Universidad de Jaén

Escuela de Doctorado

TESIS DOCTORAL

•

Análisis citogenético y molecular en especies del género *Talpa* (Insectivora; Talpidae)

PRESENTADA POR:

Juana Gutiérrez Martos

DIRIGIDA POR:

Dr. Antonio Sánchez Baca

Dr. Juan Alberto Marchal Ortega

Dr. Gaël Aleix Mata

JAÉN, octubre de 2023

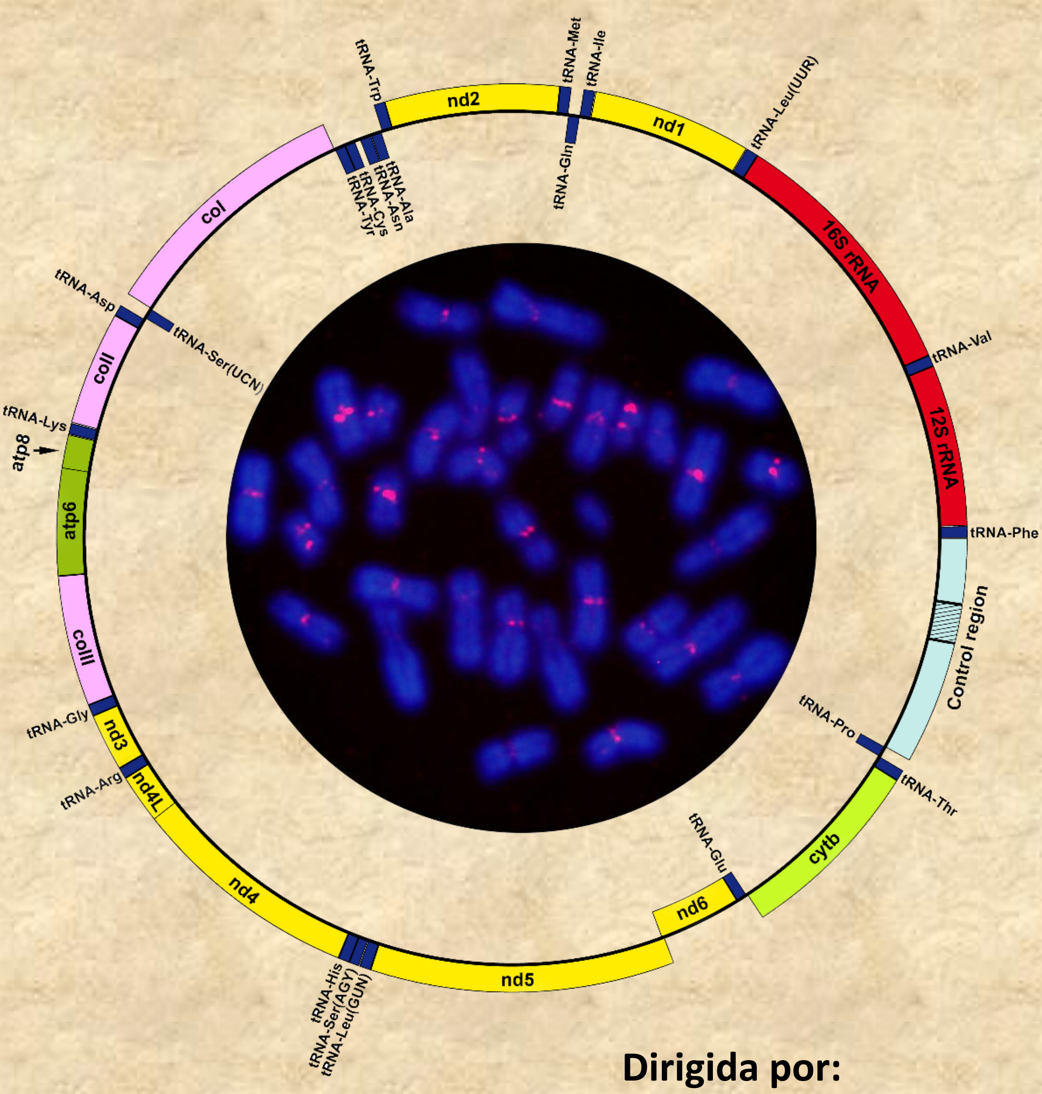
ISBN

Tesis Doctoral

Análisis citogenético y molecular en especies del género *Talpa* (Insectivora; Talpidae)

Presentada por:

Juana Gutiérrez Martos



Dirigida por:
Dr. Antonio Sánchez Baca
Dr. Juan Alberto Marchal Ortega
Dr. Gaël Aleix Mata

Universidad de Jaén

Facultad de Ciencias Experimentales

Universidad de Jaén

Escuela de Doctorado

Facultad de Ciencias Experimentales

Departamento de Biología

Experimental

Área de Genética

Los Dres. Antonio Sánchez Baca, Juan Alberto Marchal Ortega y Gaël Aleix Mata, doctores de la Universidad de Jaén, certifican que la Tesis Doctoral titulada: **"Análisis citogenético y molecular en especies del género *Talpa* (Insectivora; Talpidae)"**, que presenta Juana Gutiérrez Martos para optar al Grado de Doctora, ha sido realizada bajo su dirección, reuniendo a su juicio los requisitos exigidos para esta presentación.

Jaén, octubre de 2023

Firmado digitalmente por SANCHEZ BACA ANTONIO - 74625080Q
Fecha: 2023.05.09 11:41:34 +02'00'

MARCHAL ORTEGA JUAN ALBERTO - 77323973C

Firmado digitalmente por MARCHAL ORTEGA JUAN ALBERTO - 77323973C
Fecha: 2023.05.09 12:57:32 +02'00'

Firmado digitalmente por ALEIX MATA GAEL - Y3115302N
Fecha: 2023.05.09 12:14:12 +02'00'

Universidad de Jaén
Escuela de Doctorado

TESIS DOCTORAL

ANÁLISIS CITOGENÉTICO Y MOLECULAR EN ESPECIES DEL GÉNERO *TALPA* (INSECTIVORA; TALPIDAE)

PRESENTADA POR:

Juana Gutiérrez Martos

DIRIGIDA POR:

Antonio Sánchez Baca

Juan Alberto Marchal Ortega

Gaël Aleix Mata

AGRADECIMIENTOS

En primer lugar, quiero agradecer a mis padres su comprensión. En una época en la que la educación superior sólo estaba al alcance de algunos privilegiados, ellos entendieron que la formación era esencial para garantizar a sus hijos un porvenir más prometedor. Gracias a vuestra actitud abierta, pude desafiar la resistencia de ser de pueblo y el atávico destino reservado por aquel entonces a las niñas, que excluía la posibilidad de conseguir autonomía en lo personal.

Por otro lado, pude disfrutar de una beca, lo cual me permitió mudarme a la ciudad para ir a estudiar al instituto y después a la universidad. No haber obtenido beca habría supuesto un obstáculo económico difícilmente asumible por mis padres. Es por ello que, de alguna manera, con este trabajo quiero expresar mi gratitud por la oportunidad solidaria que la vida me ha brindado.

Gracias también a mi familia, en especial a mis tíos María y Manuel, por acogerme generosamente en vuestra casa en mis primeros años de instituto en Jaén, soportando estoicamente lo peor de mi adolescencia.

Mi más sincero agradecimiento a mis directores de tesis, los Dres. Antonio Sánchez, Juan Alberto Marchal y Gaël Aleix-Mata, porque sin vuestro apoyo, paciencia y comprensión esta tesis doctoral no habría sido posible.

Antonio, toda mi gratitud, mi reconocimiento y mi afecto. Tu empatía, tu amor por la ciencia, la pasión por tu trabajo, el compromiso y la implicación sincera en todo lo que haces te convierte en impulsor y guía de este proyecto. Y no sólo en el terreno académico, han sido tantos años y tanto tiempo compartido, que a veces no he podido distinguir cuándo hablaba con el tutor y cuándo con el amigo. Mil gracias por todo.

Al Dr. Pedro Lorite y a la Dra. Teresa Palomeque, por vuestra dedicación y acertadas observaciones a la hora de interpretar los interminables datos de gráficos y secuencias. También por vuestra buena disposición siempre que he necesitado algún reactivo o ayuda en el laboratorio.

A la Dra. Elisa Montiel, por el tiempo invertido en dilucidar algunos de los aspectos más áridos de este trabajo, como han sido el análisis y todo el manejo de programas bioinformáticos. Al Dr. Francisco Ruiz-Ruano, por tu colaboración en el ensamblaje de genomas mitocondriales. Y no puedo olvidarme de Ismael, por facilitarme la labor cuando empecé esta apasionante aventura, ni de Luz, por tu amistad y gran ayuda con las FISH y con la captura de Talpas.

A todos los compañeros del Departamento de Biología Experimental y del Área de Genética que me habéis ayudado amablemente, y en especial a Pablo Mora, Jesús Vela y María Arroyo, por vuestra disposición y el buen ambiente en el laboratorio.

Mi gratitud al Dr. Riccardo Castiglia de la Universidad la Sapienza en Roma, por el material cromosómico facilitado y por brindarme tu ayuda. En la misma línea, mi agradecimiento sincero al Dr. Diogo Cavalcanti Cabral-de-Mello de la Universidad de Estadual Paulista, Rio Claro, São Paulo, Brazil, por tu ayuda en la puesta a punto de la FISH con oligos y por tus aportaciones intelectuales.

A la Delegación Territorial de la Consejería de Agricultura, Pesca y Medio Ambiente de Granada (Junta de Andalucía) y al Servicio Territorial de Medio Ambiente de Burgos (Junta de Castilla y León), por los permisos necesarios para la captura de los especímenes de *Talpa*.

Muchas gracias a Santiago, por tantas charlas "doctorales" y tus valiosas sugerencias para dar cuerpo y formato a este texto. Y a Alicia, por tu genuino y excelente trabajo puliendo y mejorando el PowerPoint.

No quiero olvidarme del apoyo vital de las personas cuya amistad tengo el privilegio de disfrutar. Gracias por acudir y regalarme vuestro tiempo cuando lo he necesitado, por darme aliento en los baches del camino y, como no, por las alegrías compartidas. En especial, gracias de corazón María del Mar por tu paciencia y comprensión, por servirme de soporte en tantas ocasiones y por tantas horas de escucha activa.

A Pedro y Violeta por ser como sois, lo más valioso del mundo. Me siento profundamente orgullosa de los dos.

Y a Fran, por quererme.

ÍNDICE

Capítulo 1: Introducción

INTRODUCCIÓN

Especies objeto de estudio

Sistemática y Filogenia

Tradicionalmente, la denominación "Insectivora" (Cuvier 1817; Bowdich 1821) ha sido utilizada como "cajón de sastre", una categoría general para agrupar una amplia diversidad de mamíferos placentarios que comparten características morfológicas primitivas, son pequeños y se alimentan principalmente de insectos y otros invertebrados. Esta categoría también incluía formas fósiles de incierto encuadre sistemático (Simpson 1945). A lo largo del proceso evolutivo, los insectívoros se separaron tempranamente durante la diversificación de los mamíferos placentarios y, dado que conservan rasgos morfológicos considerados como primitivos, son fundamentales para comprender la evolución de este gran grupo de animales.

