UNIVERSIDADE DA CORUÑA

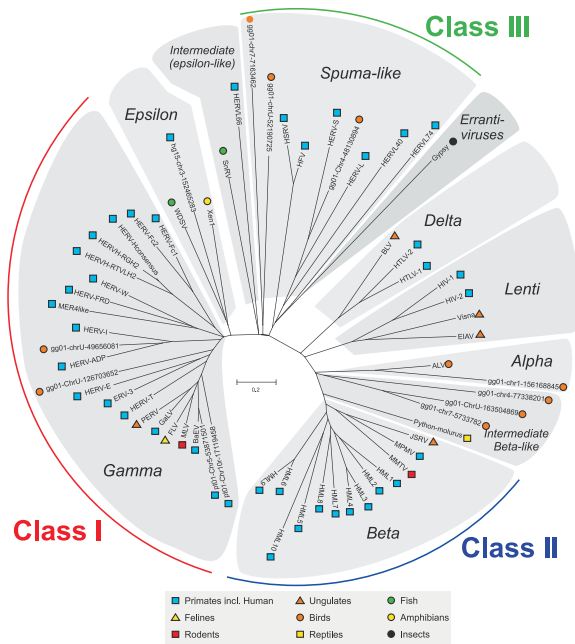
Grao en Bioloxía

Memoria do Traballo de Fin de Grao

Relacións filoxenéticas dos retrovirus endóxenos de *Petromyzon marinus*

Relaciones filogenéticas de los retrovirus endógenos de *Petromyzon marinus*

Philogenetic relationships of the endogenus retrovirus of *Petromyzon marinus*



Gonzalo Rodríguez Varela

Xullo,2021

Director académico: Horacio Naveira Fachal

Índice:

Resumo

Pese aos avances no mapeado do xenoma de animais fóra do grupo das aves e dos mamíferos, así como á localización de retrovirus endóxenos (ERVs) na gran mayoría destes xenomas, a historia evolutiva dos ERVs sigue sendo misteriosa e o seu orixe mantense descoñecido. Neste traballo realízase un estudo sobre o xenoma de *Petromyzon marinus* e o análise das insercióǹs completas destes elementos encontradas nel. Este organismo é de especial interese debido a que pertence ao grupo de vertebrados máis basal no que se atoparon ERVs. As insercióǹs retrovirais encontradas nel pertencen ao grupos dos spumavirus, o cal é considerado actualmente o máis basal dos retrovirus. Se as secuencias retrovirais de *Petromyzon marinus* son basáis ao resto de ERVs, estas deben mostrar unha erosión considerable debido as mutacións acontecidas desde o momento da inserción. Pola contra, de tratarse dunha inserción recente debido a un evento de transmisión entre especies, as secuencias estarán conservadas. Este análise confirma que a maior parte dos ERVs encontrados orixináronse a partir de insercióǹs recente, posiblemente procedentes dun evento de transmisión horizontal entre *Petromyzon marinus* e peixes de aletas lobuladas ou peixes de aletas radiadas. Sen embargo, tamén pon de manifesto a presencia dunha inserción retroviral de difícil datación que podería ser a clave para identificar un ERV especialmente basal que axude a aclarar a historia evolutiva destes elementos retrovirais.

Palabras clave: *Petromyzon marinus*, retrovirus endóxeno, evolución, paleoviroloxía.

Resumen

Pese a los avances en el mapeado del genoma de animales fuera del grupo de las aves y de los mamíferos, así como a la localización de retrovirus endógenos (ERVs) en la gran mayoría de estos genomas, la historia evolutiva de los ERVs sigue siendo un misterio y su origen se mantiene desconocido. En este trabajo se realiza un estudio sobre el genoma de Petromyzon marinus y el análisis de las inserciones completas de estos elementos encontradas en él. Este organismo es de especial interés debido a que pertenece al grupo de vertebrados más basal en el que se encontraron ERVs. Las inserciones retrovirales

encontradas en el pertenecen al grupo de los spumavirus, el cual es considerado actualmente como el más basal de los retrovirus. Si las secuencias de *Petromyzon marinus* son basales al resto de ERVs, estas deben mostrar una erosión considerable debido a las mutaciones sucedidas desde el momento de la inserción. Al contrario, de tratarse de una inserción reciente debido a una transmisión entre especies, las secuencias estarán conservadas. Este análisis confirma que la mayor parte de los ERVs encontrados se originaron a partir de inserciones recientes procedentes de un evento de transmisión horizontal, posiblemente, entre Petromyzon marinus y peces de aletas lobuladas o peces de aletas radiadas. Sin embargo, también pone de manifiesto la presencia de una inserción retroviral de difícil datación que podría ser la clave para identificar un ERV especialmente basal que ayudase a aclarar la historia evolutiva de estos elementos retrovirales.

Palabras clave: *Petromyzon marinus*, retrovirus endógeno, evolución, paleovirología.

Abstract

Despite advances in genome mapping of animals outside the avian and mammalian groups, as well as the localisation of endogenous retroviruses (ERVs) in the vast majority of these genomes, the evolutionary history of ERVs remains a mystery and their origin remains unknown. In this paper we study the genome of Petromyzon marinus and analyse the complete insertions of the elements found in it. This organism is of special interest because it belongs to the most basal group of vertebrates in which ERVs have been found. The retroviral insertions found in it belong to the spumavirus group, which is currently considered to be the most basal of the retroviruses. If the Petromyzon marinus sequences are basal to the rest of the ERVs, they must show considerable erosion due to mutations since the time of insertion. On the contrary, if it is a recent insertion due to interspecies transmission, the sequences will be conserved. This analysis confirms that most of the ERVs found originated from recent insertions from a horizontal transmission event, possibly between Petromyzon marinus and lobe-finned or ray-finned fishes. However, it also highlights the presence of a retroviral insertion that is difficult to date and could

be the key to identifying a particularly basal ERV that would help to clarify the evolutionary history of these retroviral elements.

Key-words: *Petromyzon marinus*, endogenous retrovirus, evolution, paleovirology.

1.-Introdución

Os retrovirus diferéncianse doutros virus de RNA polo seu mecanismo de replicación, o cal consiste na transcrición inversa do seu xenoma de ARN nunha copia de ADN mediante a acción da reverso transcriptasa (RT) e a integración nos cromosomas do hospedador. Os retrovirus tamén se caracterizan por integrar o seu xenoma viral no ADN cromosómico do hospedador (Gifford & Tristem, 2003). Adoitan infectar a células somáticas, pero en certas ocasións poden integrarse no xenoma das células xerminais, de forma que comezan a transmitirse verticalmente entre as xeracións do organismo hospedador, converténdose en endoretrovirus (ERV). Os ERVs non sofren ningunha selección purificadora, o que leva a acumulación de mutacións nas súas secuencias que, xunto con mecanismos de inactivación procedentes do xenoma do hospedador , leva inevitablemente a extinción da linaxe do ERV (Gifford & Tristem, 2003; Xu *et al*., 2018). Esta acumulación de mutacións pode usarse para inferir o momento de inserción do ERV, convertindo a esta insercións retrovirais en elementos fósiles de gran importancia para entender a evolución dos ERVs (Naville & Volff, 2016; Xu *et al*., 2018).

