Supplementary materials and methods

Molecular cloning of repetitive sequence

High molecular weight genomic DNA was isolated from about 200 μl of heparinized blood using the phenol-chloroform extraction method. Molecular cloning of repetitive sequences was performed as described previously.[1] Five μg of genomic DNA was digested with 26 restriction enzymes (ApaI, BamHI, BglI, BglII, DdeI, DraI, EcoRI, EcoRV, HaeII, HaeIII, HapII, HindIII, HinfI, HpaII, MspI, NcoI, NotI, PstI, PvuII, SacI, SalI, SmaI, SpeI, SphI, XbaI, and XhoI), and subjected to electrophoresis in a 1.2% agarose gel. The intense DNA bands of repetitive sequences were isolated from the gel, purified, and ligated into pBluescript SK(+) vector (Stratagene, La Jolla, CA, USA). The ligation mixture was used to transform *E. coli* XL1-Blue MRF (Agilent Technologies, Santa Clara, CA, USA). Nucleotide sequences of the DNA clones were determined using an ABI PRISM 3130 DNA sequencer (Thermo Fisher Scientific-Applied Biosystems, Carlsbad, CA, USA) after sequencing reactions with a Big Dye Terminator v3.1 Cycle Sequencing Kit (Thermo Fisher Scientific-Applied Biosystems).

Fluorescence in situ *hybridization (FISH)*

FISH analysis was performed as previously described.[2] The DNA fragments were labeled with biotin-16-dUTP or digoxigenin-11-dUTP using a nick translation kit (Roche Diagnostics, Basel, Switzerland). After hybridization, the slides were incubated with FITC-avidin (Thermo Fisher Scientific-Molecular Probes, Carlsbad, CA, USA) for biotin-labeled probes or with mouse monoclonal anti-digoxigenin clone DI-22 ascites fluid (Sigma-Aldrich, St. Louis, MO, USA) labeled with Cy3 using a Cy3 Ab labeling kit (GE Healthcare, Buckinghamshire, UK) for digoxigenin-labeled probes, and counterstained with DAPI (4', 6-diamidino-2-phenylindole-dihydrochloride) or propidium iodide (PI). The digital FISH

images were captured with the 550CW-QFISH application program (Leica Microsystems Imaging Solution, Cambridge, UK) using a cooled CCD camera (Leica DFC360 FX, Leica Microsystems, Wetzlar, Germany) mounted on a Leica DMRA microscope.

Chromosomal location of the 18S-28S ribosomal RNA genes was determined using pHr21Ab (5.8-kb for the 5' portion) and pHr14E3 (7.3-kb for the 3' portion) fragments of human 45S pre-ribosomal RNA gene (*RNA45S*), which were provided by National Institutes of Biomedical Innovation, Health and Nutrition, Osaka. For mapping of telomeric repeats, $(TTAGGG)_7$ and $(TAACCC)_7$ repeated sequence probes were used. To detect active site of 18S-28S rRNA genes, Ag-NOR staining was performed using the same chromosome slides used for FISH analysis of 18S-28S rRNA genes.[3] The 5S rDNA probe was amplified by PCR using genomic DNA of *Lethenteron camtschaticum*. A 96-bp fragment was amplified using the primers designed based on the 5S rDNA nucleotide sequences of this species (GenBank accession number D00076): 5'-acgaccatatcaccctgaat-3' and 5'-ggcggtctcccatccaagta-3'. The PCR reaction started at 95 ºC for 4 min before the reaction of 35 cycles of 95 ºC for 1 min, 62 ºC for 1 min, and 72 ºC for 1 min, and then followed by a single cycle reaction at 72 ºC for 5 min. The PCR products were subjected to electrophoresis in a 1.2% agarose gel and stained with ethidium bromide. PCR fragments were cloned into the pGEMT-easy vector (Promega, Madison, WI, USA) and sequenced.

Hybridization analysis of DNA blots

For Southern blot hybridization, the genomic DNA (5 μg) was digested with restriction endonucleases, fractionated on a 1% agarose gel, and transferred onto nylon membranes (Roche Diagnostics). DNA probes were labeled with digoxigenin-11-dUTP using a PCR DIG Labeling Mix (Roche Diagnostics) and hybridized to the membrane at 45 °C overnight in DIG

Easy Hyb solution (Roche Diagnostics). After hybridization, the membranes were washed sequentially at 45 °C in 2 × SSC, 1 × SSC, 0.5 × SSC, and 0.1 × SSC, all of which contained 0.1% SDS, for 15 min each. Chemiluminescent signals were detected with anti-digoxigenin-AP Fab fragments and CDP-Star (Roche Diagnostics) and exposed to Biomax MS-1 Autoradiography Film (Carestream Health, Rochester, NY, USA).

For slot-blot hybridization, genomic DNAs of the following species were extracted as previously documented[4] and used: Arctic lamprey (*Lethenteron camtschaticum*, Petromyzontidae), sea lamprey (*Petromyzon marinus*, Petromyzontidae), pouched lamprey (*Geotria australis*, Geotriidae), and southern lamprey (*Mordacia mordax*, Mordaciidae) of Petromyzontiformes, inshore hagfish (*Eptatretus burgeri*, Myxinidae, Myxiniformes), lesser spotted catshark (*Scyliorhinus canicula*, Scyliorhinidae, Carcharhiniformes), and medaka (*Oryzias latipes*, Adrianichthyidae, Beloniformes). Genomic DNA was denatured with 0.4N NaOH for 10 min and transferred onto nylon membranes using BIO-DOT SF blotting equipment (Bio-Rad, Hercules, CA, USA). DNA probes were labeled with digoxigenin-11-dUTP using a PCR DIG Labeling Mix (Roche Diagnostics) and hybridized to the membrane at 45 °C in DIG Easy Hyb solution. The luminescent signals were detected as described in Southern blot hybridization.

