



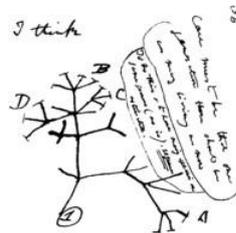
Grao en Bioloxía

Memoria do Traballo de Fin de Grao

Tráfico de ADN entre o xenoma mitocondrial e o nuclear: Estudo das insercións de NUMTs no xenoma secuenciado de *Podarcis muralis*

Tráfico de DNA entre el genoma mitocondrial y el nuclear: Estudio de las inserciones de NUMTs en el genoma secuenciado de *Podarcis muralis*

DNA traffic between mitochondrial and nuclear genomes: Study of NUMT insertions in the sequenced genome of *Podarcis muralis*



I think
98
Then between A & B. various
size of deletion. C & B. The
first prediction, B & D
rather greater distance
than former would be
formed. - binary deletion

Alexis Rodríguez Caridad

Curso: 2021-2022. Convocatoria: Xullo

Director(es) Académico(s): Horacio Naveira Fachal

Índice

Resumo.....	3
Introdución.....	5
Sobre o organismo de estudo	5
Sobre as rexións do ADN mitocondrial.....	5
Sobre o tráfico de secuencias.....	6
Sobre as secuencias transpoñibles máis frecuentes	6
Obxectivos.....	7
Material e métodos.....	7
Obtención de datos	7
Aliñamentos.....	7
Comparación entre especies.....	8
Estudo xeral dos NUMTs	8
Estudo de NUMTs e secuencias concretas	8
Resultados.....	8
Comparación entre especies.....	8
Estudo xeral dos NUMTs	11
Aliñamento das rexións 5620242 e 6758295 do cromosoma 6.....	12
Caracterización das partes da rexión de control de <i>P. muralis</i>	13
Predición da estrutura secundaria.....	15
Discusión	17
Comparativa entre especies.....	17
Orixe dos NUMTs nos cromosomas, peculiaridades	17
Estudo dos NUMTs do cromosoma 6	18
Importancia dos NUMTs na evolución	19
Conclusións.....	19
Bibliografía.....	21

Resumo

O tráfico continuo de ADN entre a mitocondria e o núcleo é unidireccional e xera numerosas copias nucleares de xenes mitocondriais (NUMTs) con discutida relevancia evolutiva. A día de hoxe existen moitas incógnitas rodeando estas secuencias, dende os procesos que as orixinan ate o xeito no que se comportan. Neste traballo realizamos numerosos aliñamentos empregando o xenoma secuenciado de *Podarcis muralis* para estudarmos a fondo estes elementos, analizando a súa importancia nos procesos evolutivos, así como a predominancia dalgunhas rexións da mitocondria nos cromosomas e discutimos a súa utilidade á hora de caracterizar elementos transpoñibles. Atopáronse un total de 23 NUMTs cun tamaño promedio de 783pb localizando o 95% destes en rexións non codificantes. Aproximadamente un 82% do xenoma mitocondrial está representado nos NUMTs, exhibindo un marcado nesgo na orixe destas insercións, con case a metade provindo da rexión de control.

Palabras chave: *Transposición, Podarcis siculus, Lacertidae*, Duplicación, ADNmt.

Resumen

El tráfico continuo de ADN entre la mitocondria y el núcleo es unidireccional y genera numerosas copias nucleares de genes mitocondriales (NUMTs) con discutida relevancia evolutiva. A día de hoy existen muchas incógnitas rodeando estas secuencias, desde los procesos que las originan hasta la manera en la que se comportan. En este trabajo realizamos numerosos alineamientos empleando el genoma secuenciado de *Podarcis muralis* para estudiar a fondo estos elementos, analizando su importancia en los procesos evolutivos, así como la predominancia de algunas regiones de la mitocondria en los cromosomas y discutimos su utilidad a la hora de caracterizar elementos transponibles. Un total de 23 NUMTs fueron encontrados, con un tamaño promedio de 783pb localizando el 95% de estos en regiones no codificantes. Aproximadamente un 82% del genoma mitocondrial está representado en los NUMTs, exhibiendo un marcado sesgo en el origen de estas inserciones, con casi la mitad proviniendo de la región de control.

Palabras clave: *Transposición, Podarcis siculus, Lacertidae*, Duplicación, ADNmt

Abstract

The continuous DNA traffic between mitochondria and the nucleus is unidirectional and generates numerous nuclear copies of mitochondrial genes (NUMTs) with controversial evolutionary relevance. To date there are many unknowns surrounding these sequences, from the processes that give rise to them to the way in which they behave. In this work we performed numerous alignments using the sequenced genome of *Podarcis muralis* to study these elements in depth, analyzing their importance in evolutionary processes, as well as the predominance of some mitochondrial regions in chromosomes and discuss their usefulness in characterizing transposable elements. A total of 23 NUMTs were found, with an average size of 783 bpb, 95% of which were located in non-coding regions. Approximately 82% of the mitochondrial genome is represented in the NUMTs, exhibiting a marked bias in the origin of these insertions, with almost half coming from the control region.

Key words: Transposition, *Podarcis siculus*, *Lacertidae*, Duplication, ADNmt

Introdución

Sobre o organismo de estudo

Podarcis muralis é unha especie que nos últimos anos está a recibir moita atención ó ser empregado como un organismo modelo, debido a que amosa altas densidades de poboación, son doados de capturar, resulta sinxelo manipularlos, viven en hábitats antropomorfizados e toleran experimentos bastante agresivos (Camargo *et al.*, 2010).

Sobre as rexións do ADN mitocondrial

O ADN mitocondrial (ADNmt) desta especie é bastante coñecido, xa que se utiliza en numerosos estudos para determinar fenómenos de radiación, invasións ou mesmo para caracterizar razas xeográficas; con todo, estes traballos non adoitan explorar todo o ADNmt, senón soamente determinadas rexións que funcionan como reloxo molecular, como pode ser a correspondente ó citocromo b (14156-15298) (Schulte *et al.* 2012; Santos *et al.* 2019; Michaelides *et al.* 2013). Estruturalmente, o xenoma mitocondrial de *P. muralis* é semellante ó doutros vertebrados a pesares do seu rápido ratio de evolución (Brown *et al.* 1979), tratándose dunha secuencia de 17311 pares de bases (pb) que contén xenes relacionados coa cadea de transporte electrónico, os ARN transferentes (ARNt) de cada aminoácido e unha rexión de control (D-Loop) dentro da maior porción non codificante (NCR) do xenoma.

A NCR está situada entre os ARNt da fenilalanina e da prolina. (15440-17311) e pode ser dividida en varias partes ben diferenciadas: 1 Repeticións en tándem que varían en número segundo a especie, 2 Grandes segmentos cun baixo grao de conservación, 3 Bloques de secuencias conservadas (CSB) separadas entre sí por rexións máis variables e, 4 Secuencias asociadas á terminación (TAS)(Brehm *et al.* 2003). Agás nas repeticións en tándem (que aparecen de xeito contiguo), o resto de partes van alternándose, aínda que o patrón que forman está bastante conservado (Nicholls e Minczuk, 2014).

O D-Loop é unha rexión non codificante caracterizada pola formación dunha triple cadea de ADN debido á inclusión dunha pequena cadea coñecida como ADN 7S, vai dende os arredores do CSB1 ate o segundo TAS e crese que participa na regulación da duplicación do material xenético durante a reprodución. A súa función parece ser a de actuar como un intermediario da replicación da cadea pesada, xa que a orixe de replicación desta coincide co extremo 5' do ADN 7S (Nicholls e Minczuk, 2014).