Os ERVs están presentes en todos os grupos de vertebrados (Herniou *et al*., 1998), sendo *Petromyzon marinus* o vertebrado máis basal no que se detectaron secuencias de ERVs (Xu *et al*., 2018). Este organismo ten unha especial importancia debido a diverxencia do seu liñaxe con respeto ao resto de vertebrados fai uns 500 millóns de anos, así como a súa utilidade como grupo externo dos mesmos(Osório & Rétaux, 2008; Smith *et al.*, 2013).

A ausencia de ERVs en *Branchiostoma floridae* (cefalocordada) parece indicar que a súa orixe encontrase entre os primeiros vertebrados, antes da aparición dos peixes con mandíbulas (Naville & Volff, 2016; Xu *et al*., 2018). Sen embargo, é erróneo concluír que debido á transmisión vertical dos ERVs estes reflexan a historia filoxenética dos seus hospedadores, xa que, contrario ao que

se adoitaba pensar, é común a transmisión horizontal de retrovirus entre especies polo que unha liñaxe evolutivamente antiga pode ser infectada por unha transmisión entre especies e presentar unha inserción recente de un ERV (Hayward, 2017; Xu *et al*., 2018).

A orixe dos ERVs está estimado mediante o uso de ortólogos, o cálculo mediante o ratio de acumulación de mutacións no xenoma viral e o estudo da coevolución co hospedador, datando esta inserción entre uns 455-473 millóns de anos, durante o Ordovícico (Hayward, 2017). Isto fai probable que esta orixe se producira no medio mariño entre os primeiros vertebrados mariños e que colonizasen o medio terrestre ao mesmo tempo que os tetrápodos ou ben se producisen varias transmisións da auga á terra mediante transmisión entre especies. Esta ultima opción é a máis probable debido a abundancia de ERVs de orixe terrestre no medio acuático e viceversa (Hayward, 2017; Xu *et al*., 2018).

Os retrovirus divídense en sete xéneros: alfa, beta, gammaretrovirus, epsilonretrovirus, lentiretrovirus, deltaretrovirus e spumavirus (tamén coñecidos como foamy virus), sen embargo, esta clasificación non é valida para os ERVs. Estes clasifícanse en ERVs clase I para os relacionados con gammaretrovirus e epsilonretrovirus; ERVs clase II para os emparentados con betaretrovirus e clase III para os relacionados con spumavirus. Con todo, a clasificación de ERVs resulta problemática, debido a incompatibilidade entre a relativa a retrovirus exóxenos e endóxenos, así como a falta de contraparte exóxena para algúns ERVs de recente endoxenización (Hayward *et al*., 2015; Naville & Volff, 2016; Wang & Han, 2021; Xu *et al*., 2018)

A distribución dos ERVs parece estar influída polo desenvolvemento do sistema inmune dos vertebrados, levándose acabo unha carreira evolutiva entre ambos dende a aparición do sistema inmune nos primeiros vertebrados (peixes sen mandíbula similares a *Petromyzon marinus* , destacando que este é o organismo do grupo máis basal no que se atoparon ERVs) (Escalera-Zamudio & Greenwood, 2016). Cabe destacar que algúns destes ERVs son beneficiosos para o seu hospedador ao codificar información para novas proteínas como é o caso das sincitinas, proteínas relacionadas ca formación da placenta e que proveñen de ERVs. Incluso hai exemplos de xens derivados de ERVs que están involucrados na resistencia contra infeccións retrovirais, como Fv1, xen que

protexe contra a infección dos virus da leucemia murina e que está derivado do dominio gag dos ERVs da familia MERV-L (Magiorkinis *et al.*, 2017; Naville & Volff, 2016).

Recentes traballos reportaron a existencia dun novo tipo de retrovirus coñecidos como lokiretrovirus, os cales forman unha liñaxe irmán aos ERVs. Estes elementos foron encontrados en todas as liñaxes dos vertebrados, estando máis relacionados co grupo dos spumavirus. Os lokiretrovirus comparten características cos ERVs pero diferéncianse pola presencia de proteínas homologas as SMC (proteínas mantedoras da estructura dos cromosomas). A súa secuencia consenso consta de xenes gag, pol e env. Outra das diferencias cos ERVs é que as proteínas do seu dominio env aseméllanse as glicoproteínas de fusión propias de virus de ARN de cadea única en sentido negativo, procedendo esta dun antepasado viral común, diferenciándose así dos ERVs (Wang & Han, 2021).

Neste traballo revisaranse as insercións de ERVs encontradas no xenoma de *Petromyzon marinus* para analizar a diverxencia entre ambas LTRs e inferir o momento de inserción destes ERVs. Este organismo presenta un interese particular ao ser un peixe sen mandíbulas e o vertebrado máis basal no que se encontraron ERVs, así como por pertencer estes ao grupo dos spumavirus, os ERVs considerados actualmente como os máis antiguos (Xu *et al.*, 2018). De esta forma, buscase clarificar se os elementos de *Petromyzon marinus* corresponden a ERVs basáis na árbore evolutiva destes virus ou se pola contra se trata de insercións recentes ocorridas por transmisión horizontal.

2.-Obxetivos:

1-Localizar e caracterizar as insercións de ERVs no xenoma de *Petromyzon marinus*

2-Estimar a antigüidade das distintas insercións provirais

3-Estudar as relacións filoxenéticas entre insercións baseadas no xen da reverso-transcriptasa

4-Analizar criticamente os resultados á luz da información actual sobre os ERVs de *Petromyzon marinus.*

3.-Material e métodos

3.1.-Secuencia consenso

A secuencia consenso de ERVs de *Petromyzon marinus* foi recuperada do traballo de Xu *et al*. 2018. Adicionalmente a secuencia de *Lethenteron camtschaticum* e o xen Cer1 de *Caenorhabditis elegans* usadas como outgroup, foron recuperadas do traballo de Wang & Han 2021 e do Genbank do NCBI (accession no. U15406) respectivamente. As secuencias foron aliñadas usando o programa BioEdit (Alzohairy, 2011).

