Supporting Information

Arriagada et al. 10.1073/pnas.1409945111

SI Materials and Methods

***Mya arenaria* Collection.** *Mya arenaria* were collected from Prince Edward Island (PEI), Canada and evaluated for leukemia during two surveys in 2009 and two in 2010 (n = 100–150 per site per survey). The clams were dug at various high- and low-intensity potato farming estuaries around PEI as previously described (1). For the second survey in 2009 and for the 2010 surveys, sample collection transects were established through the Dunk (D) and Wilmot (W) estuaries (13.6–42% potato farming) from near-field, through midfield, to far-field sites (the location of collection is reflected in the clam ID). *M. arenaria* were hand dug at low tide and transported to a field laboratory as previously described (1). All samples were processed within 24 h of collection. *M. arenaria* were collected from Larrabee Cove, ME on one date in September 2013 and kindly provided by Brian Beal (University of Maine at Machias).

Diagnoses of Leukemia. Clams from PEI were screened for disease status by withdrawing 0.1 mL of hemolymph from the posterior adductor muscle into a dry sterile 1-mL syringe fitted with a sterile 23-gauge needle. The exterior of the clam was wiped with a tissue soaked in 70% ethanol before insertion of the needle. A single drop of hemolymph was placed on a microscope slide and left to settle for 5 min before being examined using a phase-contrast microscope (Leica DMLS, 400× magnification). Visual screening was consistently conducted by the same team member, during each survey. On the basis of the apparent cell density and shape of the hemocytes (small and rounded, absence of appendages), each clam was designated as either "normal" (no leukemic hemocytes, N), "moderate" (20–50% leukemic hemocytes, M), or "heavily leukemic" (>50% leukemic hemocytes, L) (1). Clams from Maine were shipped on ice to Columbia University, New York, NY and diagnosed using a similar protocol to that used in Canada (six drops of hemolymph were placed in a well of a 96-well plate and incubated 1 h at 10 °C before microscopic inspection of cell morphology). Of 92 Maine clams scored, all 3 leukemic clams were used in the study, and 6 representative normal clams were used as controls.

Samples for Molecular Analyses. Hemocytes were pelleted in a refrigerated centrifuge for 5 min at 9,600 × g. Supernatants were discarded, and the remaining pellets were resuspended in RNAlater (Invitrogen) and stored at 4 °C for transportation to the Canada Centre for Inland Waters (CCIW), Burlington, ON, after which they were stored at −18 °C.

Hemocyte Cultures. Hemocytes from leukemic clams and normal reference clams were cultured according to the method of Walker et al. (2). Working aseptically at all times, the surface of the clam was wiped with ethanol, and the remainder of the hemolymph was removed as previously described (*Diagnoses of Leukemia*). The hemolymph was added to 10 mL of sterile Walker's medium (2) at room temperature. The hemocytes were then sedimented by centrifugation at 105 × g for 10 min at 8 °C. This "preculture supernatant" was transferred to 5-mL cryovials and flash frozen in liquid nitrogen. The hemocytes were then gently resuspended in 10 mL Walker's medium and incubated at 8 °C for 15 min in a tube inverter, after which they were sedimented by centrifugation for 8 min at 105 × g. The foregoing procedure was repeated three times for leukemic hemocytes, after which viability was assessed by Trypan Blue exclusion. The cell suspension was then counted and adjusted to 4–7 × 10^4 cells/mL by the addition of Walker's medium. Only contaminant free cell preparations with a viability of >95% were cultured. Normal hemolymph was added directly to 10 mL Walker's medium in a 15-mL tissue culture flask and incubated under stationary conditions at 8 °C. The leukemic cells were transferred to a 125-mL cell reactor/spinner flask and stirred at 32 rpm at 8–10 °C. After 12 h an aliquot of cell suspension was removed and tested for hemocyte count, viability, and evidence of microbial contamination. The foregoing procedure was repeated after 24 and 48 h. Upon completion of the incubation period the cell suspension was transferred to sterile 50-mL cell culture tubes, and the cells were sedimented by centrifugation at 67 × g for 15 min at 8 °C. The supernatant was transferred to labeled 5-mL cryovials ("post-culture supernatant"), flash frozen, and then stored in liquid nitrogen. Sufficient Walker's medium containing 10% (vol/vol) DMSO was added to the cell pellet to bring the cell count to 4 × 10^6 cells/mL. The cell suspension ("cultured cells") was then transferred to labeled 2-mL cryovials. The cryovials of cell suspension were then placed in a Nalgene "Mr. Frosty Cryo 1°C" apparatus (Thermo Scientific), which was preequilibrated to 8 °C. The loaded container was placed onto dry ice for at least 4 h, after which the frozen cell suspensions were stored in liquid nitrogen. All samples were transported from PEI to CCIW. Subsequently the frozen cultures were shipped on dry ice to Columbia University. Samples of culture medium were flash frozen and stored in liquid nitrogen until returned to CCIW, after which they were stored at −80 °C. Frozen culture medium and hemocytes in RNAlater were shipped on dry ice and ice, respectively, from CCIW to Columbia University.

