CORE-SINES can act as functional modules that enhance gene expression in mammals (Santangelo et al. 2007), thus providing an explanation for their maintenance. Additionally (Sasaki et al. 2008) demonstrated that two AmnSINE1 loci function as distal transcriptional enhancers during forebrain development of murine embryos. Thus, SINES with conserved domains can play important roles as regulators of gene expression. Bioinformatic studies supported by functional assays may reveal the reason for the deep conservation of the Nin-domain and would provide further insight into the molecular events that supported the radiation of the Metazoa.

### Supplementary Material

Supplementary figures S1–S4 are available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals.org/).

### Acknowledgments

We acknowledge the Joint Genome Institute Community Sequencing Program for providing public access to draft genome assemblies of *Amphimedon queenslandica, N. vectensis, C. teleta*, and *Lo. gigantea*. Many thanks go to Friedrich Humann for bioinformatic support. We gratefully acknowledge Hidenori Nishihara’s initial RepeatMasker searches of *Lottia, Capitella*, and *Trichoplax* genomes. This work was supported by Deutsche Forschungsgemeinschaft funding to D.J.J. and the Courant Research Centre for Geobiology (Göttingen) via the German Excellence Initiative. Public database accession numbers: Representative Nin-DC-SINE sequences have been deposited in GenBank with the following accession numbers and will be publicly available upon publication: HM595747–HM595761. See supplementary figure S2a–e, Supplementary Material online for more sequence information.

---

**Table 1. Nin Domain—Containing SINES Are Highly Represented in Their Respective Eumetazoan Genomes.**

<table>
<thead>
<tr>
<th></th>
<th>Nemastostella</th>
<th>Lottia</th>
<th>Aplysia</th>
<th>Capitella</th>
<th>Ixodes</th>
<th>Stronglylocentrotus</th>
<th>Homo</th>
</tr>
</thead>
<tbody>
<tr>
<td>Approx. genome size (Mb)</td>
<td>350</td>
<td>350</td>
<td>715</td>
<td>330</td>
<td>1,700</td>
<td>900</td>
<td>3,200</td>
</tr>
<tr>
<td>Nin-DC-SINE cons. sequence length (bp)</td>
<td>373</td>
<td>372</td>
<td>537</td>
<td>330</td>
<td>290</td>
<td>366</td>
<td>576</td>
</tr>
<tr>
<td>Nin-DC-SINES as a proportion of genome (%)</td>
<td>0.03</td>
<td>0.11</td>
<td>0.29</td>
<td>0.37</td>
<td>0.11</td>
<td>0.01</td>
<td>0.004</td>
</tr>
<tr>
<td>Approx. no. of full Nin-DC-SINES in genome</td>
<td>310a</td>
<td>1,028</td>
<td>3,841</td>
<td>3,658</td>
<td>6,608</td>
<td>330</td>
<td>66c</td>
</tr>
</tbody>
</table>

* a Based on BlastN searches with hits against >90% of a Nin-domain consensus sequence.

b Nin-DC-SINES in the genome of *N. vectensis* display six different 3′ tails and are likely mobilized by different partner LINEs.

c Based on RepeatMasker searches with hits against >60% of a Nin-domain consensus sequence. See text for details.
References


Brosius J. 1999. RNAs from all categories generate retrosequences that may be exalted as novel genes or regulatory elements. Gene 238:115–134.