Teniendo en cuenta lo anterior, los insectívoros han sido considerados un grupo polifilético compuesto por numerosas familias, algunas de las cuales fueron posteriormente excluidas. Sin embargo, los insectívoros en sentido estricto, se han clasificado dentro del orden Lipotyphla, considerado como monofilético. Este orden incluye los subórdenes Erinaceomorpha (familia Erinaceidae) y Soricomorpha con cinco familias (Talpidae, Soricidae, Solenodontidae, Tenrecidae y Chrysochloridae) (Novacek 1986; Butler 1988; Lopatin 2006). Estudios más recientemes en citogenética y genética molecular, sugieren que las familias Tenrecidae y Chrysochloridae, presentes únicamente en el continente africano, deberían separarse del resto de soricomorfos y agruparse en un nuevo orden llamado Afrosoricida. Este orden pertenecería a un clado supraordinal distinto denominado Afrotheria. De este modo, el supraorden Eulipotyphla englobaría al resto de las principales familias de insectívora (Stanhope et al. 1998; Waddell et al. 1999; Douady et al. 2002a y b) (Fig. 1).

La familia Talpidae contiene aproximadamente 43 especies descritas (He 2017) agrupadas en 17 géneros y tres subfamilias: Scalopinae, Talpinae y Uropsilinae. Algunos autores diferencian también una cuarta subfamilia, Desmaninae, que incluiría las especies de los géneros Desmana y Galemys (McKenna y Bell 1997; Wilson y Reader 2005; Cabria et al. 2006).

El género *Talpa* (subfamilia Talpinae) incluye once especies y se originó en Asia a finales del Mioceno, extendiéndose por Europa en el Plioceno. Entre estas especies tenemos seis de distribución occidental, *T. europaea* (Linnaeus 1758), *T. caeca* (Savi 1882), *T. occidentalis* (Cabrera 1907), *T. romana* (Thomas 1902), *T. stankovici* (Martino y Martino 1931) y *T.*

aquitania (Nicolas et al. 2015), y cinco de distribución oriental, *T. caucasica* (Satunin 1908), *T. davidiana* (Milne-Edwards 1884), *T. levantis* (Thomas 1906) *T. altaica* (Nikolsky 1883) y *T. martinorum* (Kryštufek et al. 2018) (Hutterer 2005; Wilson y Reeder 2005; Hugot et al. 2014; Nicolas et al. 2015, 2017b).

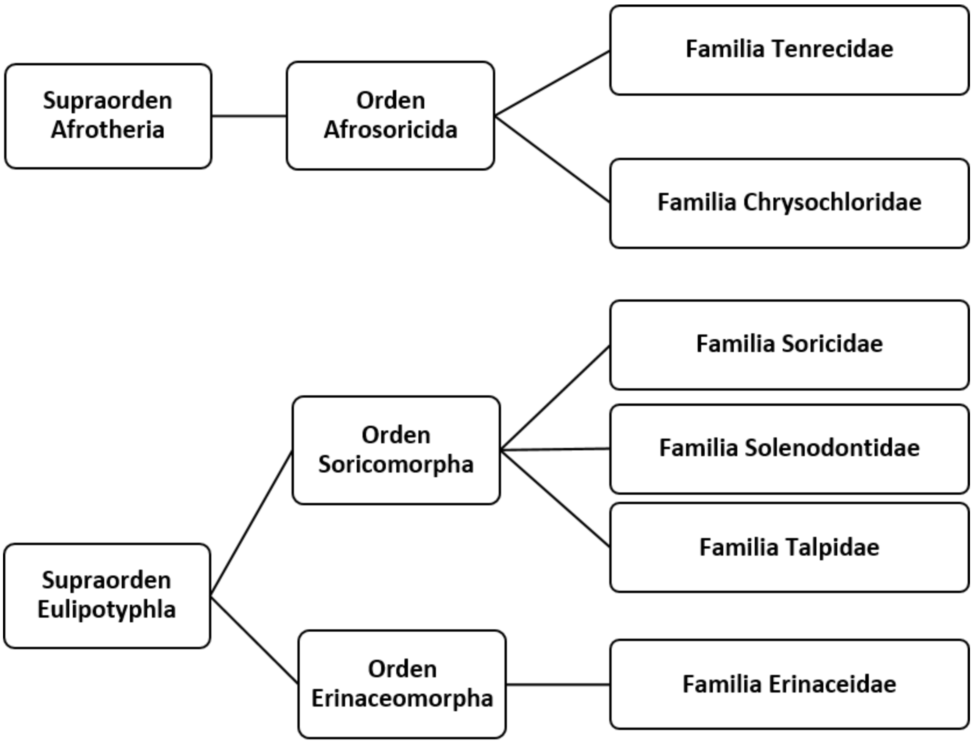


Figura 1*. Esquema que muestra las relaciones evolutivas entre las distintas familias de Insectívora*

Los datos del análisis del citocromo b de ocho especies de *Talpa* sugieren un origen monofilético del género y dos linajes diferentes, uno para las especies orientales *T. altaica* y *T. caucásica* -que divergieron más tempranamente- y otro para las especies de Europa occidental. Dentro de estas, encontramos dos sublinajes principales: un clado oriental incluye *Talpa levantis* y *T. stankovici,* y un clado occidental que incluye *T. europaea, T. caeca, T. romana* y *T. occidentalis*. Las edades estimadas de divergencia entre los linajes respaldan la propuesta de que el género se originó en Asia a finales del Mioceno (Colangelo et al. 2010).

Características morfológicas y ecológicas de Talpidae

Las especies de tálpidos se encuentran en la mayor parte de Eurasia y Norteamérica. En su mayoría excavan túneles subterráneos en los que viven y se alimentan de invertebrados, aunque también pueden ingerir partes de plantas. En contraste, las especies de la subfamilia

Desmaninae (desmanes) viven en cursos de ríos, donde se alimentan principalmente de insectos y larvas acuáticas. Los tálpidos fosoriales están bien adaptados a su peculiar modo de vida, con cuerpo fusiforme, ojos diminutos y ausencia de orejas externas. Las extremidades anteriores están altamente modificadas y adaptadas para excavar túneles, con brazos cortos y giros que hacen que los codos y las palmas de las manos miren hacia atrás. El húmero está ensanchado, lo que proporciona una gran superficie de fijación a los potentes músculos excavadores (Campbell 1939; Bickelmann et al. 2017). Los topos tienen un metabolismo relativamente alto y un apetito insaciable, lo que los hace activos tanto de día como de noche. Aunque pueden encontrarse en una gran variedad de hábitats, prefieren los suelos húmedos y fáciles de excavar (Myers 2001).

Una característica distintiva de estos animales es su desarrollo gonadal, con la presencia de ovotestes bilaterales en las hembras en lugar de ovarios normales (Jiménez et al. 1993; Sánchez et al. 1996). Recientemente se ha demostrado que esta alteración del desarrollo está relacionada con reordenaciones en genes y en sus regiones reguladoras, que afectan al control de las hormonas implicadas en el desarrollo gonadal. Esta alteración en la expresión de los genes ha podido favorecer la adaptación de estas especies, y la de las hembras en particular, a su modo de vida haciéndolas más fuertes y agresivas (Real et al. 2020).

Investigaciones realizadas en *Talpa occidentalis,* han comprobado que la actividad sexual tanto en hembras como en machos adultos coincide con los meses de septiembre a mayo, permaneciendo inactivos el resto del año (los machos se vuelven estériles y las hembras muestran cambios en el desarrollo uterino y en el suministro de sangre). Este periodo de actividad sexual, que dura casi nueve meses, es muy prolongado en comparación con otras poblaciones europeas del género *Talpa*. También se ha descrito en *T. europaea* que la temporada de cría difiere en duración entre las poblaciones del norte y las del sur de Europa, siendo más corta en las poblaciones más septentrionales y alargándose a medida que nos acercamos al ecuador (Jiménez et al. 1990).

Citogenética. Diferenciación cariotípica.

El análisis de los cariotipos en insectívoros ha indicado que existe un alto nivel de conservación en las familias Talpidae y Erinaceae, contrariamente a lo que ocurre en la familia Soricidae, donde se ha demostrado la existencia de una gran diversidad cariotípica (Graphodatsky et al. 1991; Kawada et al. 2002; Gornung et al. 2008). En concreto, las variaciones involucradas principalmente en la evolución cromosómica en insectívoros, han sido reordenaciones Robertsonianas y cambios en la cantidad y distribución de la heterocromatina

(Yong 1971; Fredga 1973; Ando et al. 1980; Obara y Miyai 1981; Yosida 1982; Searle 1984; Ruedi y Vogel 1995; Elrod et al. 1996). Así, mediante estudios de “chromosome painting”, se ha podido estimar la conservación del genoma en Eulipotyphla en función del número de cambios evolutivos en los cromosomas, a partir de un mamífero ancestral común. Los datos de este estudio ofrecen una reconstrucción de las relaciones filogenéticas en Eulipotyphla, y demuestran que la familia Talpidae muestra el mayor grado de conservación cromosómica y fue la primera en divergir del resto de grupos (Soricidae y Erinaceidae) hace unos 65 m.a. (Biltueva y Vorobieva 2012).

En las especies de la familia Talpidae no se han hallado variaciones cromosómicas intraespecíficas en la eucromatina (Yates y Moore 1990; Graphodatsky et al. 1991; Kawada et al. 2002; Gornung et al. 2008; Biltueva y Vorobieba 2012), con la excepción de la inversión pericéntrica descrita en *T. romana* por Reumer y Meylan (1986). Por otro lado, mediante técnicas de bandeo C y G, se ha podido constatar que las inversiones pericéntricas y los cambios cuantitativos en la heterocromatina constitutiva son las únicas modificaciones implicadas en la escasa diferenciación cariotípica de esta familia (Harada et al. 2001).

La mayoría de las especies de *Talpa* tienen 2n = 34 y un número fundamental autosómico (FNa) que varía entre 62 y 66. Sólo dos especies, *T. caeca* y *T. caucasica*, tienen números diploides diferentes, 2n = 36 y 2n = 38 respectivamente (Fedyk e Ivanitskaya 1972; Zima 1983; Jiménez et al. 1984; Kawada et al. 2002; Gornung et al. 2008).

Los patrones de las bandas C, G y de las regiones organizadoras del nucléolo (NOR) en los cariotipos de *Talpa* también están muy preservados en todas las especies de este género. No obstante, se han observado pequeñas diferencias en la cantidad de heterocromatina, que sugieren que algunos cromosomas podrían derivar de otros por adición o amplificación de regiones heterocromáticas. P.ej., se ha observado que *T. europaea* sólo muestra heterocromatina pericentromérica (Zima 1983), mientras que *T. altaica* y *T. occidentalis* presentan también algunas regiones heterocromáticas no relacionadas con las regiones centroméricas (Jiménez et al. 1984; Kawada et al. 2002).

Una característica común a las especies del género *Talpa* es la presencia de una constricción secundaria cerca del centrómero en el tercer par de cromosomas, en la que se localizan las regiones NOR, la cual parece estar presente también en otros géneros de la familia Talpidae (Jiménez et al. 1984; Gornung et al. 2008).

Otro rasgo distintivo en la familia es la conservación de la morfología de los cromosomas sexuales. El cromosoma X es un elemento de tamaño medio generalmente metacéntrico, aunque en algunos casos puede ser submetacéntrico. Por otro lado, el cromosoma Y suele ser de pequeño tamaño y su morfología es difícil de diferenciar (Jiménez et al. 1984; Gornung et al. 2008).