Reverse Transcriptase Assay. Reverse transcriptase (RT) activity was detected using 5 μL of precultured hemolymph or postcultured medium, as previously described (3). Reactions were performed at 20 °C, with poly(rA):oligo(dT) template and Mn^{++} as divalent cation. Images were quantitated using ImageJ software (4).

454 Sequencing. Total RNA extracts were treated with DNase I (RNase-free, Ambion). cDNA was generated by using the SuperScript II system (Invitrogen) for reverse transcription primed by random octamers that were linked to an arbitrarily defined 17-mer (5′-GTT TCC CAG TAG GTC TCN NNN NNN N-3′). The resulting cDNA was treated with RNase H, converted to double-stranded DNA template using exoKlenow (NEB) and then randomly amplified by PCR, using a primer corresponding to the defined 17-mer sequence. Products >70 bp were selected by column purification (MinElute, Qiagen) and ligated to specific linkers for sequencing on the 454 Genome Sequencer FLX (454 Life Sciences) without template fragmentation (5, 6). A total of 259,724 reads were obtained. These were clustered using cluster database at high identity with tolerance (CD-HIT) at 98% identity, resulting in 77,146 unique reads. The clustered dataset had an average read length of 170 bp and average quality score of 30. The primers and adaptors were trimmed, and reads were length-filtered and masked for low complexity regions (WU-BLAST 2.0). A database was generated from the preprocessed reads and searched with Moloney MuLV sequences and with full-length endogenous *Steamer* sequence using BLASTN.

Cloning of a Retroelement-Related RNA. One milliliter of culture medium from Dnear-HL03 cells was thawed and passed through a 0.45-μm filter, and pelletable material in the filtrate was collected by ultracentrifugation through a 3-mL 20% sucrose cushion for 2 h

at 25,000 × *g* in a SW55 rotor. Total RNA was extracted from the pellet using Trizol reagent (Invitrogen). cDNA was generated using 200 ng of RNA and the Super Script First Strand Synthesis system (Invitrogen). Five reads derived from the 454 sequencing with similarity to a retroviral *pol* gene were selected, and the following primers were designed to align with those sequences: C000504-F1 $^{5'}$gcaagtggtaccacagaggaagtgc$^{3'}$, 57O1-F2 $^{5'}$cgactgtgcttctggttattggc$^{3'}$, 57O1-F3 $^{5'}$gcgtttgtaacaccttcaggtgc$^{3'}$, WX65-F4 $^{5'}$gcggtgaaaggtgcgttatacctc$^{3'}$, WX65-R2 $^{5'}$tgactggcacgcttcacatttcc$^{3'}$, CX07-F5 $^{5'}$ccacgtacctctcgaacttgtatgc$^{3'}$, and C1Q18-R1 $^{5'}$ggcctaacatgactttgttcgg$^{3'}$. PCR reactions were performed using *PfuUltra* II fusion HS polymerase (Agilent Technologies) and the combinations of primers indicated in Fig. 1. The PCR products were TOPO cloned (Invitrogen) and sequenced.