References

1. Uno, Y., Asada, Y., Nishida, C., Takehana, Y., Sakaizumi, M., and Matsuda, Y. 2013, Divergence of repetitive DNA sequences in the heterochromatin of medaka fishes: molecular cytogenetic characterization of constitutive heterochromatin in two medaka species: *Oryzias hubbsi* and *O. celebensis* (Adrianichthyidae, Beloniformes). *Cytogenet. Genome Res.*, **141**, 212–26.
2. Matsuda, Y. and Chapman, V.M. 1995, Application of fluorescence in situ hybridization in genome analysis of the mouse. *Electrophoresis*, **16**, 261–72.
3. Howell, W.M., Hsu, T.C., and Block, B.M. 1977, Visualization of centriole-bodies using silver stain. *Chromosoma*, **65**, 9–20.
4. Kuraku, S., Qiu, H., and Meyer, A. 2012, Horizontal transfers of Tc1 elements between teleost fishes and their vertebrate parasites, lampreys. *Genome Biol. and Evol.*, **4**, 929–36.
5. Kohany, O., Gentles, A.J., Hankus, L., and Jurka, J. 2006, Annotation, submission and screening of repetitive elements in Repbase: RepbaseSubmitter and Censor. *BMC Bioinformatics*, **25**, 474.
6. Caputo, V., Giovannotti, M., Cerioni, P.N., Splendiani, A., Tagliavini, J., and Olmo, E. 2011, Chromosomal study of a lamprey (*Lampetra zanandreai* Vladykov, 1955) (Petromyzonida: Petromyzontiformes): conventional and FISH analysis. *Chromosome Res.*, **19**, 481–91.

Supplementary Table S1. The distribution of chromosome numbers in metaphase spreads of *Lethenteron camtschaticum.*

Number of chromosomes	165	166	167	168	169	170	Total
Number of cells	1	3	17	53 (65.4)[a]	6	1	81

[a]% of cells with a modal diploid chromosome number

Supplementary Table 2. List of nucleotide sequences of integrase domains of Ty3/Gypsy families used for studying the phylogenetic relationship with the LCA-ApaI sequence

Sequence name	Species name	Common species name	Accession number
Bartez	*Tetraodon nigroviridis*	Spotted green pufferfish	AJ621589
CER1	*Caenorhabditis elegans*	Roundworm	U15406
Cigr2	*Ciona intestinalis*	Yellow sea squirt	JF357717
CRR1	*Oryza sativa*	Rice	M75723
Dregg1	*Danio rerio*	Zebrafish	CR547121
GYPSY	*Drosophila melanogaster*	Fruit fly	P10401
Gypsy-26_PM-I	*Petromyzon marinus*	Sea lamprey	–[a]
HIV	human immunodeficiency virus 1	Human immunodeficiency virus 1	ADA71678
MAG	*Bombyx mori*	Domestic silkworm	X17219
LCA-ApaI	*Lethenteron camtschaticum*	Arctic lamprey	LC149811–LC149817[b]
LReO3	*Oryzias latipes*	Japanese medaka fish	Q8UUM8
RSV	Rous sarcoma virus	Rous sarcoma virus	FJ041197
Saci-2	*Schistosoma mansoni*	Flatworm	BK004069
SURL	*Tripneustes gratilla*	Hawaiian sea urchin	M75723
Sushi-ichi	*Takifugu rubripes*	Japanese pufferfish	AF030881
TED	Autographa californica nucleopolyhedrovirus	Nucleopolyhedrovirus	M32662
Ty3	*Saccharomyces cerevisiae*	Baker's yeast	Q7LHG5

[a]Gypsy-26_PM-I was taken from the Repbase (http://www.girinst.org/repbase/).[5]

[b]Seven LCA-ApaI fragments (LC149811–LC149817) isolated in this study.

Supplementary figure legends

Figure S1. Chromosome distribution of the 18S-28S rRNA and 5S rRNA genes, Ag-NORs, and telomeric TTAGGG repeats in *Lethenteron camtschaticum*. **(A)** DAPI-stained metaphase spread hybridized with biotin-labeled 18S-28S rRNA probe. Asterisks show the hybridization signals. Arranged three chromosomal pairs with FISH signals in the same metaphase spread and their C-banded patterns in a different metaphase spread are shown in the inset. **(B)** Ag-NOR staining pattern of the same metaphase spread shown in **(A)**. Arrows indicate Ag-positive regions. Ag-NOR staining patterns of the three pairs of chromosomes with FISH signals of the 18S-28S probes in **(A)** are shown in the inset. **(C)** Enlarged photographs of chromosome pairs with Ag-stained NORs in three different individuals. **(D)** FISH pattern of biotin-labeled 5S rDNA probe on PI-stained chromosomes. The 5S rDNA signals were localized to a single pair of small chromosomes that are different from chromosomes with 18S-28S rRNA signals (data not shown). Asterisks show the hybridization signals. **(E)** FISH pattern of Cy3-labeled telomeric TTAGGG repeats on a DAPI-stained metaphase spread. Scale bars represent 10 μm.

Figure S2. Ethidium bromide-stained gels of *Le. camtschaticum* genomic DNA digested with EcoRI (**A**) and ApaI (**B**). Arrowheads indicate the prominent DNA bands which were used for molecular cloning of repetitive sequences. Phi X174 DNA-HaeIII digest and ϕ X174 DNA-HincII digest were used as a molecular size marker in the left lane in (**A**) and (**B**), respectively.

Figure S3. Alignments of the fragments of LCA-EcoRIa, LCA-EcoRIb and LCA-ApaI sequence families and their consensus nucleotide sequences. **(A)** LCA-EcoRIa. **(B)** LCA-EcoRIb. **(C)** LCA-ApaI. Dots indicate identity with nucleotides in the consensus sequence

at the top; hyphens indicate gaps. Restriction enzyme recognition sites are underlined: EcoRI (dot and dash), AluI (dot), ApaI (double), HpaII/MspI (bold), PstI (wavy). Numerals at the end of lines represent the sequence length.