3.2.-ERVs

Os ERVs usados neste traballo localizáronse mediante o servidor BLAST do NCBI. O ERV da secuencia consenso do traballo de Xu *et al*.2018 foi utilizado como query para este propósito, empregando a ensamblaxe kPetMar1.pri GenBank [GCA_010993605.1]. As secuencias dos *hits* localizados descargáronse , en formato FASTA, mediante a páxina do NCBI, aumentando 10000 pares de bases a lonxitude das súas secuencias no extremo 5' e 5000 pares de bases no extremo 3'. A parte das secuencias relativa aos endoretrovirus foi identificada mediante o uso da aplicación LTRharvester (Ellinghaus *et al*., 2008)

3.3.-LTRs e inserciónss completas

As repeticións terminais longas (LTR) da secuencia consenso encontráronse usando LTRharvester (Ellinghaus et al., 2008). Así mesmo, a búsqueda das inserciónss realizouse mediante a aplicación Blast do NCBI. As secuencias recuperáronse a partir da mesma páxina. A súa vez as LTRs destas secuencias identificáronse co mesmo procedemento e foron aliñadas mediante BioEdit (Alzohairy, 2011) usando a aplicación ClustalW (Larkin et al., 2007). A reverso trancriptasa (RT) e a ribonucleasa destes elementos foron identificadas a través da aplicación Conserved Domains do NCBI. A comparación das LTRs de CM021489 foi realizada mediante o programa Blast2 (Tatusova & Madden, 1999).

En canto as RTs, obtívose a secuencia aminoacídica de cada unha. Tanto estas como as secuencias nucleotídicas foron aliñadas mediante ClustalW

(Larkin et al., 2007) no programa MEGA X (Kumar, Stecher, Li, Knyaz, and Tamura 2018).

3.4.-Datación das insercións

Dende o momento de inserción na línea xerminal do hospedador, a LTR 5' e a LTR 3' do ERV comezan a diferenciarse debido á acumulación de mutacións, podendo usarse esta diverxencia para inferir o momento de integración. As LTRs das insercións completas foron analizadas co programa MEGA X (Kumar, Stecher, Li, Knyaz, and Tamura 2018) , estimándose a distancia mediante o medida de Kimura de 2 parámetros. A data de inserción foi estimada mediante a fórmula T=d/2μ onde d equivale a distancia entre as dúas LTRs e μ ao ratio de evolución neutro. Ante a falta do ratio de evolución neutro específico para peixes usouse o dispoñible para mamíferos , $\mu=2.2\ 10^{-9}$(Wang & Han, 2021).

3.5.-Construcción da árbore filoxenética

A partir das secuencia nucleotídicas das RTs reconstruíuse a historia evolutiva mediante o método *Neighbour-Joining* no programa MegaX (Kumar, Stecher, Li, Knyaz, and Tamura 2018). A árbore consenso de bootstrap foi obtida a partir de 10000 réplicas. Para esta árbore usouse a secuencia de RTLokiLca como grupo externo en lugar de RTCer, pola gran cantidade de diferencias que esta posuía. En canto as secuencias aminoacídicas, realizouse o análise de máxima parsimonia,usando a secuencia de RTCer como grupo externo. O apoio estadístico calculouse mediante a construción da árbore condensada. Cabe destacar que a secuencia pertencente a JAAIYE01000207 non foi incluída pola ausencia de RT e as pertencentes a JAAIYE010000319 polo deterioro na súa RT, o que impedía un correcto aliñamento.

4.-Resultados e discusión

Neste traballo recuperáronse 15 secuencias de ERVs procedentes do xenoma de *Petromyzon marinus* coa búsqueda por BLAST na reconstrución do xenoma kPetMar1.pri GenBank ensamblaxe [GCA_010993605.1].

A maior parte das datas de inserción estudadas neste traballo son relativamente recentes, coas maioría orixinándose fai entre 3 millóns de anos

(ma) e 100.000 anos, cas máis antigas entre 20 e 31 ma, como se pode ver no anexo I.

```
##gff-version 3
##sequence-region seq0 1 14001
# JAAIYE010001008.1:11139000-11153000 Petromyzon marinus isolate kPetMar1 chromosome 47, whole genome shotgun sequence
seq0    LTRharvest      repeat_region   2       13286   .       ?       .       ID=RepeatReg0
seq0    LTRharvest      LTR_retrotransposon     6       13282   .       ?       .       ID=LTRret0;Parent=RepeatReg0
seq0    LTRharvest      long_terminal_repeat    6       240     .       ?       .       ID=LTR0;Parent=LTRret0
seq0    LTRharvest      long_terminal_repeat    13049   13282   .       ?       .       ID=LTR1;Parent=LTRret0
seq0    LTRharvest      target_site_duplication 2       5       .       ?       .       ID=TSD0;Parent=RepeatReg0
seq0    LTRharvest      target_site_duplication 13283   13286   .       ?       .       ID=TSD1;Parent=RepeatReg0
```

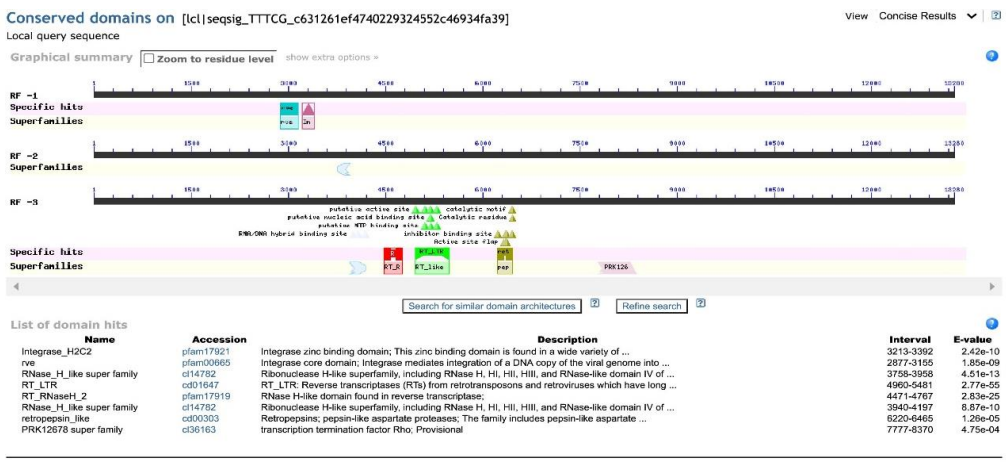


Ilustración 1: Distancia e dominios de JAAIYE010001008. A datación de 3ma indica unha inserción relativamente recente.

```
##gff-version 3
##sequence-region seq0 1 22868
# JAAIYE010001116.1:1016292-1039159 Petromyzon marinus isolate kPetMar1 chromosome 6, whole genome shotgun sequence
seq0    LTRharvest      repeat_region   8157    22838   .       ?       .       ID=RepeatReg0
seq0    LTRharvest      LTR_retrotransposon     8161    22834   .       ?       .       ID=LTRret0;Parent=RepeatReg0
seq0    LTRharvest      long_terminal_repeat    8161    10096   .       ?       .       ID=LTR0;Parent=LTRret0
seq0    LTRharvest      long_terminal_repeat    20941   22834   .       ?       .       ID=LTR1;Parent=LTRret0
seq0    LTRharvest      target_site_duplication 8157    8160    .       ?       .       ID=TSD0;Parent=RepeatReg0
seq0    LTRharvest      target_site_duplication 22835   22838   .       ?       .       ID=TSD1;Parent=RepeatReg0
```