Genome Walking. For genomic DNA extraction, frozen hemocytes of leukemic and nonleukemic animals were digested with 0.1 mg/mL of proteinase K in digestion buffer [100 mM NaCl, 10 mM Tris·HCl (pH 8.0), 25 mM EDTA, and 0.5% SDS] at 37 °C overnight, after which phenol-chloroform extraction and DNA precipitation were performed. The DNA was resuspended in buffer TE (pH 8.0) and stored at 4 °C.

Genome walking was performed using the Genome Walker Universal kit (Clontech). The primers 5′GW-1 $^{5'}$gcagcaagtccaagaagtggggcaaattcg$^{3'}$ and 5′GW-1nested $^{5'}$gtctttgcctgtgtgatctcggttttctg$^{3'}$ were designed for a first specific 5′ walk. Once PCR products were cloned and sequenced, the primers 5′GW-2 $^{5'}$ggtggaaatgggatcattgaaggaacagc$^{3'}$ and 5′GW-2 nested $^{5'}$tggctagtggtattgttgtgggtgggaaaa$^{3'}$ were designed for a second 5′ walk. For the first 3′genome walk, the primers 3′GW-1 $^{5'}$cgccaccagaagcaaagccatacttca$^{3'}$ and 3′GW-1nested $^{5'}$tcaaccgagcgcagtgtgtgttttg$^{3'}$ were designed. Once the PCR products were cloned and sequenced, the primers 3′GW-2 $^{5'}$tgctgagccagggacgagtgaccattg$^{3'}$ and 3′GW-2nested $^{5'}$tggtttcccaaacgaggccaaacaaac$^{3'}$ were designed for a second 3′ walk. All PCR products were TOPO cloned and sequenced.

Southern Blotting Analysis. *M. arenaria* genomic DNA (20 μg) was digested with the restriction endonucleases BamHI, DraI, or HindIII (5 U/μg DNA) for 2 h at 37 °C, followed by the addition of 5 more units of enzyme and incubation overnight. Digested DNA was precipitated and resuspended in 25 μL of TE buffer (pH 8.0). DNAs (15 μg per lane) were separated by electrophoresis in a 0.7% agarose gel. After ethidium bromide staining, DNAs were denatured in alkaline transfer buffer (0.4 M NaOH and 1 M NaCl) and transferred to a nylon membrane. The membrane was neutralized by incubation with neutralization solution [0.5 M Tris·HCl (pH 7.2) and 1 M NaCl] and prehybridized for 1 h at 42 °C in ULTRAhyb (Ambion). The probe was obtained by PCR from heavily leukemic genomic DNA using the primers Clamprobe-F $^{5'}$cctgccgatcattgaagatttactacc$^{3'}$ and Clamprobe-R $^{5'}$agttgccaagaaactttgtgagg$^{3'}$, and 30 ng of the probe were labeled using {α-^{32}P}dCTP and the Prime-It II Random Primer Labeling Kit (Agilent Technologies). Hybridization in ULTRAhyb with the labeled probe was performed at 42 °C for 20 h. After two washes with 2× SSC, 0.1% SDS for 5 min at 42 °C and two washes with 0.1× SSC, 0.1% SDS for 15 min at 42 °C, the membrane was exposed to X-ray film or to Typhoon plate, exposing for 3 h.

Phylogenetic Analysis. The amino acid sequences of the conserved regions of the Gag, Protease, RT, RNase H, and Integrase (IN) domains of *Steamer* were added to an alignment of representative sequences from a database of retrotransposon sequences (7). PhyML 3.0 (8) was used to generate a maximum likelihood phylogenetic tree using the LG substitution model with 100 replicates for bootstrap analysis.

