Supporting Information Appendix for

Expansions, diversification and inter-individual copy number variations of AID/APOBEC family cytidine deaminase genes in lampreys

Stephen J. Holland, Lesley M. Berghuis, Justin J. King, Lakshminarayan M. Iyer, Katarzyna Sikora, Heather Fifield, Sarah Peter, Emma M. Quinlan, Fumiaki Sugahara, Prashant Shingate, Inês Trancoso, Norimasa Iwanami, Elena Temereva, Christine Strohmeier, Shigeru Kuratani, Byrappa Venkatesh, Guillaume Evanno, L. Aravind, Michael Schorpp, Mani Larijani, and Thomas Boehm

Lampetra planeri/Lampetra fluviatilis

L. planeri

Lp#236

CDA1
CDA1L1_1
CDA1L1_2
CDA1L1_3
CDA1L1_4
CDA1L2_1
CDA1L2_2

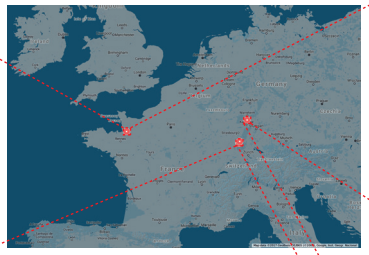
L. fluviatilis

CDA1L1_1
CDA1L1_2
CDA1L1_3
CDA1L1_4

Lf#29 (PCR)

CDA1L1_1
CDA1L1_2
CDA1L1_3
CDA1L1_4

Lf#33 (PCR)



Lampetra planeri

Lp#173

Lp#196 (PCR)

CDA1
CDA1L1_1
CDA1L1_2
CDA1L1_3
CDA1L1_4
CDA1L2_1
CDA1L2_2

Lp#242

CDA1
CDA1L1_1
CDA1L1_2
CDA1L1_3
CDA1L1_4
CDA1L2_1
CDA1L2_2

Lp#175

CDA1
CDA1L1_1
CDA1L1_2
CDA1L1_3
CDA1L1_4
CDA1L2_1
CDA1L2_2

Petromyzon marinus

Pm#144

CDA1
CDA1L1_1
CDA1L1_2
CDA1L1_3
CDA1L1_4
CDA1L2_1
CDA1L2_2

Pm#1

CDA1
CDA1L1_1
CDA1L1_2
CDA1L1_3
CDA1L1_4
CDA1L2_1
CDA1L2_2

Lethenteron japonicum

Lj#1

CDA1
CDA1L1_1
CDA1L1_2
CDA1L1_3
CDA1L1_4
CDA1L2_1
CDA1L2_2

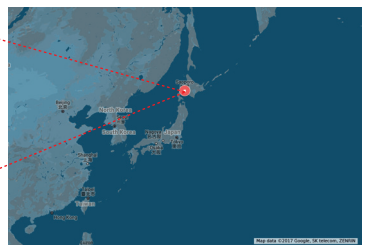
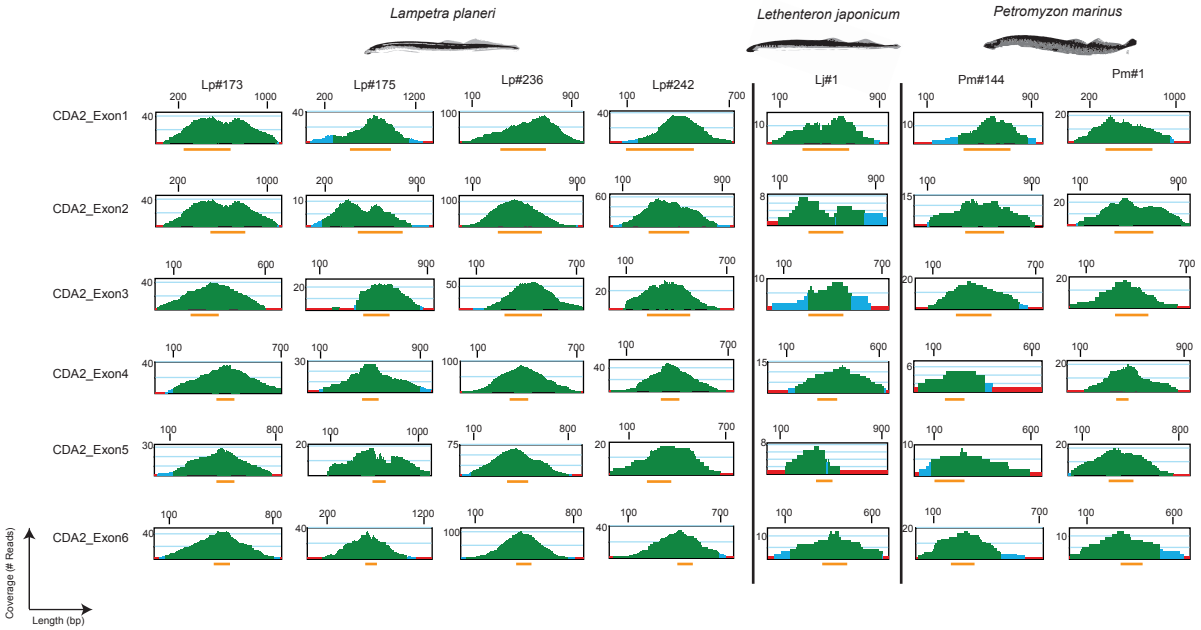


Fig. S1. *CDA1*-like gene copy number in lamprey varies based on geographic location. Geographic locations of capture site for *Lampetra planeri* and *Lampetra fluviatilis* specimens analyzed here. The presence of individual *CDA1L1* genes is indicated by green color, the presence of *CDA1L2* genes by blue color; absence (or presence of pseudogenized versions) of a gene is indicated by grey color. For comparison, the *CDA* gene complements of *Petromyzon marinus* and *Lethenteron japonicum* specimens are also shown. Further details are given in Table S1.

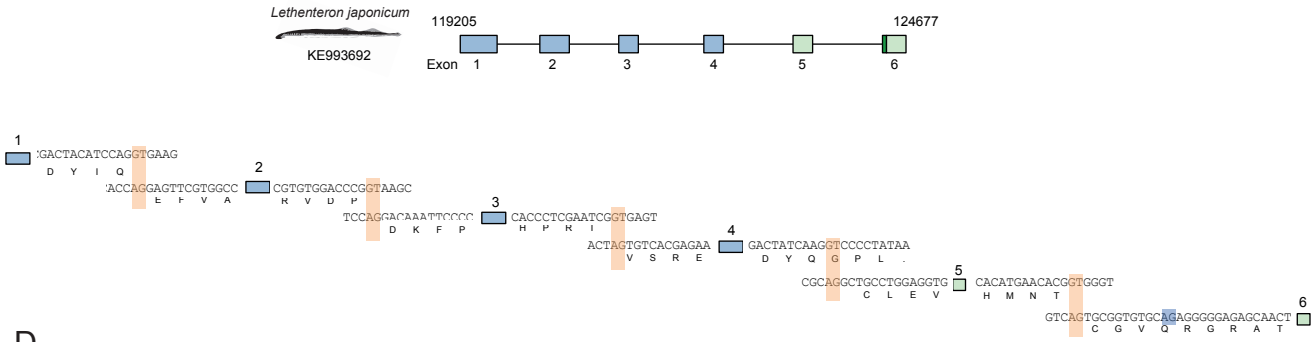
A

```
Lp_CDA2_Exon1 MELREVVDCALGSCVRHEPLGRAAFLRCFAAPSRKPRGTVILFDVDGAGRGLSGGHVVNYNKQGTSIHAEVLLLSAVRAALPQR--CEGDAEEAPRGCTVHCYSTYSPCRDCVDYIQ
Lj_CDA2_Exon1 MELREVVDCALGSCVRHEPLGRAAFLRCFAAPSRKPRGTVILFDVDGAGRGLSGGHVVNYNKQGTSIHAEVLLLSAVRAALPRR--CEGDAEEAPRDCTVHCYSTYSPCRDCVDYIQ
Pm_CDA2_Exon1 MELREVVDCALASCVRHEPLSRVAFLRCFAAPSQKPRGTVILFYVEGAGRGVTGGHAVNYNKQGTSIHAEVLLLSAVRAALLRRRRCE-DGEEATRGCTLHCYSTYSPCRDCVEYIQ

Lp_CDA2_Exon2 EFVASTGVRVAMRCCRLYELDVTRRRPEAEGVLRSLSSLGRDFRLMGPRDAIALLLGGRLA---DGESGASG-------GDAEPLVEMAGFGDEQLHAQVQRNRQIVEAYARYAGAVSLVLGELRVDP
Lj_CDA2_Exon2 EFVASTGVRVAMRCCRLYELDVTRRRPEAEGVLRSLSSLGRDFRLMGPRDAIALLLGGRLA---DGESGASG-------GNAEPLVEMAGFGDEQLHAQVQRNRQIVEAYARYAGAVSLVLGELRVDP
Pm_CDA2_Exon2 EFGASTGVRVVIHCCRLYELDVNRRRSEAEGVLRSLSRLGRDFRLMGPRDAIALLLGGRLANTADGESGASGNAWVTETNVVEPLVDMTGFGDEDLHAQVQRNKQIREAYANYASAVSLMLGELHVDP

Lp_CDA2_Exon3 DKFPFLADFLAQTSVEPSGTPRGARGRPRGASSRGPGIGRQRPADFERALGAYGLFLHPRI
Lj_CDA2_Exon3 DKFPFLADFLAQTSVEPSGTPRGARGRPRGASSRGPGIGRQRPADFERALGAYGLFLHPRI
Pm_CDA2_Exon3 DKFPFLAEFLAQTSVEPSGTPRETRGRPRGASSRGPEIGRQRPADFERALGAYGLFLHPRI

Lp_CDA2_Exon4 VSREADREEIKRDLVAVMRKHDYQGPL.
Lj_CDA2_Exon4 VSREADREEIKRDLIAVMRKHDYQGPL.
Pm_CDA2_Exon4 VSREADREEIKRDLIVVMRKHNYQGP.
```