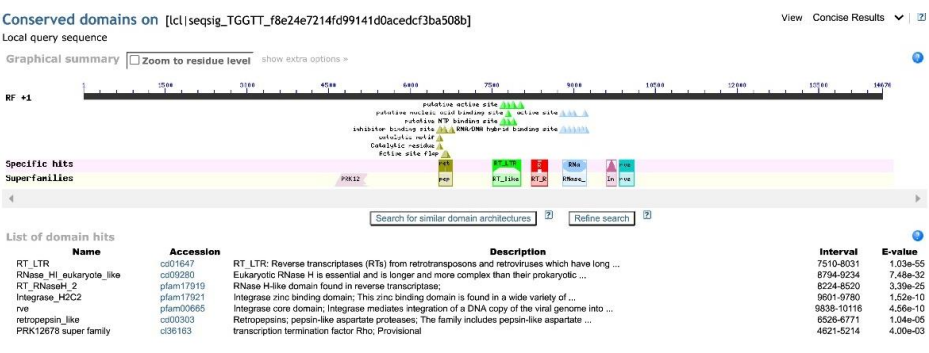
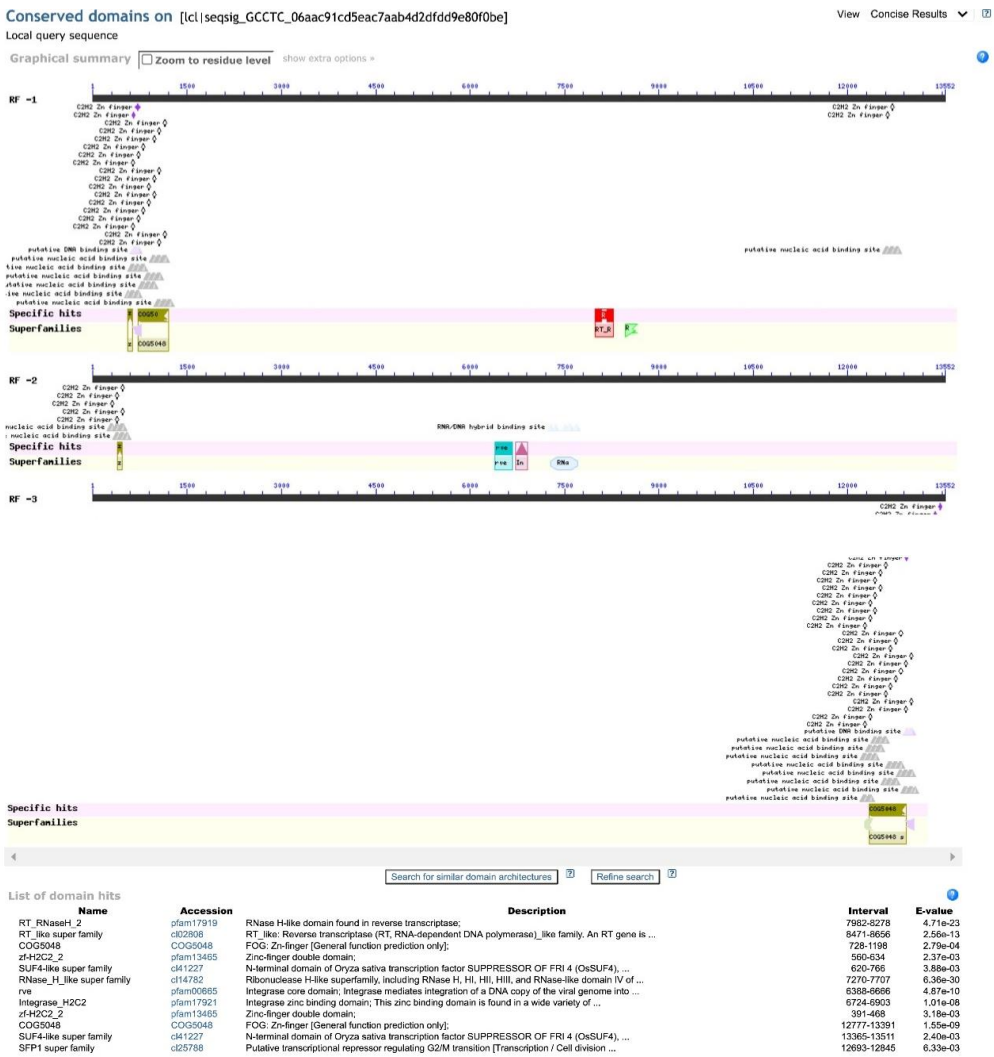


Ilustración 2:Distancia e dominios de JAAIYE010001116. Das secuencias estudiadas, esta é a que presenta a data de inserción máis recente.

```
##gff-version 3
##sequence-region seq0 1 17867
# JAAIYE010000737.1:360156-378022 Petromyzon marinus isolate kPetMar1 scaffold_292_arrow_ctg1, whole genome shotgun sequence
seq0    LTRharvest      repeat_region   4212    17768   .       ?       .       ID=RepeatReg0
seq0    LTRharvest      LTR_retrotransposon     4216    17764   .       ?       .       ID=LTRret0;Parent=RepeatReg0
seq0    LTRharvest      long_terminal_repeat    4216    5451    .       ?       .       ID=LTR0;Parent=LTRret0
seq0    LTRharvest      long_terminal_repeat    16546   17764   .       ?       .       ID=LTR1;Parent=LTRret0
seq0    LTRharvest      target_site_duplication 4212    4215    .       ?       .       ID=TSD0;Parent=RepeatReg0
seq0    LTRharvest      target_site_duplication 17765   17768   .       ?       .       ID=TSD1;Parent=RepeatReg0
```



Name	Accession	Description	Interval	E-value
RT_RNaseH_2	pfam17919	RNase H-like domain found in reverse transcriptase;	7982-8278	4.71e-23
RT_like super family	cl02808	RT_like: Reverse transcriptase (RT, RNA-dependent DNA polymerase)_like family. An RT gene is ...	8471-8656	2.56e-13
COG5048	COG5048	FOG: Zn-finger [General function prediction only];	728-1198	2.79e-04
zf-H2C2_2	pfam13465	Zinc-finger double domain;	560-634	2.37e-03
SUF4-like super family	cl41227	N-terminal domain of Oryza sativa transcription factor SUPPRESSOR OF FRI 4 (OsSUF4), ...	620-766	3.88e-03
RNase_H_like super family	cl14782	Ribonuclease H-like superfamily, including RNase H, HI, HII, HIII, and RNase-like domain IV of ...	7270-7707	6.36e-30
rve	pfam00665	Integrase core domain; Integrase mediates integration of a DNA copy of the viral genome into ...	6388-6666	4.87e-10
Integrase_H2C2	pfam17921	Integrase zinc binding domain; This zinc binding domain is found in a wide variety of ...	6724-6903	1.01e-08
zf-H2C2_2	pfam13465	Zinc-finger double domain;	391-468	3.18e-03
COG5048	COG5048	FOG: Zn-finger [General function prediction only];	12777-13391	1.55e-09
SUF4-like super family	cl41227	N-terminal domain of Oryza sativa transcription factor SUPPRESSOR OF FRI 4 (OsSUF4), ...	13365-13511	2.40e-03
SFP1 super family	cl25788	Putative transcriptional repressor regulating G2/M transition [Transcription / Cell division ...	12693-12845	6.33e-03