El “chromosome painting”mediante el uso de sondas de cromosomas humanos, ha demostrado la conservación de segmentos sinténicos en varias especies de insectívoros. Esto ha puesto de relieve los reordenamientos cromosómicos que se han producido en los distintos grupos desde su divergencia a partir de un ancestro común, y el alto grado de conservación carotípica de la familia Talpidae con respecto al resto de insectívoros (Ye et al. 2006; Biltueva y Vorobieva 2012). La asociación sinténica de segmentos cromosómicos se interpreta como la conservación de rasgos ancestrales compartidos. En concreto, se ha demostrado que los 22 autosomas humanos se conservan hasta en 54 segmentos en el cariotipo de *T. europaea*. La sonda del cromosoma X humano también produce hibridación, siendo el cromosoma Y el único que no hibrida (Volleth y Müller 2006).

ADN repetitivo.

Consideraciones generales

La cromatina está compuesta por una doble hélice de ADN asociada a histonas, la cual adopta un elevado grado de empaquetamiento en los cromosomas eucarióticos. Este empaquetamiento es mayor en determinadas zonas del cromosoma, como en la heterocromatina constitutiva, que se encuentra principalmente en los centrómeros y los telómeros. Por otro lado, las zonas con menor grado de empaquetamiento, donde se localizan la mayoría de los genes, se denominan eucromatina (Charlesworth et al. 1994). Generalmente, se asume que el ADN que forma la heterocromatina constitutiva está compuesto principalmente por secuencias altamente repetidas, las cuales presentan una gran variabilidad, tienen una función principalmente estructural y son, por lo general, no codificantes (Singer 1982; John y Milkos 1988) (Fig. 2).

El genoma de los eucariotas es extremadamente complejo y altamente variable en contenido y organización, a veces incluso entre especies muy próximas. La cantidad y disposición de los diferentes tipos de secuencias que componen el genoma puede variar por diversos mecanismos. Así, regiones de ADN de tamaño variable pueden ser suprimidas, duplicadas, mutadas o translocadas. Además, los cromosomas pueden sufrir cambios estructurales (p.ej. translocación, deleción, duplicación) o numéricos (p.ej. fusión, fisión, poliploidía). La fijación de este tipo de alteraciones puede crear barreras genéticas, contribuyendo a reducir la fertilidad en los heterocigóticos y a una posible especiación.

El ADN repetitivo se divide en dos clases en función de su organización genómica: ADN interdisperso y ADN organizado en tándem.

a) *ADN interdisperso*: cuando las secuencias repetidas están dispersas por todo el genoma. Dentro de este se encuentran los elementos genéticos móviles y las secuencias repetidas no organizadas en tándem.

- *Elementos genéticos móviles*: con capacidad para movilizarse por todo el genoma. Pueden llegar a representar una fracción muy importante del genoma en mamíferos, p.ej. 45% en el caso del genoma humano (Nurk et al. 2022). Entre ellos tenemos al menos cuatro clases de elementos, que se pueden agrupar en retrotransposones (LINEs, SINEs y retrotransposones tipo LTR) y transposones de ADN, en función de si en su mecanismo de replicación tienen o no un intermediario de ARN, respectivamente. El número de copias en el genoma de estos elementos es muy elevado (p.ej. $1.5x10^6$ copias de secuencias SINEs en el genoma humano).

LINEs (Long Interdispersed Nuclear Elements): elementos interdispersos largos, con un tamaño mayor de 5 kb, con capacidad de retrotransponerse de forma autónoma.

SINEs (Short Interdispersed Nuclear Elements): elementos nucleares interdispersos cortos, con un tamaño que varía entre 80-400 pb, sin capacidad de retrotransposición autónoma; estos elementos dependen de las LINEs.

Retrotransposones LTR (Long Terminal Repeat): también conocidos como retrotransposones de tipo retrovirus porque presentan similitudes en su genoma y en su mecanismo de replicación con los retrovirus.

Transposones de ADN: se replican mediante un mecanismo de copia y pegado, moviéndose directamente de una posición a otra del genoma. Su tamaño oscila en torno a 0,3 Kb.

- *Secuencias repetidas no organizadas en tándem*: multitud de secuencias repetidas existentes en los genomas no relacionadas con la (retro)transposición. Su origen y naturaleza son muy diversas, si bien tienen en común no estar organizadas en tándem, lo que las diferencia del ADN satélite.

b) *ADN organizado en tándem:* formados por una unidad monomérica que se repite en tándem. Se pueden clasificar en cuatro categorías, en función de su tamaño y de su localización:

Microsatélite: repeticiones en tándem cuyo motivo base es menor de 10 pb. Se localizan en numerosas regiones del genoma, pudiendo alcanzar tamaños de hasta 400 pb.

Minisatélite: repeticiones en tándem de una unidad que contiene entre 10-100 pb. Se localizan en numerosos lugares del genoma, como en las regiones intersticiales de los brazos de los cromosomas y en los telómeros. Tienen una longitud muy variable, desde pocos centenares a decenas de miles de pb.

Megasatélite: se denominan así a las repeticiones en tándem más largas, definidas inicialmente como repeticiones directas y contiguas de secuencias de ADN de tres motivos o más, cada una de una longitud individual de al menos 90 pb. La longitud de su motivo base es siempre un múltiplo de tres, y siempre se encuentran dentro de un marco de lectura abierta. Se denominan megasatélites para distinguirlos de los minisatélites, formados por motivos más pequeños y que se encuentran principalmente en regiones intergénicas (Descorps-Declère y Richard 2022).

ADN satélite (ADNsat): compuesto por unidades monoméricas de alta repetición, generalmente con un tamaño de 100-500 pb, que pueden repetirse miles de veces formando bloques de varias megabases.

Los ADNsat se localizan generalmente en las regiones heterocromáticas de los cromosomas y forman parte normalmente de los telómeros (Pons et al. 1993; Hartley y Davidson 1994) y especialmente de los centrómeros (Ugarkovic y Plohl 2002; Galián y Vogler 2003), aunque también pueden tener una localización intersticial, formando regiones o bloques de heterocromatina constitutiva (Reed y Phillips 1995; Marchal et al. 2003; Lamelas et. 2018). También se han descrito secuencias de ADNsat que se encuentran distribuidas a lo largo de la eucromatina de los cromosomas (Ruiz-Ruano et al. 2016; Pita et al. 2017).

Los monómeros de este ADNsat suelen ser variables, pudiendo existir polimorfismos en sus secuencias (Ugarkovic 2005). La existencia de estos polimorfismos y la rápida tasa evolutiva de los ADNsat, hace que estos sean de gran utilidad como marcadores moleculares en medicina forense, así como en estudios poblacionales y de filogenias (Charlesworth et al. 1994; Garrido-Ramos 2017).

Tradicionalmente, el estudio de los ADNsat se ha abordado mediante análisis de restricciones enzimáticas sobre ADN genómico, al objeto de poner de manifiesto bandas de ADN repetitivo; o bien, mediante la construcción de genotecas de ADN Cot (Zwick 1997; Vicari et al. 2010; Garrido-Ramos 2017). Actualmente, las nuevas tecnologías de secuenciación genómica y el desarrollo de programas bioinformáticos específicos como RepeatExplorer o TAREAN (Novák et al. 2010, 2013, 2017), permiten caracterizar el conjunto de secuencias de ADN altamente repetidas a partir de una muestra al azar de lecturas o “reads”, obtenidos de los datos de secuenciación masiva. Esta aproximación experimental se ha convertido hoy día en una herramienta indispensable y de gran utilidad para la caracterización y el análisis de las secuencias repetidas presentes en el genoma, definidas como "repeatoma" por Maumus y Quesneville (2014), y de la fracción correspondiente a ADN satélite, hoy llamado "satelitoma" (Ruiz-Ruano et al. 2016). Por otro lado, el análisis de su organización y distribución cromosómica, requiere de la aplicación posterior de técnicas de citogenética molecular como es la FISH, algo fundamental en aquellas especies en las cuales el genoma no está aún secuenciado o bien no tiene una cobertura o grado de ensamblaje elevado.

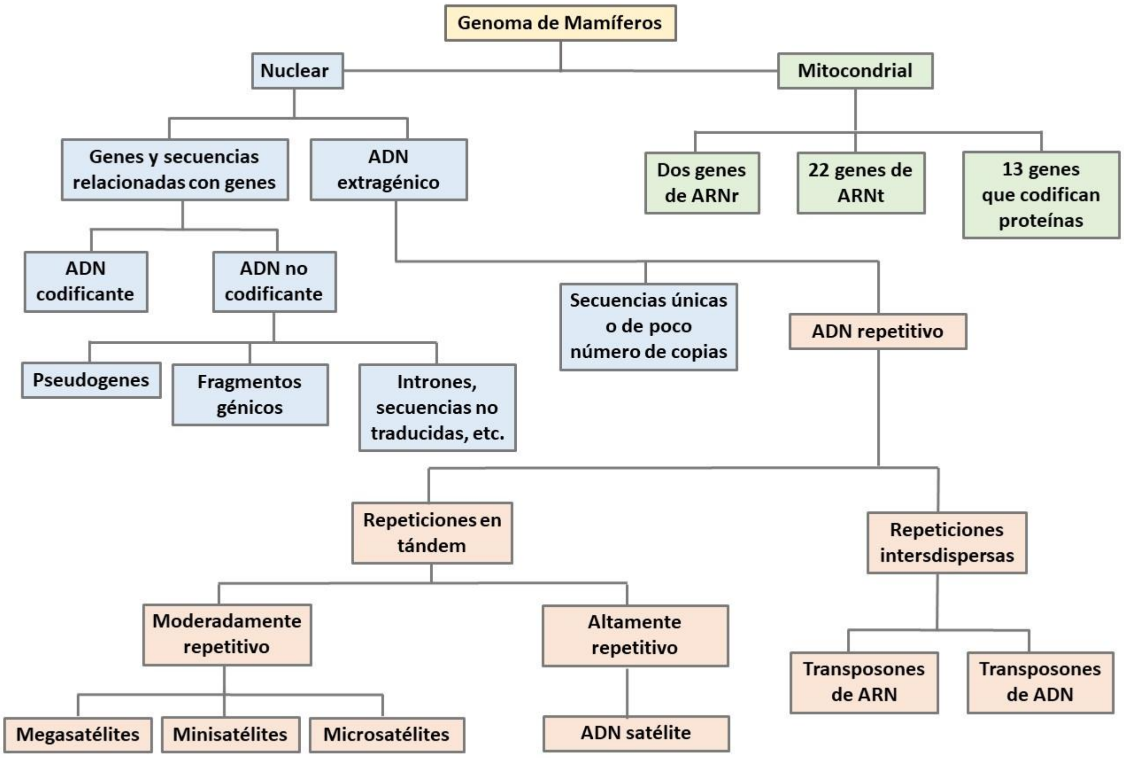


Figura 2. *Organización del genoma eucariota mostrando los diferentes tipos de ADN*

Significado biológico y análisis funcional del ADN repetitivo

El significado biológico del ADN repetitivo y la heterocromatina es un tema de gran interés científico en la actualidad. Esto se debe a que, tradicionalmente, se ha considerado a la heterocromatina constitutiva como un mecanismo defensivo del genoma frente a la expansión de diferentes tipos de secuencias repetidas, consideradas "ADN basura" y sin función específica. Los estudios recientes sobre el significado funcional de la heterocromatina y el ADN repetitivo están revelando un nuevo e interesante resultado. Así, aunque tradicionalmente se ha concebido como "ADN silenciado", cada vez son más los ejemplos, en especies tanto de insectos, plantas o mamíferos, de secuencias repetidas centroméricas o teloméricas que son transcritas (Ugarkovic y Plohl 2002; Dimitri et al. 2005; Palomeque y Lorite 2008; Plohl et al. 2008; Garrido-Ramos 2017; Shatskikh et al. 2020). La transcripción de secuencias repetidas parece ser un fenómeno bastante generalizado: así, los datos del proyecto ENCODE muestran que la mayoría de regiones del genoma humano analizadas en dicho estudio son transcritas, existiendo una gran cantidad y variedad de ARNs no codificantes cuyo significado funcional se desconoce (Mighell et al. 2000; Balakirev et al. 2003; Zheng et al. 2005; The ENCODE Consortium 2007).