Quantitative RT-PCR. RNA was extracted from hemocytes conserved in RNAlater using Trizol reagent according to the manufacturer's instructions and treated with RNase-free DNaseI (Invitrogen). cDNA was generated using 500 ng of RNA and the SuperScriptIII First-Strand Synthesis SuperMix for quantitative RT-PCR (qRT-PCR) kit (Invitrogen) according to the instructions. One microliter of cDNA was used in each of the qPCR reactions to detect *Steamer* RNA with the FastStart Universal SYBR Green Master (Rox) kit (Roche) using the primers clamRT-F $^{5'}$tgcgtcggaaaccggtcttgg$^{3'}$and clamRT-R $^{5'}$caaccactcggcgcccgtat$^{3'}$, or to detect EF1 mRNA using the primers clamEF1F $^{5'}$gaaggatgagggaaaagaggg$^{3'}$ and clamEF1R $^{5'}$cacattttcctgctatggtgc$^{3'}$ (9). The levels of *Steamer* mRNA were calculated using a standard curve and expressed as relative to the *EF1* mRNA levels. The levels of *Steamer* RNA in normal and leukemic clams were compared using a two-tailed *t* test and the GraphPad Prism6 program.

qPCR. qPCR of genomic DNA was conducted using the same primer pairs as in qRT-PCR. Twenty-five nanograms of genomic DNA was used per reaction in triplicate. Copy number of *Steamer RT* and *EF1* was determined by a standard curve using a single linearized plasmid containing both a full length copy of *Steamer* and the clam *EF1* fragment cloned from WfarNM01 DNA.

Inverse PCR. Genomic DNA from mantle tissue (WfarNM01) or leukemic hemocytes (Dnear08 and DnearHL03) was extracted (DNeasy Kit, Qiagen) from PEI clams, and for Maine clams genomic DNA was extracted from both hemocytes and siphon tissue (a tissue with relatively low levels of hemocyte infiltration). One hundred or 125 ng was first digested overnight with 2.5 U of MfeI-HF (NEB) at 37 °C, which does not cut in the *Steamer* element. Digested DNA was ligated with T4 DNA ligase (NEB) in a 25-μL reaction for 20 min at room temperature, heat inactivated for 10 min at 65 °C, and digested for 4 h at 37 °C with 5 U of NsiI (NEB), which cuts four times in the *Steamer* element. DNA was purified (PCR purification kit, Qiagen), and integration junctions were amplified with *PfuUltra* II Fusion HS polymerase using primers in the *Steamer* LTRs (ClamLTR-F2, $^{5'}$acatgcacattaaaagttatcg$^{3'}$ and ClamLTR-R1, $^{5'}$ttagtatagccaatactgttac$^{3'}$). The PCR protocol consisted of incubations at 95 °C for 2 min, followed by 35 cycles of 95 °C for 20 s, 50 °C for 20 s, and 68 °C for 5 min, with a final extension at 72 °C for 5 min. Inverse PCR products were analyzed on an agarose gel, isolated by gel extraction of specific bands or PCR purification of the whole PCR product (Qiagen) and cloned using the Zero Blunt TOPO cloning kit (Invitrogen). DNA sequences of the inserts in individual cloned plasmids were determined using flanking M13F and M13R primers. The integration sites were confirmed by a diagnostic PCR using ClamLTR-F2 and a reverse primer in the genomic DNA flanking the corresponding integration site (enSR6 $^{5'}$tccagccatgtgttcctgct$^{3'}$; IMDL8c1R $^{5'}$aactccaataccttcaatt$^{3'}$; IMDL8c6R $^{5'}$agctgtctagattggaagtg$^{3'}$; IMHL03c2R $^{5'}$attgtcccagattcacagat$^{3'}$; and IMHL03c3R $^{5'}$gtaggtcttatacatttgag$^{3'}$). For these reactions 10 ng of DNA was used with Taq polymerase at 95 °C for 5 min, followed by 35 cycles of 95 °C for 30 s, 50 °C for 30 s, and 72 °C for 30 s, with a final extension of 72 °C for 5 min (products are ~150 bp each).

The complete endogenous *Steamer* sequence (enS6) was amplified from normal clam genomic DNA (WfarNM01) with primers enSR6 and enSF1 $^{5'}$cgcagggatcaatagacgacac$^{3'}$ (Fig. S1 and GenBank accession no. KF319019).

1. Muttray A, et al. (2012) Haemocytic leukemia in Prince Edward Island (PEI) soft shell clam (*Mya arenaria*): Spatial distribution in agriculturally impacted estuaries. *Sci Total Environ* 424:130–142.