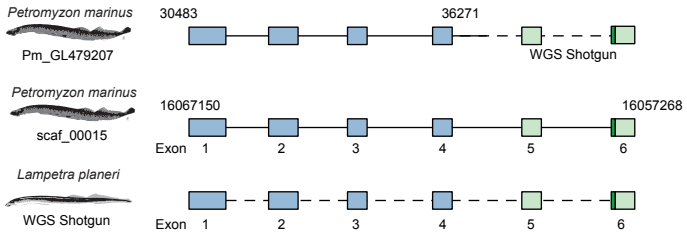
B



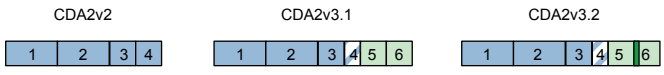
C



D



E

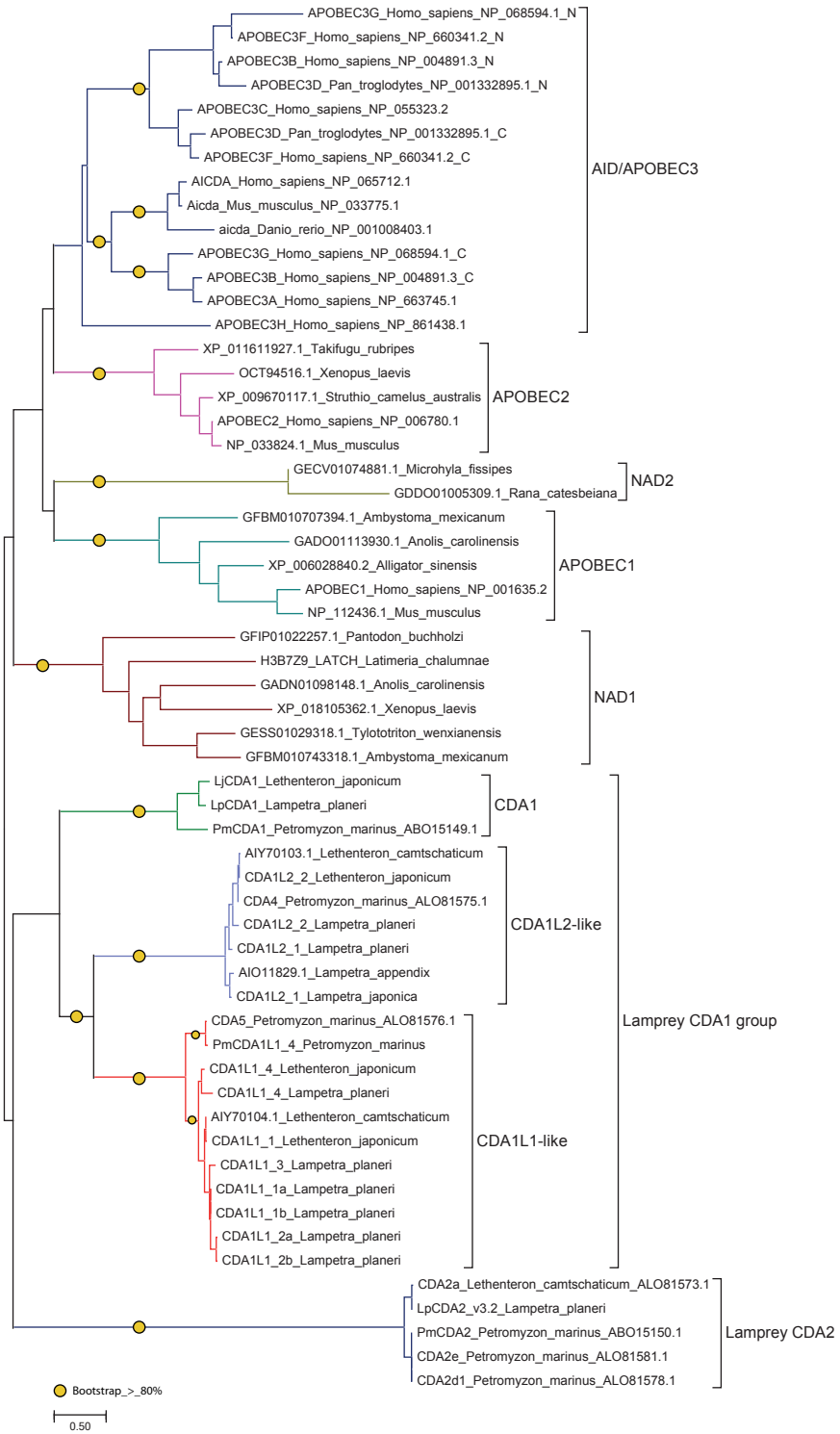



```
PmCDA2      MELREVVDCALASCVRHEPLSRVAFLRCFAAPSQKPRGTVILFYVEGAGRGVTGGHAVNYNKQGTSIHAEVLLLSAVRAALLRRRRCE-DGEEATRGCTLHCYSTYSPCRDCVEYIQEFGASTGVRVVIHCCRLYELDVN
LpCDA2_v2   MELREVVDCALGSCVRHEPLGRAAFLRCFAAPSRKPRGTVILFDVDGAGRGLSGGHVVNYNKQGTSIHAEVLLLSAVRAALPRR--CEGDAEEAPRGCTVHCYSTYSPCRDCVDYIQEFVASTGVRVAMRCCRLYELDVT
LpCDA2_v3.1 MELREVVDCALGSCVRHEPLGRAAFLRCFAAPSRKPRGTVILFDVDGAGRGLSGGHVVNYNKQGTSIHAEVLLLSAVRAALPRR--CEGDAEEAPRGCTVHCYSTYSPCRDCVDYIQEFVASTGVRVAMRCCRLYELDVT
LpCDA2_v3.2 MELREVVDCALGSCVRHEPLGRAAFLRCFAAPSRKPRGTVILFDVDGAGRGLSGGHVVNYNKQGTSIHAEVLLLSAVRAALPRR--CEGDAEEAPRGCTVHCYSTYSPCRDCVDYIQEFVASTGVRVAMRCCRLYELDVT

PmCDA2      RRRSEAEGVLRSLSRLGRDFRLMGPRDAIALLLGGRLANTADGESGASGNAWVTETNVVEPLVDMTGFGDEDLHAQVQRNKQIREAYANYASAVSLMLGELHVDPDKFPFLAEFLAQTSVEPSGTPRETRGRPRGASSRG
LpCDA2_v2   RRRPEAEGVLRSLSSLGHDFRLMGPRDAIALLLGGRLA---DGESGASG-------GDAEPLVEMAGFGDEQLHAQVQRNRQIVEAYARYAGAVSLVLGELRVDPDKFPFLADFLAQTSVEPSGTPRGARGRPRGASSRG
LpCDA2_v3.1 RRRPEAEGVLRSLSSLGRDFRLMGPRDAIALLLGGRLA---DGESGASG-------GDAEPLVEMAGFGDEQLHAQVQRNRQIVEAYARYAGAVSLVLGELRVDPDKFPFLADFLAQTSVEPSGTPRGARGRPRGASSRG
LpCDA2_v3.2 RRRPEAEGVLRSLSSLGRDFRLMGPRDAIALLLGGRLA---DGESGASG-------GDAEPLVEMAGFGDEQLHAQVQRNRQIVEAYARYAGAVSLVLGELRVDPDKFPFLADFLAQTSVEPSGTPRGARGRPRGASSRG

PmCDA2      PEIGRQRPADFERALGAYGLFLHPRIVSREADREEIKRDLIVVMRKHNYQGP.
LpCDA2_v2   PGIGRQRPADFERALGAYGLFLHPRIVSREADREEIKRDLVAVMRKHDYQGPL.
LpCDA2_v3.1 PGIGRQRPADFERALGAYGLFLHPRIVSREADREEIKRDLVAVMRKHDYQGCLEVSPCVCSGACVRACPCRQAGRMCIPLCGAHRATCTRGGAVGGAKSHMNT----RGRATLQLTAKTSQDGAGEKTQLHCSQSHEDDI
LpCDA2_v3.2 PGIGRQRPADFERALGAYGLFLHPRIVSREADREEIKRDLVAVMRKHDYQGCLEVSPCVCSGACVRACPCRQAGRMCIPLCGAHRATCTRGGAVGGAKSHMNTCGVQRGRATLQLTAKTSQDGAGEKTQLHCSQSHEDDI

PmCDA2
LpCDA2_v2
LpCDA2_v3.1 P.
LpCDA2_v3.2 P.
```

Fig. S2. Identification of novel splice variants of *CDA2* genes in *Lampetra planeri* and *Lethenteron japonicum*. (*A*) Comparison of translated *CDA2* exon sequences identified in the genomes of *Lampetra planeri* and *Lethenteron japonicum* with those previously characterized in *Petromyzon marinus*. Conserved APOBEC catalytic HxE and PCxxC motifs are highlighted. (*B*) Read coverage plots retrieved from whole genome sequences using the constituent *Petromyzon marinus CDA2* exon sequences as queries. Green color indicates a coverage by >5 reads; blue, 2-5 reads, and red, single read. Orange bars correspond to region of the contigs containing the open reading frames of the exons. Sequences corresponding to the 4 exons of *PmCDA2* were readily identifiable in the genomic sequence collections of all four individuals. Notably, the exon sequences of *CDA2* genes of Pm#1 and Pm#144 were identical to the previously described *Petromyzon marinus* sequence and clearly distinguishable from the corresponding *Lampetra planeri* and *Lethenteron japonicum CDA2* gene sequences (see panel *A*). Similarly, the exon sequences of *CDA2* genes from Lp#236 and Lp#242 were identical to those identified in Lp#173 and Lp#175. (*C*) Schematic of *CDA2* exon structure as established by the *Lethenteron japonicum* genome assembly (version LetJap 1.0) (top panel); the exon/intron junctions for each exon are indicated (lower panel). Splice donor/acceptor sites are highlighted in orange; the alternative splice acceptor site inside exon 6 is highlighted in blue. (*D*) Schematic of *CDA2* exon structure in *Petromyzon marinus*; in the genome assembly (version 7.0)(47) the first 4 exons of *CDA2* are present on a single scaffold (Pm_GL479207, solid line), whereas sequences corresponding to exons 5 and 6 are not contained within this scaffold, but could be identified in our sequence collection obtained with a different *Petromyzon marinus* individual, Pm#1 (dashed line). A recent genome assembly from germline DNA of *P. marinus* (62) confirmed the presence of all 6 exons on a single scaffold (scaf_00015). In the shotgun genome libraries of *Lampetra planeri* individuals, all exons could be identified but were not assembled into a contig (dashed line); the order of exons was established by cDNA sequences. (*E*) Alternatively spliced variants of *CDA2* genes are shown as schematic (top panel) and as amino acid alignments (lower panel). The striped part of exon 4 is lacking in *CDA2v3.1* and *CDA2v 3.2* variants as a result of intra-exonic splicing between exons 4 and 5. The solid dark green box in *CDA2v3.2* refers to the presence of an additional segment encoding 4 amino

acids as a result of the use of a different splice acceptor signal (see panel *C*). Pm, *Petromyzon marinus*, Lp, *Lampetra planeri*.

A


APOBEC3G_Homo_sapiens_NP_068594.1_N
APOBEC3F_Homo_sapiens_NP_660341.2_N
APOBEC3B_Homo_sapiens_NP_004891.3_N
APOBEC3D_Pan_troglodytes_NP_001332895.1_N
APOBEC3C_Homo_sapiens_NP_055323.2
APOBEC3D_Pan_troglodytes_NP_001332895.1_C
APOBEC3F_Homo_sapiens_NP_660341.2_C
AICDA_Homo_sapiens_NP_065712.1
Aicda_Mus_musculus_NP_033775.1
aicda_Danio_rerio_NP_001008403.1
APOBEC3G_Homo_sapiens_NP_068594.1_C
APOBEC3B_Homo_sapiens_NP_004891.3_C
APOBEC3A_Homo_sapiens_NP_663745.1
APOBEC3H_Homo_sapiens_NP_861438.1
AID/APOBEC3
XP_011611927.1_Takifugu_rubripes
OCT94516.1_Xenopus_laevis
XP_009670117.1_Struthio_camelus_australis
APOBEC2_Homo_sapiens_NP_006780.1
NP_033824.1_Mus_musculus
APOBEC2
GECV01074881.1_Microhyla_fissipes
GDDO01005309.1_Rana_catesbeiana
NAD2
GFBM010707394.1_Ambystoma_mexicanum
GADO01113930.1_Anolis_carolinensis
XP_006028840.2_Alligator_sinensis
APOBEC1_Homo_sapiens_NP_001635.2
NP_112436.1_Mus_musculus
APOBEC1
GFIP01022257.1_Pantodon_buchholzi
H3B7Z9_LATCH_Latimeria_chalumnae
GADN01098148.1_Anolis_carolinensis
XP_018105362.1_Xenopus_laevis
GESS01029318.1_Tylototriton_wenxianensis
GFBM010743318.1_Ambystoma_mexicanum
NAD1
LjCDA1_Lethenteron_japonicum
LpCDA1_Lampetra_planeri
PmCDA1_Petromyzon_marinus_ABO15149.1
CDA1
AIY70103.1_Lethenteron_camtschaticum
CDA1L2_2_Lethenteron_japonicum
CDA4_Petromyzon_marinus_ALO81575.1
CDA1L2_2_Lampetra_planeri
CDA1L2_1_Lampetra_planeri
AIO11829.1_Lampetra_appendix
CDA1L2_1_Lampetra_japonica
CDA1L2-like
CDA5_Petromyzon_marinus_ALO81576.1
PmCDA1L1_4_Petromyzon_marinus
CDA1L1_4_Lethenteron_japonicum
CDA1L1_4_Lampetra_planeri
AIY70104.1_Lethenteron_camtschaticum
CDA1L1_1_Lethenteron_japonicum
CDA1L1_3_Lampetra_planeri
CDA1L1_1a_Lampetra_planeri
CDA1L1_1b_Lampetra_planeri
CDA1L1_2a_Lampetra_planeri
CDA1L1_2b_Lampetra_planeri
CDA1L1-like
Lamprey CDA1 group
CDA2a_Lethenteron_camtschaticum_ALO81573.1
LpCDA2_v3.2_Lampetra_planeri
PmCDA2_Petromyzon_marinus_ABO15150.1
CDA2e_Petromyzon_marinus_ALO81581.1
CDA2d1_Petromyzon_marinus_ALO81578.1
Lamprey CDA2
Bootstrap_>_80%
0.50