La función del ADNsat es uno de los temas más discutidos (Biscotti et al. 2015a y b). La transcripción de estas secuencias podría estar asociada a algunas de las funciones que se han propuesto para el ADNsat. Así, parece estar relacionado con el correcto funcionamiento del mecanismo de ensamblaje de la heterocromatina en levaduras (Ekwall et al. 1995; Cam et al. 2005). También podría estar implicado en el mantenimiento de la cromatina y la formación del centrómero, así como la del cinetocoro y en el mantenimiento de la integridad de los genomas (Ugarkovic 2005; Plohl et al. 2012; Rosic et al. 2014; Perea-Resa y Blower 2017; Louzada et al. 2020). En ratón y en marsupiales, está implicado en la regulación de la correcta función del centrómero a través de la interacción con proteínas centroméricas y del cinetocoro (Bernard et al. 2001; Carone et al. 2009). Además, se ha propuesto la posible participación de este tipo de secuencias en diversos fenómenos tales como la condensación y la función centromérica, la cohesión entre cromátidas hermanas, la arquitectura nuclear, la especiación y la evolución cariotípica e incluso en la regulación de la expresión génica (Guenatri et al. 2004; Grewal et al. 2007; Josse et al. 2007). Probablemente, donde más se ha estudiado este y otros aspectos, es en el satélite α humano, donde se ha establecido que su transcripción y los correspondientes ARNs no codificantes, desempeñan funciones distintas e imprescindibles en el centrómero y en la región pericentromérica a lo largo del ciclo celular (McNulty et al. 2018; Gambogi et al. 2020; Ohzeki et al. 2020).

En la mayoría de las especies, el ADNsat se transcribe diferencialmente en un tipo particular de célula, tejido u órgano, o se expresa temporalmente en etapas concretas del desarrollo. Además, la transcripción de secuencias repetidas heterocromáticas es un fenómeno que puede dar lugar a una gran diversidad de ARNs. Por tanto, esta transcripción parece estar regulada de forma muy compleja y variable. Así, se ha propuesto que está influida por el ciclo celular, por el tipo celular, por factores externos causantes de estrés celular, por la etapa del desarrollo y por el mecanismo del ARN de interferencia (Rudert et al. 1995; Biamonti 2005; Usakin et al. 2007; Lu y Gilbert 2008; Martínez-Guitarte et al. 2008).

La regulación de la transcripción de secuencias repetidas a lo largo el ciclo celular es un fenómeno interesante y poco estudiado. Así, en ratón se ha demostrado que los transcritos del SatIII presentan picos dependiendo del estadio del ciclo en que se encuentran las células (Zhang et al. 2002; Lu y Gilbert 2007; Shi et al. 2007). Estos transcritos podrían ser importantes para la correcta segregación cromosómica (Prasanth et al. 2003), para el mantenimiento de la estructura centromérica (Grady et al. 1992; Wong et al. 2007), o para el mantenimiento de las conexiones entre cromosomas durante la mitosis (Craig et al. 1999; Abad et al. 2000).

Finalmente, también hay que considerar la posible implicación del mecanismo del ARN de interferencia en la regulación de la transcripción de las secuencias heterocromáticas, y en el procesamiento postranscripcional de los transcritos (Djupedal et al. 2009).

Estudios sobre ADN repetitivo en insectívoros

Telomérico y ribosómico

Los estudios de ADN repetitivo en el género *Talpa* son escasos. Se han realizado algunos trabajos sobre la distribución y localización de los principales genes ribosómicos y las secuencias teloméricas (Zurita et al. 1997). Las secuencias teloméricas están generalmente situadas en las regiones terminales de los cromosomas, aunque en *T. romana* y *T. europaea* también se encuentran como secuencias teloméricas intersticiales (ITSs) en las regiones pericentroméricas de algunos cromosomas (Gornung et al. 2008). Los genes ribosómicos 5S sólo se han estudiado en *T. romana* y *T. europaea*, mostrando pocas diferencias en su localización cromosómica en ambas especies (Gornung et al. 2008). Sobre la base de las diferencias en las ITSs y el ribosómico 5S observadas en *T. romana* y *T. europaea*, se ha propuesto que la evolución cariotípica en *Talpa* podría estar asociada principalmente a cambios/eventos genómicos que no implicarían grandes reordenaciones del cariotipo (Gornung et al. 2008).

Elementos genéticos móviles

También se ha descrito la presencia de elementos genéticos móviles en el genoma de varias familias de insectívoros. En particular, se han analizado los retrotransposones SINE, que tendrían características específicas para cada familia, habiéndose utilizado para establecer relaciones filogenéticas (Borodulina y Kramerov 2001; Bannikova et al. 2005).

ADN satélite

Los estudios sobre las secuencias de ADN satélite en estas especies también son muy escasos. Se han descrito microsatélites en especies del género *Sorex* (Balloux et al. 1998; Wyttenbach et al. 1999), y también hay secuencias disponibles en bases de datos de ADN satélite de *Crocidura sibirica* (GenBanK_nºde acceso: Y14805 y Y14806), si bien no está publicado y se desconocen su organización y su localización cromosómica. Por último, en *Urotrichus talpoides* y *Dymecodon pilirostris*, ambos de la familia Talpidae, se han localizado mediante FISH, bandas de secuencias altamente repetidas de 0,7, 0,9 y 1,4 Kb, obtenidas por restricción del ADN genómico. La repetición de 0,7 Kb se encuentra en las regiones pericentroméricas de la mayoría de los cromosomas, mientras que las otras dos se localizan preferentemente en las regiones heterocromáticas no pericentroméricas del cromosoma 13. Estas secuencias no han sido clonadas ni secuenciadas y se desconoce su composición y organización molecular (Nakata et al. 2005).

Recientemente se ha secuenciado y anotado el genoma de la especie *T. occidentalis*, el cual tiene un tamaño de unas 2.099 gigabases (Real et al. 2020). Aproximadamente el 61% del genoma está compuesto por secuencias únicas y el 39% por secuencias repetidas. De estas secuencias repetidas, aproximadamente el 30% son elementos transponibles, al igual que ocurre en otras especies del clado. Las LINEs son las más abundantes, representando ellas solas el 15% del total del ensamblaje, seguidos de las secuencias ALUs y las SINEs (Real et al. 2020).

Secuenciación de genomas mitocondriales en Talpidae

Tradicionalmente, los genomas mitocondriales o mitogenomas completos se han secuenciado utilizando amplificaciones por PCR largas. Para ello, se utilizaba una batería de primers, que permitía amplificar el genoma mitocondrial completo en varios fragmentos solapantes (Simon et al. 1994, 2006; Cabria et al. 2006; Kim y Park 2015). Actualmente, el analisis de los resultados de secuenciación masiva permite ensamblar los mitogenomas a partir de los datos de la secuenciación del genoma completo, y esto incluso cuando la cobertura de la secuenciación es baja debido a la alta porción de copias del mitogenoma en las muestras de ADN genómico (Straub et al. 2012; Ye et al. 2014; Fernández-Pérez et al. 2017). Los genomas mitocondriales completos son útiles para las reconstrucciones filogenéticas; sin embargo, en la familia Talpidae, el número de especies con mitogenomas completos descritos es aún reducido. P.ej., en el género *Talpa,* solo se había descrito el mitogenona de *T. europaea* (Mouchaty et al. 2000).

Capítulo 2: Objetivos

OBJETIVOS

A la vista de lo anteriormente expuesto, así como al reciente descubrimiento de la nueva especie *T. aquitania,* distribuida en el norte de la Península Ibérica y el sur de Francia (Nicolas et al. 2015, 2017b), de la que no hay estudios a nivel de citogenética y genética molecular, y dado que no hay datos de la mayoría de insectivoros sobre secuencias repetidas y mitogenomas, en esta Tesis Doctoral nos hemos propuesto los siguientes objetivos:

1- Descripción del cariotipo de la nueva especie *Talpa aquitania* y análisis comparativo con *Talpa europaea* y *Talpa occidentalis.*

2- Análisis de la distribución cromosómica de secuencias repetidas en varias especies del género *Talpa.*

3- Caracterización del satelitoma de *Talpa aquitania* y análisis comparativo con otras especies del género *Talpa*.

4. Descripción de los mitogenomas *de Talpa aquitania* y *Talpa occidentalis*.

Capítulo 3: Material y métodos

MATERIAL Y MÉTODOS

MATERIAL

Las especies con las que hemos trabajado para llevar a cabo nuestra investigación pertenecen al género *Talpa* (Insectivora, Talpidae). En concreto, ejemplares de *T. occidentalis* (Granada), *T. aquitania* (Torme, Burgos) *T. europaea* (Pavia, Italia) y *T. romana* (Roma, Italia).

Los permisos para la captura de los ejemplares de *T. occidentalis* se obtuvieron de la Consejería de Medio Ambiente de la Junta de Andalucía (Referencias: SGYB-AFR-CMM y SGMN/GyB/JMIF) y el de *T. aquitania* de la Delegación Territorial de Medio Ambiente de Burgos, Junta de Castilla-León (Referencia: EN/082/18). Los protocolos de captura, manejo y sacrificio de los animales han sido aprobados por el Comité de Ética de Experimentación Animal de la Junta de Andalucía (code: 22/05/2018/094).

El material de *T. europaea* pertenece a ejemplares capturados por el Dr. Antonio Sánchez durante una estancia postdoctoral en 1994 en Pavia (Italia) y el de *T. romana* fue suministrado por el Dr. Riccardo Castiglia de la Universidad La Sapienza (Roma).

MÉTODOS

Extracciones de ADN

Para extraer el ADN total de los tejidos (almacenados en etanol al 100% a -20 °C) se han utilizado el método tradicional de fenol-cloroformo y diferentes kits comerciales siguiendo las instrucciones de los fabricantes (DNeasyBlood&Tissue Kit (Qiagen), GentraPuregeneTissue Kit (Qiagen) y el Quick-DNA™ Tissue/InsectMiniprep Kit (ZymoResearch)).

La concentración y pureza del ADN genómico extraído fue determinada mediante la lectura con NanoDropd con la absorbancia a 260 y 280 nm.

Caracterización de los ejemplares de T. aquitania

Tradicionalmente, los topos del norte de la Península Ibérica y del sur de Francia se han considerado como pertenecientes a la especie *T. europaea.* Recientemente se ha descrito la nueva especie *T. aquitania*, emparentada con *T. occidentalis*, que se encuentra distribuida en estas mismas regiones (Nicolas et al. 2015, 2017b). Por tanto, para determinar si los ejemplares capturados en Torme (Burgos) pertenecían a *T. europaea* o a la nueva especie *T. aquitania,* se ha hecho necesaria la caracterización morfológica y molecular de los mismos.