Figure S4. Molecular phylogenetic trees of LCA-EcoRIb sequences and LCA-ApaI sequences. **(A)** Molecular phylogenetic tree of LCA-EcoRIb and EcoRI satellite DNAs isolated from three lamprey species, *La. planeri*,[6] *La. zanandreai*,[6] and *P. marinus* (X92515). **(B)** Molecular phylogenetic tree of integrase domains of the LCA-ApaI sequence and Ty3/Gypsy families of LTR retrotransposons. The nucleotide sequences used for this analysis are shown in Supplementary Table S2. Species name are represented in parentheses. The neighbor-joining tree is rooted with two vertebrate retroviruses, RSV and HIV1. Bootstrap values (out of 1000 replicates) are shown for branches with >50% support.

Figure S5. Genomic organization of LCA-EcoRIb and LCA-ApaI sequences. **(A, B)** Southern blot hybridization probed with LCA-EcoRIb **(A)** and LCA-ApaI **(B)** fragments. The *Le. camtschaticum* genomic DNA was digested with six restriction enzymes, AluI, ApaI, EcoRI, HpaII, MspI, and PstI. A mixture of λ DNA-HindIII and ϕ X174 DNA-HaeIII digests was used as a molecular size marker.

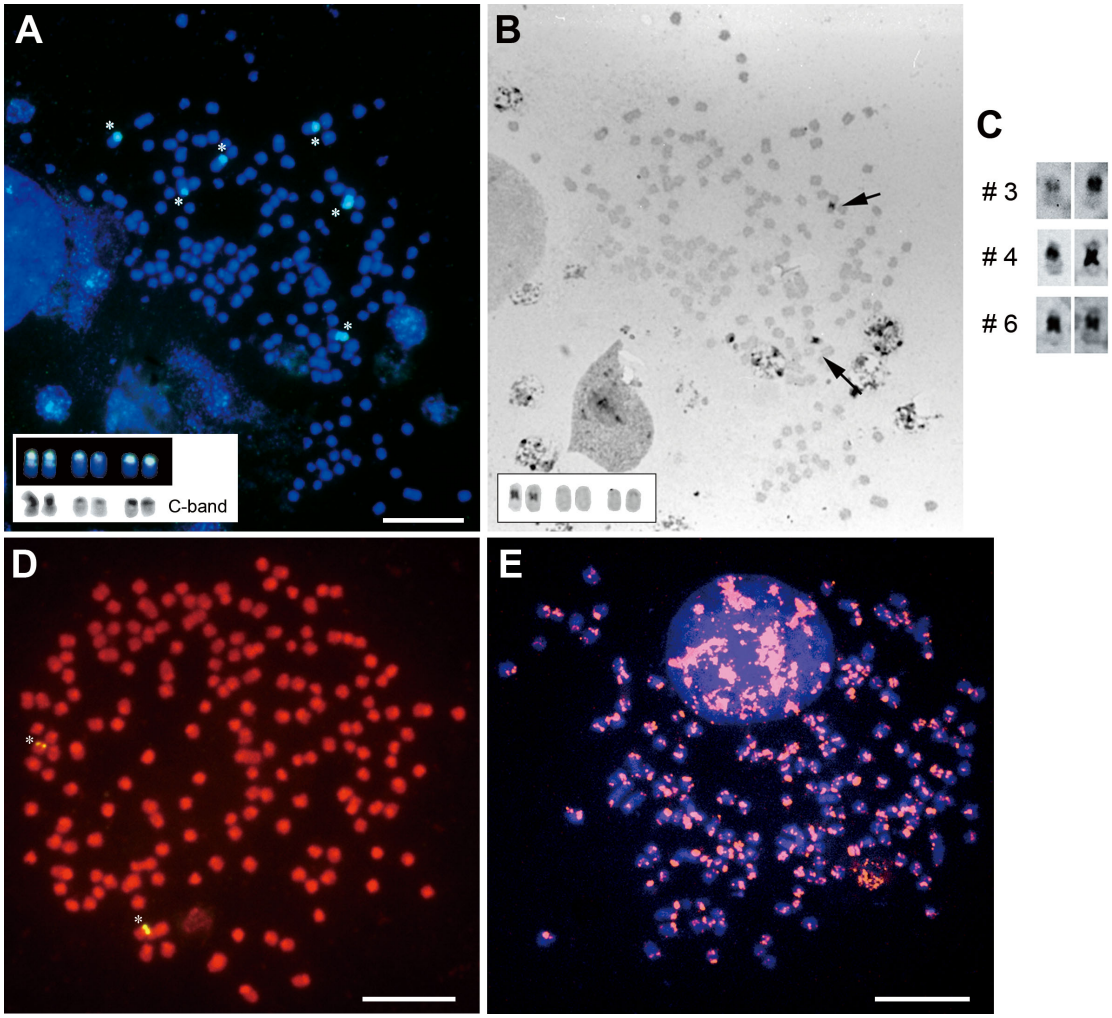

Figure S1.

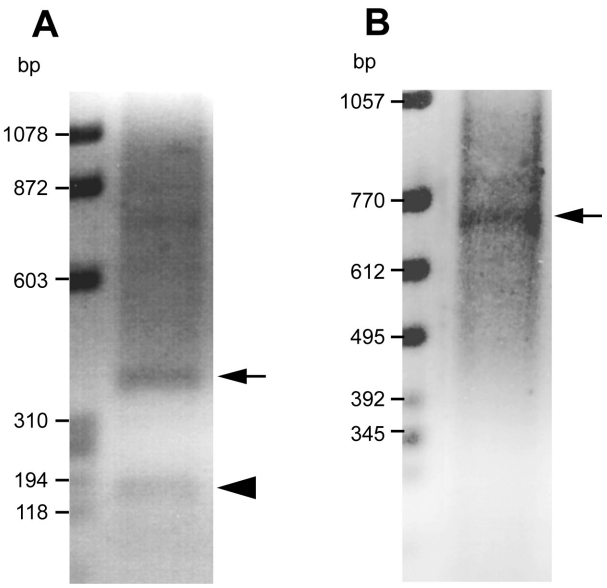

Figure S2.