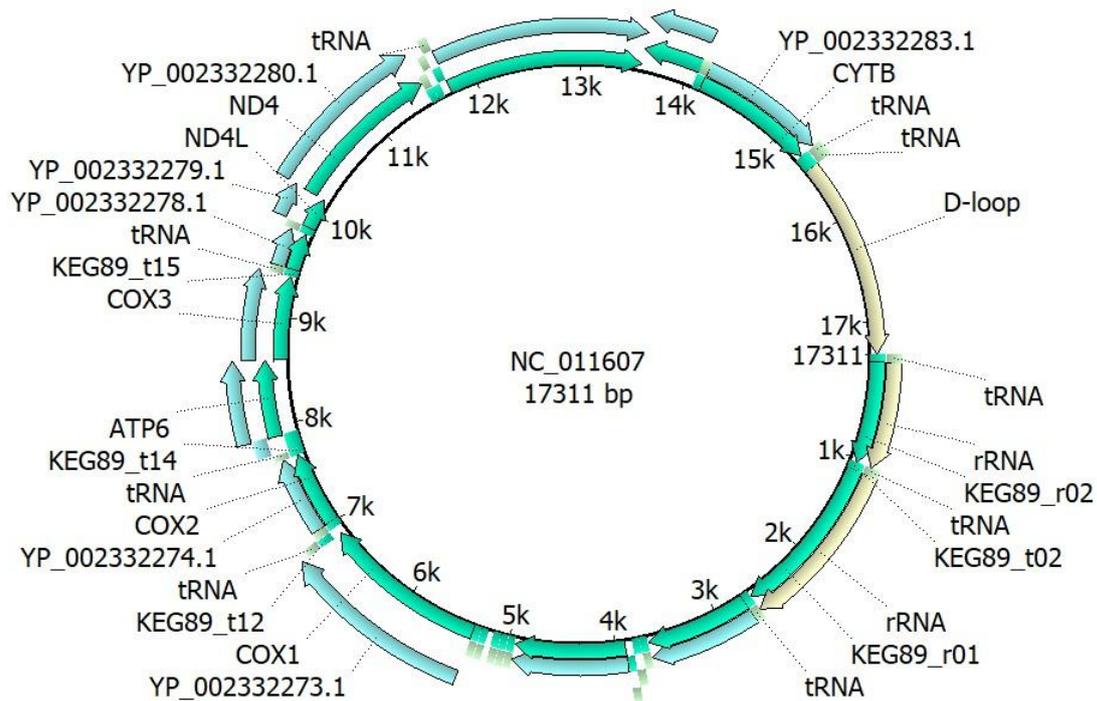


Figura 1: Organización do ADN mitocondrial de *P. muralis*, as frechas indican o sentido da lectura nas secuencias. As frechas azuis e verdes representan a colocación dos principais xenos e ARNt, mentres que as grises representan a rexión de control e os ARNr.

Sobre o tráfico de secuencias

Os NUMTs (Nuclear paralog of mtADN genes) son secuencias cromosómicas orixinarias da mitocondria das que se ten pouca información. Aínda que os mecanismos polos cales acontecen estes eventos son descoñecidos, sábese que este tráfico de secuencias é exclusivamente unidireccional, da mitocondria ó núcleo (Adams e Palmer, 2003), probablemente porque o máis probable é que este tráfico implique a liberación ó citoplasma do contido xenético da mitocondria.

Estas secuencias resultan de gran interese, xa que están envoltas en numerosas incógnitas como os procesos que median as transferencias ou se existe algún tipo de regulación (Blanchard e Lynch, 2000). Fóra do eido da xenética molecular tamén poden aportar información sobre a orixe dunha especie ou variedade, así como revelar eventos de hibridación entre especies (Podnar *et al.*, 2007).

Sobre as secuencias transpoñibles máis frecuentes

Os elementos curtos intercalados (SINEs) son pequenos fragmentos de ADN que aparecen en gran número ó longo do xenoma de moitos organismos, a diferenza doutros elementos transpoñibles estes non teñen capacidade de movemento propia, senón que “parasitan” a elementos longos intercalados (LINEs); ámbolos dous son retrotransposóns. Un SINE non ten capacidade

propia para traspoñerse, senón que requiren da actuación da retrotransposasa codificada por un LINE (Kramerov e Vassetzky, 2011).

Obxectivos

Neste traballo buscamos estudar os NUMTs presentes en *Podarcis muralis*, a súa orixe, número, tamaño e localización no xenoma con respecto ás rexións codificantes. Tamén tentaremos relacionalos con outras secuencias/elementos e valorar a súa importancia.

Material e métodos

Obtención de datos

Descargamos a secuencia do ADN mitocondrial tanto de *Podarcis muralis* (NC_011607.1) como de *Podarcis siculus* (NC_011609.1) dende GeneBank genomes nos formatos FASTA (.fasta) e GenBank (.gb), trátase de secuencias de 17Kb. Usándoas como *query*, realizamos un BLAST (*Basic Local Alignment Search Tool*) de alta similitude (*megablast*) sobre a base de datos *refseq_genomes* e quedámonos cos resultados pertencentes ós cromosomas caracterizados, dos que imos obter a táboa de *hits* e as secuencias. Repetimos este proceso para o megablast discontiguo.

Aliñamentos

Os aliñamentos deste traballo foron realizados grazas ó programa UGene versión 43.0, mediante o procedemento Smith-Waterman (SSE2, *Gap penalty: Open=10, Extension=1*) no caso de seren aliñamentos por pares, mentres que no caso de aliñamentos con múltiples secuencias o algoritmo escollido foi MUSCLE (*default, stable*). Para os aliñamentos con máis secuencias involucradas efectuáronse correccións manuais sobre os resultados. Aliñouse o xenoma mitocondrial cos *hits* que se atopan na mesma situación dentro dos cromosomas, no caso dos cromosomas 4 e 6, un dos resultados iniciais foi complementado cunha mostra máis grande da rexión cromosómica na que estaban localizados para comprobar continuidade:

- Cromosoma 4: os *hits* correspondentes á rexión 105456864 foron aliñados entre si, co mitocondrial e ca secuencia cromosómica ± 100 pb.
- Cromosoma 6: aliñamos todos os *hits* pertencentes ás rexións 5620242 e 6758295 sobre as secuencias correspondentes á rexión onde están localizadas ± 1000 pb. Os ocios atopados entre as rexións duplicadas do cromosoma 6 foron clasificados en macrogaps e minigaps segundo fosen maiores ou menores a 50pb respectivamente. As secuencias correspondentes ós ocios foron empregadas como *query* para facer un megablast sobre a base de datos do NCBI (*National Center for Biotechnology Information*) *nucleotide collection nr/nt*. Descargamos as táboas de *hits* resultantes.

Comparación entre especies

As táboas de *hits* obtidas foron empregadas na elaboración dun documento en microsoft excel (versión 16.0) que incluía a cada unha delas nunha folla, ademais e seguindo co estudo de Podnar *et al.* (2007) construímos unha nova folla na que elaboramos unha táboa comparativa entre os NUMTs de ambas especies empregando primariamente os resultados do megablast, incluíndo os do megablast discontinuo, unicamente de non atoparmos equivalencias. Para comparar se existían diferenzas significativas realizamos un test de rangos signados de Wilcoxon para datos aparellados, cun intervalo de confianza do 95% e supoñendo que a distribución non é normal (o programa empregado foi R v4.2.1).

Por último tamén realizamos un aliñamento por pares entre o ADNmt de *P. muralis* e *P. siculus*.

Estudo xeral dos NUMTs

Empregando o xenoma mitocondrial de *P. muralis* facemos un BLAST en Ensembl para obter representacións gráficas que amosan os *hits* dentro da mitocondria e dos cromosomas (a base de datos escollida foi *Genomic secuencia*, todos os demais valores mantivéronse por defecto agás o *repeat masker* que foi desactivado). Dentro destes buscáronse as rexións onde se localizaban os *hits* para anotar se pertencían a xenes ou non.

Estudo de NUMTs e secuencias concretas

Existen varios *hits* idénticos de 34 pb repartidos por varios cromosomas, empregamos a súa secuencia para facer un megablast (base de datos *nucleotide collection nr/nt*). Para estudar máis a fondo este fragmento de ADN, empregamos o programa RNAeval web server que nos permite predicir a estrutura secundaria para ver se o lugar onde se sitúa está relacionado coas súas características, para isto utilizamos como *query* un fragmento que inclúe a nosa secuencia ± 300 pb (16123-16756).