B

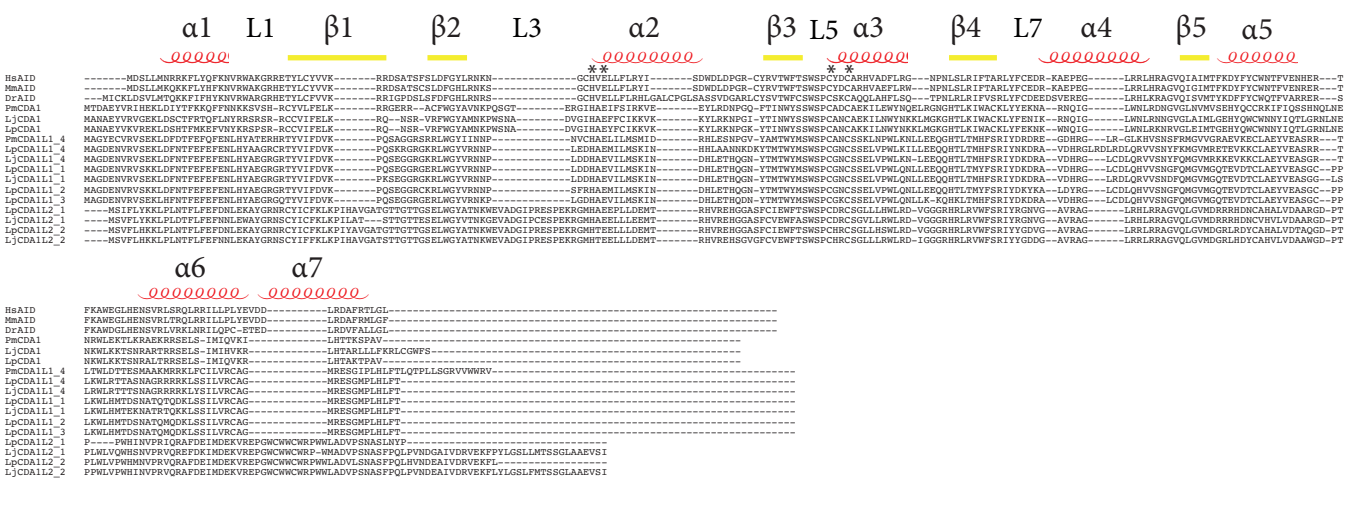
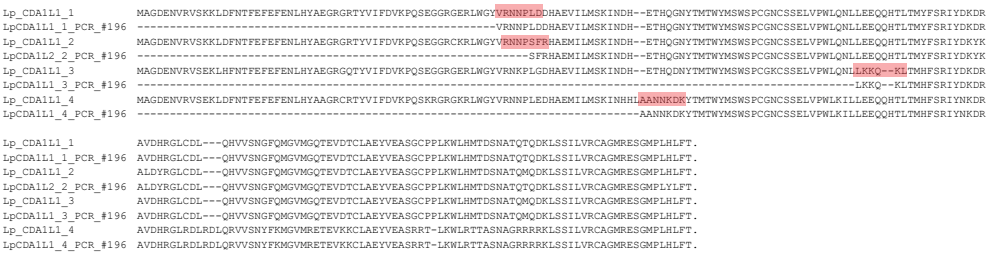
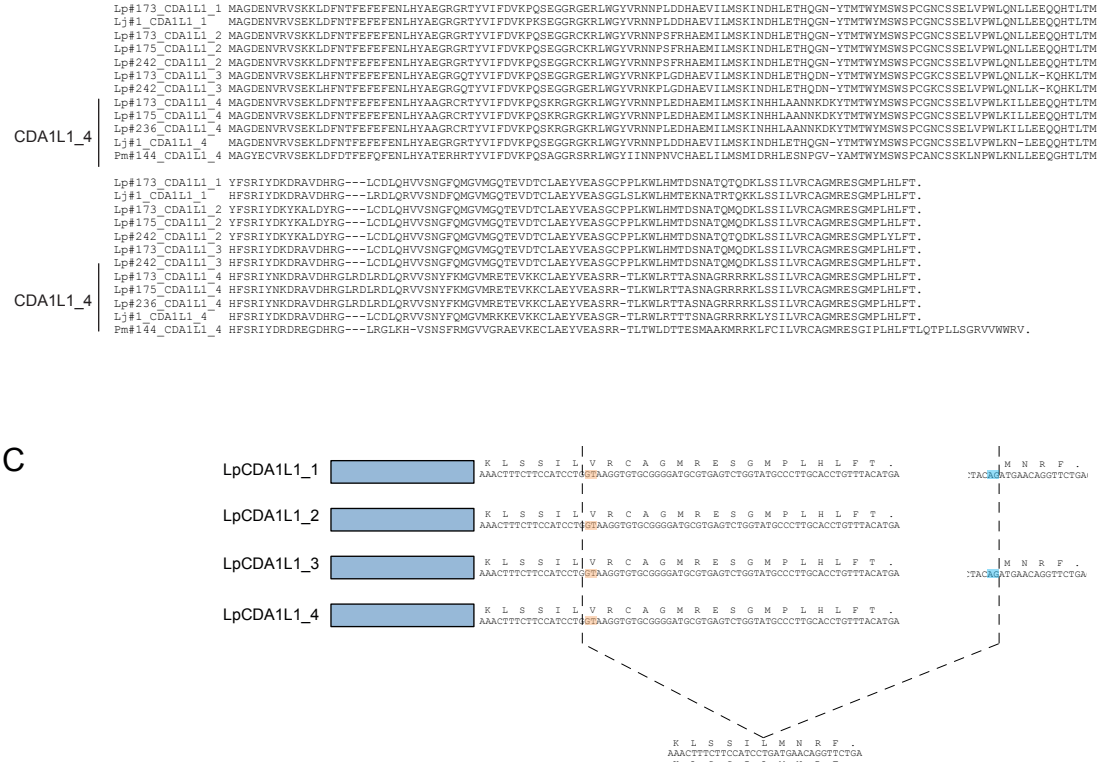

α1 L1 β1 β2 L3 α2 β3 L5 α3 β4 L7 α4 β5 α5
α6 α7

Fig. S3. Sequence comparisons of AID/APOBEC deaminases. (*A*) Maximum-likelihood tree of lamprey CDAs and jawed vertebrate AID/APOBEC deaminases. Nodes with bootstrap support greater than 80% are marked. NAD1 and NAD2 are novel AID/APOBEC-like deaminases, which are discussed in the companion paper. For lamprey *CDA1*-like genes, related Genbank database entries were also used; small letters at the end of gene names correspond to presumptive allelic variants. (*B*) Sequence alignment of predicted protein sequences encoded by *CDA1*, *CDA1L1* and *CDA1L2* genes in *Lampetra planeri* (Lp), *Lethenteron japonicum* (Lj), and *Petromyzon marinus* (Pm). Conserved APOBEC catalytic glutamic acid and zinc-coordinating and cysteine and histidine residues are highlighted with an asterisk (*) (c.f., Fig. 2A).

A


Lp_CDA1L1_1 MAGDENVRVSKKLDFNTFEFEFENLHYAEGRGRTYVIFDVKPQSEGGRGERLWGYVRNNPLDDHAEVILMSKINDH--ETHQGNYTMTWYMSWSPCGNCSSELVPWLQNLLEEQQHTLTMYFSRIYDKDR
LpCDA1L1_1_PCR_#196 ----------VRNNPLDDHAEVILMSKINDH--ETHQGNYTMTWYMSWSPCGNCSSELVPWLQNLLEEQQHTLTMYFSRIYDKDR
Lp_CDA1L1_2 MAGDENVRVSKKLDFNTFEFEFENLHYAEGRGRTYVIFDVKPQSEGGRCKRLWGYVRNNPSFRHAEMILMSKINDH--ETHQGNYTMTWYMSWSPCGNCSSELVPWLQNLLEEQQHTLTMYFSRIYDKYK
LpCDA1L2_2_PCR_#196 ----------SFRHAEMILMSKINDH--ETHQGNYTMTWYMSWSPCGNCSSELVPWLQNLLEEQQHTLTMYFSRIYDKYK
Lp_CDA1L1_3 MAGDENVRVSEKLHFNTFEFEFENLHYAEGRGQTYVIFDVKPQSEGGRGERLWGYVRNKPLGDHAEVILMSKINDH--ETHQDNYTMTWYMSWSPCGKCSSELVPWLQNLLKKQ--KLTMHFSRIYDKDR
LpCDA1L1_3_PCR_#196 ----------LKKQ--KLTMHFSRIYDKDR
Lp_CDA1L1_4 MAGDENVRVSEKLDFNTFEFEFENLHYAAGRCRTYVIFDVKPQSKRGRGKRLWGYVRNNPLEDHAEMILMSKINHHLAANNKDKYTMTWYMSWSPCGNCSSELVPWLKILLEEQQHTLTMHFSRIYNKDR
LpCDA1L1_4_PCR_#196 ----------AANNKDKYTMTWYMSWSPCGNCSSELVPWLKILLEEQQHTLTMHFSRIYNKDR
Lp_CDA1L1_1 AVDHRGLCDL---QHVVSNGFQMGVMGQTEVDTCLAEYVEASGCPPLKWLHMTDSNATQTQDKLSSILVRCAGMRESGMPLHLFT.
LpCDA1L1_1_PCR_#196 AVDHRGLCDL---QHVVSNGFQMGVMGQTEVDTCLAEYVEASGCPPLKWLHMTDSNATQTQDKLSSILVRCAGMRESGMPLHLFT.
Lp_CDA1L1_2 ALDYRGLCDL---QHVVSNGFQMGVMGQTEVDTCLAEYVEASGCPPLKWLHMTDSNATQMQDKLSSILVRCAGMRESGMPLHLFT.
LpCDA1L2_2_PCR_#196 ALDYRGLCDL---QHVVSNGFQMGVMGQTEVDTCLAEYVEASGCPPLKWLHMTDSNATQTQDKLSSILVRCAGMRESGMPLYLFT.
Lp_CDA1L1_3 AVDHRGLCDL---QHVVSNGFQMGVMGQTEVDTCLAEYVEASGCPPLKWLHMTDSNATQMQDKLSSILVRCAGMRESGMPLHLFT.
LpCDA1L1_3_PCR_#196 AVDHRGLCDL---QHVVSNGFQMGVMGQTEVDTCLAEYVEASGCPPLKWLHMTDSNATQMQDKLSSILVRCAGMRESGMPLHLFT.
Lp_CDA1L1_4 AVDHRGLRDLRDLQRVVSNYFKMGVMRETEVKKCLAEYVEASRRT-LKWLRTTASNAGRRRRKLSSILVRCAGMRESGMPLHLFT.
LpCDA1L1_4_PCR_#196 AVDHRGLRDLRDLQRVVSNYFKMGVMRETEVKKCLAEYVEASRRT-LKWLRTTASNAGRRRRKLSSILVRCAGMRESGMPLHLFT.

B

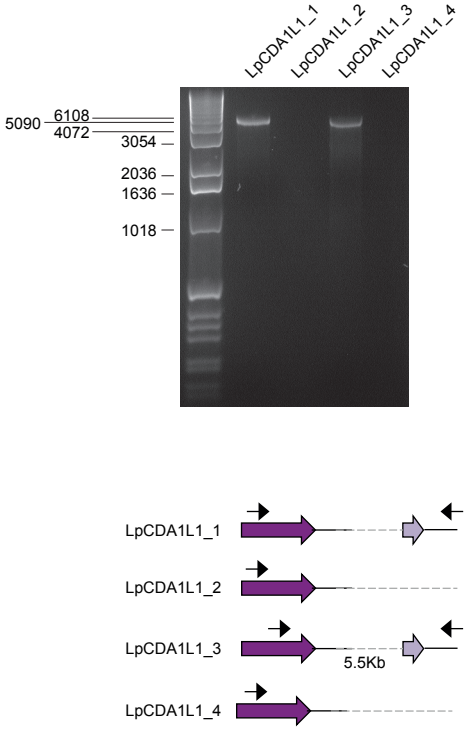

Lp#173_CDA1L1_1 MAGDENVRVSKKLDFNTFEFEFENLHYAEGRGRTYVIFDVKPQSEGGRGERLWGYVRNNPLDDHAEVILMSKINDHLETHQGN-YTMTWYMSWSPCGNCSSELVPWLQNLLEEQQHTLTM
Lj#1_CDA1L1_1 MAGDENVRVSEKLDFNTFEFEFENLHYAEGRGRTYVIFDVKPKSEGGRGKRLWGYVRNNPLDDHAEVILMSKINDHLETHQGN-YTMTWYMSWSPCGNCSSELVPWLQNLLEEQQHTLTM
Lp#173_CDA1L1_2 MAGDENVRVSKKLDFNTFEFEFENLHYAEGRGRTYVIFDVKPQSEGGRCKRLWGYVRNNPSFRHAEMILMSKINDHLETHQGN-YTMTWYMSWSPCGNCSSELVPWLQNLLEEQQHTLTM
Lp#175_CDA1L1_2 MAGDENVRVSKKLDFNTFEFEFENLHYAEGRGRTYVIFDVKPQSEGGRCKRLWGYVRNNPSFRHAEMILMSKINDHLETHQGN-YTMTWYMSWSPCGNCSSELVPWLQNLLEEQQHTLTM
Lp#242_CDA1L1_2 MAGDENVRVSKKLDFNTFEFEFENLHYAEGRGRTYVIFDVKPQSEGGRCKRLWGYVRNNPSFRHAEMILMSKINDHLETHQGN-YTMTWYMSWSPCGNCSSELVPWLQNLLEEQQHTLTM
Lp#173_CDA1L1_3 MAGDENVRVSEKLHFNTFEFEFENLHYAEGRGQTYVIFDVKPQSEGGRGERLWGYVRNKPLGDHAEVILMSKINDHLETHQDN-YTMTWYMSWSPCGKCSSELVPWLQNLLK-KQHKLTM
Lp#242_CDA1L1_3 MAGDENVRVSEKLHFNTFEFEFENLHYAEGRGQTYVIFDVKPQSEGGRGERLWGYVRNKPLGDHAEVILMSKINDHLETHQDN-YTMTWYMSWSPCGKCSSELVPWLQNLLK-KQHKLTM
Lp#173_CDA1L1_4 MAGDENVRVSEKLDFNTFEFEFENLHYAAGRCRTYVIFDVKPQSKRGRGKRLWGYVRNNPLEDHAEMILMSKINHHLAANNKDKYTMTWYMSWSPCGNCSSELVPWLKILLEEQQHTLTM
Lp#175_CDA1L1_4 MAGDENVRVSEKLDFNTFEFEFENLHYAAGRCRTYVIFDVKPQSKRGRGKRLWGYVRNNPLEDHAEMILMSKINHHLAANNKDKYTMTWYMSWSPCGNCSSELVPWLKILLEEQQHTLTM
Lp#236_CDA1L1_4 MAGDENVRVSEKLDFNTFEFEFENLHYAAGRCRTYVIFDVKPQSKRGRGKRLWGYVRNNPLEDHAEMILMSKINHHLAANNKDKYTMTWYMSWSPCGNCSSELVPWLKILLEEQQHTLTM
Lj#1_CDA1L1_4 MAGDENVRVSEKLDFNTFEFEFENLHYAEGRGRTYVIFDVKPQSEGGRGKRLWGYVRNNPLDDHAEVILMSKINDHLETHQGN-YTMTWYMSWSPCGNCSSELVPWLKN-LEEQQHTLTM
Pm#144_CDA1L1_4 MAGYECVRVSEKLDFDTFEFQFENLHYATERHRTYVIFDVKPQSAGGRSRRLWGYIINNPNVCHAELILMSMIDRHLESNPGV-YAMTWYMSWSPCANCSSKLNPWLKNLLEEQGHTLTM
CDA1L1_4