La caracterización morfológica se realizó comparando los mesostilos de los molares superiores (M1, M2 y M3) de los ejemplares de Torme, con los de las especies *T. europaea* y *T. occidentalis.* Para ello, se hirvieron los cráneos y se les extrajeron los tejidos, posteriormente se blanquearon por inmersión en una solución de peróxido de hidrógeno al 10%. Finalmente, las mandíbulas superiores se fotografiaron en un estereomicroscopio.

La caracterización molecular se realizó mediante identificación de 16 posiciones diagnósticas del gen del citocromo b del genoma mitocondrial, descritas por Nicolas et al. (2015, 2017b). Este gen se amplificó por PCR a partir de ADN genómico con los primers L14723 y H15915, tal y como describieron Ducroz et al. (2001). Los fragmentos amplificados se aislaron del gel de agarosa con el QIAquick Gel Extraction Kit (Qiagen) y se clonaron en la bacteria JM109 utilizando el vector PGEMT-easy (Promega). Los clones positivos se secuenciaron en ambas direcciones.

Preparaciones cromosómicas y bandeo G y C

Los cromosomas de *T. occidentalis, T. aquitania,* y *T. romana* se prepararon a partir de células de médula ósea siguiendo el método descrito por Burgos et al. (1986). En nuestro estudio no hemos podido disponer de preparaciones cromosómicas de *T. europaea.* El bandeo G y C de los cromosomas de *T. aquitania* se realizó en portaobjetos envejecidos durante una noche a 60°C siguiendo los protocolos de Burgos et al. (1986) y Sumner (1972), respectivamente.

FISH y Chromosome Painting

Se han realizado FISH indirectas, tal y como describieron Fernández et al. (2001), Rovatsos et al. (2011) y Cabral-de-Mello y Marec (2021). Se han utilizado sondas marcadas con biotin-16-dUTP o digoxigenin-11-dUTP mediante amplificación de las secuencias repetidas a analizar por PCR, Nick translation y transferasa terminal.

El "chromosome painting" se realizó con una sonda preparada a partir del cromosoma Y de *T. occidentalis siguiendo* el protocolo de FISH directo descrito por Marchal et al. (2004). Para ello, se microdiseccionaron y amplificaron 15 cromosomas Y y se marcaron mediante DOP-PCR utilizando Spectrum-Orange dUTP (Abbott).

Las imágenes de FISH se capturaron con un microscopio de fluorescencia (Olympus BX51) equipado con una cámara CCD (Olympus DP70).

Identificación de ADN repetitivo

a) Mediante digestión con enzimas de restricción

La primera aproximación para abordar el estudio del ADN repetitivo fue la tradicional, mediante digestiones con enzimas de restricción. De trabajos previos de nuestro grupo de investigación conocíamos que la digestión del ADN genómico de los *Talpa* con *Hind*III producía en la electroforesis varias bandas de secuencias de ADN repetitivo (1, 0,8, 0,6 y 0,4 kb). Así pues, para poder clonar y caracterizar estas secuencias se realizaron digestiones del ADN genómico de *T. occidentalis, T. europaea, T. aquitania* y *T. romana* con la endonucleasa de restricción *Hind*III. Primeramente nos centramos en la banda de 1 kb, aislando esta banda de *T. europaea del* gel de agarosa y clonándola en el vector pGEM-T easy (Promega), como describieron anteriormente Sánchez et al. (1996). Finalmente se transformaron bacterias JM109 competentes (Mix&GoCompetentCells; ZymoResearch) y los clones positivos se identificaron utilizando la misma banda marcada con digoxigenina como sonda**.**

SouthernBlot

El ADN genómico fue digerido con la endonucleasa de restricción *Hind*III. Los fragmentos resultantes se separaron en geles de agarosa al 1% y se colocaron en membranas de nylon (Amersham), según método descrito por Bullejos et al. (1997). Las membranas se incubaron durante la noche a 55 °C utilizando como sonda un clon positivo marcado con digoxigenina por PCR. La detección de la fosfatasa alcalina se realizó según las recomendaciones del proveedor (Roche).

Análisis de la secuencia de los clones

Los plásmidos de los clones bacterianos se secuenciaron con el método Sanger utilizando los cebadores universales T7 y SP6 y se revisaron y analizaron con el programa Bioedit (versión 7.0.9.0) (Hall 1999) (http://www.mbio.ncsu.edu/BioEdit/bioedit.html). Se realizaron alineamientos de secuencias con el programa Clustal-Omega (Sievers et al. 2011) y con Bioedit. Las secuencias de ADN repetidas se examinaron con el programa RepeatMasker versión 4.0.6 (http://repeatmasker.org). Se realizaron búsquedas de secuencias en GenBank con BLASTN (Zhang et al. 2000) y BLASTP (Altschul et al. 1997) (https://blast.ncbi.nlm.nih.gov/Blast.cgi).

b) *Mediante Repeat-Explorer con datos de Secuenciación Illumina*

El ADN genómico de *T. aquitania* y *T. occidentalis* se secuenció usando la plataforma Illumina® Hiseq ™ 2000 o 2500, con librerías de 350 pb y 750 pb, respectivamente, y lecturas de 100 o 150 pb.

La determinación de secuencias de ADN repetitivo a partir de los datos de secuenciación masiva se realizó usando el Programa RepeatExplorer (www.repeatexplorer.org) (Novák et al. 2010, 2013, 2020). Una muestra tomada al azar de secuencias pareadas se usó para agruparlas en clústeres de RepeatExplorer. Se utilizaron las condiciones por defecto, seleccionando que se analizasen clústeres que tuvieran una representación superior al del 0,001% del genoma. El análisis con el programa TAREAN (Novák et al. 2017) implementado en RepeatExplorer, identifica directamente algunos clústeres que contienen ADN satélites. Todos los clústeres fueron analizados utilizando diferentes programas bioinformáticos que nos permitieron determinar qué tipo de secuencias repetidas contenían. Para estos análisis se han utilizado los siguientes programas: Geneious (Kearse et al. 2012); RepeatMasker versión 4.0.9 (http://repeatmasker.org;https://www.dfam.org/) (Storer et al. 2021); CENSOR (http://www.girinst.org/); Dotmatcher (http://emboss.bioinformatics.nl/cgi-bin/emboss/dotmatcher/); Clustal-Omega (https://www.ebi.ac.uk/Tools/msa/clustalo/) (Sievers et al. 2011; Madeira et al. 2019); Bioedit (versión 7.2.5) (http://www.mbio.ncsu.edu/BioEdit/bioedit.html) (Hall 1999); TandemRepeatsFinder (Benson 1999) (https://tandem.bu.edu/trf/trf.html). Finalmente, todas las secuencias consenso obtenidas se usaron como entrada en las bases de datos del GenBank/NCBI DNA con la herramienta BLAST (http://www.ncbi.nlm.nih.gov/), así como con la base de datos del EMBL (Altschul et al. 1997).

La secuencia consenso de cada familia de ADNsat se utilizó para diseñar uno (secuencias consenso pequeñas) o dos oligonucleótidos (secuencias consenso grandes), utilizando el software Primer3 (Untergasser et al. 2012). Para la localización cromosómica por FISH de las familias de ADNsat, los primers de las familias de monómero pequeño fueron marcados por transferasa terminal y usados como sonda, mientras que para las familias de monómero grande se generaron sondas amplificando y marcando por PCR cada ADNsat con las parejas de primers correspondientes.

ADN mitocondrial

Dada la alta frecuencia de las lecturas del ADN mitocondrial (ADNmt) en la secuenciación Illumina, RepeatExplorer las agrupa en clústeres conteniendo generalmente un contig consenso. Para ensamblar los genomas mitocondriales, se seleccionaron los clústeres que contenían secuencias correspondientes al ADNmt de *T. occidentalis* y *T. aquitania*, utilizando el mitogenoma completo de *T. europaea* como referencia (Mouchaty et al. 2000; número de acceso: Y19192) y posteriormente se ensamblaron y revisaron manualmente. En el caso del mitogenoma de *T. aquitania* también realizamos el ensamblaje con el programa MITObim v1.8 (Hahn et al. 2013) con una selección aleatoria de un millón de pares de lecturas. Utilizando como referencia los mitogenomas de *T. europaea* y *T. occidentalis* (Mouchaty et al. 2000; Gutiérrez et al. 2018) en ejecuciones independientes, se generaron dos secuencias similares, que fueron alineadas con la secuencia del mitogenoma procedente del ensamblaje de RepeatExplorer.

Una parte de la región control (D-loop) del genoma mitocondrial no se pudo ensamblar correctamente y tuvo que completarse mediante amplificaciones por PCR. Se hizo a partir de los mismos ADNs que se utilizaron para la secuenciación masiva, utilizando los cebadores Pro+ (5'-ACCATCAGCACCCAAAGCTG-3') y Phe- (5'-AAGCATTTTCAGTGCTTTGCTT-3'), según lo descrito anteriormente por Haring et al. (2000). La PCR se realizó con 50 µl de mezcla de reacción que contenía 100 ng de ADN genómico, 10 pmol de cada cebador, 1 µl de dNTPs 10 mM, 1 µl de DMSO, 5 µl de tampón de reacción 10× NH4, 2,5 µl de MgCl2 50 mM y 5 U de BIOTAQ™ DNA Polymerase (Bioline). El programa de PCR utilizado fue de 5 min a 95 °C y 30 ciclos: 30 s a 95 °C, 30 s a 55 °C y 90 s a 72 °C, con una extensión final de 5 min a 72 °C. Los productos de la PCR se visualizaron en geles de agarosa teñidos con bromuro de etidio al 1%; las bandas correspondientes se clonaron con pGEM-T easy en JM109 como se ha descrito anteriormente. Las secuencias se analizaron con el programa Bioedit (versión 7.0.9.0) (http://www.mbio.ncsu.edu/BioEdit/bioedit.html) y finalmente se ensamblaron con el resto del mitogenoma. Las versiones circularizadas finales se confirmaron mediante inspección manual y alineamiento con Clustal-Omega con otros mitocondriales disponibles en GenBank.

Anotación y análisis del ADNmt

La anotación del ADNmt de *T. occidentalis* y *T. aquitania* se realizó utilizando las herramientas de la web MITOS (http://mitos.bioinf.uni-leipzig.de/help.py) (Bernt et al. 2013) y tRNAscan-SE (http://lowelab.ucsc.edu/tRNAscan-SE/) (Lowe y Eddy 1997). Las anotaciones de los genes codificadores de proteínas (PCG), los ARN de transferencia (ARNt) y los genes de ARNr se refinaron comparándolos manualmente con el mitogenoma de *T. europaea*. La composición de bases y el uso de codones se analizaron utilizando la versión 7 de MEGA (Kumar et al. 2016). El dibujo circularizado del mitogenoma se realizó utilizando las herramientas OrganellarGenomeDRAW (http://ogdraw.mpimp-golm.mpg.de/) (Lohse et al. 2013).