2. Walker C, et al. (2009) Mass culture and characterization of tumor cells from a naturally occurring invertebrate cancer model: Applications for human and animal disease and environmental health. *Biol Bull* 216(1):23–39.

3. Goff S, Traktman P, Baltimore D (1981) Isolation and properties of Moloney murine leukemia virus mutants: Use of a rapid assay for release of virion reverse transcriptase. *J Virol* 38(1):239–248.
4. Schneider CA, Rasband WS, Eliceiri KW (2012) NIH Image to ImageJ: 25 years of image analysis. *Nat Methods* 9(7):671–675.
5. Margulies M, et al. (2005) Genome sequencing in microfabricated high-density picolitre reactors. *Nature* 437(7057):376–380.
6. Cox-Foster DL, et al. (2007) A metagenomic survey of microbes in honey bee colony collapse disorder. *Science* 318(5848):283–287.
7. Llorens C, et al. (2011) The Gypsy Database (GyDB) of mobile genetic elements: Release 2.0. *Nucleic Acids Res* 39(Database issue):D70–D74.
8. Guindon S, et al. (2010) New algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the performance of PhyML 3.0. *Syst Biol* 59(3): 307–321.
9. Siah A, McKenna P, Danger JM, Johnson GR, Berthe FC (2011) Induction of transposase and polyprotein RNA levels in disseminated neoplastic hemocytes of soft-shell clams: *Mya arenaria*. *Dev Comp Immunol* 35(2):151–154.