Ilustración 3: Distancia e dominios de JAAIYE010000737

Gran cantidade dos ERVs datados na actualidade superan con creces esa idade, chegando a encontrarse xa só en euterios, foamy virus cunha datación de entre 104 e 110 ma. As estimas para a ampla maioría de retrovirus superan os 30ma (Aiewsakun & Katzourakis, 2015; Lee et al., 2013). Esta datación coincide coa hipótese de que estas inserciónsvirais son posteriores a diverxencia entre

a lamprea mariña e a lamprea ártica, suceso acontecido fai uns 30-38 ma. Isto pon de manifesto que maioría das inserciónsde Petromyzon marinus non pertencen a un antepasado dos ERVs de vertebrados nin a un grupo especialmente antigo, se non que posiblemente se orixinaron a través da transmisión horizontal entre especies, procedendo de inserciónsrecentes de retrovirus de peixes de aletas radiadas e lobuladas, un suceso bastante habitual en ERVs (Chalopin et al., 2015.; Escalera-Zamudio & Greenwood, 2016; Hayward et al., 2015; Wang & Han, 2021; Xu et al., 2018). Esta hipótese está reforzada polo feito de que tanto lampreas como peixes teleósteos comparten hábitat, ademais, a dieta de Petromyzon marinus basease na depredación de peixes teleósteos, aos cales se adhire usando a súa boca como unha ventosa, alimentándose do sangue e tecidos que obten ao raspar o tecido da presa coa súa língua cornea. Isto favorece a transmisión horizontal debido ao intercambio de fluídos corporais como o sangue, facendo máis probable a transmisión entre especies (Gifford & Tristem, 2003; Osório & Rétaux, 2008).

Cabe destacar dúas excepción entre estas secuencias, a primeira sendo JAAIYE010001210 a cal foi datada en 27millóns de anos pero nas súa LTR 5' encontrouse unha secuencia repetida dispersa e corta (SINE coas siglas en inglés), que orixinalmente perturbara a datación facendo moito mais antiga. A aparición deste elemento é inusual debido a súa escasez no xenoma de Petromyzon marinus, o cal presenta unha maior cantidade de secuencia repetidas dispersas e largas (LINE coas siglas en inglés) retrotransposóns e transposóns de DNA (Chalopin et al., 2015.) ao contrario que en humanos, onde son moito máis abundantes (Dunker et al., 2017). En canto a unión ao ERV, estes son elementos non autónomos que precisan da reverso transcriptasa doutro elemento para replicarse (Deininger, 2011; Göke & Ng, 2016).

A súa vez, a secuencia de CM021489 presenta una antigüidade moi elevada, cunha datación de 277 millóns de anos. Aínda así, o estado das súas LTRs encóntrase excepcionalmente deteriorado, polo que non pode asegurarse completamente a exactitude de inserción mediante a comparación das LTRS. Sen embargo, a súa datación si que supera a antigüidade da maioría dos ERVs datados ata o momento. Para confirmar ou descartar esta datación, é necesario cambiar o método de empregado, xa que a comparación de LTRs presenta este

tipo de problemas cando os elementos son especialmente antigos debido ao ratio de mutacións que presentan, facendoo máis recomendable para datacións de insercións recentes. Así, con este método hai que prestar especial atención a procesos como a conversión xénica (Kijima & Innan, 2010), os procesos de recombinación (Stoye, 2001) e as diferencias nos ratios de mutación neutra e o ratio de evolución vírica (Aiewsakun & Katzourakis, 2015). De este modo, a data de 277 ma pode estar influenciada polo uso do ratio de evolución neutra específico para peixes (N=2.2*10-9). Un método máis fiable para lograr unha data de inserción mínima consiste na búsqueda de ortólogos. Debido a baixa probabilidade de que dúas insercións se localicen no mesmo espazo en especies diferentes, a aparición dun ortólogo indica que o antepasado común de ambas especies xa presentaba esa inserción retroviral, sendo anterior ao evento de especiación (Lee et al., 2013).

```
##gff-version 3
##sequence-region seq0 1 18001
# CM021489.1:158000-176000 Petromyzon marinus isolate kPetMar1 chromosome 49, whole genome shotgun sequence
seq0    LTRharvest      repeat_region   3815    13988   .       ?       .       ID=RepeatReg0
seq0    LTRharvest      LTR_retrotransposon     3819    13984   .       ?       .       ID=LTRret0;Parent=RepeatReg0
seq0    LTRharvest      long_terminal_repeat    3819    3945    .       ?       .       ID=LTR0;Parent=LTRret0
seq0    LTRharvest      long_terminal_repeat    13851   13984   .       ?       .       ID=LTR1;Parent=LTRret0
seq0    LTRharvest      target_site_duplication 3815    3818    .       ?       .       ID=TSD0;Parent=RepeatReg0
seq0    LTRharvest      target_site_duplication 13985   13988   .       ?       .       ID=TSD1;Parent=RepeatReg0
```

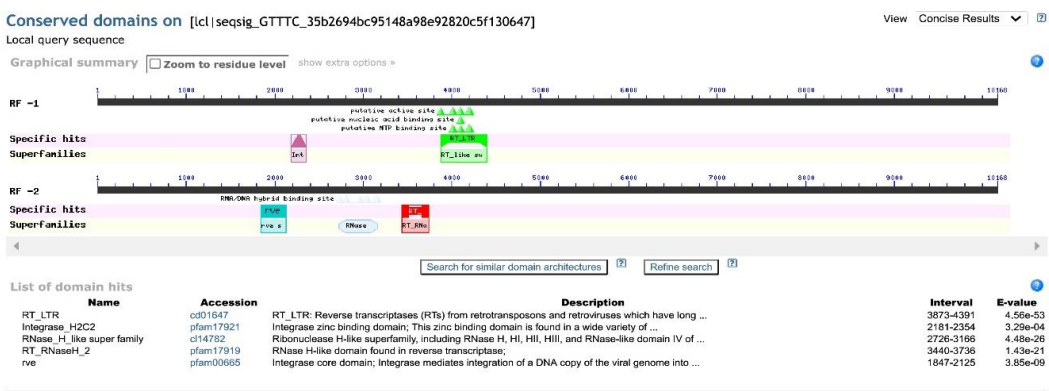


Ilustración 4: Distancia e dominios de CM021489. A datación extremadamente antiga, aínda que puidese deberse a un ERV realmente basal, pode estar relacionada con dificultades do método usado para datar elementos tan antigos