Identificamos os distintos compoñentes da rexión de control empregando como referencia o pertencente a *Lacerta dugesii* anotado por Brehm *et al.* (2003). Para isto realizamos un aliñamento entre o xenoma mitocondrial de *P. muralis* e as secuencias de *L. dugesii* que eles proporcionan (AY147872.1), buscando e caracterizando as distintas partes guiándonos pola a figura 4 que aparece no seu artigo.

Resultados

Comparación entre especies

As táboas de *hits* obtidas de *P. muralis* e *P. siculus* foron comparadas, creando unha nova (táboa 1). Todos os NUMTs que atopamos nunha especie teñen o seu equivalente na outra (ocupan a mesma posición dentro dos cromosoma), ben sexa na táboa do BLAST de alta similitude ou no descontiguo. A única

excepción é un *hit* do cromosoma 4 onde non atopamos equivalente directo, mais si outros NUMTs que ocupan a mesma posición.

A través desta comparativa podemos apreciar como existen uns NUMTs máis semellantes ó ADNmt de *P. siculus* e outros máis semellantes ó ADNmt de *P. muralis*, máis as diferenzas nas porcentaxes de identidade son reducidas (2,908 a máis grande por parte de *P. siculus* e 4,480 por parte de *P. muralis*). En cómputo global o ADNmt de *P. muralis* é máis semellante, aínda que as diferenzas non son significativas (figura 2), xa que, o test de Wilcoxon devolveunos un p-valor de 0.09563.

Ó aliñarmos ambos xenomas mitocondriais vemos que a equivalencia é moi alta (89% de identidade) sendo a diferenza máis notable a existencia dun oco entre a posición 15863 e 15904, contigua ás repeticións en tándem da rexión de control, máis non tendo a secuencia dunha destas.

	Ubicación do NUMT		% identidade P. muralis	% identidade P. siculus	Diferencia
Cromosoma 1	125268539	125268444	93,750	94,792	1,042
	125268491	125268448	90,909	92,308	1,399
Cromosoma 10	8334476	8334958	77,092	80,000	2,908
	34020092	34019625	75,309	75,820	0,511
Cromosoma 11	62006094	62007062	80,600	79,055	-1,545
Cromosoma 13	54237320	54237608	90,378	90,690	0,312
Cromosoma 14	10173512	10172769	79,221	77,821	-1,400
Cromosoma 17	7131504	7131144	85,319	84,746	-0,573
Cromosoma 18	27459061	27458923	80,723	81,275	0,552
Cromosoma 2	126962466	126962137	96,108	93,413	-2,695
	22293450	22292353	75,766	77,360	1,594
	28191581	28191949	88,095	87,399	-0,696
	28191830	28191863	100,000	100,000	0,000
Cromosoma 3	30103465	30103681	92,627	93,868	1,241
	30103429	30103498	95,714	92,000	-3,714
	37094743	37094357	80,303	78,426	-1,877
	105456864	105457857	91,046	88,531	-2,515
Cromosoma 4	105456867	105456824	100,000		-100,000
	56574735	56573857	87,401	86,682	-0,719
	56575673	56575070	79,734	81,219	1,485
	11931695	11931872	83,799	82,222	-1,577
Cromosoma 5	66082526	66082924	91,542	89,552	-1,990
	66083182	66083471	91,724	89,583	-2,141
	71797844	71797581	90,189	89,811	-0,378
	71797676	71797643	100,000	100,000	0,000
Cromosoma 6	5627463	5625052	80,590	80,560	-0,030
	5624678	5623248	79,336	80,669	1,333
	6759527	6760653	80,345	79,628	-0,717

	6758295	6759126	83,353	81,053	-2,300
	5621021	5620242	84,371	83,621	-0,750
	6765088	6765857	83,503	82,512	-0,991
	5621697	5621238	85,652	83,000	-2,652
	6763663	6764179	81,973	83,366	1,393
	6764441	6764868	82,851	80,723	-2,128
	30948652	30948872	86,425	87,783	1,358
	83643772	83644071	86,424	85,855	-0,569
	83647106	83647197	86,957	89,130	2,173
	83643987	83644020	100,000	100,000	0,000
Cromosoma 7	85631076	85630493	94,017	89,537	-4,480
Cromosoma 8	17715292	17715506	85,714	87,963	2,249
Cromosoma 9	55854119	55853859	90,038	89,922	-0,116
Cromosoma Z	50256337	50256733	80,000	77,970	-2,030

Táboa 1: Comparación dos ADN mitocondriais de *P. muralis* e *P. Siculus* cos NUMTs atopados nos cromosomas de *P. muralis*. Cómpanse os *hits* obtidos trala realización dun megablast, agás cando nunha das partes non se atopa correspondencia, recurriendo entón ós resultados do megablast discontiguo (en laranxa).

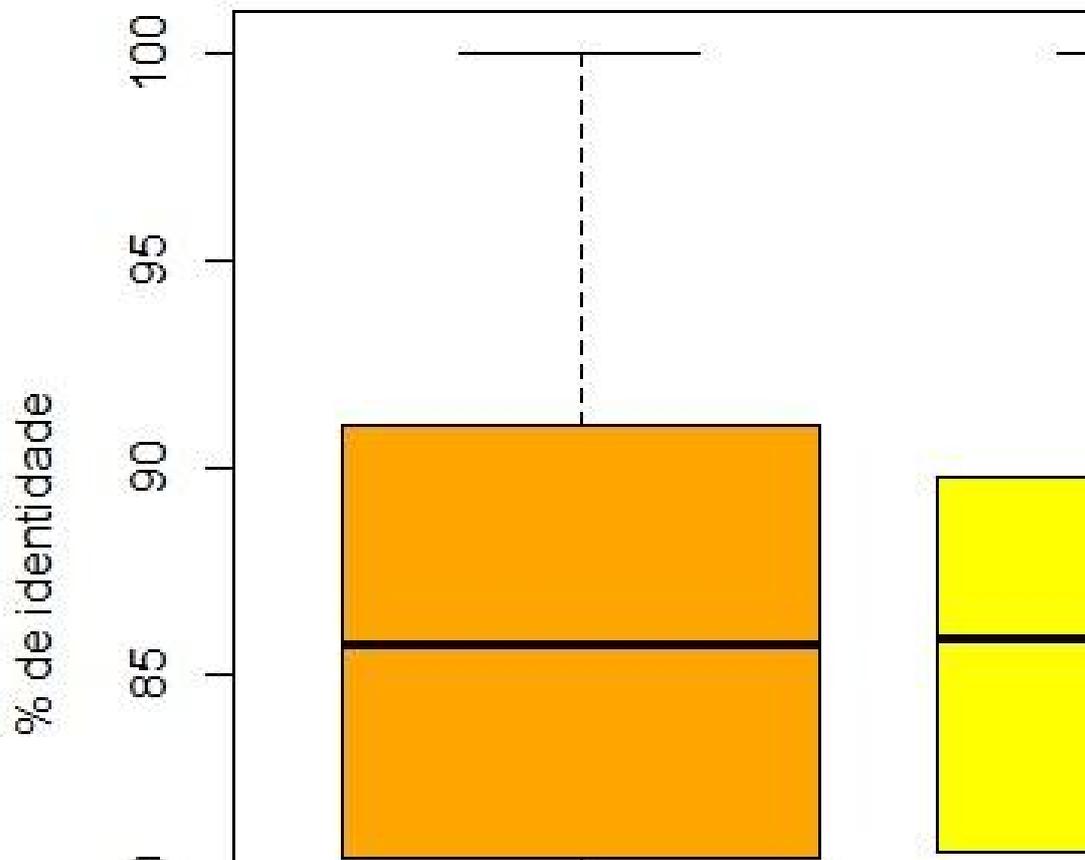


Figura 2: Boxplot comparado da porcentaxe de identidade dos *hits* de *P. muralis* e *P. siculus*. A liña negra grosa indica a mediana que é semellante en ambas especies. A

caixa principal abrangue os datos situados entre os percentiles 25 e 75%, sendo neste último onde *P. muralis* acada valores máis altos.