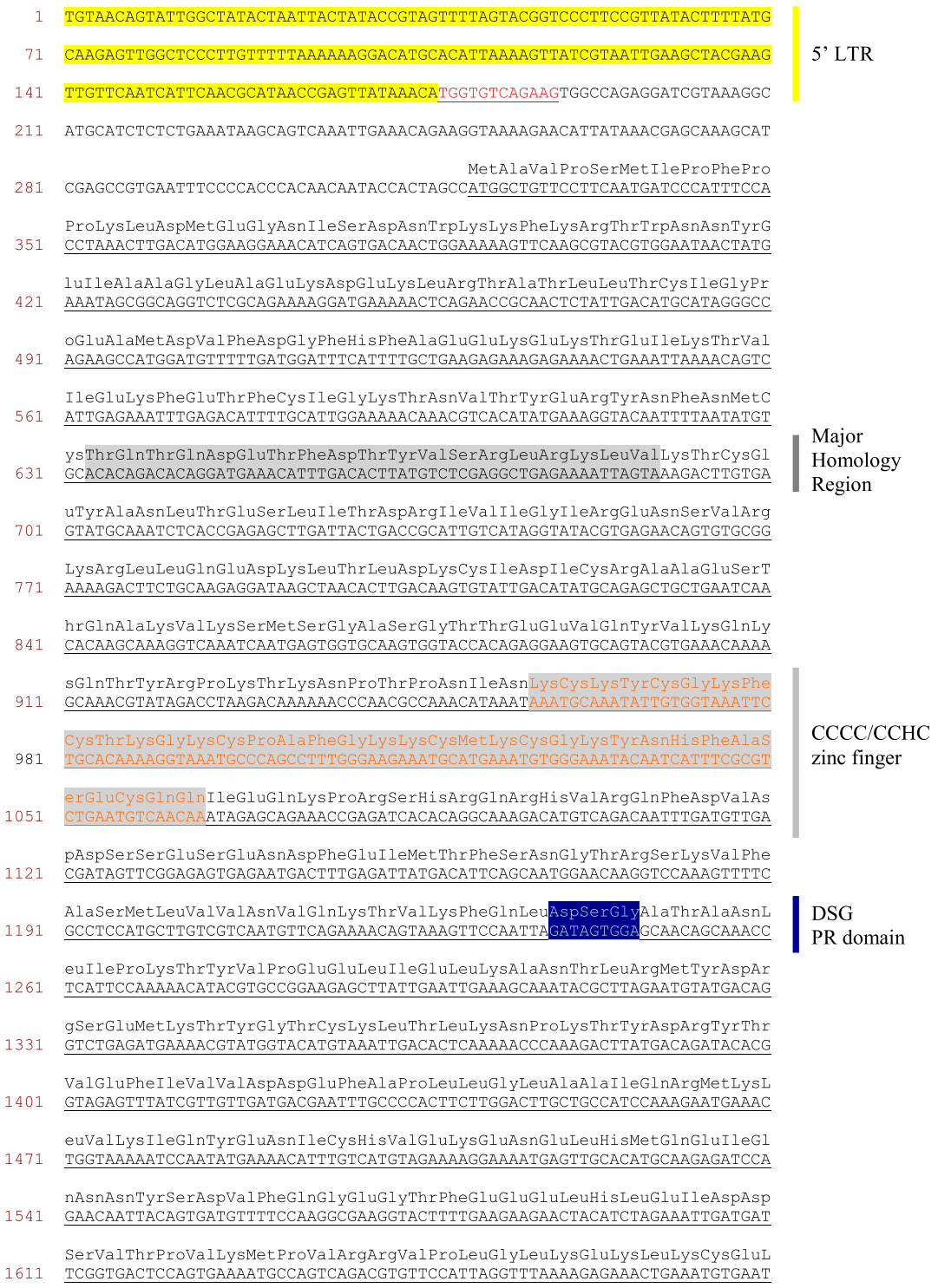

Fig. S1. (Continued)

Fig. S1. (Continued)