A

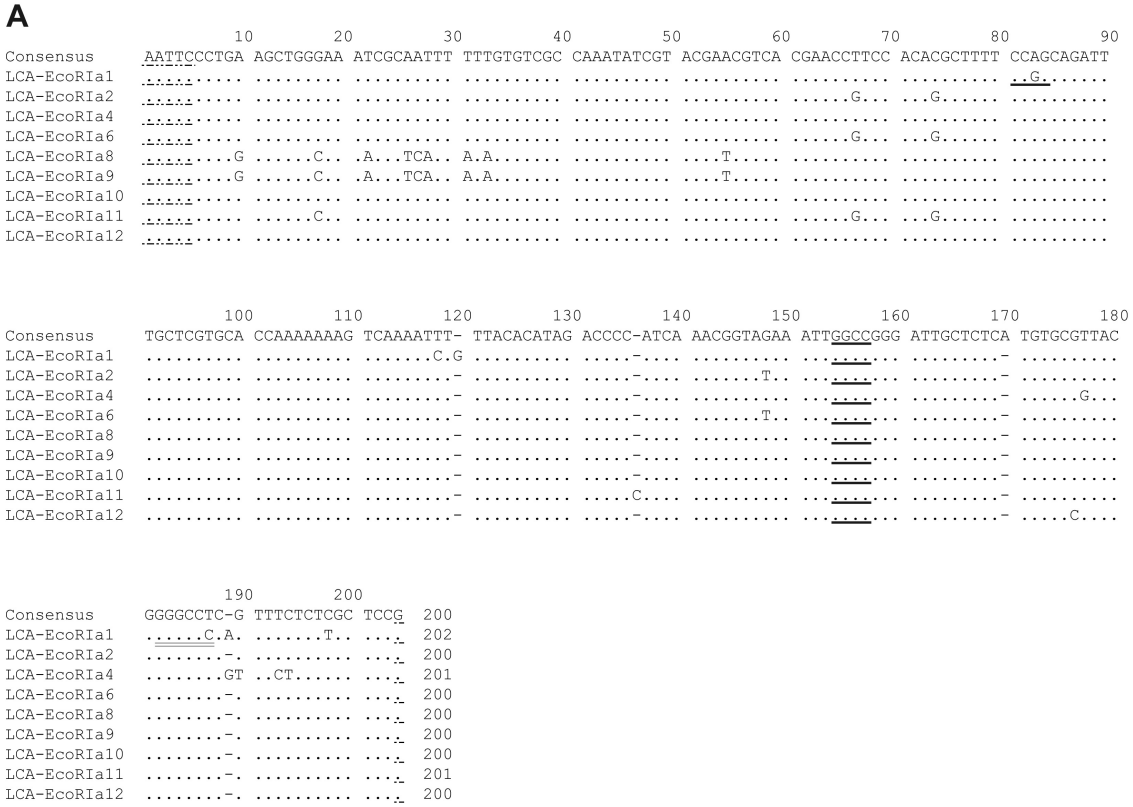
10 20 30 40 50 60 70 80 90
Consensus AATTCCCTGA AGCTGGGAA ATCGCAATTT TTTGTGTCGC CAAATATCGT ACGAACGTCA CGAACCTTCC ACACGCTTTT CCAGCAGATT
LCA-EcoRIa1
LCA-EcoRIa2
LCA-EcoRIa4
LCA-EcoRIa6
LCA-EcoRIa8
LCA-EcoRIa9
LCA-EcoRIa10
LCA-EcoRIa11
LCA-EcoRIa12
100 110 120 130 140 150 160 170 180
Consensus TGCTCGTGCA CCAAAAAAAG TCAAAATTT- TTACACATAG ACCCC-ATCA AACGGTAGAA ATTGGCCGGG ATTGCTCTCA TGTGCGTTAC
LCA-EcoRIa1
LCA-EcoRIa2
LCA-EcoRIa4
LCA-EcoRIa6
LCA-EcoRIa8
LCA-EcoRIa9
LCA-EcoRIa10
LCA-EcoRIa11
LCA-EcoRIa12
190 200
Consensus GGGGCCTC-G TTTCTCTCGC TCCG 200
LCA-EcoRIa1 202
LCA-EcoRIa2 200
LCA-EcoRIa4 201
LCA-EcoRIa6 200
LCA-EcoRIa8 200
LCA-EcoRIa9 200
LCA-EcoRIa10 200
LCA-EcoRIa11 201
LCA-EcoRIa12 200

Figure S3A.