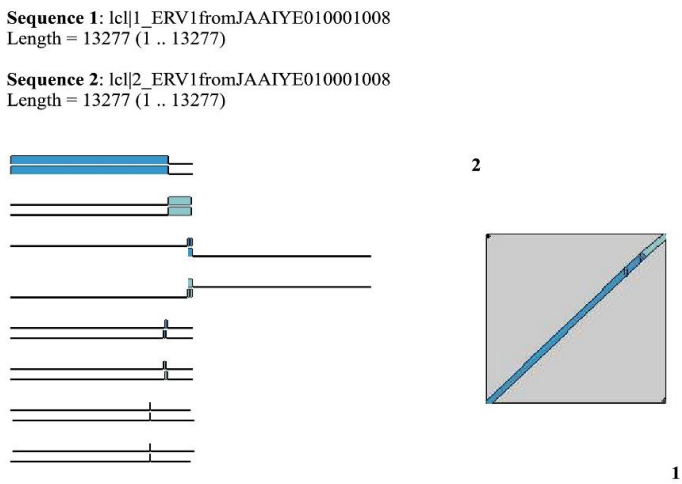


Ilustración 5: Blast2 do ERV from JAAIYE010001008 sobre si mesmo

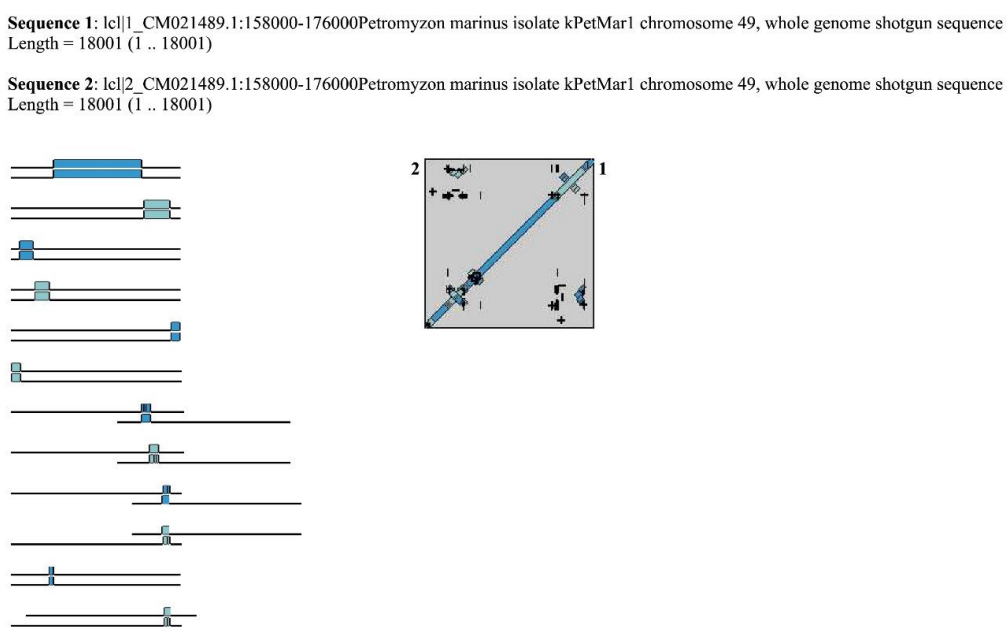


Ilustración 6: Blast2 do ERV de CM021489 sobre si mesmo. A diferencia con JAAIYE010001008 indica un deterioro considerable nas súas LTRs

Todos os ERVs, coa excepción de JAAIYE010001207 contan con dominios conservados de reverso transcriptasa (RT), ribonuclesa (RNasa) e integrasa.

A reverso transcriptasa de cada inserción foi identificado e usada para reconstruír as relacións filoxenéticas destas insercións. Para isto usouse o método *Neighbour-Joining*. A súa vez, a relación das secuencia aminoacídicas foi analizada mediante o método de máxima parsimonia.

En canto a filoxenia, o análise das árbores construídas en base as secuencias aminoacídicas (ilustración 7) e nucleotídicas (ilustración 8) das RTs pertencentes as insercións, indica que as analizadas neste traballo pertencen a dúas liñaxes distintas de ERVs, distinguíndose pola diferencia en tres aminoácidos do grupo de CM021489,CM021441 e JAAIYE01000724 con respecto á outra liñaxe. Hai certa discordancia con JAAIYE010001210 e o resto do seu grupo, debéndose a que presenta un dos tres diferentes aminoácidos do grupo anterior. Os valores de apoio das ramas que xustifican esta interpretación son altos, superiores ao 70% nas arbores das secuencias nucleótidicas, e relativamente altos nas aminoacídicas, entre un 40% e 50%. Sen embargo, os valores de apoio con respecto as secuencias aminoacídicas son realmente baixos para o segundo grupo, o que impide unha clara diferenciación dos integrantes do mesmo.

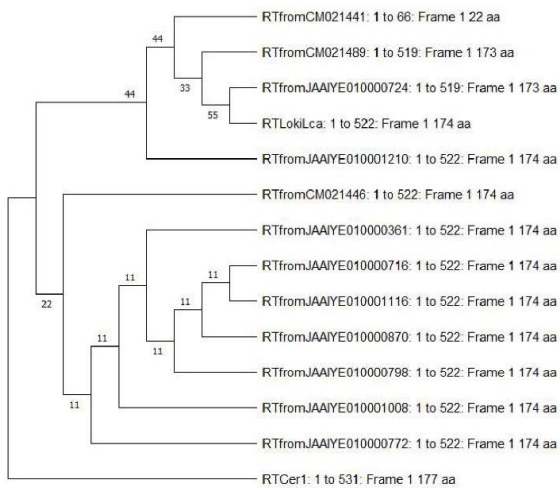


Ilustración 7: Relacións filoxenéticas das secuencias aminoacídicas dos ERVs analizados. A árbore filoxenética reconstruíuse a partir das secuencias aminoacídicas pertencentes as RTs dos ERVs analizados mediante o método de máxima parsimonia

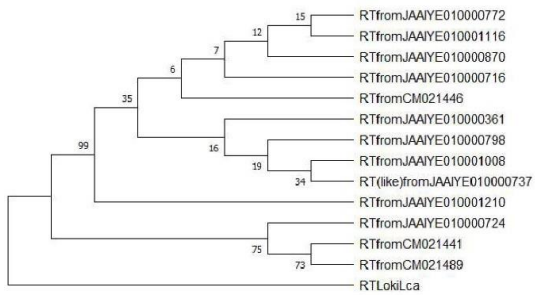


Ilustración 8: Relacións filoxenéticas das secuencia nucleotídicas dos ERVs analizados. A historia evolutiva reconstruíuse mediante o método de Neighbour-Joining. Poden apreciarse dous grupos ben definidos cuns valores de apoio altos (superiores a 70%)

5.1.-Conclusións

1.-Foron localizadas 15 inserción de ERVs no xenoma de *Petromyzon marinus*. Coa excepción de JAAIYE010001207, todas poseen dominios conservados de reverso transcriptasa, ribonucleasa e integrasa.