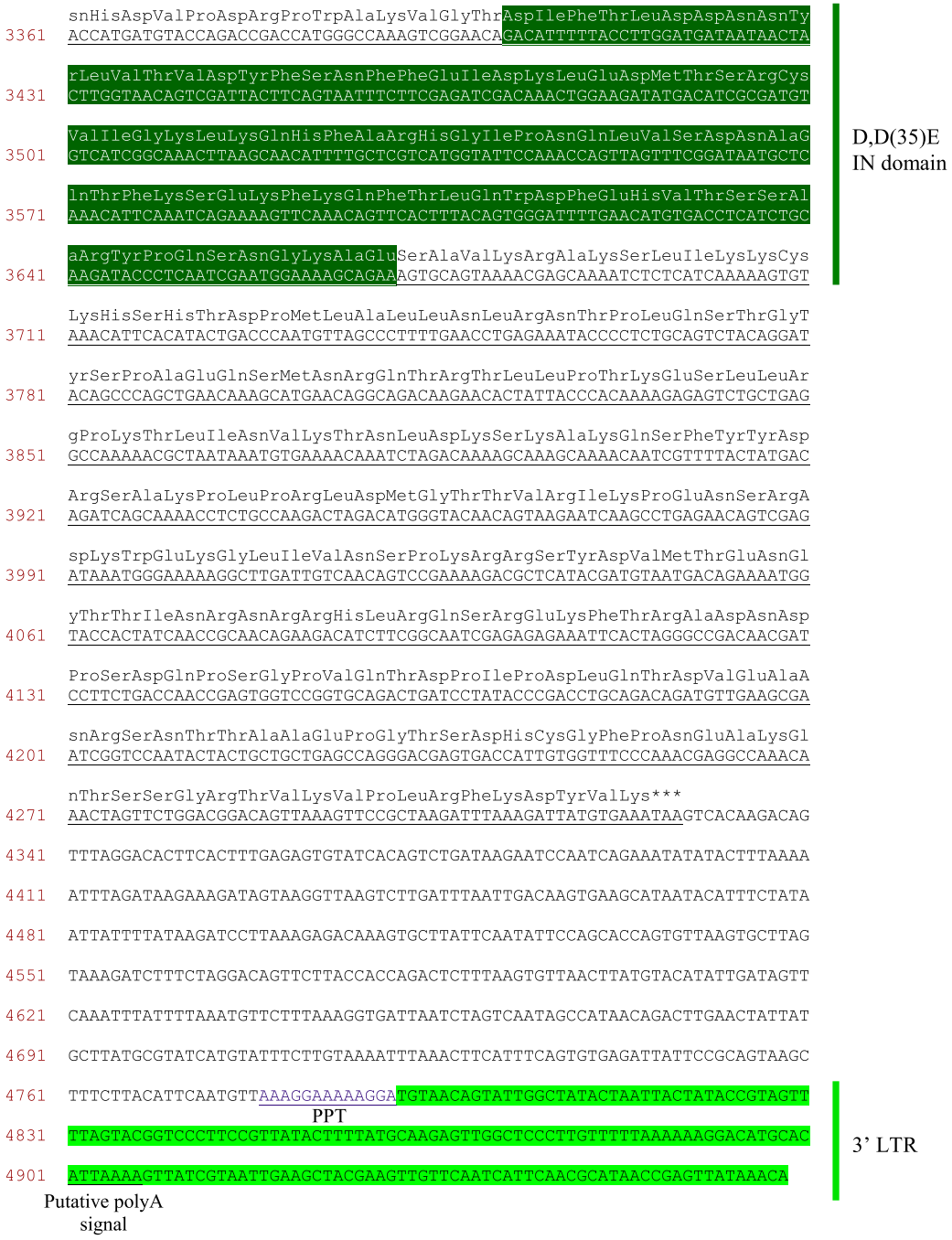

Fig. S1. Nucleotide sequence of complete *Steamer* retroelement. Specific features are highlighted.