Estudo xeral dos NUMTs

O primeiro BLAST (megablast) deu como resultado 38 *hits*, pero só ocupan 23 lugares ben diferenciados. Teñen un tamaño relativamente variable, cunha lonxitude media de 783pb, sendo o máis longo de 5196pb e o máis curto de 179pb (táboa 2).

Localizáronse NUMTs en todos os cromosomas autosómicos, agás no 12, 15, 16 e 18, mais se empregamos os datos do BLAST descontiguo atopamos novos en 12 e 18. No que respecta ós cromosomas sexuais, no Z tamén existe un *hit*, e do W non existe información dispoñible. Non parece existir unha predisposición por ocupar uns cromosomas fronte a outros sendo os que máis NUMTs conteñen o 4 e o 2 con 3.

Dos 38 *hits* case a metade (16) están dentro da rexión de control (figura 3), ademais de dous que a tocan parcialmente, é dicir, pouco menos do 11% de todo o ADNmt contén máis do 42% dos *hits*, mais curiosamente, non atopamos ningún nas repeticións en tándem. Non foi posible atopar unha preferencia do mesmo tipo por inserirse nunhas rexións cromosómicas fronte a outras (figura 4). Cómpre destacar que as figuras 3 e 4 non coinciden exactamente coa táboa de hits, debido a que a base de datos empregada foi outra.

Todos os *hits* foron localizados en rexións interxénicas (56,52%) ou en intróns (39,13%), agás un do cromosoma 4 que contén un xene completo, que é mitocondrial, posto que o NUMT é continuo en toda a súa extensión (non ten insercións/deleccións).

Código do cromosoma	Cromosoma	% identidade	Lonxitude	Erros	Ocos	Inicio na mitocondria	Fin na mitocondria	Inicio no cromosoma	Fin no cromosoma	E valor	Puntuación	Ubicación en xenes
NC_041312.1	1	93,750	96	6	0	16371	16466	125268539	125268444	5,24E-31	145	Non
NC_041312.1	1	90,909	44	4	0	16357	16400	125268491	125268448	1,95E-05	60,2	Non
NC_041321.1	10	77,092	502	80	26	14993	15478	8334476	8334958	6,45E-65	257	Non
NC_041321.1	10	75,309	486	87	22	16148	16618	34020092	34019625	3,07E-48	202	>PPP1R12A: Subunidade reguladora da proteína fosfatasa 1 (intrón 1)
NC_041322.1	11	80,600	1000	152	21	40	1028	62006094	62007062	0	734	Non
NC_041324.1	13	90,378	291	26	2	476	766	54237320	54237608	3,67E-102	381	Proteína GYF 1 interactuante con GRB10 (intrón 20)
NC_041325.1	14	79,221	770	121	29	14715	15471	10173512	10172769	9,71E-138	499	Non
NC_041328.1	17	85,319	361	52	1	2637	2996	7131504	7131144	2,21E-99	372	Non
NC_041313.1	2	96,108	334	9	4	16655	16988	126962466	126962137	1,59E-150	542	Dedos de cinc e proteína 22-like contadora do dominio SCAN (intrón 1)
NC_041313.1	2	75,766	1143	203	43	4385	5498	22293450	22292353	1,61E-140	508	Non
NC_041313.1	2	88,095	378	34	6	16163	16538	28191581	28191949	2,15E-119	438	Non
NC_041313.1	2	100,000	34	0	0	16357	16390	28191830	28191863	1,51E-06	63,9	α-glicosidasa ácida (intrón 3)
NC_041314.1	3	92,627	217	11	2	16419	16630	30103465	30103681	6,32E-80	307	Non
NC_041314.1	3	80,303	396	61	16	15930	16317	37094743	37094357	1,06E-72	283	Non
NC_041314.1	3	95,714	70	3	0	16321	16390	30103429	30103498	1,48E-21	113	Non
NC_041315.1	4	91,046	994	89	0	8091	9084	105456864	105457857	0	1343	Integra completamente ó xene ENSPMRG00000000404
NC_041315.1	4	87,401	881	109	2	11057	11937	56574735	56573857	0	1011	Non
NC_041315.1	4	83,799	179	26	3	16171	16347	11931695	11931872	1,12E-37	167	Non
NC_041315.1	4	100,000	44	0	0	8102	8145	105456867	105456824	4,17E-12	82,4	Integra completamente ó xene ENSPMRG00000000404
NC_041316.1	5	91,542	402	31	1	12812	13213	66082526	66082924	2,64E-153	551	Subunidade catalítica β da fosfatidilinositol-4,5-bisfosfato-3-kinasa (intrón 19)
NC_041316.1	5	91,724	290	23	1	13213	13501	66083182	66083471	2,82E-108	401	Non
NC_041316.1	5	90,189	265	25	1	16250	16514	71797844	71797581	4,82E-91	344	Non
NC_041316.1	5	100,000	34	0	0	16357	16390	71797676	71797643	1,51E-06	63,9	Delta/notch like EGF repeat containing (intrón 9)
NC_041317.1	6	80,590	2473	403	30	10765	13221	5627463	5625052	0	1836	Non
NC_041317.1	6	79,336	1476	252	30	3322	4789	5624678	5623248	0	987	ENSPMRG000000003325 (intrón 4)
NC_041317.1	6	80,345	1160	175	25	2747	3886	6759527	6760653	0	830	Non
NC_041317.1	6	83,353	859	114	8	11520	12376	6758295	6759126	0	767	ENSPMRG000000004075 (intrón 3)
NC_041317.1	6	84,371	787	114	6	7024	7808	5621021	5620242	0	763	ENSPMRG000000003325 (intrón 4)
NC_041317.1	6	83,503	788	109	14	7024	7808	6765088	6765857	0	715	ENSPMRG000000004075 (intrón 3)
NC_041317.1	6	85,652	460	66	0	6538	6997	5621697	5621238	2,72E-133	484	ENSPMRG000000003325 (intrón 4)
NC_041317.1	6	82,851	449	56	7	6546	6994	6764441	6764868	1,02E-102	383	ENSPMRG000000004075 (intrón 3)
NC_041317.1	6	86,424	302	38	3	16204	16504	83643772	83644071	4,85E-86	327	Non
NC_041317.1	6	86,957	92	11	1	16483	16573	83647106	83647197	3,20E-18	102	Proteína morfóxena de óso 7 (intrón 1)
NC_041317.1	6	100,000	34	0	0	16357	16390	83643987	83644020	1,51E-06	63,9	Non
NC_041318.1	7	94,017	585	32	3	8989	9571	85631076	85630493	0	883	Non
NC_041319.1	8	85,714	217	29	2	2413	2629	17715292	17715506	5,06E-56	228	Molécula de adhesión das células neurais 1-B-like (intrón 2)
NC_041320.1	9	90,038	261	25	1	15920	16179	55854119	55853859	8,06E-89	337	Non
NC_041330.1	Z	80,000	405	59	15	15889	16279	50256337	50256733	1,38E-71	279	Non

Táboa 2: Táboa de *hits* obtida dun megablast empregando como *query* o ADNmt, engadindo as columnas do cromosoma no que están e a que xene pertencen se fose o caso. No Anexo deste traballo pode consultarse a táboa con maior calidade.