Lp#173_CDA1L1_1 YFSRIYDKDRAVDHRG---LCDLQHVVSNGFQMGVMGQTEVDTCLAEYVEASGCPPLKWLHMTDSNATQTQDKLSSILVRCAGMRESGMPLHLFT.
Lj#1_CDA1L1_1 HFSRIYDKDRAVDHRG---LRDLQRVVSNDFQMGVMGQTEVDTCLAEYVEASGGLSLKWLHMTEKNATRTQKKLSSILVRCAGMRESGMPLHLFT.
Lp#173_CDA1L1_2 YFSRIYDKYKALDYRG---LCDLQHVVSNGFQMGVMGQTEVDTCLAEYVEASGCPPLKWLHMTDSNATQMQDKLSSILVRCAGMRESGMPLHLFT.
Lp#175_CDA1L1_2 YFSRIYDKYKALDYRG---LCDLQHVVSNGFQMGVMGQTEVDTCLAEYVEASGCPPLKWLHMTDSNATQMQDKLSSILVRCAGMRESGMPLHLFT.
Lp#242_CDA1L1_2 YFSRIYDKYKALDYRG---LCDLQHVVSNGFQMGVMGQTEVDTCLAEYVEASGCPPLKWLHMTDSNATQTQDKLSSILVRCAGMRESGMPLYLFT.
Lp#173_CDA1L1_3 HFSRIYDKDRAVDHRG---LCDLQHVVSNGFQMGVMGQTEVDTCLAEYVEASGCPPLKWLHMTDSNATQMQDKLSSILVRCAGMRESGMPLHLFT.
Lp#242_CDA1L1_3 HFSRIYDKDRAVDHRG---LCDLQHVVSNGFQMGVMGQTEVDTCLAEYVEASGCPPLKWLHMTDSNATQMQDKLSSILVRCAGMRESGMPLHLFT.
Lp#173_CDA1L1_4 HFSRIYNKDRAVDHRGLRDLRDLQRVVSNYFKMGVMRETEVKKCLAEYVEASRR-TLKWLRTTASNAGRRRRKLSSILVRCAGMRESGMPLHLFT.
Lp#175_CDA1L1_4 HFSRIYNKDRAVDHRGLRDLRDLQRVVSNYFKMGVMRETEVKKCLAEYVEASRR-TLKWLRTTASNAGRRRRKLSSILVRCAGMRESGMPLHLFT.
Lp#236_CDA1L1_4 HFSRIYNKDRAVDHRGLRDLRDLQRVVSNYFKMGVMRETEVKKCLAEYVEASRR-TLKWLRTTASNAGRRRRKLSSILVRCAGMRESGMPLHLFT.
Lj#1_CDA1L1_4 HFSRIYDKDRAVDHRG---LCDLQRVVSNYFQMGVMRKKEVKKCLAEYVEASGR-TLRWLRTTTSNAGRRRRKLYSILVRCAGMRESGMPLHLFT.
Pm#144_CDA1L1_4 HFSRIYDRDREGDHRG---LRGLKH-VSNSFRMGVVGRAEVKECLAEYVEASRR-TLTWLDTTESMAAKMRRKLFCILVRCAGMRESGIPLHLFTLQTPLLSGRVVWWRV.
CDA1L1_4

C

LpCDA1L1_1
LpCDA1L1_2
LpCDA1L1_3
LpCDA1L1_4
K L S S I L V R C A G M R E S G M P L H L F T .
M N R F .
AAACTTTCTTCCATCCTGGTAAGGTGTGCGGGGATGCGTGAGTCTGGTATGCCCTTGCACCTGTTTACATGA
K L S S I L M N R F .
AAACTTTCTTCCATCCTGATGAACAGGTTCTGA

D


LpCDA1L1_1
LpCDA1L1_2
LpCDA1L1_3
LpCDA1L1_4
5090
6108
4072
3054
2036
1636
1018
LpCDA1L1_1
LpCDA1L1_2
LpCDA1L1_3
5.5Kb
LpCDA1L1_4

E

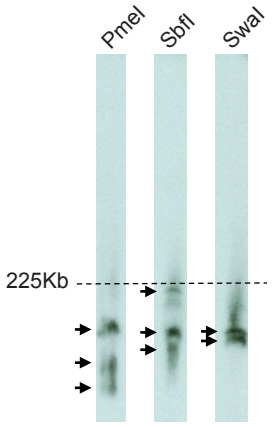
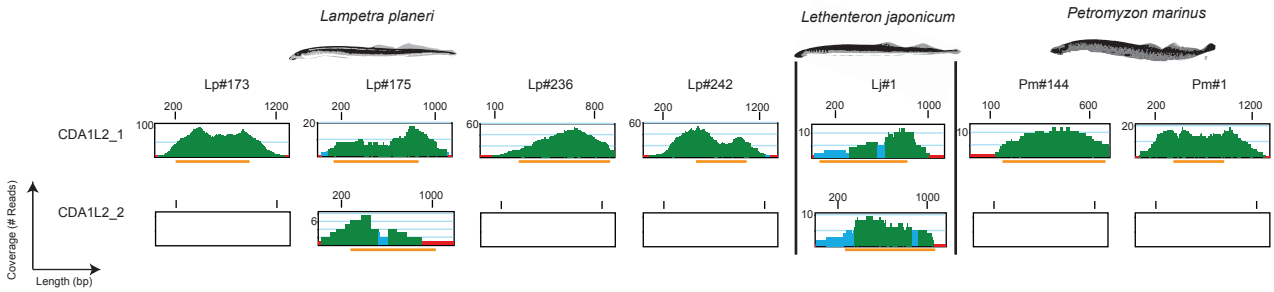

PmeI
SbfI
SwaI
225Kb

Fig. S4. Conservation of *CDA1L1_4* gene and alternative splicing of *CDA1L1_1* and *3*. (*A*) Validation of gene-specific PCR assays. Amino acid sequences deduced from whole genome sequences of *Lampetra planeri* #173 correspond perfectly to sequences derived from PCR products generated using version specific forward primers designed for each gene using genomic DNA of individual Lp#196. Primer binding sites are highlighted in red. (*B*) Comparison of deduced CDA1L1 amino acid sequences derived from whole genome sequences of *Lampetra planeri* (Lp#173, 175, 236, 242), and *Lethenteron japonicum* (Lj#1) individuals. A *CDA1L1_4* gene was also identified in a *Petromyzon marinus* individual (Pm#144) that lacked the canonical *PmCDA1* gene (c.f., Fig. 1). As is the case for many *Lampetra planeri* individuals, the Lp#236 shotgun genome sequences lacked detectable *PmCDA1* sequences, but contained a single *CDA1L1* gene, *CDA1L1_4*; this result was confirmed by Southern filter hybridization and PCR analyses (c.f., Fig. 1). Indeed, additional variations in *CDA1L1* gene content were revealed upon further study; the genomes of *Lampetra planeri* individuals Lp#8a and Lp#8b contained three and two *CDA1L1* genes, respectively see (c.f., Fig. 1). One additional combination of three *CDA1L1* genes was observed in a specimen of another lamprey species, *Lampetra fluviatilis* (Lf#33) (Table S3); *Lampetra planeri* and *Lampetra fluviatilis* are considered to be highly related paired species (46). (*C*) Schematic depicting intra-exonic splicing of *CDA1L1_1* and *CDA1L1_3* genes. Splice donor and acceptor sequences are highlighted in pink and blue, respectively. (*D*) Long-range PCR of genomic DNA using *CDA1L1* gene-specific forward primers (see panel A) and a reverse primer designed to anneal to the 3'-UTR of the exon. The identity of the PCR products was confirmed by sequencing the 5'- and 3'-ends, resulting in the deduced genomic arrangement shown at the bottom of the panel. (*E*) Representative Southern filter hybridizations after separation of genomic restriction digests of genomic DNA of *Lampetra planeri* on pulsed-field gels, hybridized with the *CDA1L1_4* probe (Accession MG495256). The position of the lowest size marker (225Kb) is indicated with a dashed line; fragments hybridizing to the probe are indicated by arrows. The autoradiographic images shown are taken from different parts of the same film.