B

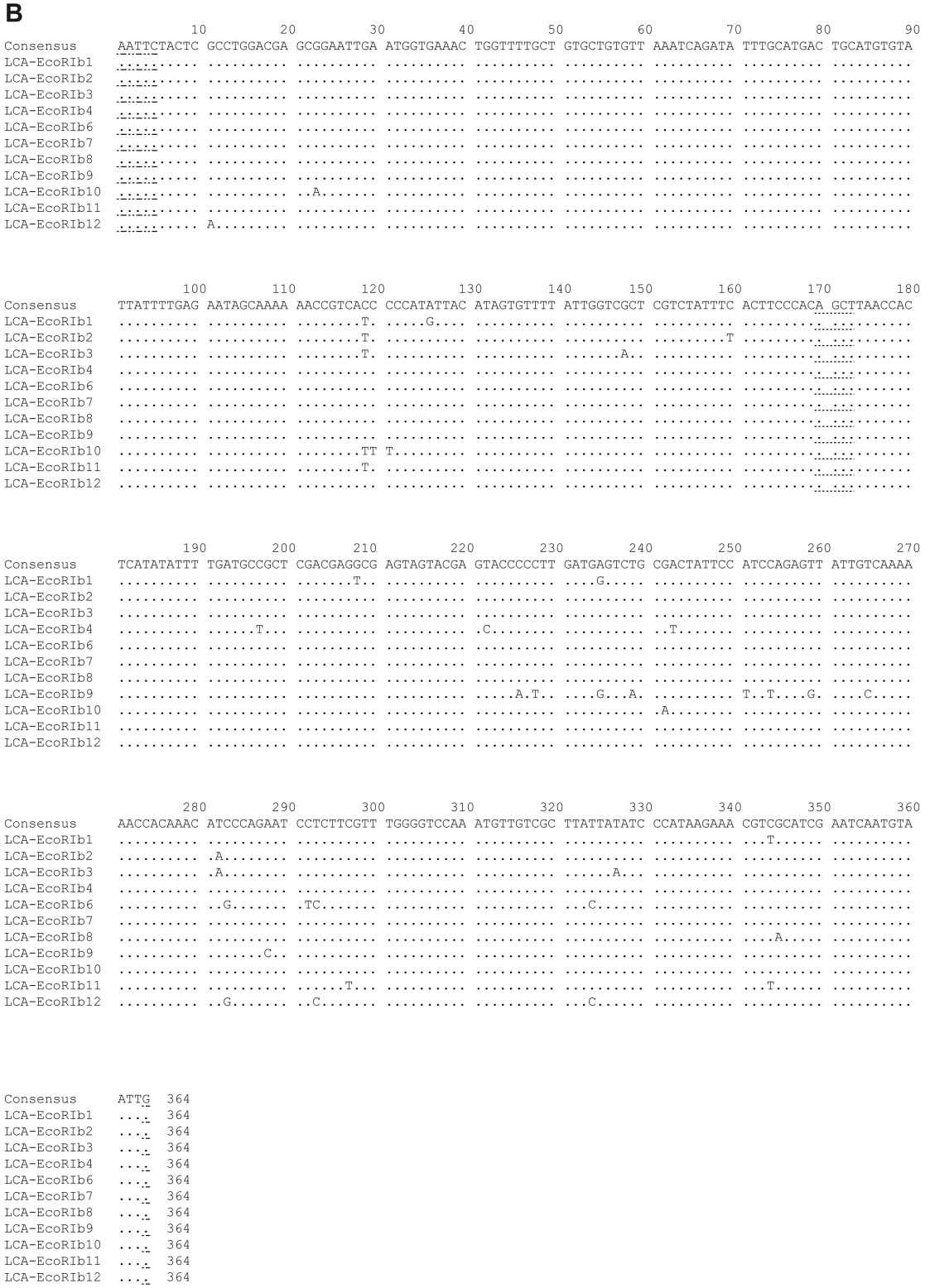
10 20 30 40 50 60 70 80 90
Consensus AATTCTACTC GCCTGGACGA GCGGAATTGA ATGGTGAAAC TGGTTTTGCT GTGCTGTGTT AAATCAGATA TTTGCATGAC TGCATGTGTA
LCA-EcoRIb1 .....
LCA-EcoRIb2 .....
LCA-EcoRIb3 .....
LCA-EcoRIb4 .....
LCA-EcoRIb6 .....
LCA-EcoRIb7 .....
LCA-EcoRIb8 .....
LCA-EcoRIb9 .....
LCA-EcoRIb10 .....
LCA-EcoRIb11 .....
LCA-EcoRIb12 .....
100 110 120 130 140 150 160 170 180
Consensus TTATTTTGAG AATAGCAAAA AACCGTCACC CCCATATTAC ATAGTGTTTT ATTGGTCGCT CGTCTATTTC ACTTCCCACA GCTTAACCAC
LCA-EcoRIb1
LCA-EcoRIb2
LCA-EcoRIb3
LCA-EcoRIb4
LCA-EcoRIb6
LCA-EcoRIb7
LCA-EcoRIb8
LCA-EcoRIb9
LCA-EcoRIb10
LCA-EcoRIb11
LCA-EcoRIb12
190 200 210 220 230 240 250 260 270
Consensus TCATATATTT TGATGCCGCT CGACGAGGCG AGTAGTACGA GTACCCCCTT GATGAGTCTG CGACTATTCC ATCCAGAGTT ATTGTCAAAA
LCA-EcoRIb1
LCA-EcoRIb2
LCA-EcoRIb3
LCA-EcoRIb4
LCA-EcoRIb6
LCA-EcoRIb7
LCA-EcoRIb8
LCA-EcoRIb9
LCA-EcoRIb10
LCA-EcoRIb11
LCA-EcoRIb12
280 290 300 310 320 330 340 350 360
Consensus AACCACAAAC ATCCCAGAAT CCTCTTCGTT TGGGGTCCAA ATGTTGTCGC TTATTATATC CCATAAGAAA CGTCGCATCG AATCAATGTA
LCA-EcoRIb1
LCA-EcoRIb2
LCA-EcoRIb3
LCA-EcoRIb4
LCA-EcoRIb6
LCA-EcoRIb7
LCA-EcoRIb8
LCA-EcoRIb9
LCA-EcoRIb10
LCA-EcoRIb11
LCA-EcoRIb12
Consensus ATTG 364
LCA-EcoRIb1 364
LCA-EcoRIb2 364
LCA-EcoRIb3 364
LCA-EcoRIb4 364
LCA-EcoRIb6 364
LCA-EcoRIb7 364
LCA-EcoRIb8 364
LCA-EcoRIb9 364
LCA-EcoRIb10 364
LCA-EcoRIb11 364
LCA-EcoRIb12 364

Figure S3B.