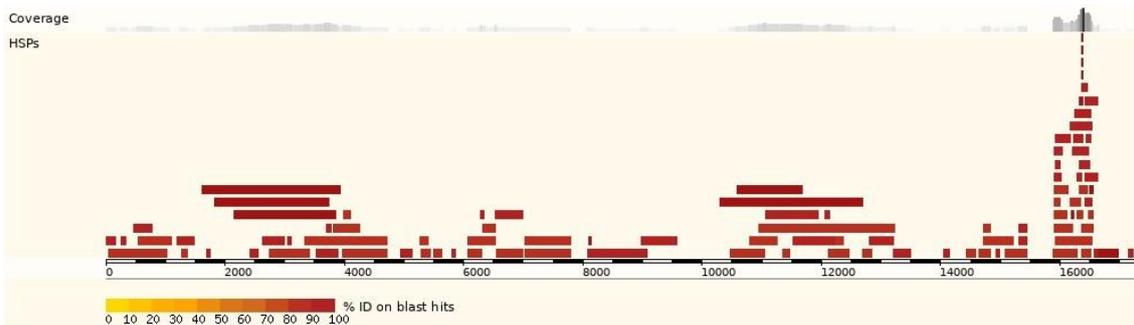


Figura 3: Localización da secuencia correspondente ós NUMTs dentro da mitocondria, a cobertura total sobre o ADNmt é do 81,47%. Existe unha clara sobrerepresentación de secuencias pertencentes á CR, a pesar disto chama a atención a ausencia de NUMTs na porción correspondente ás repeticións en tándem 15439-15900. (figura obtida da base de datos de Ensembl).

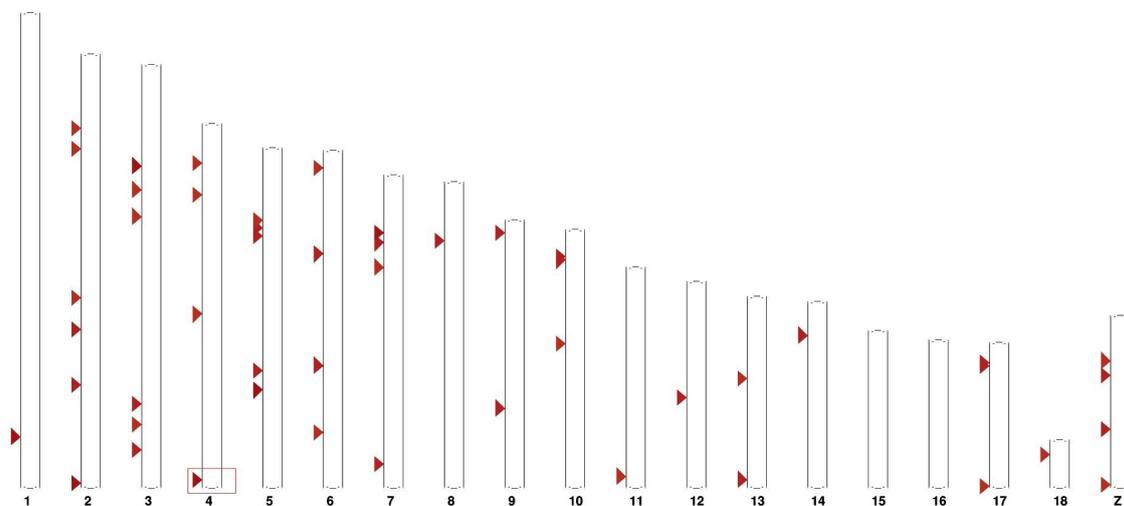


Figura 4: Ubicación dos NUMTs dentro dos cromosomas de *P. muralis*, en frechas vermellas. Datos obtidos de Ensembl.

Aliñamento das rexións 5620242 e 6758295 do cromosoma 6



Figura 5: Resumo do aliñamento entre os NUMTs localizados no cromosoma 6 de *P. muralis* e as secuencias adxacentes (± 1000 pb).

Cando comparamos a rexión 5620242 coa 6758295 podemos observar tódolos escenarios posibles, que responden a diferentes eventos: rexións duplicadas con cadanseu NUMT (duplicación normal), rexións sen duplicar onde só unha delas está representada co seu NUMT (duplicación de secuencias orixinais ou insercións anteriores á gran duplicación) e rexións. que están duplicadas pero carecen dun ou de ambos NUMTs (deleccións e insercións post-duplicación). O resultado do BLAST sobre a secuencia dos ocós amósase a continuación, numerados de esquerda a dereita, tal e como aparece na figura 6.

Minigaps:

1. (1392-1404): Falta un pequeno fragmento na rexión 5620242, máis podemos atopalo máis ou menos completo dentro da mitocondria, polo que o máis probable é que se perdesse unha copia por deleción.
2. (5878-5090): Este fragmento non aparece no ADNmt nin na 5620242. O máis probable é que corresponda á duplicación da secuencia contigua TCTCATGTATTTT.
3. (9954-9974): Falta na 6758295, aparecendo tanto na 5620242 coma no ADNmt, polo que semellante ó minigap 1 o máis probable é que se trate dunha deleción.
4. (10768-10778): Aparece na 5620242 e no ADNmt, polo que tamén podemos concluír que é debido a unha deleción.

Macrogaps:

- A. (1754-2120): Esta secuencia aparece en todos os cromosomas, un total de 80751 veces, mediante un BLAST comprobamos que se trata de SQUAM1, un dos SINEs máis extendidos entre lacértidos (Piskurek *et al.*, 2009).
- B. (2950-3266): Esta secuencia non aparece en ningún outro punto do cromosoma, aínda que si se observaron certas semellanzas con determinadas secuencias mitocondriais doutras especies de lacértidos.
- C. (3416-4220): Esta secuencia aparece ó longo de todos os cromosomas unhas 2709 veces, máis só unha pequena porción desta.
- D. (4221-4963): Unha pequena porción desta secuencia corresponde un Bov-B LINE, que é o asociado a SQUAM 1 (Piskurek *et al.*, 2009), aparecendo 35886 veces.
- E. (5534-6144): Esta secuencia que aparece unhas 294 veces, pertence a un xene que codifica para unha proteína asociada á membrana nuclear interna con moitas isoformas (NURIM).
- F. (7190-8352): Aparece 321 veces e tamén a atopamos en especies moi distanciadas como *Oryzias latipes* con gran identidade, pese a todo isto non está anotada.
- G. (8833-9245): Aparece 670 veces ó longo dos cromosomas, mais o BLAST sobre nucleótidos non amosa ningún resultado.
- H. (11615-11970): Aparece 12754 veces, mais os *hits* unicamente corresponden a unha pequena porción desta secuencia e tampouco está anotada.



Figura 6: Representación resumida da posición dos minigaps (números en azul) e dos macrogaps (números en vermello). Estes ocios corresponden a seccións onde a rexión 6758295 non atopa equivalente na 5620242 e viceversa.

Caracterización das partes da rexión de control de *P. muralis*

Na rexión de control (15440-17311) o primeiro elemento recoñecible é a secuencia CACCACCTGCCGCTCCCAGCGGCTTTTTTGCCTC repetida nove veces, situada entre as 15472 e as 15777 bases.

Os CSB aparecen de xeito alternante, sendo o primeiro o máis distante dos tres

Primeiro CSB: (16631-16652).

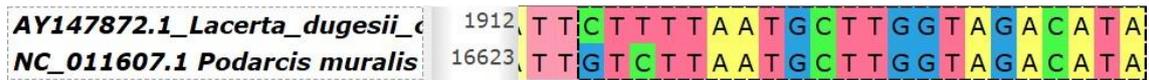


Figura 7: Ubicación do primeiro CSB nun aliñamento entre o ADNmt de *P. muralis* e *L. Dugesii*. Este CSB (marcado por unha liña descontinua) é o menos conservado dos tres.

Segundo CSB: (16902-16919).

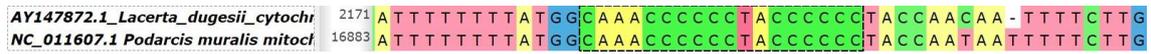


Figura 8: Ubicación do segundo CSB nun aliñamento entre o ADNmt de *P. muralis* e *L. Dugesii*. O segmento conservado exténdese moito máis aló das lindes do CSB (marcado por unha liña descontinua).

Terceiro CSB: (16957-16976).



Figura 9: Ubicación do terceiro CSB (marcado por unha liña descontinua) nun aliñamento entre o ADNmt de *P. muralis* e *L. Dugesii*.

Predición da estrutura secundaria.