A



B

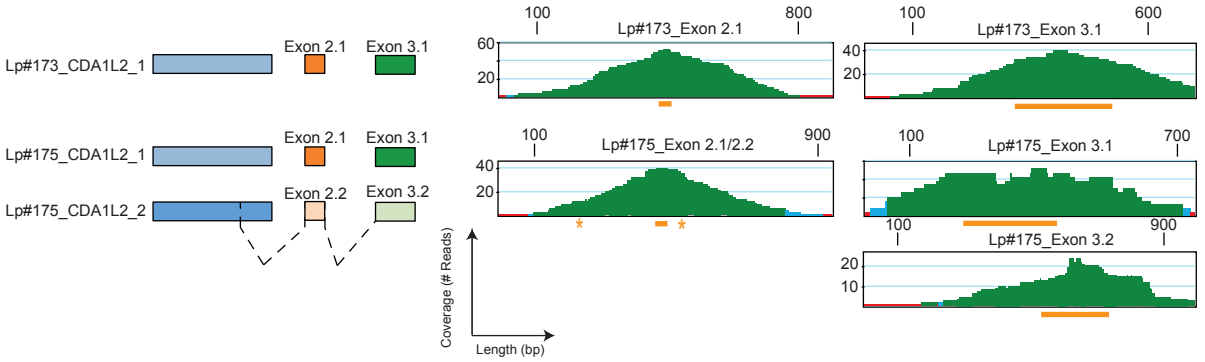
```
#173_CDA1L2_1_Lp ATGTCGATCTTCCTTTACAAGAAGCTGCCCCTCAACACGTTTCTCTTCGAGTTCGACAACCTCGAGAAGGCGTACGGAAGGAACAGATGCTACATTTGCTTCAAGCTCAAACCCATCCAC
#175_CDA1L2_1_Lp ATGTCGATCTTCCTTTACAAGAAGCTGCCCCTCAACACGTTTCTCTTCGAGTTCGACAACCTCGAGAAGGCGTACGGAAGGAACAGATGCTACATTTGCTTCAAGCTCAAACCCATCCAC
#236_CDA1L2_1_Lp ATGTCGATCTTCCTTTACAAGAAGCTGCCCCTCAACACGTTTCTCTTCGAGTTCGACAACCTCGAGAAGGCGTACGGAAGGAACAGATGCTACATTTGCTTCAAGCTCAAACCCATCCAC
#175_CDA1L2_2_Lp ATGTCGGTCTTCCTCCACAAGAAGCTGCCCCTCAACACGTTTCTCTTCGAGTTCGACAACCTCGAGAAGGCGTACGGAAGGAACAGATGCTACATTTGCTTCAAGCTCAAACCCATCTAC
#242_CDA1L2_1_Lp ATGTCGATCTTCCTTTACAAGAAGCTGCCCCTCAACACGTTTCTCTTCGAGTTCGACAACCTCGAGAAGGCGTACGGAAGGAACAGATGCTACATTTGCTTCAAGCTCAAACCCATCCAC
Lj#1_CDA1L2_1_Lj ATGTCGGTCTTCCTTCACAAGAAGCTGCCCCTCGACACGTTTCTCTTCGAGTTCAACAACCTCGAGTGGGCGTACGGAAGGAACAGCTGCTACATTTGCTTCAAGCTCAAACCCATCC--
Lj#1_CDA1L2_2_Lj ATGTCGGTCTTCCTCCACAAGAAGCTGCCCCTCAACACGTTTCTCTTCGAGTTCAACAACCTCGAGAAGGCGTACGGAAGGAACAGCTGCTACATTTTCTTCAAGCTCAAACCCATCCAC
#144_CDA1L2_1_Pm ATGTCGGTCTCCCTCCGCAAGAAGCTGCCCCTCAACACATTTCCCCTCGAGTTGCACAACGTCGAGAGGGCGTACGGAAGGAACAGCTGCTACATCCGCTTCGTGCTCAAATCCACCCAC
Pm#1_CDA1L2_1_Pm ATGTCGGTCTCCCTCCGCAAGAAGCTGCCCCTCAACACATTTCCCCTCGAGTTGCACAACGTCGAGAGGGCGTACGGAAGGAACAGCTGCTACATCCGCTTCGTGCTCAAATCCACCCAC

#173_CDA1L2_1_Lp GCCGTCGGCGCCACCGGCACCACCGGCACCACAGGATCCGAGCTCTGGGGCTACGCCACCAACAAGTGGGAGGTCGCCGACGGCATCCCACGCGAGAGCCCGGAGAAGCGCGGCATGCAC
#175_CDA1L2_1_Lp GCCGTCGGCGCCACCGGCACCACCGGCACCACAGGATCCGAGCTCTGGGGCTACGCCACCAACAAGTGGGAGGTCGCCGACGGCATCCCACGCGAGAGCCCGGAGAAGCGCGGCATGCAC
#236_CDA1L2_1_Lp GCCGTCGGCGCCACCGGCACCACCGGCACCACAGGATCCGAGCTCTGGGGCTACGCCACCAACAAGTGGGAGGTCGCCGACGGCATCCCACGCGAGAGCCCGGAGAAGCGCGGCATGCAC
#175_CDA1L2_2_Lp GCCGTCGGCGCCACCGGCACCACCGGCACCACAGGATCCGAGCTCTGGGGCTACGCCACCAACAAGTGGGAGGTCGCCGACGGCATCCCACGCGAGAGCCCGGAGAAGCGCGGCATGCAC
#242_CDA1L2_1_Lp GCCGTCGGCGCCACCGGCACCACCGGCACCACAGGATCCGAGCTCTGGGGCTACGCCACCAACAAGTGGGAGGTCGCCGACGGCATCCCACGCGAGAGCCCGGAGAAGCGCGGCATGCAC
Lj#1_CDA1L2_1_Lj ----TCG---CCACCAGCACCACCGGCACCACAGAATCCGAGCTCTGGGGCTACGTCACTAACAAGGGGGAGGTCGCCGACGGCATCCCATGCGAGAGCCCGGAGAAGCGCGGCATGCAC
Lj#1_CDA1L2_2_Lj GCCGTCGGCGCCACCAGCACCACCGGCACCACAGGATCCGAGCTCTGGGGCTACGCCACCAACAAGTGGGAGGTCGCCGATGGCATCCCACGCGAGAGCCCGGAGAAGCGCGGCATGCAC
#144_CDA1L2_1_Pm GCCGTCGGCGCCACCGGCGCCGCC---------GGATMCGAGCTCTGGGGCTACGCCACCGGCGAGCGCGAGGTCGCCGGCGGCAGCCCACGCGAGAGCCCCGAGGGGGGCGGCCCGCAC
Pm#1_CDA1L2_1_Pm GCCGTCGGCGCCACCGGCGCCGCC---------GGATCCGAGCTCTGGGGCTACGCCACCGGCGAGCGCGAGGTCGCCGGCGGCAGCCCACGCGAGAGCCCCGAGGGGGGCGGCCCGCAC

#173_CDA1L2_1_Lp GCCGAGGAGCCGCTGCTGGATGAGATGACACGCCACGTCCGCGAGCACGGCGG-CGCCAGCTTCTGCATCGAGTGGT-TCACGTCGTGGAGCCCCTGCGACCGCTGCTCGGGGCTGCTGC
#175_CDA1L2_1_Lp GCCGAGGAGCCGCTGCTGGATGAGATGACACGCCACGTCCGCGAGCACGGCGG-CGCCAGCTTCTGCATCGAGTGGT-TCACGTCGTGGAGCCCCTGCGACCGCTGCTCGGGGCTGCTGC
#236_CDA1L2_1_Lp GCCGAGGAGCCGCTGCTGGATGAGATGACACGCCACGTCCGCGAGCACGGCGG-CGCCAGCTTCTGCATCGAGTGGT-TCACGTCGTGGAGCCCCTGCGACCGCTGCTCGGGGCTGCTGC
#175_CDA1L2_2_Lp ACCGAGGAGCTGCTGCTGGATGAGATGACACGCCACGTCCGCGAGCACGGCGG-CGCCAGCTTCTGCATCGAGTGGT-TCACGTCGTGGAGCCCCTGCGACCGCTGCTCGGGGCTGCTGC
#242_CDA1L2_1_Lp GCCGAGGAGCCGCTGCTGGATGAGATGACACGCCACGTCCGCGAGCACGGCGG-CGCCAGCTTCTGCATCGAGTGGT-TCACGTCGTGGAGCCCCTGCGACCGCTGCTCGGGGCTGCTGC
Lj#1_CDA1L2_1_Lj GCCGAGGAGCTGCTGCTGGAAGAGATGACACGCCACGTCCGCGAGCACGGCGG-CGCCAGCTTCTGCGTCGAGTGGT-TCGCGTCGTGGAGCCCCTGCGACCGCTGCTCGGGGGTGCTGC
Lj#1_CDA1L2_2_Lj ACCGAGGAGCTGCTGCTGGATGAGATGACACGCCACGTCCGCGAGCACAGTGG-CGTGGGCTTCTGCGTCGAGTGGT-TCACGTCGTGGAGCCCCTGCCACCGCTGCTCGGGGTTGCTGC
#144_CDA1L2_1_Pm GCCGGGCAGCTGCCGCTGGACGAGGTGACGCGCCACGTCCGCGAGCACGGCGGACGCGAGCTTCTGCGTCGAGCGGTATCCCGTCGTGGAGCCCCAGCGACCGCCGCTCGGGGCTGCCGC
Pm#1_CDA1L2_1_Pm GCCGGGCAGCTGCCGCTGGACGAGGTGACGCGCCACGTCCGCGAGCACGGCGGACGCGAGCTTCTGCGTCGAGCGGTATCCCGTCGTGGAGCCCCCGCGACCGCCGCTCGGGGCTGCCGC

#173_CDA1L2_1_Lp TCCACTGGCTGCGCGACGTCGGCGGCGGCCGACACCGGCTGCGCGTCTGGTTCTCCCGAATCTACCGCGGAAACGTCGGGGCCGTGCGGGCCGGGCTGCGTCACCTACGCCGCGCGGGGG
#175_CDA1L2_1_Lp TCCACTGGCTGCGCGACGTCGGCGGCGGCCGACACCGGCTGCGCGTCTGGTTCTCCCGAATCTACCGCGGAAACGTCGGGGCCGTGCGGGCCGGGCTGCGTCACCTACGCCGCGCGGGGG
#236_CDA1L2_1_Lp TCCACTGGCTGCGCGACGTCGGCGGCGGCCGACACCGGCTGCGCGTCTGGTTCTCCCGAATCTACCGCGGAAACGTCGGGGCCGTGCGGGCCGGGCTGCGTCACCTACGCCGCGCGGGGG
#175_CDA1L2_2_Lp ACAGCTGGCTGCGCGACGTCGGCGGCGGCCGACACCGGCTGCGCGTCTGGTTCTCCCGAATCTACTACGGAGACGTCGGGGCCGTGCGGGCCGGGCTGCGTCGCCTGCGCCGCGCGGGGG
#242_CDA1L2_1_Lp TCCACTGGCTGCGCGACGTCGGCGGCGGCCGACACCGGCTGCGCGTCTGGTTCTCCCGAATCTACCGCGGAAACGTCGGGGCCGTGCGGGCCGGGCTGCGTCACCTACGCCGCGCGGGGG
Lj#1_CDA1L2_1_Lj TCCGATGGCTGCGCGACGTCGGCGGCGGCCGACACCGGCTGCCGCGTCTGGTTCTCCCGAATCTACCGCGGAAACGTCGGGGCCGTGCGGGCCGGGCTGCGTCACCTACGCCGCGCGGGGG
Lj#1_CDA1L2_2_Lj TCCGTTGGCTGCGCGACATCGGCGGCGGCCGACACCGGCTGCGCGTCTGGTTCTCCCGAATCTACTACGGAGACGACGGCGCCGTGCGGGCCGGGCTGCGTCGCCTGCGCCGCGCGGGGG
#144_CDA1L2_1_Pm GCCGCTGGCTGCGCGACGTCGCCGGCGCCGCCGGCACCGGCTGCGCGTCTGGTTCTCCCGCATCGACC----------
Pm#1_CDA1L2_1_Pm GCCGCTGGCTGCGCGACGTCGCCGGCGCCGCCGGCACCGGCTGCGCGTCTGGTTCTCCCGCATCGACC----------

#173_CDA1L2_1_Lp TGCAGCTAGGCGTGATGGACAGGAGGCGGCACGATAACTGCGCGCACGCCCTGGTGGACGCGGCGCGGGGCGATCCCACGCCGC----------CGTGGCACATAAACGTCCCCCGCA
#175_CDA1L2_1_Lp TGCAGCTAGGCGTGATGGACAGGAGGCGGCACGATAACTGCGCGCACGCCCTGGTGGACGCGGCGCGGGGCGATCCCACGCCGC----------CGTGGCACATAAACGTCCCCCGCA
#236_CDA1L2_1_Lp TGCAGCTAGGCGTGATGGACAGGAGGCGGCACGATAACTGCGCGCACGCCCTGGTGGACGCGGCGCGGGGCGATCCCACGCCGC----------CGTGGCACATAAACGTCCCCCGCA
#175_CDA1L2_2_Lp TGCAGCTAGGCGTGATGGACGGGAGGCTGCGCGATTACTGCGCGCACGCCCTGGTGGACACGGCGCAGGGCGATCCCACGCCGCTGTGGCTCGTGCCGTGGCACATGAACGTCCCCCGCG
#242_CDA1L2_1_Lp TGCAGCTAGGCGTGATGGACAGGAGGCGGCACGATAACTGCGCGCACGCCCTGGTGGACGCGGCGCGGGGCGATCCCACGCCGC----------CGTGGCACATAAACGTCCCCCGCA
Lj#1_CDA1L2_1_Lj TGCAGCTCGGCGTGATGGACAGGAGGCGGCACGATAACTGCGTGCACGTCCTGGTGGACGCGGCGCGGGGCGATCCCACGCCGCGTGGCTCGTGCAGTGGCACAGTAACGTCCCCCGCG
Lj#1_CDA1L2_2_Lj TGCAGCTCGGCGTGATGGACGGGAGGCTGCACGATTACTGCGCTCACGTCCTGGTGGACGCGGCGTGGGGCGATCCCACGCCGCCGTGGCTCGTGCCGTGGCACATAAACGTCCCCCGCG
#144_CDA1L2_1_Pm ----------CCCGCGCCGCCCTGGCTCTTGCCGTGGCACAGAATCGTCCCCCGCG
Pm#1_CDA1L2_1_Pm ----------CCCGCGCCGCCCTGGCTCTTGCCGTGGCACAGAATCGTCCCCCGCG

#173_CDA1L2_1_Lp TCC-AACGGGCCTTCGACGAGATCATGGACGAGAAGGTGAGAGAGCCAGGATGGTGCTGGTGGTGCTGGCGGCCGTGGTG------GTTGGCT-GATGTGCCATCGAATGCTTCC----C
#175_CDA1L2_1_Lp TCC-AACGGGCCTTCGACGAGATCATGGACGAGAAGGTGAGAGAGCCAGGATGGTGCTGGTGGTGCTGGCGGCCGTGGTG------GTTGGCT-GATGTGCCATCGAATGCTTCC----C
#236_CDA1L2_1_Lp TCC-AACGGGCCTTCGACGAGATCATGGACGAGAAGGTGAGAGAGCCAGGATGGTGCTGGTGGTGCTGGCGGCCGTGGTG------GTTGGCT-GATGTGCCATCGAATGCTTCC----C
#175_CDA1L2_2_Lp TCC-AACGGGCCTTCGACGAGATCATGGACGAGAAGGTGAGAGAGCCAGGATGGTGCTGGTGGTGCTGGCGGCCGTGGTG------GTTGGCT-GATGTGCTATCGAATGCTTCCTTCCC
#242_CDA1L2_1_Lp TCC-AACGGGCCTTCGACGAGATCATGGACGAGAAGGTGAGAGAGCCAGGATGGTGCTGGTGGTGCTGGCGGCCGTGGTG------GTTGGCT-GATGTGCCATCGAATGCTTCC----C
Lj#1_CDA1L2_1_Lj TCC-AACGGGAATTCGACAAGATCATGGACGAGAAGGTGAGAGAGCCAGGATGGTGCTGGTGGTGCTGGCGGCCGTG---------GATGGCT-GATGTGCCATCGAATGCTTCCTTCCC
Lj#1_CDA1L2_2_Lj TCC-AACGGGCCTTCGACGAGATCATGGACGAGAAGGTGAGAGAGCCAGGATGGTGCTGGTGGTGCTGGCGGCCGTGGTG------GTTGGCT-GATGTGCCATCGAATGCTTCCTTCCC
#144_CDA1L2_1_Pm TCCCAACGGGCCTCCGACGAGATCGTGGACGAGGCGA-GAGAGCACAAGGACGGTGCTGGCGGTGCTGGCGGCCGTGGTGACGGTGGTTGGCCCGACGTGCCGTCGAATGCTTCCTTCCC
Pm#1_CDA1L2_1_Pm TCCCAACGGGCCTCCGACGAGATCGTGGACGAGGCGA-GAGAGCACAAGGACGGTGCTGGCGGTGCTGGCGGCCGTGGTGACGGTGGTTGGCCCGACGTGCCGTCGAATGCTTCCTTCCC

#173_CDA1L2_1_Lp TCAATTACCCGTGA
#175_CDA1L2_1_Lp TCAATTACCCGTGA
#236_CDA1L2_1_Lp TCAATTACCCGTGA
#175_CDA1L2_2_Lp TCAATTACACGTGAATGATGAGGCAATTGTGGACAGAGTGGAGAAGTTTCTGTA
#242_CDA1L2_1_Lp TCAATTACCCGTGA
Lj#1_CDA1L2_1_Lj TCAATTACCCGTGAATGATGGGGCAATTGTGGACAGAGTGGAGAAGTTTCCGTACCTGGGCAGTTTGCTTATGACCAGCTCTGGTTTGGCAGCGGAGGTCTCAATCTGA
Lj#1_CDA1L2_2_Lj TCAATTACCCGTGAATGATGAGGCAATTGTGGACAGAGTGGAGAAGTTTCTGTACCTGGGCAGTTTGTTTATGACCAGCTCTGGTTTGGCAGCGGAGGTCTCAATCTGA
#144_CDA1L2_1_Pm TCAATTACCCGTGAATGATGGGGCAATTGTGGACAGAGTGGAGAAGTTTTCGTACTTGGGCAGTTTGCTCATGACCAGCTCTGGTTTGGCA
Pm#1_CDA1L2_1_Pm TCAATTACCCGTGAATGATGGGGCAATTGTGGACAGAGTGGAGAAGTTTTCGTACTTGGGCAGTTTGCTCATGACCAGCTCTGGTTTGGC
```