C

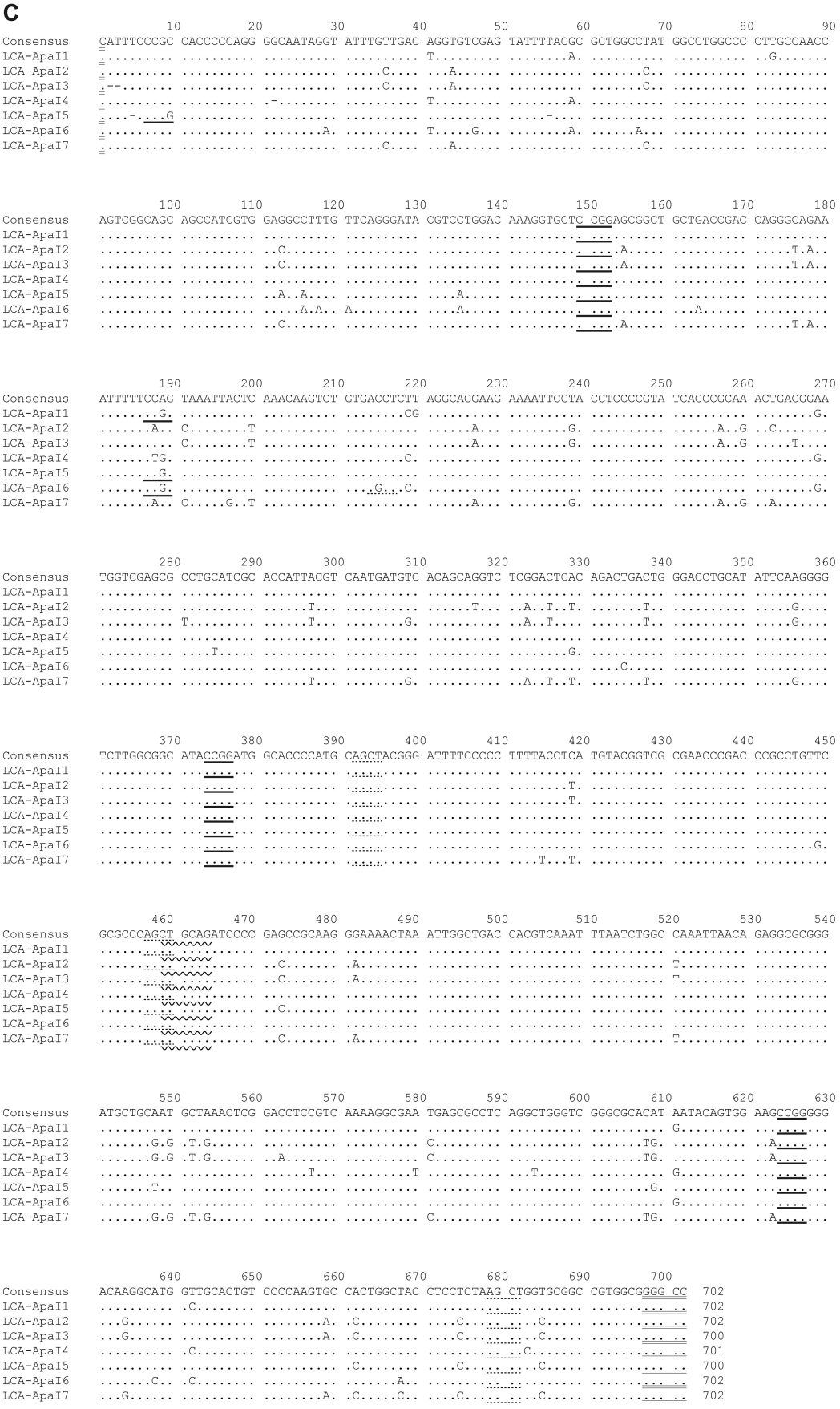
10 20 30 40 50 60 70 80 90
Consensus CATTTCCCGC CACCCCCAGG GGCAATAGGT ATTTGTTGAC AGGTGTCGAG TATTTTACGC GCTGGCCTAT GGCCTGGCCC CTTGCCAACC
LCA-ApaI1 . T. A. ..G.....
LCA-ApaI2 . C.... ...A..... .....C.. .....C.. .....
LCA-ApaI3 .-- .C.... ...A..... .....C.. .....
LCA-ApaI4 .- T. A. .....
LCA-ApaI5 .---.---G .....-.... .....
LCA-ApaI6 . A. T.....G... A. .....A... .....
LCA-ApaI7 . .C.... ...A..... .....C.. .....
100 110 120 130 140 150 160 170 180
Consensus AGTCGGCAGC AGCCATCGTG GAGGCCTTTG TTCAGGGATA CGTCCTGGAC AAAGGTGCTC CGGAGCGGCT GCTGACCGAC CAGGGCAGAA
LCA-ApaI1 .....
LCA-ApaI2 .C..... .....A.... .....T.A..
LCA-ApaI3 ..C..... .....A.... .....T.A..
LCA-ApaI4 .....
LCA-ApaI5 ..A..A.... ....A..... .....
LCA-ApaI6 .....A.A.. A..... ....A..... ...A.....
LCA-ApaI7 ..C..... .....A.... .....T.A..
190 200 210 220 230 240 250 260 270
Consensus ATTTTTCCAG TAAATTACTC AAACAAGTCT GTGACCTCTT AGGCACGAAG AAAATTCGTA CCTCCCCGTA TCACCCGCAA ACTGACGGAA
LCA-ApaI1 ..G. CG ....G.
LCA-ApaI2 .A.. C.....T ....A... .....G. ....A..G ..C.....
LCA-ApaI3 ..... C.....T ....A... .....G. ....A..G .....T....
LCA-ApaI4 .....TG. .....C. ....G.
LCA-ApaI5 ..G. ....
LCA-ApaI6 ..G. ...G...C. ....G.
LCA-ApaI7 .....A.. C.....G..T ....A... .....G. ....A..G ..A.....
280 290 300 310 320 330 340 350 360
Consensus TGGTCGAGCG CCTGCATCGC ACCATTACGT CAATGATGTC ACAGCAGGTC TCGGACTCAC AGACTGACTG GGACCTGCAT ATTCAAGGGG
LCA-ApaI1 ....
LCA-ApaI2 ....T... ....T... ..A..T..T. .....T.. ....G....
LCA-ApaI3 ..... T..... .....T... .....G. ....A..T.... .....T.. ....G....
LCA-ApaI4 ....
LCA-ApaI5 ....T..... ....G. ....
LCA-ApaI6 ....C..... ....