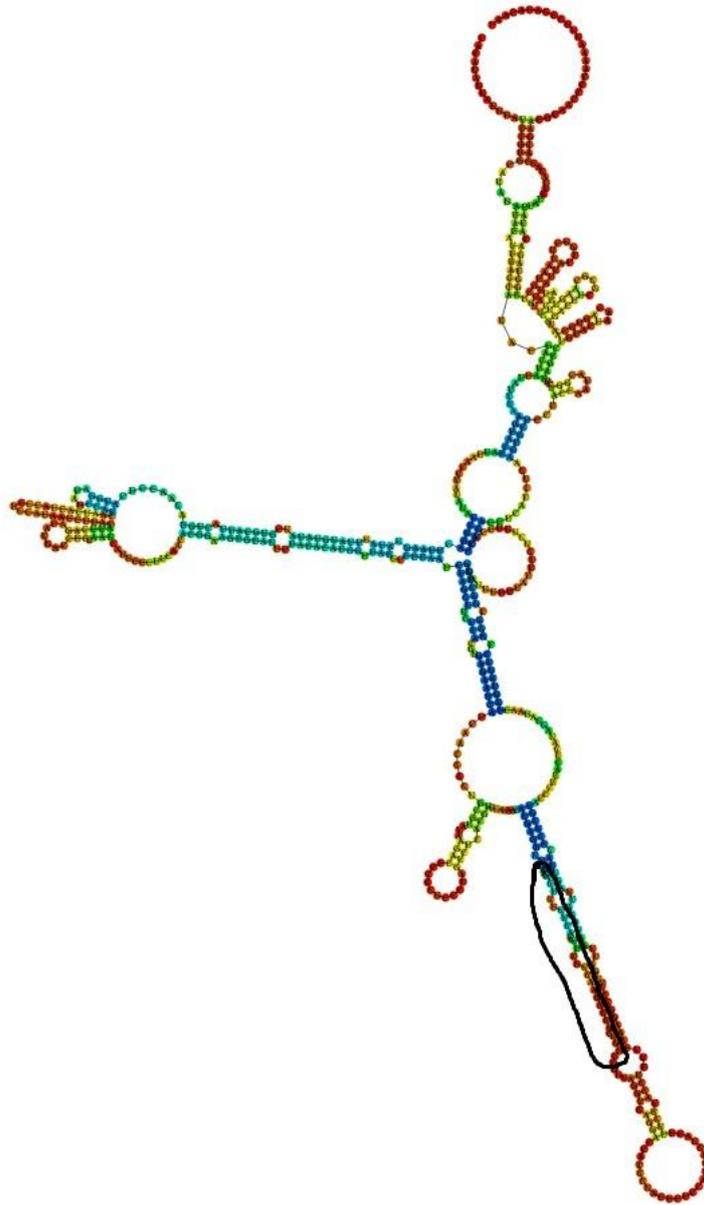


Figura 10: Predición da estrutura secundaria polo modelo de mínima enerxía libre do segmento 16123-16756 do ADNmt. A secuencia altamente conservada e sobrerrepresentada como NUMT de 34 pb indicase rodeada por unha liña negra. Cómpre destacar que no D-loop hai formación de triple cadea, polo que o modelo pode non axustarse á realidade. As gráficas de montaña amosan un val entrópico a carón da secuencia de 34pb que podería estar relacionado coa súa capacidade migratoria.

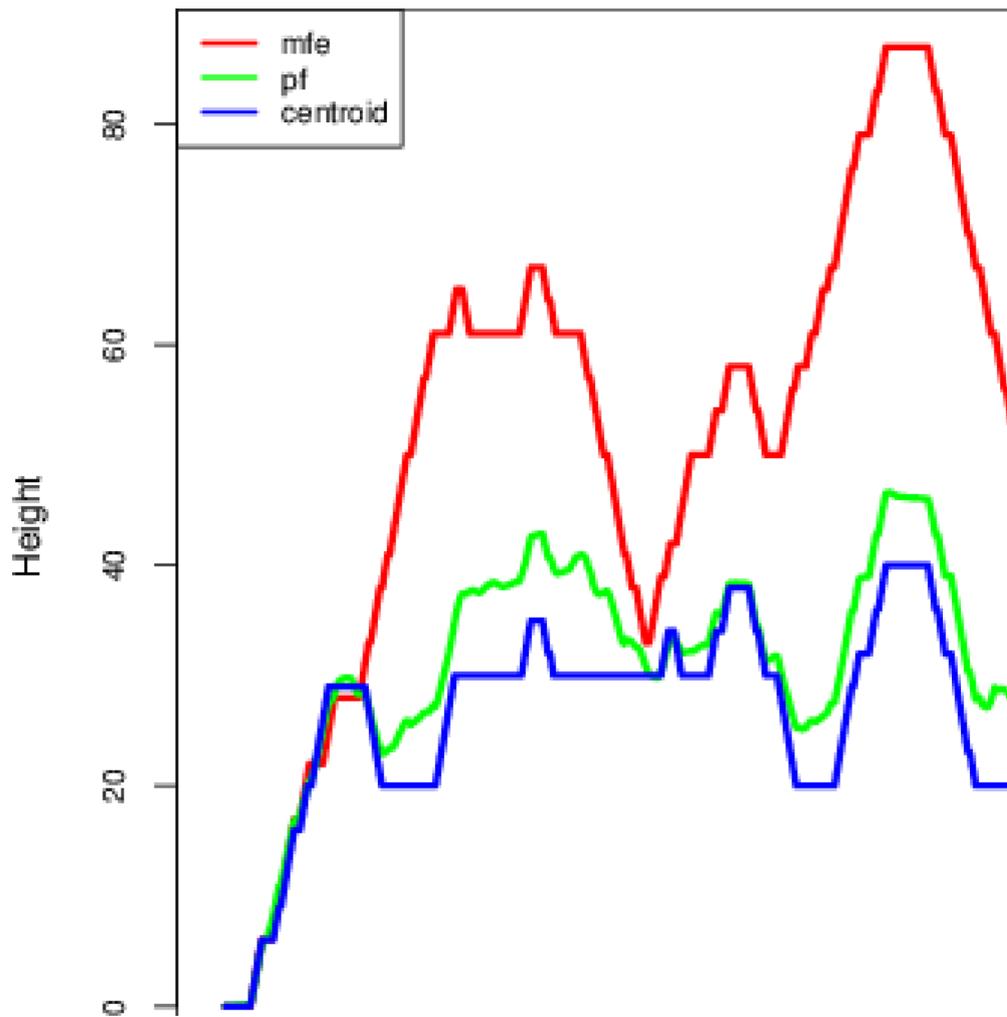


Figura 11: Gráficas de montaña e entropía das posicións que ocupan as distintas bases. No *mountain plot* os vales representan

e os picos horquillas. A gráfica da entropía revela a estabilidade estrutural ó longo do segmento, nótase a presenza dun val entrópico tras a secuencia de interese (300-334pb).

Discusión

Comparativa entre especies

A comparativa feita na táboa 1 indica que nalgúns secuencias o xenoma mitocondrial de *P. siculus* é máis semellante ós NUMTs atopados en *P. muralis* que ó do propio *P. muralis*. Isto pode ter dúas posibles explicacións, que se deba a un fenómeno de hibridación entre ambas especies como xa confirmou Podnar *et al.* (2007) ou que os NUMTs migrasen antes de que estas especies foran separadas e por casuística variase máis o mitocondrial de *P. muralis*. Os nosos resultados parecen apoiar máis a segunda opción, xa que non existe unha diferenza na porcentaxe de identidade en absoluto comparable á atopada por Podnar *et al.* (2007) sendo a meirande dun 2,908% non habendo diferenzas significativas entre ambas (figura 1). Aínda que algunhas secuencias parezan moi recentes debemos ter coidado á hora de tirar conclusións, xa que o grao de variación das secuencias varía segundo a rexión de onde procedan e pode confundirse coa antigüidade (Grechko, 2013).

Orixe dos NUMTs nos cromosomas, peculiaridades

Nos cromosomas 7, 8, 9, 11, 13, 14, 17, Z non hai dúbida de que temos fenómenos de inserción individuais, porque só se produciu un evento en cada un deles.