C

```
LpCDA1L2_1_Unspliced MSIFLYKKLPLNTFLFEFDNLEKAYGRNRCYICFKLKPIHAVGATGTTGTTGSELWGYATNKWEVADGIPRESPEKRGMHAEEPLLDEMTRHVREHGGASFCIEWFTSWS
LpCDA1L2_1_Spliced   ---------------FEFDNLEKAYGRNRCYICFKLKPIHAVGATGTTGTTGSELWGYATNKWEVADGIPRESPEKRGMHAEEPLLDEMTRHVREHGGASFCIEWFTSWS
LpCDA1L2_2_Unspliced MSVFLHKKLPLNTFLFEFDNLEKAYGRNRCYICFKLKPIYAVGATGTTGTTGSELWGYATNKWEVADGIPRESPEKRGMHTEELLLDEMTRHVREHGGASFCIEWFTSWS
LpCDA1L2_2_Spliced   MSVFLHKKLPLNTFLFEFDNLEKAYGRNRCYICFKLKPIYAVGATGTTGTTGSELWGYATNKWEVADGIPRESPEKRGMHTEELLLDEMTRHVREHGGASFCIEWFTSWS
                                                                                                                    <> Exon 2
LpCDA1L2_1_Unspliced PCDRCSGLLLHWLRDVGGGRHRLRVWFSRIYRGNVGAVRAGLRHLRRAGVQLGVMDRRRHDNCAHALVDAARGDPTP----PWHINVPRIQRAFDEIMDEKVR------E
LpCDA1L2_1_Spliced   PCDRCSGLLLHWLRDVGGGRHRLRVWFSRIYRGNVGTVRAGLRHLRRAGVQLGVMDRRRHDNCAHALVDAARGDPTP----PWHINVPRIQRAFDEIMDEKDDNGSDNSD
LpCDA1L2_2_Unspliced PCHRCSGLLHSWLRDVGGGRHRLRVWFSRIYYGDVGAVRAGLRRLRRAGVQLGVMDGRLRDYCAHALVDTAQGDPTPLWLVPWHMNVPRVQRAFDEIMDEKVR------E
LpCDA1L2_2_Spliced   PCHRCSGLLHSWLRDVGGGRHRLRVWFSRIYYGDVGAVRAGLRRLRRAGVQLGVMDGRLRDYCAHALVDTAQGDPTPLWLVPWHMNVPRVQRAFDEIMDEKDDNGSDNSD
                     <>                  Exon 3
LpCDA1L2_1_Unspliced PG---------WCWWCWRPWWL------------A-----DVPSNASLNYP
LpCDA1L2_1_Spliced   PGKLSESASGGHESWHDDDLHLPLDDLTVVVECTPSKQGPPEATAAPTLPRKRQQEDPVDALTAKRALF
LpCDA1L2_2_Unspliced PG---------WCWWCWRPWWL------------A-----DVLSNASFPQLHVNDEAIVDRVEKFL
LpCDA1L2_2_Spliced   PGKLSESISGGHESWHDDDLHLPLEDLTVVVECTPSKQGPPEATAAPTLPRKRQQEDPVDALTAKRALF
```

Fig. S5. Alternative splicing of *CDA1L2* genes. (*A*) Read coverage plots from whole genome sequences (WGS) of several individuals, indicated above the plots. Green color indicates a coverage by >5 reads; blue, 2-5 reads, and red, single read. Orange bars correspond to region of contig containing open reading frame (ORF) of the gene. (*B*) Nucleotide alignment of ORFs of *CDA1L2* genes derived from whole genome sequences of *Lampetra planeri* (Lp#173, Lp#175, Lp#236, Lp#242), *Lethenteron japonicum* (Lj#1), and *Petromyzon marinus* (Pm#144 and Pm#1). *CDA1L2* genes in *Petromyzon marinus* are predicted to be pseudogenized based on numerous single nucleotide insertions (green asterisk) or deletions (red asterisk), and a large-scale deletion in the middle of the predicted ORF (blue asterisks). (*C*) Amino acid alignments of spliced and unspliced *CDA1L2* gene products from *Lampetra planeri*. Conserved APOBEC catalytic HxE and PCxxC motifs are highlighted in red. Exon 2 sequences are highlighted in orange, exon 3 sequences in green. Unspliced sequences are derived from whole genome shotgun sequences of lamprey Lp#175, the spliced sequences from RT-PCR or *de novo* transcriptome assembly.

A



B

```
Lp173_CDA1L2_Exon2.1 -------TGGTTTGTTTGAGGAGCTCACAGATCCTTGAAAGATCTGGAAACAGTCACATCAGTGAACTCCAGCAGGCGAGGCTATGTTGGCTGGGTCATGTAGCCTGCAT
Lp175_CDA1L2_Exon2.1 ACCACACTGGTTTGTTTGAGGAGCTCACAGATCCTTGAAAGATCTGGAAACAGTCACATCAGTGAACTCCAGCAGGCGAGGCTATGTTGGCTGGGTCATGTAGCCTGCAT
Lp175_CDA1L2_Exon2.2 ACCACACTGGTTTGTTTGAGGAGCTCACAGATCCTTGAAAGATCTGGAAACAGTCACATCAGTGAACTCCAGCAGGCGAGGCTATGTTGGCTGGGTCATGTAGCCTGCAT

Lp173_CDA1L2_Exon2.1 GCCCAGCCACCGCATGCCCATGCAGCTTCTATTTGGCTGGATAAAGGAGGGCTAAACAGGCACGGCATGGGCTTCAGAAGAGAATGGGCTGATGTGGTATAGGAGAATTT
Lp175_CDA1L2_Exon2.1 GCCCAGCCACCGCATGCCCATGCAGCTTCTATTTGGCTGGATAAAGGAGGGCTAAACAGGCACGGCATGGGCTTCAGAAGAGAATGGGCTGATGTGGTATAGGAGAATTT
Lp175_CDA1L2_Exon2.2 GCCCAGCCACCGCATGCCCATGCAGCTTCTATTTGGCTGGATAAAGGAGGGCTAAACAGGCACGGCATGGGCTTCAGAAGAGAATGGGCTGATGTGGTATAGGAGAATTT

Lp173_CDA1L2_Exon2.1 GGGAGTGTGTGGACTTGCGAAGGATAAGGTAGTAGCGGTGTCAATATTGGACTCTGTGAAGGAGGCTCGTGAAAGACGCTACCAGGCAGTTGGAGGCAACCGCCCTCCAT
Lp175_CDA1L2_Exon2.1 GGGAGTGTGTGGACTTGCGAAGGATAAGGTAGTAGCGGTGTCAATATTGGACTCTGTGAAGGAGGCTCGTGAAAGACGCTACCAGGCAGTTGGAGGCAACCGCCCTCCAT
Lp175_CDA1L2_Exon2.2 GGGAGTGTGTGGACTTGCGAAGGATAAGGTAGTAGCGGTGTCAATATTGGACTCTGTGAAGGAGGCTCGTGAAAGACGCTACCAGGCAGTTGGAGGCAACCGCCCTCCAT

Lp173_CDA1L2_Exon2.1 CACCAGCGGACCATCGCCGCTGCATGCGCAGTGTCTCCCCGTGTTTCACTGCAATGCCACTACTGCCCAAGTACATTGCTAACAAAATCACATCTCTTTTTGTCTTTGCA
Lp175_CDA1L2_Exon2.1 CACCAGCGGACCATCGCCGCTGCATGCGCAGTGTCTCCCCGTGTTTCACTGCAATGCCACTACTGCCCAAGTACATTGCTAACAAAATCACATCTCTTTTTGTCTTTGCA
Lp175_CDA1L2_Exon2.2 CACCAGCGGACCATCGCCGCTGCATGCGCAGTGTCTCCCCGTGTTTCACTGCAATGCCACTACTGCCCAAGTACATTGCTAACAAAATCACATCTCTTTTTGTCTTTGCA

Lp173_CDA1L2_Exon2.1 GGACGACAACGGGAGTGATAACTCAGGTAAATGGATGGCCAATGTATACTTTAAATATGCAGTATTTACATGAGAGGGATAATCCAGTCAATTGGAATGTGAACCCAGTA
Lp175_CDA1L2_Exon2.1 GGACGACAACGGGAGTGATAACTCAGGTAAATGGATGGCCAATGTATACTTTAAATATGCAGTATTTACATGAGAGGGATAATCCAGTCAATTGGAATGTGAACCCAGTA
Lp175_CDA1L2_Exon2.2 GGACGACAACGGGAGTGATAACTCAGGTAAATGGATGGCCAATGTATACTTTAAATATGCAGTATTTCCATGAGAGGGATAATCCAGTCAATTGGAATGTGAACCCAGTA

Lp173_CDA1L2_Exon2.1 AACATGCATTGTCGGGCCAAATATTGGTCTGTTGTTGGAATGTTTGCACCAATGTTGGTTCAATATTGGTCCAACATTTCCTGTTGACTAGGGAGAGTCTTTAGTTTGCT
Lp175_CDA1L2_Exon2.1 AACATGCATTGTCGGGCCAAATATTGGTCTGTTGTTGGAATGTTTGCACCAATGTTGGTTCAATATTGGTCCAACATTTCCTGTTGACTAGGGAGAGTCTTTAGTTTGCT
Lp175_CDA1L2_Exon2.2 AACATGCATTGTCGGGCCAAATATTGGTCTGTTGTTGGAATGTTTGCACCAATGTTGGTTCAATATTGGTCCAACATTTCCTGTTGACTAGGGAGAGTCTTTAGTTTGCT

Lp173_CDA1L2_Exon2.1 TGGAGCAACATATTTCAAAAGGAAAAATTGTAATAGTAAAAATGCCCTTTATTTACCAAATAATTATCTCACTGATCTCAAGTACTCTTTTTCGGGAGGGCCGTGCATTG
Lp175_CDA1L2_Exon2.1 TGGAGCAACATATTTCAAAAGGAAAAATTGTAATAGTAAAAATGCCCTTTATTTACCAAATAATTATCTCACTGATCTCAAGTACTCTTTTTCGGGAGGGCCGTGCATTG
Lp175_CDA1L2_Exon2.2 TGGAGCAACATATTTCAAAAGGAAAAATTGTAATAGTAAAAATGCCCTTTATTTACCAAATAATTATCTCACTGATCTCAAGTACTCTTTTTCGGGAGGGCCGTGCATTG

Lp173_CDA1L2_Exon2.1 GGACGTTTGACCAGTCCCATGAATGAGTGTTAAGTATGCCCACTGTCGCCTAGCCAGTCCGCCATGAGGCGGAGAAGAAGTGCAGTACAGTAGTTGTATCGTTCCACTTT
Lp175_CDA1L2_Exon2.1 GGACGTTTGACCAGTCCCATGAATGAGTGTTAAGTATGCCCACTGTCGCCTAGCCAGTCCGCCATGAGGCGGAGAAGAAGTGCAGTACAGTAGTTGTATCGTTCCACTTT
Lp175_CDA1L2_Exon2.2 GGACGTTTGACCAGTCCCATGAATGAGTGTTAAGTATGCCCACTGTCGCCTAGCCAGTCCGCCATGAGGCGGAGAAGAAGTGCAGTACAGTAGTTGTATCGTTCCACTTT

Lp173_CDA1L2_Exon2.1 TGCAAGAACGAGTA
Lp175_CDA1L2_Exon2.1 TGCAAGAACGAGTAGTGACATGTGGTTTCACATCAATCCCAACAGACGCGGCACTGTTATCATTGTGTAA
Lp175_CDA1L2_Exon2.2 TGCAAGAACGAGTAGTGACATGTGGTTTCACATCAATCCCAACAGACGCGGCACTGTTATCATTGTGTAA
```