LCA-ApaI7 ....T... .....G. ....A..T..T. .....T.. ....G....
370 380 390 400 410 420 430 440 450
Consensus TCTTGGCGGC ATACCGGATG GCACCCCATG CAGCTACGGG ATTTTCCCCC TTTTACCTCA TGTACGGTCG CGAACCCGAC CCGCCTGTTC
LCA-ApaI1 ....
LCA-ApaI2 ....T. ....
LCA-ApaI3 ....T. ....
LCA-ApaI4 ....
LCA-ApaI5 ....
LCA-ApaI6 ....G.
LCA-ApaI7 ....T...T. ....
460 470 480 490 500 510 520 530 540
Consensus GCGCCCAGCT GCAGATCCCC GAGCCGCAAG GGAAAACTAA ATTGGCTGAC CACGTCAAAT TTAATCTGGC CAAATTAACA GAGGCGCGGG
LCA-ApaI1 ....
LCA-ApaI2 ....C..... .A..... .... T..... .....
LCA-ApaI3 ....C..... .A..... .... T..... .....
LCA-ApaI4 ....
LCA-ApaI5 ....C..... ....
LCA-ApaI6 ....
LCA-ApaI7 ....C..... .A..... .... T..... .....
550 560 570 580 590 600 610 620 630
Consensus ATGCTGCAAT GCTAAACTCG GACCTCCGTC AAAAGGCGAA TGAGCGCCTC AGGCTGGGTC GGGCGCACAT AATACAGTGG AAGCCGGGGG
LCA-ApaI1 .... G..... .....
LCA-ApaI2 .....G.G .T.G..... .... C..... ....TG. ....A.....
LCA-ApaI3 .....G.G .T.G..... ..A..... ..... C..... ....TG. ....A.....
LCA-ApaI4 ....T... ....T ....T..... ..... G..... .....
LCA-ApaI5 .....T.. ....G. ....
LCA-ApaI6 .... G..... .....
LCA-ApaI7 .....G.G .T.G..... .... C..... ....TG. ....A.....
640 650 660 670 680 690 700
Consensus ACAAGGCATG GTTGCACTGT CCCCAAGTGC CACTGGCTAC CTCCTCTAAG CTGGTGCGGC CGTGGCGGGG CC 702
LCA-ApaI1 ..... .C..... 702
LCA-ApaI2 ...G..... ....A. .C..... ....C..... .....C..... 702
LCA-ApaI3 ...G..... ....A. .C..... ....C..... .....C..... 700
LCA-ApaI4 ..... .C..... ....C..... 701
LCA-ApaI5 .... .C..... ....C..... .....C..... 700
LCA-ApaI6 .....C.. .C..... ....A.. 702
LCA-ApaI7 ...G..... ....A. .C.....C.. ....C..... .....C..... 702