2.- A maior parte dos ERVs localizados pertencen a inserción recentes, producidas hai entre 100.000 anos e 31 millóns de anos.

3.-O análise das relacións filoxenéticas destes ERVs indica a existencia de dúas liñaxes claramente diferenciadas (valores de apoio bootstrap do 99%).

4.-Traballos anteriores sobre os ERV de *Petromyzon marinus* relacionaron estas inserción coa habitual transmisión horizontal acontecida nestes elementos entre especies distintas(Chalopin *et al*., 2015.; Escalera-Zamudio & Greenwood, 2016; Gifford & Tristem, 2003; Xu *et al*., 2018), descartando así a posibilidade de que a posición evolutivamente basal desta especie estivese relacionada cunha especial antigüidade dos ERVs encontrados no seu xenoma. Se ben na maior parte das ERVs analizadas isto é demostrado polas datas de inserción, o caso de CM021489 posibilita a existencia de fósiles de elemento retrovirais moito máis basais no xenoma de *Petromyzon marinus.* Unha investigación máis específica sobre estes elementos será necesaria para achegar máis luz a complexa historia evolutiva dos ERVs, co obxetivo de aumentar a compresión sobre os retrovirus, a súa relación en enfermidades como o cáncer e a sua expresión en diversos

tecidos tanto normais como canceríxenos (Bustamante Rivera *et al*., 2018; Wang-Johanning *et al*., 2003)

5.2-Conclusiones

1.-Fueron localizadas 15 inserciones en el genoma de *Petromyzon marinus*. Con la excepción de JAAIYE010001207, todas poseen dominios conservados de reverso transcriptasa, ribonucleasa y integrasa.

2.-La mayor parte de los ERVs localizados pertenecen a inserciones recientes, producidas hace entre 100.000 años y 31 millones de años.

3.-El análisis de las relaciones filogenéticas de estos ERVs indica la existencia de dos linajes claramente diferenciados (valores de apoyo Bootstrap del 99%).

4.-Trabajos anteriores sobre los ERVs de *Petromyzon marinus* relacionaron estas inserciones con la habitual transmisión horizontal de estos elementos entre individuos de distinta especie (Chalopin *et al.*, 2015; Escalera-Zamudio & Greenwood, 2016; Gifford & Tristem, 2003; Xu *et al*., 2018), descartando así la posibilidad de que la posición evolutivamente basal de esta especie estuviese relacionada con una especial antigüedad de los ERVs encontrados en su genoma. Si bien la mayor parte de las ERVs analizados esto se demuestra por las datas de inserción, el caso de CM021489 posibilita la existencia de fósiles de elementos retrovirales mucho más basales en el genoma de *Petromyzon marinus*. Una investigación más específica sobre estos elementos será necesaria para aportar más luz a la compleja historia evolutiva de los ERVs, con el objetico de aumentar la comprensión sobre los retrovirus, su relación con enfermedades como el cáncer y a su expresión tanto en tejidos normales como cancerígenos (Bustamante Rivera *et al*., 2018; Wang-Johanning *et al.*, 2003)

5.3-Conclusions

1.-15 insertions were located in the genome of Petromyzon marinus. Except for JAAIYE010001207, all of them had conserved reverse transcriptase, ribonuclease and integrase domains.

2.-Most of the ERVs located belong to recent insertions, produced between 100,000 years ago and 31 million years ago.

3.-The analysis of the phylogenetic relationships of these ERVs indicates the existence of two clearly differentiated lineages (Bootstrap support values of 99%).

4.-Previous work on Petromyzon marinus ERVs related these insertions to the usual horizontal transmission of these elements between individuals of different species (Chalopin *et al*., 2015; Escalera-Zamudio & Greenwood, 2016; Gifford & Tristem, 2003; Xu *et al*., 2018), thus ruling out the possibility that the evolutionarily basal position of this species was related to a particular age of the ERVs found in its genome. While for most of the ERVs analysed this is demonstrated by the insertion dates, the case of CM021489 makes possible the existence of much more basal retroviral element fossils in the genome of Petromyzon marinus. More specific research on these elements will be needed to shed more light on the complex evolutionary history of ERVs, with the aim of increasing understanding of retroviruses, their relationship with diseases such as cancer and their expression in both normal and cancerous tissues (Bustamante Rivera *et al*., 2018; Wang-Johanning *et al*., 2003).

6.-Bibliografía:

Aiewsakun, P., & Katzourakis, A. (2015). Endogenous viruses: Connecting recent and ancient viral evolution. Virology, 479–480, 26–37. https://doi.org/10.1016/J.VIROL.2015.02.011

Alzohairy, A. M., & Kulkarni-Kale, U. (1991). Mfold©: RNA modeling program. http://www.mbio.ncsu.edu/bioedit/page2.html. https://bioedit.software.informer.com/download/

Bustamante Rivera, Y. Y., Brütting, C., Schmidt, C., Volkmer, I., & Staege, M. S. (2018). Endogenous retrovirus 3 − History, physiology, and pathology. *Frontiers in Microbiology*, *8*, 2691. https://doi.org/10.3389/fmicb.2017.02691

Chalopin, D., Naville, M., Plard, F., Galiana, D., & Volff, J.-N. (2015). Comparative analysis of transposable elements highlights mobilome diversity and evolution in vertebrates. *Genome Biology and Evolution, 7*(2), 567−580. https://doi.org/10.1093/gbe/evv005

Deininger, P. (2011). Alu elements: Know the SINEs. *Genome Biology, 12*, 236. https://doi.org/10.1186/GB-2011-12-12-236

Dunker, W., Zhao, Y., Song, Y., & Karijolich, J. (2017). Recognizing the SINEs of infection: Regulation of retrotransposon expression and modulation of host cell processes. *Viruses, 9*(12), 386. https://doi.org/10.3390/V9120386

Ellinghaus, D., Kurtz, S., & Willhoeft, U. (2008). LTRharvest, an efficient and flexible software for de novo detection of LTR retrotransposons. BMC Bioinformatics, 9. https://doi.org/10.1186/1471-2105-9-18. http://tools.bat.infspire.org/ltrharvester/

Escalera-Zamudio, M., & Greenwood, A. D. (2016). On the classification and evolution of endogenous retrovirus: Human endogenous retroviruses may not be 'human' after all. *APMIS*, *124*(1–2), 44–51. https://doi.org/10.1111/apm.12489

Gifford, R., & Tristem, M. (2003). The evolution, distribution and diversity of endogenous retroviruses. *Virus Genes*, *26*(3), 291–315. https://doi.org/10.1023/A:1024455415443

Göke, J., & Ng, H. H. (2016). CTRL+INSERT: Retrotransposons and their contribution to regulation and innovation of the transcriptome. *EMBO Reports*, *17*(8), 1131–1144. https://doi.org/10.15252/embr.201642743