C

```
Lp173_CDA1L2_Exon3.1 ----------------------------------------------------------------------------------------------------------------------------
Lp175_CDA1L2_Exon3.1 ----------------------------------------------------------------------------------------------------------------------------
Lp175_CDA1L2_Exon3.2 AGAAGACCCCGTGGATGCCTTGACGGCCAAGAGGGCCCTCTTTTAAATTTTGTACCTCTTCATGCGTATACTTGATCAGGTGCGGCACCGCTACATTTGCTAAACGTATGTACGGCATCG

Lp173_CDA1L2_Exon3.1 ----------------------------------------------------------------------------------------------------GTAACCC-CACACCACCCGACACAGCCAA
Lp175_CDA1L2_Exon3.1 ----------------------------------------------------------------------------------------------------------------------------
Lp175_CDA1L2_Exon3.2 TTTTTAATGTTGTGTATGTAAAAAGAAACCAAGTACAATTGTTCATAATCMAGGTTTTTGGGTTAGATCGCGTAAGTATAAATGTTGCGTCGTAACCCTCACACCACCCGACACAGCCAA

Lp173_CDA1L2_Exon3.1 TGAGTAAGGGATTTGTCGTGTAAACAAAACGCCAATGAGTATTGTAAATTGTAAATGTTGTGGATTTACTCTCCAGCACACCGATGGGCTGAAATAGTAACTAATTAGCGTTTTGTTTAC
Lp175_CDA1L2_Exon3.1 ------------------------------------------------------------GGATTTACTCTCCAGCACACCGATGGGCTGAAATAGTAACTAATTAGCGTTTTGTTTAC
Lp175_CDA1L2_Exon3.2 TGAGTAAGGGATTTGTCGTGTAAACAAAACGCCAATGAGTATTGTAAATTGTAAATGTTGTGGATTTACTCTCCAGCACACCGATGGGCTGAAATAGTAACTAATTAGCGTTTTGTTTAC

Lp173_CDA1L2_Exon3.1 GTGACGAATCCCTTACTAATTGGCTGTGTCGGGTGGTGGAGGGTTACGACACAACATTTATACTTACGCGATCTAACCCAAAAAA-CTTTATTCTGAATAATATTTGTCGCTAAAGCCTA
Lp175_CDA1L2_Exon3.1 GTGACGAATCCCTTACTAATTGGCTGTGTCGGGTGGTGGAGGGTTACGACACAACATTTATACTTACGCGATCTAACCCAAAAAA-CTTTATTCTGAATAATATTTGTCGCTAAAGCCTA
Lp175_CDA1L2_Exon3.2 GTGACGAATCCCTTACTAATTGGCTGTGTCGGGTGGTGGAGGGTTACGAGCAACATTTATACTTACGCGATCTAACCCAAAAAAACTTTATTCTGAATAATATTTGTCGCTAAAGCCGA

Lp173_CDA1L2_Exon3.1 TGCGTTAAACTGATGATCGAACGCGCGTTTCCGCTCCCCCCAGACCCCGGGAAGCTGTCGGAGTCCGCCAGCGGCGGGCACGAGAGTTGGCATGACGATGATTTGCACCTGCCGCTTGAT
Lp175_CDA1L2_Exon3.1 TGCGTTAAACTGATGATCGAACGCGCGTTTCCGCTCCCCCCAGACCCCGGGAAGCTGTCGGAGTCCGCCAGCGGCGGGCACGAGAGTTGGCATGACGATGATTTGCACCTGCCGCTTGAT
Lp175_CDA1L2_Exon3.2 TGCGTTAAACTGATGATCGAACGCGCGTTTCCGCTCCCCCCAGACCCCGGGAAGCTGTCGGAGTCCATCAGCGGCGGGCACGAGAGTTGGCATGACGATGATTTGCACCTGCCGCTTGAG

Lp173_CDA1L2_Exon3.1 GATCTGACGGTGGTCGTGGAGTGCACGCCCAGCAAGCAAGGACCTCCCGAAGCCACGGCCGCCCCCACGCTGCCCAGGAAACGACAACAAGAAGACCCCGTGGATGCCTTGACGGCCAAG
Lp175_CDA1L2_Exon3.1 GATCTGACGGTGGTCGTGGAGTGCACGCCCAGCAAGCAAGGACCTCCCGAAGCCACGGCCGCCCCCACGCTGCCCAGGAAACGACAACAAGAAGACCCCGTGGATGCCTTGACGGCCAAG
Lp175_CDA1L2_Exon3.2 GATCTGACGGTGGTCGTGGAGTGCACGCCCAGCAAGCAAGGACCTCCCGAAGCCACGGCCGCCCCCACGCTGCCCAGGAAACGACAACAAGAAGACCCCGTGGATGCCTTGACGGCCAAG

Lp173_CDA1L2_Exon3.1 AGGGCCCTCTTTTAAATTTTGTACCTCTTCATGCGTATACTTGATCAGGTGCGGCACCGCTACATTTGCTAAACGTATGTACGGCATCGTTTTTAATGTTGTGTATGTAAAAAGAAACCA
Lp175_CDA1L2_Exon3.1 AGGGCCCTCTTTTAAATTTTGTACCTCTTCATGCGTATACTTGATCAGGTGCGGCACCGCTACATTTGCTAAACGTATGTACGGCATCGTTTTTAATGTTGTGTATGTAAAAAGAAACCA
Lp175_CDA1L2_Exon3.2 AGGGCCCTCTTTTAAATTTTGTACCTCTTCATGCGTATACTTGATCAGGTGCGGCACCGCTACATTTGCTAAACGTATGTACGGCATCGTTTTTAATGTTGTGTATGTAAAAAGAAACCA

Lp173_CDA1L2_Exon3.1 AGTACAATTGTTCATAATCAAGGTTTTTGGATCAGATCGCGTTATTTATAAATGTGTCGTAACCCTCCACCACCCGACACGGCCAAACGGTAAAAAAATGTGTCACGTTTTTCAAAGGACA
Lp175_CDA1L2_Exon3.1 AGTACAATTGTTCATAATCAAGGTTTTTGGATCAGATCGCGTTATTTATAAATGTGTCGTAACCCTCCACCACCCGACACGGCCAAACGGTAAAAAAATGTGTCACGTTTTTCAAAGGACA
Lp175_CDA1L2_Exon3.2 AGTACAATTGTTCATAATCAAGGTTTTTGGATCAGATCGCGTTATTTATAAATGTGTCGTAACCCTCCACCACCCGACACGGCCAAACGGTAAAAAATGTGTCACGTTTTTCAAAGGACA

Lp173_CDA1L2_Exon3.1 AAATTATTCACTTGTTTAGCCCACCGGTGTGTTGAAGAGAGAATCAACAGCATTTT
Lp175_CDA1L2_Exon3.1 AAATTATTCACTTGTTTAGCCCACCGGTGTGTTGAAGAGAGAATCAACAGCATTTTAAAATTTGAGTACGACGTTTCGCTGATAATTACAATCAGCTTCATCGGGCAAAAATGGTGAAGC
Lp175_CDA1L2_Exon3.2 AA-TAATTCACTTGTTTAGCCCACCGGTGTGTTGAAGAGAGAATCAACAGCATTTTAAAATTTGAGTACGACGTTTCGCT

Lp173_CDA1L2_Exon3.1
Lp175_CDA1L2_Exon3.1 TCTCCTATATTATGAATAATATTGGTCGCGAAAGCCTACGCGTTAAAAAG
Lp175_CDA1L2_Exon3.2
```

Fig. S6. Genomic organization of *CDA1L2* exons. (*A*) Schematic (left) and read coverage plots (right) for individuals containing either one (Lp#173) or two (Lp#175) *CDA1L2* genes. Green color indicates a coverage by >5 reads; blue, 2-5 reads, and red, single read. Orange bars correspond to region of contig containing open reading frame (ORF) of the gene. Due to the high degree of sequence similarity between exons 2.1 and 2.2 in Lp#175 whole genome sequences, only one single contig was assembled containing both exon variants. Orange asterisks indicate the position of SNPs contained within exon 2. (*B, C*) Nucleotide alignments of sequences containing exons 2 (*B*) and 3 (*C*) of *CDA1L2* genes. ORFs highlighted in red, SNPs in purple and splice signal site in orange.

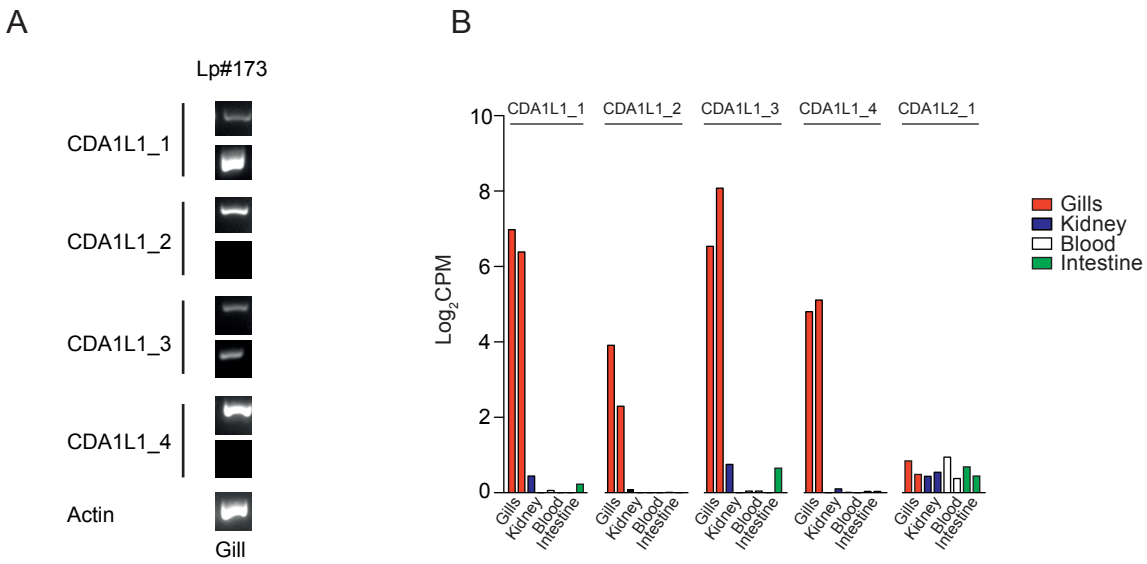
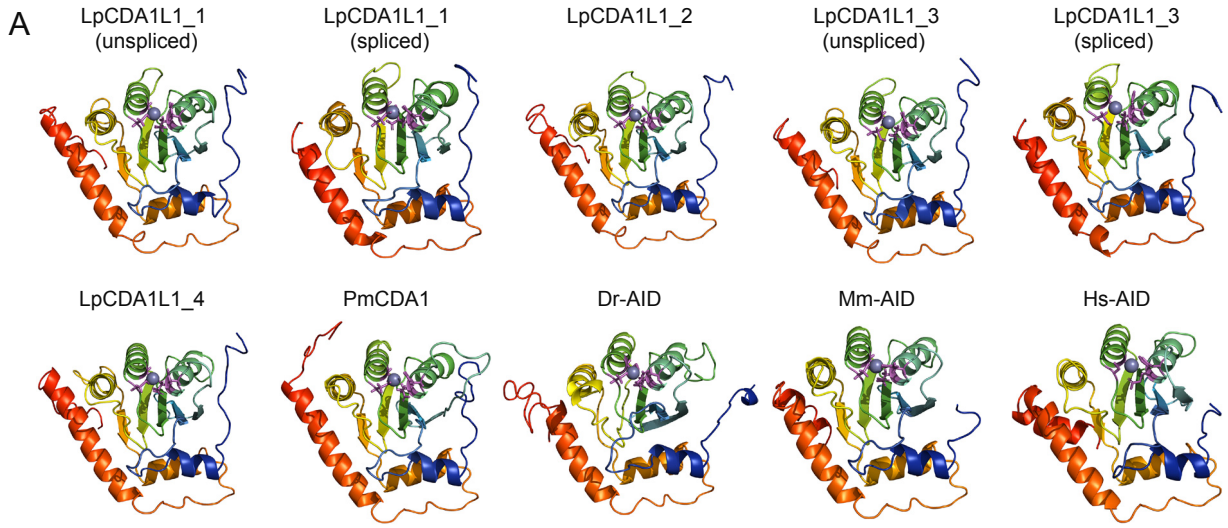

A
Lp#173
CDA1L1_1
CDA1L1_2
CDA1L1_3
CDA1L1_4
Actin
Gill
B
CDA1L1_1
CDA1L1_2
CDA1L1_3
CDA1L1_4
CDA1L2_1
Log2CPM
10
8
6
4
2
0
Gills
Kidney
Blood
Intestine

Fig. S7. Expression of *CDA1*-like genes. (*A*) RT-PCR analysis of *CDA1L1* gene expression using cDNA prepared from RNA of gills Lp#173, whose genome was confirmed to contain all 4 *CDA1L1* genes (c.f., Fig. S4). Gene-specific forward primers were used with universal reverse primers for unspliced (upper panels) or spliced (lower panels) gene products. Actin-specific primers were used as a control for cDNA integrity. (*B*) Expression of *CDA1*-like genes in immune organs of lamprey larvae. RNAseq-derived expression levels are expressed as $\log_2$ of counts per million ($\log_2$CPM) in two *Lampetra planeri* individuals (Lp#131; Lp#132).