Nos cromosomas 1, 2, 3, 5, 6 e 10 atopamos unha secuencia de 34pb (16357-16390) que ten un 100% de identidade (ás veces aparece como un *hit* un pouco máis grande, baixando a súa porcentaxe de identidade) coa rexión equivalente do mitocondrial e que adoita estar dentro doutro *hit* máis grande. Por algún motivo esta secuencia ten unha capacidade migradora moi alta (aparece reflectido na figura 2) e atópase extremadamente conservada, xa que non só aparece en *P. muralis*, senón que foi atopada en moitos organismos de todos os xéneros da familia *Lacertidae* de xeito exacto. A estrutura secundaria analizada neste fragmento pode ter que ver con estas características, xa que aparece formando unha horquilla cunha secuencia complementaria, se a formación desta estrutura tivese un papel moi relevante para a replicación, explicaría a razón de que estea tan conservada. A maiores o val entrópico podería explicar a maior predisposición desta rexión a desprenderse.

Cromosoma 2: Alén da secuencia xa comentada, incluída dentro do NUMT 16163-16538, as outras semellan pertencer a eventos de migración diferentes, xa que aínda que son grandes, diverxen enormemente na porcentaxe de identidade e pertencen a distintas ubicacións, tanto na mitocondria como no cromosoma.

Cromosoma 3: Aquí atopamos outra vez a secuencia de 34 pares de bases facendo parte dunha máis grande de 70, mais o *hit* non aparece dúas veces, en troques aparece unha secuencia que comeza 30 pares de bases despois e que presumiblemente migrou canda ela, xa que están no mesmo lugar do cromosoma. Ante estes datos podemos asumir que o evento de migración foi

único e que incluía o espazo intermedio, pero este foi eliminado do cromosoma. O NUMT restante parece ser independente, xa que aínda que é contiguo ós outros na mitocondria, non é así no cromosoma e tendo en conta a porcentaxe de identidade parece moi anterior.

Cromosoma 4: Semellante ó que acontece coa secuencia de 34 pb en tantos outros cromosomas, atopamos aquí unha secuencia que parece comportarse igual, isto é, ten un número reducido de pares de bases (44), está incluída dentro dunha maior e aparece cunha porcentaxe de identidade do 100%, máis non está presente noutros cromosomas e non aparece como *hit* en *P. siculus*. No resto de NUMTs as porcentaxes de identidade non coinciden e as rexións son diferentes tanto no cromosoma como na mitocondria polo que podemos concluír que pertencen a eventos diferentes.

Cromosoma 5: Neste cromosoma houbo en total dous eventos de inserción. O primeiro é o que inclúe a secuencia de 34 pares de bases. O segundo aparece na táboa como dous *hits* porque o NUMT orixinal quedou dividido en dous debido á inserción dunha secuencia entre ambos.

Cromosoma 6: Desenvólvese máis adiante, no seu propio apartado.

Cromosoma 10: Os eventos de inserción que aquí aconteceron semellan diferentes, debido a que na mitocondria están de xeito descontinuo e no cromosoma están en puntos moi separados. Aínda que as súas porcentaxes de identidade sexan semellantes, non necesariamente tiveron que acontecer no mesmo período, xa que poden pertencer a rexións con variabilidade dispar dentro da CR.

Estudo dos NUMTs do cromosoma 6

Na táboa 1 atopamos 11 *hits* asociados ó cromosoma 6, 4 na rexión 5620242, 4 na 6758295 e 3 na 83643772 que en total inclúen un 38% do ADNmt; pese a todo isto o máis probable é que só se producisen dous eventos, un que deu lugar ós *hits* da rexión 83643772 e outro que orixinou os *hits* restantes.

Teoría da formación dos *hits* da rexión 83643772: Aínda que só houbo un evento de inserción, na táboa 1 aparecen como *hits* diferentes, semellante ó que acontece con outros NUMTs.

Demostración: Dous dos aliñamentos solápanse e o terceiro forma parte da secuencia de 34 pares de bases xa comentada.

Teoría da formación dos *hits* das rexións 5620242 e 6758295: Inicialmente foi inserido un gran fragmento de ADN mitocondrial nunha das dúas rexións (resulta imposible afirmar cal foi a primeira), que logo foi fragmentado e separado, por último toda a rexión foi duplicada e cada copia seguiu un camiño evolutivo diferente.

Demostración: Algunhas das rexións do propio cromosoma que non corresponden a NUMTs tamén foron duplicadas (pódese apreciar nalgunhas rexións que marcan o valor 2 na figura 4), isto por sí só xa confirma a teoría. Como veremos máis adiante, en moitos casos as secuencias non duplicadas pertencen a elementos transpoñibles, polo que podemos supoñer que foron inseridos tras a duplicación. Ademais as porcentaxes de identidade son semellantes e cada NUMT 5620242 ten a súa copia máis ou menos equivalente en 6758295.

Dificultades: A orde en como están colocados os NUMTs non coincide co seu equivalente mitocondrial. Existen uns fragmentos que non aliñan en non parecen ter equivalente.

Resulta rechamante o achado da proteína NURIM no macrogap E, xa que poderían ser indicios de creación de novas

Importancia dos NUMTs na evolución

Nunha primeira instancia, a importancia dos NUMTs na evolución semella ben escasa por diferentes motivos. Para comezar este fenómeno é relativamente infrecuente, sendo que *P. muralis* e *P. siculus* distanciáronse entre 18,60 e 6,24 millóns de anos (Yang *et al.*, 2021) e comparten a práctica totalidade dos *hits* (todos se atendemos só ós eventos de inserción), ademais todos agás un foron inseridos en intróns ou rexións sen xenes. Porén, esta excepción que corresponde a un NUMT situado no cromosoma 4 é a que resulta máis interesante, xa que todo o xene que codifican está dentro del (probablemente corresponda ó xene do ATP 6 mitocondrial tendo en conta a súa posición nesta), o que ven a demostrar que estes poden orixinar novos puntos de iniciación da transcripción dentro dos cromosomas e suporía un exemplo de dun xene funcionalmente activo transferido dende a mitocondria ó núcleo. A raíz deste achado no cromosoma 4 podemos teorizar que os NUMTs puideron ter nunha orixe maior importancia da que teñen a día de hoxe, sendo un dos principais axentes que modificou o xenoma mitocondrial ate o punto de só conter información que debe ser sintetizada na mitocondria (hipótese hidrofóbica) (Popot e de Vitry, 1990). De ser así o máis plausible é que nunha orixe os NUMTs serviron para trasladar a meirande parte dos xenes dende a mitocondria ó núcleo e unha vez alí, aqueles que non requirían que a súa síntese fose especificamente na mitocondria foron perdéndose debido ó mecanismo de Muller's ratchet (Lynch, 1996).

Conclusiones

- Un só NUMT pode aparecer como varios *hits*, atopando 23 NUMTs en *P. muralis* a partires dunha táboa de 38 *hits* iniciais. Non se atopou un nesgo por determinados cromosomas, aparecendo en 15 dos 19 e sendo os que máis tiñan o 4 e o 2 con 3.
- Os NUMTs teñen un tamaño moi variable, oscilando entre as 179 e as 5196pb, cun tamaño promedio de 783pb.
- O cromosoma 6 de *P. muralis* sufriu un proceso de duplicación e como consecuencia atopamos dous NUMTs derivados dun só evento de inserción.
- Os NUMTs poden ser empregados para identificar secuencias transpoñibles, así como para estimar a súa frecuencia, xa que grazas ós seus parálogos mitocondriais podemos aproximar como era a secuencia orixinal, especialmente se pertence a unha zona moi conservada.
- A meirande parte dos NUMTs foron atopados en rexións interxénicas (59%) ou en intróns (36%), mais non foi posible confirmar se existe algún tipo de predisposición a ocupar unha destas.