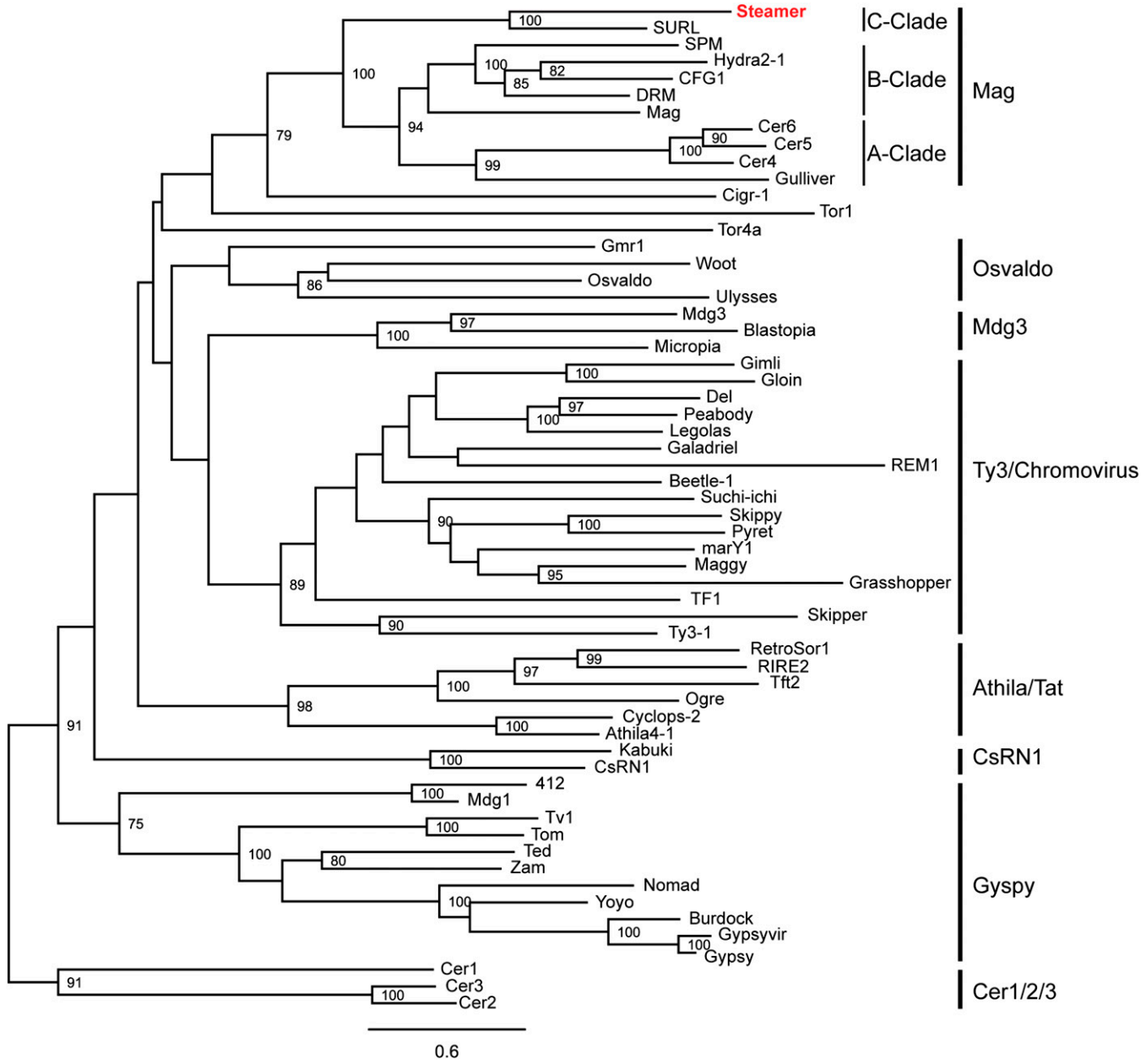

Fig. S2. *Steamer* phylogenic tree. A maximum likelihood tree generated by PhyML using the amino acid sequences of the conserved regions of the Gag, Protease, RT, RNase H, and IN domains of *Steamer* and representative sequences from a database of retrotransposon sequences. Bootstrap values above 75 are shown.

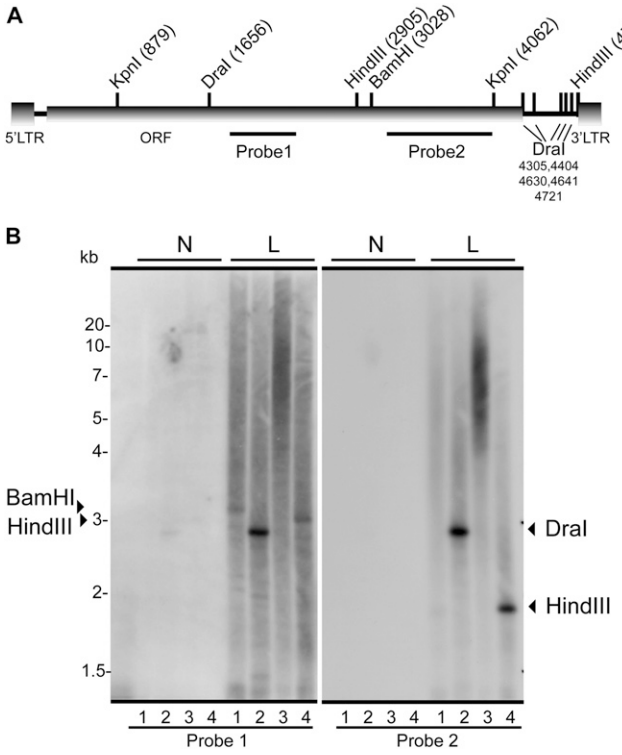

Fig. S3. Organization of amplified *Steamer* elements is internally homogeneous with heterogeneous flanking DNA. The organization of *Steamer* DNAs from both normal and diseased animals were analyzed by Southern blot with several restriction enzyme digests and two hybridization probes. (*A*) Schematic of the retrotransposon. Positions of selected restriction enzyme digestion sites and two hybridization probes are indicated. (*B*) DNA from hemocytes of a normal (N, clam nor1) and highly leukemic (L, clam Dnear-HL03) were digested with enzymes: lanes 1: BamHI; lanes 2: DraI; lanes 3: EcoRI; lanes 4: HindIII. Blots were hybridized with probe 1 (*Left*) or probe 2 (*Right*) as indicated. Positions of major internal fragments released from the L DNA by BamHI, HindIII, and DraI are indicated with arrows. The "non-cutter" EcoRI only releases a large smear of DNAs of heterogeneous sizes. In all cases, digests predicted to release internal fragments yielded DNA fragments of the expected sizes, suggesting general homogeneity of sequence and close identity to the cloned *Steamer* DNA. Digests probed so as to detect junction fragments produced small number of bands in normal DNA, and an intense smear indicative of heterogeneous integrations of many copies of the element in diseased DNA.