Análisis filogenético

Para el análisis filogenético de los ANDmt obtenidos empleamos los genomas mitocondriales completos y las secuencias concatenadas de los 13 genes que codifican para proteínas (PCGs) de especies representativas de insectívoros disponibles en GenBank. Las secuencias se alinearon con ClustalW-Omega y las posiciones mal alineadas y las regiones divergentes se eliminaron con el programa Gblocks v 0.91b (http://molevol.cmima.csic.es/castresana/Gblocks_server.html) (Talavera y Castresana 2007). Los modelos de sustitución de nucleótidos fueron evaluados con la versión X de MEGA (Kumar et al. 2018). Los modelos con bajas puntuaciones BIC (Criterio de Información Bayesiano) fueron los que mejor describían el patrón de sustitución. En todos los los alineamientos el mejor modelo de sustitución de nucleótidos fue GTR+G+I. Los métodos de Máxima Verosimilitud y Neighbor-Joining fueron implementados en el programa MEGA versión X (Kumar et al. 2018), y la inferencia bayesiana (BI) fue implementada en MrBayes versión 3.1 (Ronquist y Huelsenbeck 2003).

Capítulo 4: Resultados y discusión

Karyotype analysis of the new Talpa species Talpa aquitania (Talpidae; Insectivora) from northern Spain

Authors: Juana Gutiérrez, Gaël Aleix-Mata, Luz Lamelas, María Arroyo, Juan Alberto Marchal, Antonio Sánchez

Departamento de Biología Experimental, Área de Genética, Universidad de Jaén, Paraje de las Lagunillas s/n., 23071 Jaén, Spain.

Cytogenet Genome Res (10.1159/000502599)

Cytogenetic and Genome Research

Original Article

Cytogenet Genome Res 2019;159:26–31
DOI: 10.1159/000502599

Accepted: August 3, 2019
Published online: September 18, 2019

Karyotype Analysis of the New *Talpa* Species *Talpa aquitania* (Talpidae; Insectivora) from Northern Spain

Juana Gutiérrez Gaël Aleix-Mata Luz Lamelas María Arroyo Juan A. Marchal Antonio Sánchez

Departamento de Biología Experimental, Área de Genética, Universidad de Jaén, Jaén, Spain

Abstract

Karyotypes of three male specimens of *Talpa* from northern Spain were analyzed. The mesostyles of upper molars and cytochrome b sequenced analysis identified these specimens as belonging to *Talpa aquitania*, a new *Talpa* species recently described from northern Spain and southern France. We describe here for the first time the karyotype of *T.aquitania*. Its diploid number is 2n = 34 and NFa = 64 and all chromosomes including the sex chromosomes are biarmed, either metacentric or submetacentric. G-bands demonstrated that the karyotypes of *T. aquitania* and *T. occidentalis* (the most closely related species) are almost identical. However, the karyotype of *T. aquitania* differs from the karyotypes of both *T. europaea* and *T. occidentalis* in that it has a medium-sized biarmed Y chromosome rather than a dot-like chromosome, and that the chromosome 16 is submetacentric in *T. aquitania* but have a small p-arm in both *T. europaea* and *T. occidentalis*. Pericentromeric C-bands were scarce and only clearly visible in a few chromosomal pairs. In addition, C-banding demonstrated that half of the 14p, the 16p, and the Y chromosome are all heterochromatic. rDNA genes were located at the secondary constriction in autosomal pair 3, a common feature in the karyotypes of all *Talpa* species. Hybridization signals for telomeric repeats were found on the telomeres and the pericentric regions of some chromosomes, and co-localized in the secondary constriction of pair 3 with the rDNA genes. In conclusion, the karyotype of *T. aquitania* from northern Spain is very similar to the karyotype of other species belonging to the genus *Talpa*.

Keywords: C-bands; G-bands; heterochromatin; Insectivora; karyotype; rDNA genes; *Talpa aquitania*; Talpidae; telomeric sequences.

Introduction

Until recently, the genus *Talpa* (Insectivora; Talpidae) included nine species of fossorial insectivores (Hutterer 2005; Wilson and Reeder 2005): *T. altaica* (Nikolasky 1883), *T. caucasica* (Satunin 1908), *T. levantis* (Thomas 1906), *T. davidiana* (Milne-Edwards 1884), *T. caeca* (Savi 1882), *T. europaea* (Linnaeus 1758), *T. occidentalis* (Cabrera 1907), *T. romana* (Thomas 1965) and *T. stankovici* (Martino and Martino 1931). Of the five *Talpa* species occurring in Europe, *T. europaea* is widely distributed in the centre, while *T. caeca, T. occidentalis, T. romana* and *T. stankovici* are restricted to the southern part of the continent. Recently, a combination of morphological and molecular methods has led to the description of two new *Talpa* species in Europe, *T. aquitania* (Hugot et al. 2014; Nicolas et al. 2015, 2017a y b) and *T. martinorum* (Kryštufek et al. 2018). *T. aquitania* is restricted to the west and south of the river Loire and reaches northern Spain (Nicolas et al. 2015). Before the description of *T. aquitania* as a new species, all *Talpa* specimens from this area were regarded as *T. europaea*.

The karyotype of *Talpa* species has been well studied and is extremely well conserved. Most species have 2n = 34 and a FNa of 62–66. Only two species, *T. caeca* and *T. caucasica*, have different diploid numbers, 2n = 36 and 2n = 38, respectively. The X chromosome is typically medium-sized and biarmed, while the Y chromosome is a dot-like chromosome with no distinguishable morphology, although it is described as metacentric, submetacentric or acrocentric depending on the species (see review in Gornung et al. 2008). Karyotype evolution of *Talpa* is characterized by Robertsonian translocations and by variations in the morphology and size of the sex chromosomes (Gornung et al. 2008).

Here we describe for the first time the karyotype of specimens of *T. aquitania* from northern Spain using a combination of Giemsa staining, C-banding and FISH with telomeric and rDNA sequences.

Materials and methods

Specimens analyzed

In this study, three male *T. aquitania* moles (TA-1, TA-2 and TA-3) captured in Torme (42°59'36.7"N 3°33'47.2"W) (Burgos, Spain) were analyzed. Permission for captures was granted by the Servicio Territorial de Medio Ambiente de Burgos (Junta de Castilla-León). We also used material from *T. occidentalis* specimens captured in Granada (southern Spain) with the permission from the Delegación Territorial de la Consejería de Agricultura, Pesca y Medio

Ambiente de Granada (Junta de Andalucía). All capture and sacrifice protocols were approved by the Junta de Andalucía Ethics Committee for Animal Experimentation (code: 22/05/2018/094).

Identification of specimens

Morphological (mesostyles of upper molars) and molecular (Cytb sequences) markers were used in the specific identification of the moles.

Molar identification

The skulls of the analyzed specimens were boiled in water and tissue was removed with tweezers and a scalpel. Finally, skulls were bleached by immersion in a solution of 10% hydrogen peroxide, and the upper mandibles were photographed under a stereomicroscope.

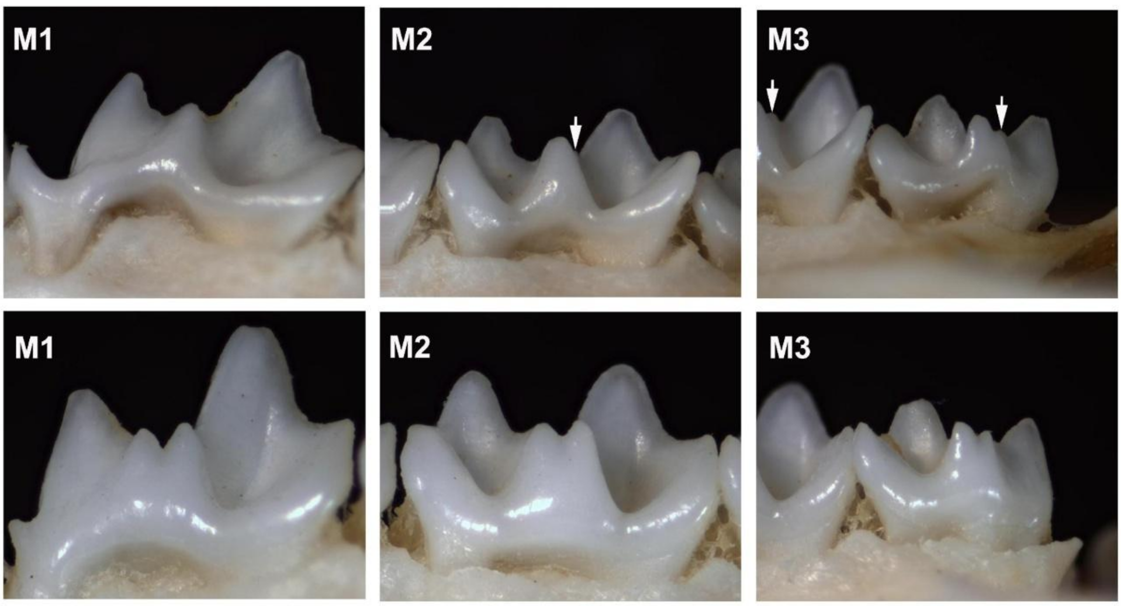


Figure 1. *The three upper molars (M1, M2 and M3) of T. aquitania (above) and T. occidentalis (below). The mesostyles of the upper molars in T. occidentalis are double, while in T. aquitania these molars are simple, although with an additional small cusp on M2 and M3 (arrows).*

Cytb sequences

Total DNA was extracted from liver tissues (stored in 100% ethanol at -20°C) using the DNeasy Blood & Tissue Kit (Qiagen). The Cytb gene of the mitochondrial genome was amplified using the PCR of one of the male specimen (TA-3) and the primer pairs L14723 and H15915, as previously described by Ducroz et al., (2001). Amplified fragments were resolved in ethidium bromide-stained 1% agarose gels; the appropriate bands were isolated from the gel with QIAquick gel extraction kit (Qiagen) and cloned in JM109 bacteria using PGEMT-easy vector (Promega). Positive clones were sequenced in both directions (GenBank accession numbers: MK747361 and MK747362). Sequences were analyzed using the Bioedit program (version 7.0.9.0) (http://www.mbio.ncsu.edu/BioEdit/bioedit.html).

```
                                                 11
                                   122333344788900
                                  8778025669412847
                                  7162981952538174
Taqu-Torme-c10 (Spain)            CGCGTTAATCTGCATT
Taqu-Torme-c11 (Spain)            ................
Taqu-Montseny (Spain)             ................
Taqu-Sant-Pere-De-Vilamajor (Spain) ................
Taqu-Petritegui (Spain)           ................
Taqu-Gironde (France)             ................
Taqu-Creste (France)              ................
Taqu-Avensan (France)             ................
Tocc-Granada (Spain)              AATACAGGCTCATTCC
Tocc-Fajera (Spain)               AATACAGGCTCATTCC
Tocc-Vegadali (Spain)             AATACAGGCTCATTCC
Tocc-Merón (Spain)                AATACAGGCTCATTCC
Teur-Grieges (France)             AATACAGGCTCATTCC
Teur-Saint-Malo (France)          AATACAGGCTCATTCC
Teur-Pont-sur-Veyle (France)      AATACAGGCTCATTCC
Teur-Kiev (Ukraine)               AATACAGGCTCATTCC
```

***Figure 2**. Alignment of the 16 key base positions of the Cytb sequence differentiating T. aquitania from T. europaea and T. occidentalis. This alignment includes the two T. aquitania clones from Torme, six T. aquitania sequences, of which three were from Spain (KF801509; KF801507; KU189592) and three from France (KU189684; KU189628; KU189687), four of T. occidentalis (NC_039630; KU189721; KU189719; KU189715) and four of T. europaea (KU189567; KU189480; KU189561; KF801573). The place of capture of each specimen is shown in the figure.*

Chromosome preparations, banding and FISH

Chromosomes were prepared from bone marrow cells following the method described by Burgos et al. (1986).G-banding and C-banding were performed on slides aged overnight at 60°C following the Burgos et al. (1986) and Sumner's (1972) protocols, respectively. FISH was performed with an rDNA probe (plasmid pDmra.51#1 containing rDNA of *Drosophila melanogaster* (Endow 1982) labelled with the biotin-nick-translation kit (Roche)) and with a telomeric probe (generated and labelled with biotin by PCR as described by Ijdo et al. 1991) following the method used by Rovatsos et al. (2011).