Figure S3C.

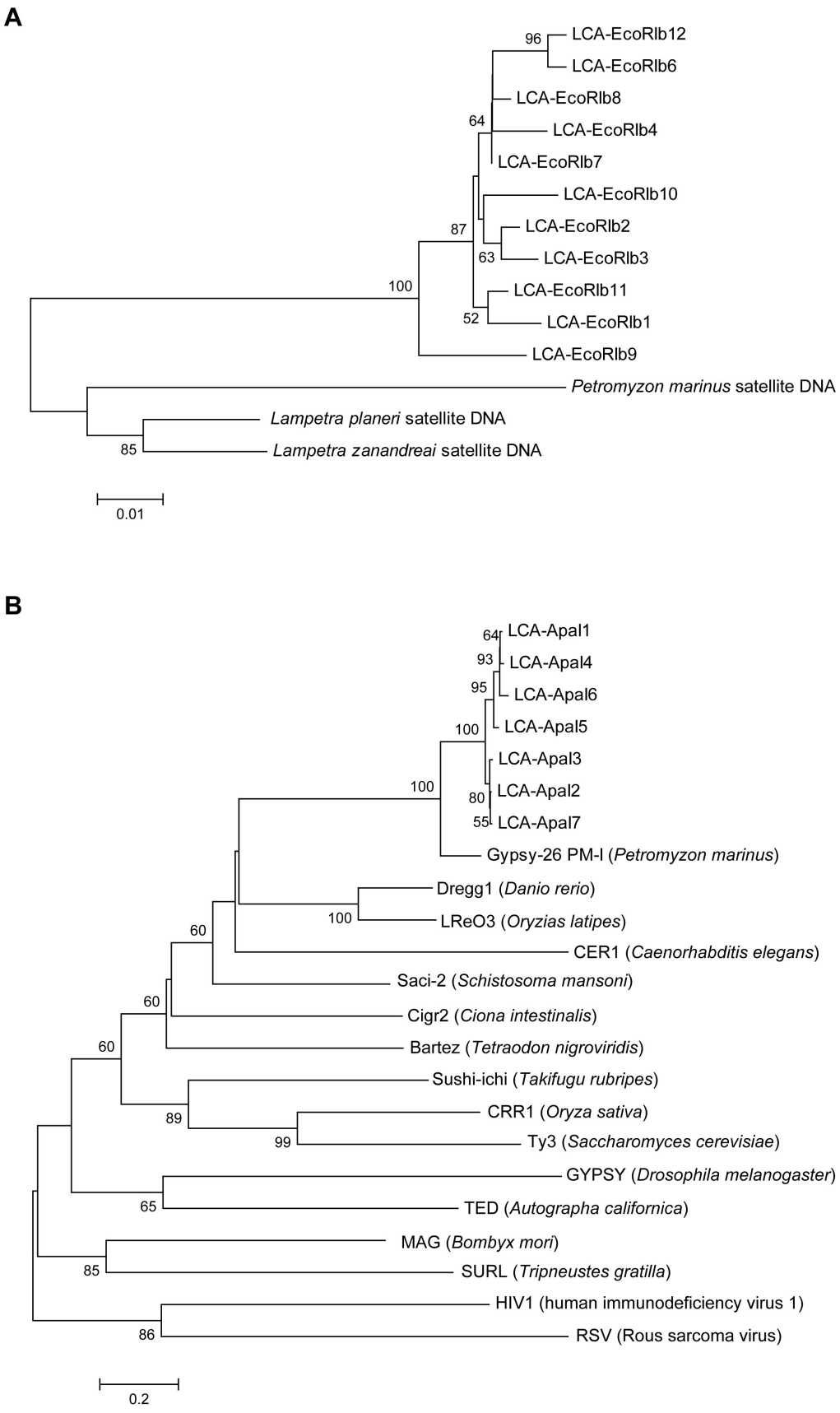
A
96
LCA-EcoRlb12
LCA-EcoRlb6
LCA-EcoRlb8
64
LCA-EcoRlb4
LCA-EcoRlb7
LCA-EcoRlb10
87
LCA-EcoRlb2
63
LCA-EcoRlb3
100
LCA-EcoRlb11
52
LCA-EcoRlb1
LCA-EcoRlb9
Petromyzon marinus satellite DNA
Lampetra planeri satellite DNA
85
Lampetra zanandreai satellite DNA
0.01
B
64
LCA-ApaI1
93
LCA-ApaI4
95
LCA-ApaI6
100
LCA-ApaI5
LCA-ApaI3
100
80
LCA-ApaI2
55
LCA-ApaI7
Gypsy-26 PM-I (Petromyzon marinus)
Dregg1 (Danio rerio)
100
LReO3 (Oryzias latipes)
60
CER1 (Caenorhabditis elegans)
Saci-2 (Schistosoma mansoni)
60
Cigr2 (Ciona intestinalis)
60
Bartez (Tetraodon nigroviridis)
Sushi-ichi (Takifugu rubripes)
89
CRR1 (Oryza sativa)
99
Ty3 (Saccharomyces cerevisiae)
GYPSY (Drosophila melanogaster)
65
TED (Autographa californica)
MAG (Bombyx mori)
85
SURL (Tripneustes gratilla)
HIV1 (human immunodeficiency virus 1)
86
RSV (Rous sarcoma virus)
0.2

Figure S4.

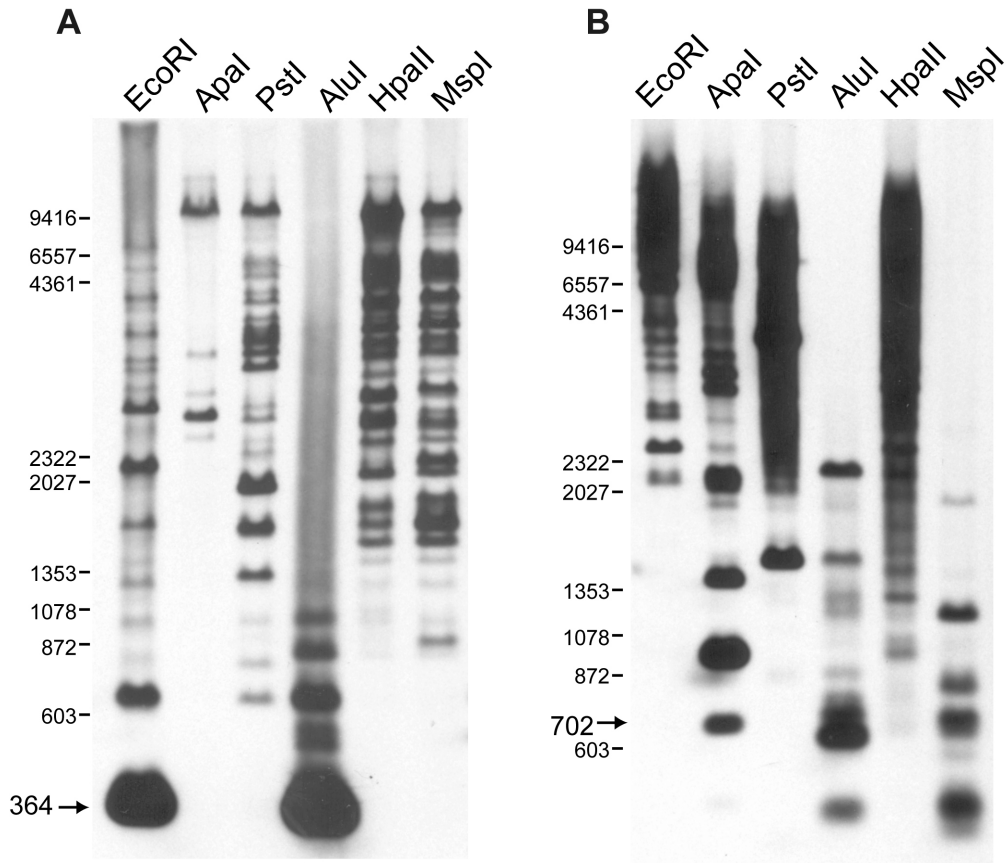
A
EcoRI
ApaI
PstI
AluI
HpaII
MspI
9416
6557
4361
2322
2027
1353
1078
872
603
364
B
EcoRI
ApaI
PstI
AluI
HpaII
MspI
9416
6557
4361
2322
2027
1353
1078
872
702
603

Figure S5.