- O NUMT atopado no cromosoma 4 foi o único que contiña un exón, demostrando que os NUMTs poden orixinar xenes funcionais no núcleo.
- Parece ser que actualmente o impacto dos NUMTs na evolución é máis ben escaso, mais no pasado puideron ser a causa da redución do xenoma mitocondrial ate o seu tamaño actual.
- En lacértidos o aliñamento dos NUMTs sobre o ADNmt cobre o 81,47% e amosa un marcado sesgo, onde case a metade destes proveñen da rexión de control.

Conclusiones

- Un solo NUMT puede aparecer como varios *hits*, encontrando 23 NUMTs en *P. muralis* a partir de una tabla de 38 *hits* iniciales. No se encontró un sesgo por determinados cromosomas, apareciendo en 15 de los 19 y siendo los que más tenían el 4 y el 2 con 3.
- Los NUMTs tienen un tamaño muy variable, oscilando entre las 179 y las 5196pb, con un tamaño promedio de 783pb.
- El cromosoma 6 de *P. muralis* sufrió un proceso de duplicación y como consecuencia encontramos dos NUMTs derivados de un solo evento de inserción.
- Los NUMTs pueden ser usados para identificar secuencias transponibles, así como para estimar su frecuencia, ya que gracias a sus parálogos mitocondriales podemos aproximar como era la secuencia original, especialmente si pertenece a una zona muy conservada.
- La mayor parte de los NUMTs fueron encontrados en regiones intergénicas (59%) o en intrones (36%), pero no fué posible confirmar si existe algún tipo de predisposición a ocupar una de estas.
- El NUMT encontrado en el cromosoma 4 fué el único que contenía un exón, demostrando que los NUMTs pueden originar genes funcionales en el núcleo.
- Parece ser que actualmente el impacto de los NUMTs en la evolución es más bien escaso, mas en el pasado, pudieron ser la causa de la redución del genoma mitocondrial hasta su tamaño actual.
- En lacértidos el alineamiento de los NUMTs sobre el ADNmt cubre el 81,47% de este y muestra un marcado sesgo, donde casi la mitad de estos provienen de la región de control.

Conclusions

- A single NUMT can appear as several hits, finding 23 NUMTs in *P. muralis* from a table of 38 initial hits. No bias for particular chromosomes was found, appearing in 15 of the 19 and with the most having 4 and 2 with 3.
- The NUMTs are highly variable in size, ranging from 179 to 5196bp, with an average size of 783bp.
- Chromosome 6 of *P. muralis* underwent a duplication process and as a consequence we found two NUMTs derived from a single insertion event.

- NUMTs can be used to identify transposable sequences, as well as to estimate their frequency, since thanks to their mitochondrial paralogs we can approximate what the original sequence was like, especially if it belongs to a highly conserved region.
- Most of the NUMTs were found in intergenic regions (59%) or in introns (36%), but it was not possible to confirm whether there is any predisposition to occupy one of these.
- The NUMT found on chromosome 4 was the only one that contained an exon, demonstrating that NUMTs can originate functional genes in the nucleus.
- It seems that currently the impact of NUMTs in evolution is rather low, but in the past, they could have been the cause of the reduction of the mitochondrial genome to its current size.
- In lizards, the alignment of NUMTs on the mtDNA covers 81.47% of the mtDNA and shows a marked bias, with almost half of the NUMTs coming from the control region.

Bibliografía

- Adams, K. e Palmer, J. 2003. Evolution of mitochondrial gene content: Gene loss and transfer to the nucleus. *Molecular Phylogenetics and Evolution* 29(3): 380–395. [https://doi.org/10.1016/S1055-7903\(03\)00194-5](https://doi.org/10.1016/S1055-7903(03)00194-5)
- Blanchard, J. e Lynch, M. 2000. Organellar genes why do they end up in the nucleus. *Trends in Genetics* 16(7): 315–320.
- Brehm, A., James Harris, D., Alves, C., Jesus, J., Thomarat, F. e Vicente, L. 2003. Structure and evolution of the mitochondrial DNA complete control region in the lizard *Lacerta dugesii* (*Lacertidae*, *Sauria*). *Journal of Molecular Evolution* 56(1): 46–53. <https://doi.org/10.1007/s00239-002-2379-0>
- Brown, W., George, M. e Wilson, A. 1979. Rapid evolution of animal mitochondrial DNA. *Proceedings of the National Academy of Sciences* 76(4): 1967-1971. <https://www.pnas.org>
- Camargo, A., Sinervo, B. e Sites, J. 2010. Lizards as model organisms for linking phylogeographic and speciation studies. *Molecular Ecology* 19(16): 3250–3270. <https://doi.org/10.1111/j.1365-294X.2010.04722.x>
- Grechko, V. 2013. The problems of molecular phylogenetics with the example of squamate reptiles: Mitochondrial DNA markers. *Molecular Biology* 47(1): 55–74. <https://doi.org/10.1134/S0026893313010056>
- Kramerov, D. e Vassetzky, N. 2011. Origin and evolution of SINEs in eukaryotic genomes. *Heredity* 107(6): 487–495. <https://doi.org/10.1038/hdy.2011.43>
- Lynch, M. 1996. Mutation accumulation in transfer RNAs: Molecular evidence for Muller's ratchet in mitochondrial genomes. *Molecular Biology and Evolution* 13(1): 209-220. <https://academic.oup.com/mbe/article/13/1/209/1055494>
- Mathews, D., Disney, M., Childs, J., Schroeder, S., Zuker, M., e Turner, D. 2004. Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure.

- Proceedings of the National Academy of Sciences 101(19): 7287-7292.
www.pnas.org/cgi/doi/10.1073/pnas.0401799101
- Michaelides, S., While, G., Bell, C. e Uller, T. 2013. Human introductions create opportunities for intra-specific hybridization in an alien lizard. *Biological Invasions* 15(5): 1101–1112. <https://doi.org/10.1007/s10530-012-0353-3>
- Nicholls, T. e Minczuk, M. 2014. In D-loop: 40 years of mitochondrial 7S DNA. *Experimental Gerontology* 56: 175–181. <https://doi.org/10.1016/j.exger.2014.03.027>
- Piskurek, O., Nishihara, H. e Okada, N. 2009. The evolution of two partner LINE/SINE families and a full-length chromodomain-containing Ty3/Gypsy LTR element in the first reptilian genome of *Anolis carolinensis*. *Gene* 441(1–2): 111–118. <https://doi.org/10.1016/j.gene.2008.11.030>
- Podnar, M., Haring, E., Pinsker, W. e Mayer, W. 2007. Unusual origin of a nuclear pseudogene in the Italian wall lizard: Intergenomic and interspecific transfer of a large section of the mitochondrial genome in the genus *Podarcis* (*Lacertidae*). *Journal of Molecular Evolution* 64(3): 308–320. <https://doi.org/10.1007/s00239-005-0259-0>
- Popot, J., e de Vitry, C. 1990. On the microassembly of integral membrane proteins. *Annual Reviews* 19: 369-403. www.annualreviews.org/aronline
- Santos, J., Žagar, A., Drašler, K., Rato, C., Ayres, C., Harris, D., Carretero, M., e Salvi, D. 2019. Phylogeographic evidence for multiple long-distance introductions of the common wall lizard associated with human trade and transport. *Amphibia-Reptilia* 40(1): 121–127. <https://doi.org/10.1163/15685381-20181040>
- Schulte, U., Gassert, F., Geniez, P., Veith, M. e Hochkirch, A. 2012. Origin and genetic diversity of an introduced wall lizard population and its cryptic congener. *Amphibia-Reptilia* 33(1): 129–140. <https://doi.org/10.1163/156853812X626160>
- Yang, W., Feiner, N., Salvi, D., Laakkonen, H., Jablonski, D., Pinho, C., Carretero, M., Sacchi, R., Zuffi, M., Scali, S., Plavos, K., Pafilis, P., Poulakakis, N., Lymberakis, P., Jandzik, D., Schulte, U., Aubret, F., Badiane, A., Perez, G., Avalos, J., While, G. e Uller, T. 2021. Population genomics of wall lizards reflects the dynamic history of the Mediterranean basin. *Molecular Biology and Evolution* 39(1). <https://doi.org/10.1093/molbev/msab311>