Hayward, A. (2017). Origin of the retroviruses: When, where, and how? *Current Opinion in Virology, 25*, 23–27. https://doi.org/10.1016/j.coviro.2017.06.006

Hayward, A., Cornwallis, C. K., & Jern, P. (2015). Pan-vertebrate comparative genomics unmasks retrovirus macroevolution. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(2), 464–469. https://doi.org/10.1073/pnas.1414980112

Herniou, E., Martin, J., Miller, K., Cook, J., Wilkinson, M., & Tristem, M. (1998). Retroviral diversity and distribution in vertebrates. *Journal of Virology*, *72*(7), 5955–5966. https://doi.org/10.1128/JVI.72.7.5955-5966.1998

Kijima, T. E., & Innan, H. (2010). On the estimation of the insertion time of LTR retrotransposable elements. *Molecular Biology and Evolution*, *27*(4), 896–904. https://doi.org/10.1093/MOLBEV/MSP295

Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., Mcgettigan, P. A., McWilliam, H., Valentin, F., Wallace, I. M., Wilm, A., Lopez, R., Thompson, J. D., Gibson, T. J., & Higgins, D. G. (2007). Clustal W and Clustal X version 2.0. Bioinformatics, 23(21), 2947–2948. https://doi.org/10.1093/bioinformatics/btm404

Lee, A., Nolan, A., Watson, J., & Tristem, M. (2013). Identification of an ancient endogenous retrovirus, predating the divergence of the placental mammals. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *368*(1626), 20120503. https://doi.org/10.1098/RSTB.2012.0503

Magiorkinis, G., Katzourakis, A., & Lagiou, P. (2017). Roles of endogenous retroviruses in early life events. *Trends in Microbiology, 25*(11), 876–877. https://doi.org/10.1016/j.tim.2017.09.002

Naville, M., & Volff, J.-N. (2016). Endogenous retroviruses in fish genomes: From relics of past infections to evolutionary innovations? *Frontiers in Microbiology*, *7*, 1197. https://doi.org/10.3389/fmicb.2016.01197

Osório, J., & Rétaux, S. (2008). The lamprey in evolutionary studies. *Development Genes and Evolution*, *218*(5), 221–235. https://doi.org/10.1007/S00427-008-0208-1

Smith, J. J., Kuraku, S., Holt, C., Sauka-Spengler, T., Jiang, N., Campbell, M. S., Yandell, M. D., Manousaki, T., Meyer, A., Bloom, O. E., Morgan, J. R., Buxbaum, J. D., Sachidanandam, R., Sims, C., Garruss, A. S., Cook, M., Krumlauf, R., Wiedemann, L. M., Sower, S. A., … Li, W. (2013). Sequencing of the sea lamprey (*Petromyzon marinus*) genome provides insights into vertebrate evolution. *Nature Genetics*, *45*(4), 415–421. https://doi.org/10.1038/ng.2568

Stoye, J. P. (2001). Endogenous retroviruses: Still active after all these years? *Current Biology*, *11*(22), R914–R916. https://doi.org/10.1016/S0960-9822(01)00553-X

Sudhir Kumar, Glen Stecher, Michael Li, Christina Knyaz, and Koichiro Tamura (2018) MEGA X: Molecular Evolutionary Genetics Analysis across computing platforms. Molecular Biology and Evolution 35:1547-1549. https://www.megasoftware.net/

Tatusova, T. A., & Madden, T. L. (1999). BLAST 2 Sequences, a new tool for comparing protein and nucleotide sequences. FEMS Microbiology Letters, 174(2), 247–250. https://doi.org/10.1111/j.1574-6968.1999.tb13575.x http://polyp.biochem.uci.edu/blast/wblast2.html

Wang, J., & Han, G. Z. (2021). A sister lineage of sampled retroviruses corroborates the complex evolution of retroviruses. *Molecular Biology and Evolution*, *38*(3), 1031–1039. https://doi.org/10.1093/molbev/msaa272

Wang-Johanning, F., Frost, A. R., Jian, B., Azerou, R., Lu, D. W., Chen, D.-T., & Johanning, G. L. (2003). Detecting the expression of human endogenous retrovirus E envelope transcripts in human prostate adenocarcinoma. *Cancer*, *98*(1), 187–197. https://doi.org/10.1002/CNCR.11451

Xu, X., Zhao, H., Gong, Z., & Han, G.-Z. (2018). Endogenous retroviruses of non-avian/mammalian vertebrates illuminate

diversity and deep history of retroviruses. *PLOS Pathogens*, *14*(6), e1007072. https://doi.org/10.1371/journal.ppat.1007072

7.-Anexo I

Anexo I: Datos das ERVs analizadas. Posición no scaffold: intervalo comprendido entre o extremo inicial da LTR5' e o extremo final da LTR3'. LTR5': intervalo comprendido pola LTR5'. LTR3': intervalo comprendido pola LTR3'. TSD: target site duplication. Distancia K-2P: distancia de Kimura de dous parámetros

JAAIYE010001210	JAAIYE010001207	JAAIYE010000870	JAAIYE010001116	JAAIYE010000798	JAAIYE010000716	JAAIYE010000772	JAAIYE010000361	JAAIYE010001008	**Posición no scaffold (inicio-fin)**
Petromyzon marinus isolate kPetMar1 scaffold 790 arrow ctg1, whole genome shotgun sequence	Petromyzon marinus isolate kPetMar1 scaffold 787 arrow ctg1, whole genome shotgun sequence	Petromyzon marinus isolate kPetMar1 scaffold 459 arrow ctg1, whole genome shotgun sequence	Petromyzon marinus isolate kPetMar1 chromosome 6, whole genome shotgun sequence	Petromyzon marinus isolate kPetMar1 chromosome 74 S74unloc.7 whole genome shotgun sequence	Petromyzon marinus isolate kPetMar1 scaffold 255 arrow ctg1 whole genome sequence	Petromyzon marinus isolate kPetMar1 scaffold 345 arrow ctg1 whole genome shotgun sequence	Petromyzon marinus isolate kPetMar1 scaffold 133 arrow ctg1 1, whole genome sequence	Petromyzon marinus isolate kPetMar1 chromosome 47, whole genome sequence	**Descrición**
58000-81000	35000-55000	116964-149833	1016292-1039159	16906-36051	106000-126000	12646-36512	215530-239360	11139000-11153000	**Inserción proviral**
5252-5540	12-992	18069-19988	8161-10096	3714-4705	6481-8321	4353-6326	8160-9871	6-240	**LTR5'**
18017-18286	4415-5398	30934-32852	20941-22834	16302-17273	18208-19994	17150-19141	19396-211333	13049-13282	**LTR3'**
AACA	CAGA	CGGC	TCTA	ATAA	TTCT	TTGA	CCCC	AGAT	**TSD**
0,120	0,114	0.,026	0,001	0,092	0,026	0,004	0,017	0,013	**Distancia K-2P**
27,27	25,91	0,59	0,11	20,95	5,84	0,82	3,95	3	**Idade (ma)**