A
LpCDA1L1_1
(unspliced)
LpCDA1L1_1
(spliced)
LpCDA1L1_2
LpCDA1L1_3
(unspliced)
LpCDA1L1_3
(spliced)
LpCDA1L1_4
PmCDA1
Dr-AID
Mm-AID
Hs-AID

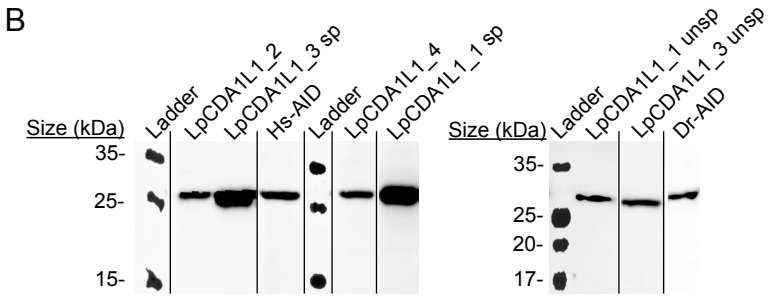

B
Size (kDa)
35-
25-
15-
Ladder
LpCDA1L1_2
LpCDA1L1_3 sp
Hs-AID
Ladder
LpCDA1L1_4
LpCDA1L1_1 sp
Size (kDa)
35-
25-
20-
17-
Ladder
LpCDA1L1_1 unsp
LpCDA1L1_3 unsp
Dr-AID

Fig. S8. Predicted structures of CDA1L1 proteins. (*A*) Representative ribbon models of LpCDA1L1 variants, PmCDA1 and DrAID, MmAID, and HsAID. All structures were predicted to have the same overall conserved tertiary structure apart from the C-terminal α7 domain of HsAID, which was not predicted in the LpCDA1L1 variants or PmCDA1. N- to C-terminal progression is shown from blue to red, with the catalytic residues colored purple and the coordinated Zn denoted by a grey sphere. (*B*) Western blot analysis confirming the expression of CDA variants in 293T cells. SDS-PAGE electrophoresis on a 10% polyacrylamide gel was conducted on lysates of 293T cells transfected with expression constructs encoding each variant. Proteins were transferred onto nitrocellulose membranes and probed with polyclonal rabbit anti-V5 Tag antibody. Expression was confirmed by a specific band present at the expected size ~27kDa. sp= spliced, unsp= unspliced.

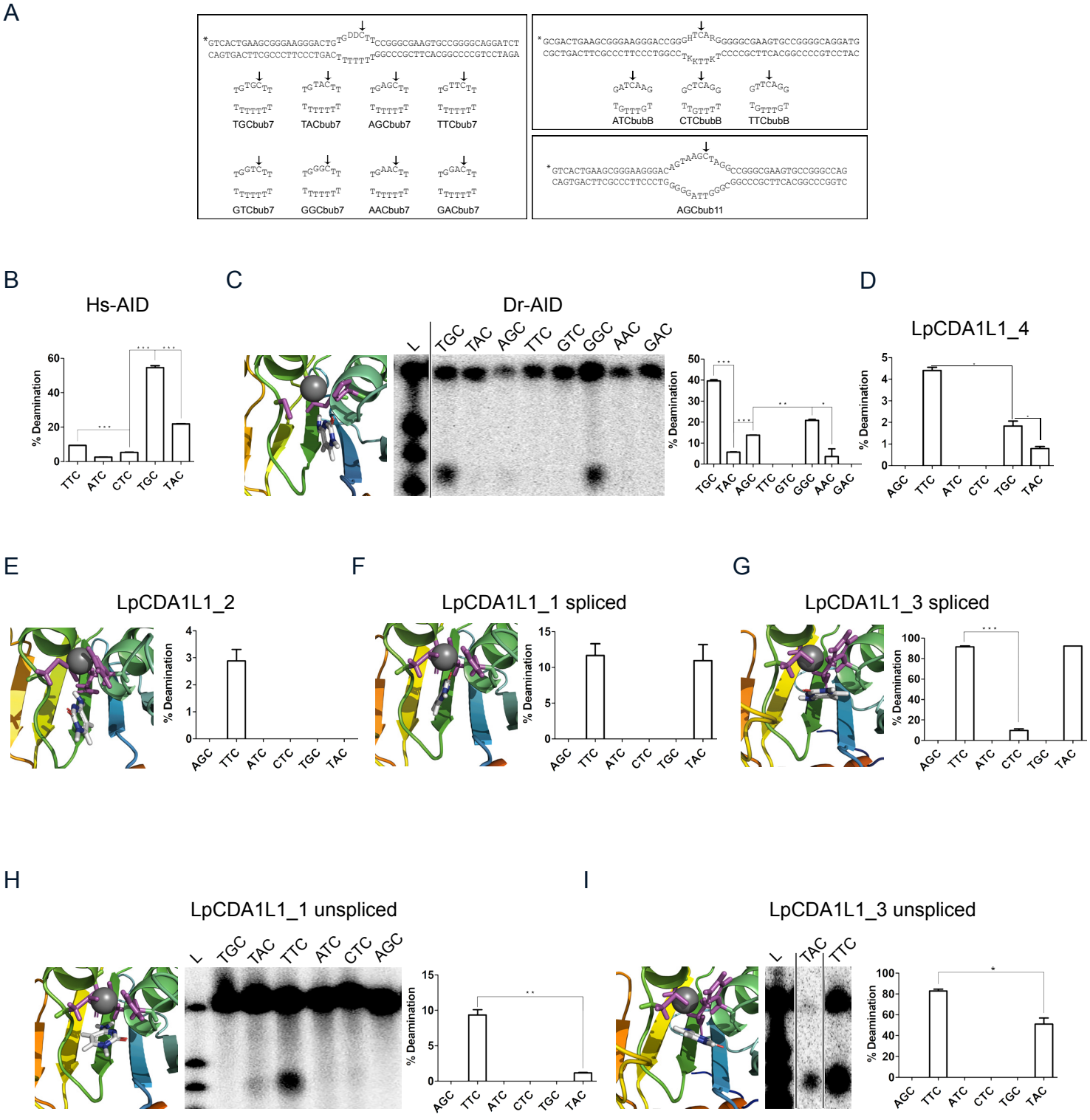

A
TGCbub7
TACbub7
AGCbub7
TTCbub7
GTCbub7
GGCbub7
AACbub7
GACbub7
ATCbubB
CTCbubB
TTCbubB
AGCbub11
B
Hs-AID
% Deamination
TTC
ATC
CTC
TGC
TAC
C
Dr-AID
L
TGC
TAC
AGC
TTC
GTC
GGC
AAC
GAC
D
LpCDA1L1_4
AGC
E
LpCDA1L1_2
F
LpCDA1L1_1 spliced
G
LpCDA1L1_3 spliced
H
LpCDA1L1_1 unspliced
I
LpCDA1L1_3 unspliced

Fig. S9. Cytidine deamination activity of CDA1L1 variants. (*A*) Oligonucleotide substrates for cytidine deamination by alkaline cleavage experiments. The target (d)C is indicated by an arrow. The size of cleaved product at the target (d)C indicative of cytidine deamination activity is 28 nt. Left panel, DD(d)Cbub7 set, where D= A or T or G. All substrates in this set are identical except for the two nucleotides immediately upstream of the target (d)C in the bubble. Top right panel, bubB set. Substrates in this set have identical double-stranded arms, with different 7 nucleotide single-stranded bubbles. H= A or C or T, R= A or G, K= T or G. Bottom right panel, AGCbub11. This substrate has an AGC motif in the center of an 11-nucleotide bubble. (*B*) Substrate specificity profile of Hs-AID measured in the alkaline cleavage assay. (*C*) Catalytic pocket of Dr-AID docked with (d)C in a deamination-feasible configuration, showing the interactions between the Zn-coordinating triad and target (d)C; the coordinated Zn is depicted as a grey sphere, with the Zn-coordinating and catalytic glutamic acid residues colored purple (left), representative alkaline cleavage gel (middle), and substrate specificity profile graphs (right). (*D*) Substrate specificity profile for LpCDA1L1_4. (*E*-*G*) Substrate specificity profiles for LpCDA1L1_2, LpCDA1L1_1, and LpCDA1L1_3 proteins, the latter two in spliced configuration (c.f., Fig. 3); catalytic pockets of proteins docked with (d)C in a deamination-feasible configuration are shown in the left panels. (*H,I*). Catalytic pockets of unspliced LpCDA1L1_1, and LpCDA1L1_3 proteins docked with (d)C in a deamination-feasible configuration (left panels), representative alkaline cleavage gels (middle panels), and substrate specificity profile graphs (right panels),*P $\leq$ 0.05, **P< 0.01, ***P< 0.001. Error bars represent standard error of the mean. Some display items were combined from different parts of the same autoradiographic films; the splice sites are indicated by solid lines. Ladder size as indicated in Figure 3.

Table S1: Summary of lamprey usage and tissue sources used in this study

Animal ID	Species	Stage	River	gDNA				RNA	
				Genome Sequencing	Southern Blot	PFGE	PCR	Transcriptome	RT-PCR
Lp#8a	*Lampetra planeri*	Larva	La Sélune				WB		
Lp#8b	*Lampetra planeri*	Larva	La Sélune				WB		
#159	*Lampetra planeri*	Larva	Rhine			B			
#130	*Lampetra planeri*	Larva	Rhine						K
#131	*Lampetra planeri*	Larva	Rhine					K G B T	G
#132	*Lampetra planeri*	Larva	Rhine					K G B T	
#144	*Petromyzon marinus*	Larva	Rhine	B					
#173	*Lampetra planeri*	Larva	Rhine	K G B T					G
#175	*Lampetra planeri*	Larva	Rhine	K G B T					
#196	*Lampetra planeri*	Larva	Rhine		WB		WB		
#236	*Lampetra planeri*	Larva	La Sélune	WB	WB				
#242	*Lampetra planeri*	Larva	Rhine	WB					
Lf#29	*Lampetra fluviatilis*	Larva	La Sélune				F		
Lf#33	*Lampetra fluviatilis*	Larva	La Sélune				F		
Pm#1	*Petromyzon marinus*	Larva	Rhine	WB					
Lj#1	*Lethenteron japonicum*	Adult	Ishikari	Te					

B=Blood; F=Fin Clip G=Gills; K=Kidney; T=Typhlosole (Intestine); Te=Testes; WB=Whole Body

Table S2: Summary of *de novo* transcriptome assembly statistics

L. planeri	**Tissue**	**Total # Read Pairs**	**Total # Reads after trimming**		**Total # of ORFs**	**# Unique ORFs**
			R1	**R2**		
#131	Blood	186590513	183786365	155576954	675747	**153032**
	Gills	202230155	199814264	173904091		
	Kidney	209200560	206317660	180153510		
	Intestine	169020866	166889768	138656896		
	Total	767042094	756808057	648291451		
#132	Blood	173750048	171219322	148579163	645503	**152284**
	Gills	207350277	204725620	177506558		
	Kidney	199057420	196587887	169331118		
	Intestine	132513609	131013341	103726338		
	Total	712671354	703546170	599143177		

Table S3: Summary of CDA1L1 genotype of different lamprey species.

WGS or PCR	P	W	P	W	P	W	W	P	P	W	P	W	W
CDA1L1_1											?		
CDA1L1_2											?		
CDA1L1_3											?		
CDA1L1_4											?		
CDA1													
	#196	#173	#8a	#175	#8b	#236	#242	#29	#33	Lj#1	Lj#2	Pm#1	#144
	Lp							*Lf*		*Lj*		*Pm*	

Lp = Lampetra planeri; Lf = Lampetra fluviatilis; Lj = Lethenteron japonicum; Pm = Petromyzon marinus

References

1. Smith JJ, et al. (2013) Sequencing of the sea lamprey (Petromyzon marinus) genome provides insights into vertebrate evolution. *Nat Genet* 45: 415-421.

2. Smith JJ, et al. (2018) The sea lamprey germline genome provides insights into programmed genome rearrangement and vertebrate evolution. *Nat Genet* 50: 270-277.

3. Rougemont Q, et al. (2015) Low reproductive isolation and highly variable levels of gene flow reveal limited progress towards speciation between European river and brook lampreys. *J Evol Bio.* 28: 2248-2263.