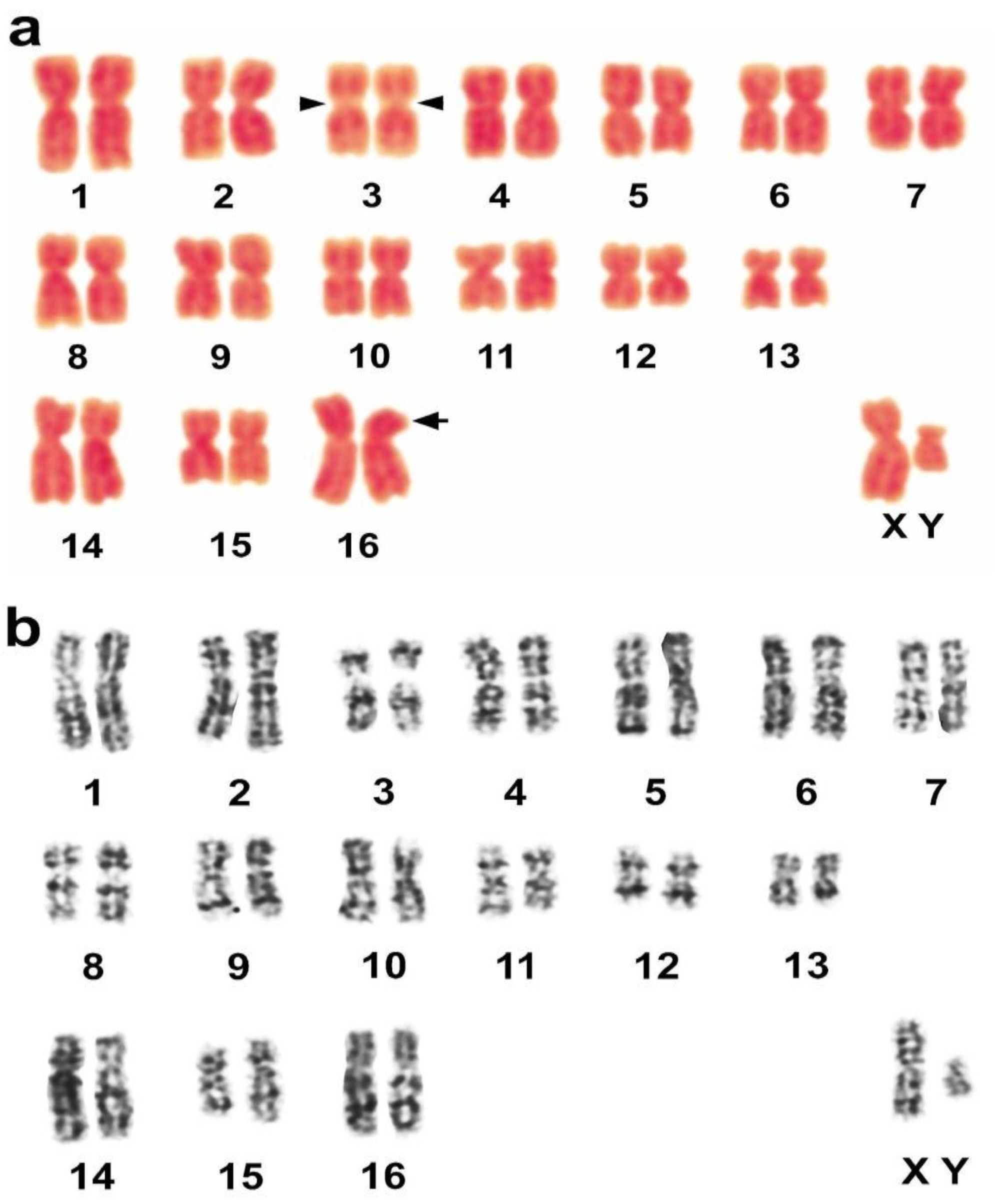


Figure 3. *Karyotypes of T. aquitania males with 2n = 34:* ***a)*** *Giemsa staining Karyotype (TA-3 specimen) and* ***b)*** *G-banded Karyotype (TA-1 specimen)***.** *Arrow heads: secondary constriction of pair 3 of the chromosomes; arrow: polymorphism of chromosome pair 16 showing size reduction of the heterochromatic block in the short arm.*

Results and discussion

Identification of specimens

The mesostyles of the three upper molars of the analyzed moles were compared with those of *T. occidentalis* moles from Granada (Spain). The mesostyles of the upper molars in *T. occidentalis* are double but were simple in the specimens from Torme; these moles' molars also had an additional small cusp on M2 and M3 (Fig. 1) on the crest that connects the mesostyle to the metacone, as described Nicolas et al. (2017a).

In *T. aquitania* the mesostyle of the upper first molar is simple, as in *Talpa europaea*, although in this latter species the mesostyles of the upper second (M2) and third (M3) molars are double (Nicolas et al. 2017a).

The Cytb sequences obtained from the TA-3 *Talpa* sp. captured in Torme were aligned with sequences from GenBank belonging to *T. europaea* (four specimens), *T. occidentalis* (four specimens) and *T. aquitania* (3 sequences from specimens from Spain and 3 more from France), as described previously by Nicolas et al. (2015, 2017a) (Fig. 2). The alignment of the sequences demonstrated that the 16 key base positions of the Cytb sequence that differentiate *T. aquitania* from *T. europaea* and *T. occidentalis* (Nicolas et al. 2015, 2017b) are all conserved in the moles from Torme and clearly differ from those found in the other two species.

Hence, both the morphology of the molars and Cytb sequences analysis clearly demonstrate that the moles from Torme in northern Spain belong to the new species *T. aquitania*.

Karyotype

To date, the moles from the north of the Iberian Peninsula were identified as belonging to *T. europaea* (Feuda et al. 2015), even though no chromosomal analyses had ever been performed. Since the recent description of *T. aquitania* in northern Spain as a new sister species of *T. occidentalis* (Nicolas et al. 2017a), there has been a need to analyse the karyotype of moles from northern Spain formally described as *T. aquitania.* The karyotype, arranged according to Jiménez et al. 1984, of the three specimens analyzed in this study have 2n= 34 and NFa= 64. All the chromosomes, including the sex chromosomes, are bi-armed, and either metacentric or submetacentric (Fig. 3a). The chromosome number 2n= 34 is very common in the genus *Talpa*, being present in eight species including *T. aquitania*. Only *T. caeca* and *T.*

caucasica differ, having 2n= 36 and 2n= 38, respectively (Gornung et al. 2008), due to the presence in *T. caeca* of a very small pair of chromosomes (Meylan 1966) and five acrocentric pairs in *T. caucasica* (Arslan and Zima 2014). The NFa varies between 62 (*T. romana* and *T. davidiana*) and 66 (*T. caeca*) (Fedyk and Ivanitskaya 1972; Gornung et al. 2008; Arslan and Zima 2014), although the most abundant fundamental number is 64, which is found in seven species, including *T. aquitania*.

In *T. aquitania*, one of the largest chromosomes (pair 3) has a secondary constriction (Fig. 3a), a common feature in the karyotypes of all *Talpa* species (Jiménez et al. 1984; Gornung et al. 2008; Arslan and Zima 2014).

The karyotypes of *T. aquitania*, *T. europaea* and *T. occidentalis* are very similar despite size and shape of short arm of chromosome 16 and the Y chromosome. Chromosomes on the pair 16 are clear submetacentric elements in *T. aquitania*, while in *T. europaea* (pair 9) and *T. occidentalis* (pair 16) presented a small short arm (Jiménez et al. 1984; Volleth and Müller 2006; Gornung et al. 2008; Arslan and Zima 2014). The karyotype of specimen TA-3 has size polymorphism in this chromosomal pair in the short arm (Fig. 3a). Chromosome pair 16 of the *Talpa* karyotype is one of the most variable chromosomal pairs within this group of species due to variations in the length of the heterochromatic short arm. This pair corresponds to the large metacentric chromosome 1 (the largest in the karyotype) in *T. altaica,* which has arisen as a consequence of the enlargement of the heterochromatin in the short arm (Kawada et al. 2002).

In *T. aquitania* the Y chromosome is a submetacentric biarmed chromosome of medium size, while in *T. europaea* and *T. occidentalis* the Y is a small dot-like chromosome of probably metacentric conformation. A dot-like Y chromosome of variable morphology is common in moles and is found in all species except *T. levantis*, in which it has been identified as either a medium-sized metacentric or a dot-like chromosome (Arslan and Zima 2014).

G and C-banding

G-banding allowed the identification of the two homologous chromosomes of the same pair in the karyotype. On the basis of the G-banding patterns we could assume that the karyotypes of this species and *T. occidentalis* (Jiménez et al. 1984) are almost identical, with the exception of the differences denoted above (Fig. 3b).

C-banding showed that the centromeric heterochromatin in *T. aquitania* is scarce, being visible only in a few pairs of chromosomes (Fig. 4a). However, in other *Talpa* sp. the centromeric C-bands are more prominent and are observed in most chromosomes of the karyotype including *T. romana* and *T. europaea* (Gornung et al. 2008). The Y chromosome in *T. aquitania* is completely heterochromatic, while in other mole species such as *T. romana* and *T. altaica* it is euchromatic (Kawada et al. 2002; Gornung et al. 2008). The proximal part to the centromere of the short arm of pair 14 is heterochromatic; the entire short arm of chromosome pair 16 is heterochromatic, but is polymorphic in mole TA-3 (Fig. 4a). Pair 14 in *T. occidentalis* is heterochromatic within the centromeric proximal region of both arms, and the short arm is almost completely heterochromatic. The minimal short arm of pair 16 in *T. occidentalis* is fully heterochromatic (Jiménez et al. 1984). The amplification of the heterochromatin in the short arm of this chromosome could give rise to the largest short arm of this pair in *T. aquitania*; alternatively, the deletion of heterochromatin in the short arm of *T. aquitania* could generate the shortest short arm of this pair in *T. occidentalis*. The amplification/deletion of heterochromatin in the short arm of this chromosome pair is one of the characteristic variations found in *Talpa* karyotypes. For those species where no G-banding pattern has been studied, chromosome 16 is supposed to be telocentric (*T. davidina*) (Sözen et al. 2012), acrocentric (*T. europaea*) (Arslan and Zima 2014), submetacentric (*T. aquitania*) or even the largest metacentric of the karyotype, as occurs in *T. altaica* (Kawada et al. 2002). Independently of the morphology of this chromosome, the G-banding pattern of the long arm is identical in all the *Talpa* karyotypes with available data (Jiménez et al. 1984; Volleth and Müller 2006; Gornung et al. 2008; Selçuk and Kefelioğlu 2017).

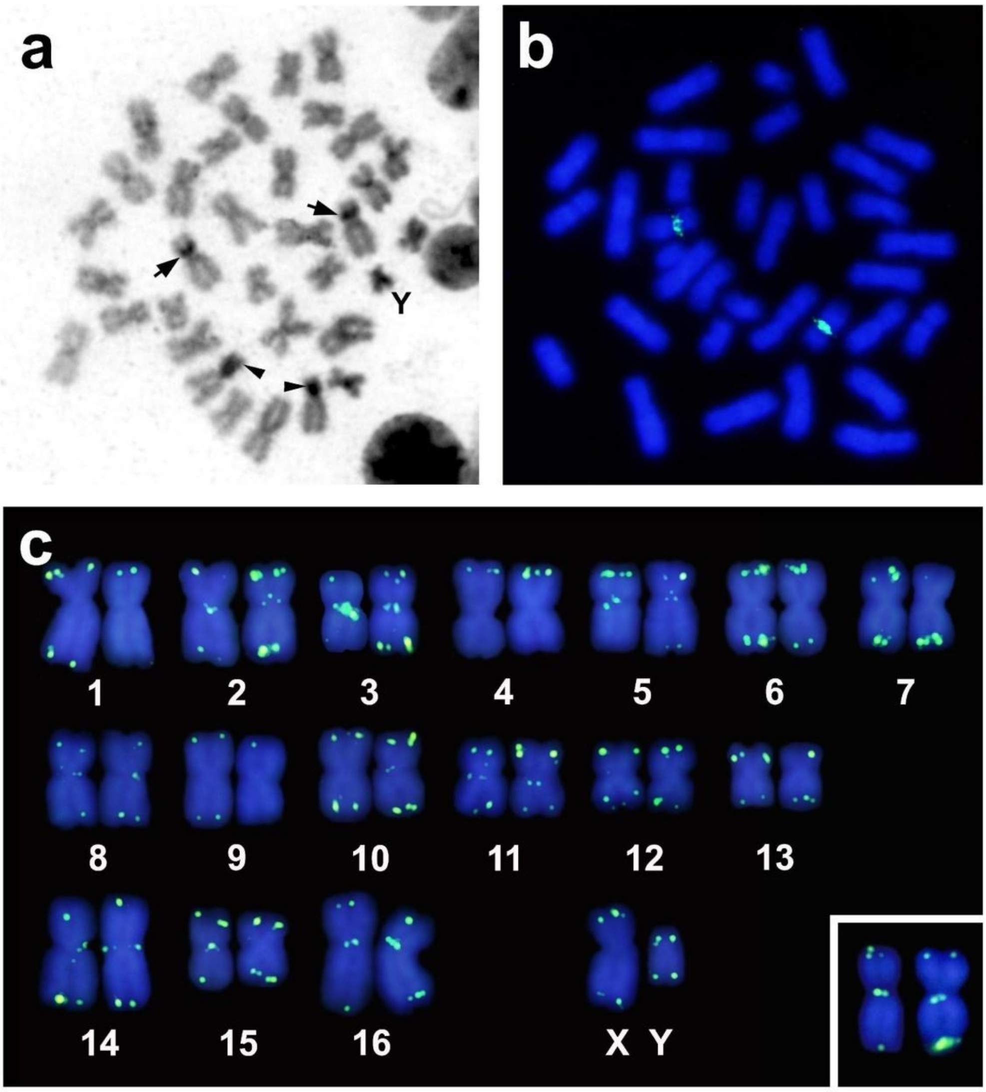


Figure 4. M**etaphases and Karyotype of T. aquitania (TA-3 male specimen).* ***a) *C-banded chromosomes. In addition to centromeric C-bands, the most prominent bands are located on the short arms of chromosome pairs 14 (bands proximal to the pericentromeric area; shown by arrows) and 16 (the entire short arm is heterochromatic; shown by arrowheads). The Y chromosome is also heterochromatic.* ***b)*** *Localization of rDNA genes on the secondary constriction of pair 3 by fluorescence in situ hybridization.* ***c)*** *DAPI staining karyotype of chromosomes after performing FISH with the telomeric probe. Some autosomal pairs have pericentromeric interstitial telomeric sequences (ITSs). Insert: two selected chromosomes of pair 3 in which the localization of ITSs on the secondary constrictions is clearly visible.*

rDNA and telomeric sequences FISH

In the species *T. aquitania* the rDNA genes are located in the secondary constriction of a medium-sized metacentric chromosome pair (Fig. 4b). The existence of a chromosome pair with the secondary constriction carrying the rDNA genes is a common feature in all karyotypes of the *Talpa* species analyzed with either silver staining or FISH (Jiménez et al. 1984; Zurita et al. 1997, 1998; Gornung et al. 2008; Selçuk and Kefelioğlu 2017).

The telomeric probe hybridized on telomeres, however, the signal intensity is highly variable, probably due to the different amount of repeat sequences included (Fig. 4c). Moreover, interstitial telomeric sequences (ITSs) are clearly visible on the pericentomeric regions of some chromosomes. One of these autosomes is the pair with the secondary constriction on which the telomeric sequences and rDNA genes co-localize (Fig. 4b and c). This co-localization of both rDNA genes and the telomeric sequences was also noted in the chromosome pair with the secondary constriction in *T. romana* and *T. europaea* (Gornung et al. 2008), suggesting that this could be a feature shared by all *Talpa* karyotypes. Additionally, it is interesting to note that pair 16 in *T. aquitania* also has centromeric ITSs, as occurs in the corresponding pair 9 in *T. romana* and *T. europaea* (Gornung et al. 2008).

Centromeric ITSs were also present in several autosomal pairs in the *T. romana* and *T. europaea* karyotypes (Gornung et al. 2008). The non-telomeric distribution of telomere repeats has been reported in other insectivores (Zhdanova et al. 2005) and is also common in a number of mammal species and groups (Rovatsos et al. 2011). ITSs are also present in other vertebrate groups and could arise after chromosomal rearrangements during karyotype evolution. However, there are many examples in which their origin is unrelated to this process (for a revision, see Ruíz-Herrera et al. 2008).

In conclusion, we present here for the first time the karyotype description of *T. aquitania* from northern Spain. With the exception of the remarkable differences in the organization of the Y chromosome and the autosomal pair 16, the observed features underline its similarity with the karyotype of other species of the genus *Talpa*.

Acknowledgments

This work was funded by the Consejería de Innovación, Ciencia y Empresa de la Junta de Andalucía (researchgroup RNM-924), and by Universidad de Jaén (via the program Plan de Apoyo a la Investigación 2017-2019, Acción 1). The authors express their gratitude to the Servicio Territorial de Medio Ambiente de Burgos (Junta de Castilla y León) and to the Delegación Territorial de la Consejería de Agricultura, Pesca y Medio Ambiente de Granada (Junta de Andalucía) for permission to capture *T. aquitania* and *T. occidentalis,* respectively.

References

Arslan A, Zima J: Karyotypes of the mammals of Turkey and neighbouring regions: review. Folia Zool 63:1-62 (2014).

Burgos M, Jiménez R, Díaz de la Guardia R: A rapid, simple and reliable combined method for G-banding mammalian and human chromosomes. Stain Technol 61:257-260 (1986).

Ducroz JF, Volobouev V, Granjon L: An assessment of the systematics of arvicanthine rodents using mitochondrial DNA sequences: evolutionary and biogeographical implications. J Mamm Evol 8:173–206 (2001).

Endow SA: Polytenization of the ribosomal genes on the X and Y chromosomes of *Drosophila melanogaster*. Genetics 100:375-385 (1982).

Fedyk S, Ivanitskaya EY: Chromosomes of Siberiain shrews. Acta theriol 27:475-492 (1972).

Feuda R, Bannikova AA, Zemlemerova ED, Febbraro M, Loy A, Hutterer R, Aloise G, Zykov AE, Annesi F, Colangelo P: Tracing the evolutionary history of the mole, *Talpa europaea*, through mitochondrial DNA phylogeography and species distribution modelling. Biol J Linn Soc 114:495–512 (2015).

Gornung E, Volleth M, Capanna E, Castiglia R: Comparative cytogenetics of moles (Eulipotyphla, Talpidae): chromosomal differences in *Talpa romana* and *Talpa europaea*. Cytogenet Genome Res 121:249-254 (2008).

Hugot JP, Gu SH, Feliu C, Ventur J, Ribas A, Dormion J, Yanagihara R, Nicolas V: Genetic diversity of *Talpa europaea* and Nova Hanta Virus (NVAV) in France. Bull Acad Vet Fr. 167 (2014).

Ijdo JW, Wells RA, Baldini A, Reeders ST: Improved telomere detection using a telomere repeat probe (TTAGGG)n generated by PCR. Nucleic Acids Res 19:4780 (1991).

Hutterer R: Order Soricomorpha. In: Wilson, D.E, Reeder, D.M. (Eds.), Mammals Species of the World. A Taxonomic and Geographic Reference, third ed. The Johns Hopkins University Press, Baltimore, pp 220–311 (2005).

Jiménez R, Burgos M, Díaz de la Guardia R: Karyotype and chromosome banding in the mole (*Talpa occidentalis*) from the south-east of the Iberian Peninsula. Implications on its taxonomic position. Caryologia 37:253–258 (1984).

Kawada S, Harada M, Grafodatsky AS, Oda S: Cytogenetic study of the Siberian mole, *Talpa altaica* (Insectivora: Talpidae) and karyological relationships within the genus *Talpa*. Mammalia 66:53-62 (2002).

Meylan A: Données nouvelles sur les chromosomes des insectivores européens. Rev Suiss Zool 75:691–696 (1966).

Nicolas V, Martínez-Vargas J, Hugot JP: *Talpa aquitania* nov. sp. (Talpidae, Soricomorpha) a new mole species from southwest France and north Spain. Bull Acad Vet Fr. 168:329–334 (2015).

Nicolas V, Martínez-Vargas J, Hugot JP: Molecular data and ecological niche modelling reveal the evolutionary history of the common and Iberian blind moles (Talpidae) in Europe. Zoologica Scripta 46:12–26 (2017a).

Nicolas V, Martínez-Vargas J, Hugot JP: *Talpa aquitania* sp. nov. (Talpidae, Soricomorpha), a new mole species from SW France and N Spain. Mammalia 81:641–642 (2017b).

Rovatsos MTh, Marchal JA, Romero-Fernández I, Fernández FJ, Giagia-Athanosopoulou EB, Sánchez A: Rapid, independent, and extensive amplification of telomeric repeats in pericentromeric regions in karyotypes of arvicoline rodents M. Th. Chromosome Res 19:869–882 (2011).

Ruíz-Herrera A, Nergadze SG, Santagostino M, Giulotto E: Telomeric repeats far from the ends: mechanisms of origin and role in evolution. Cytogenet Genome Res 122:219–228 (2008).

Selçuk AY, Kefelioğlu H: Cytogenetic characteristic of the Caucasian pygmy shrew (*Sorex volnuchini*) and Levant mole (*Talpa levantis*) (Mammalia: Eulipotyphla) in northern Anatolia, Turkey. Turk J Zool 41:963-969 (2017).

Sözen M, Matur F, Çolak F, Irmak S: Karyological characteristics, morphological peculiarities, and a new distribution locality for *Talpa davidiana* (Mammalia: Soricomorpha) in Turkey. Turk J Zool 36:806-813 (2012).

Sumner AT: A simple technique for demonstrating centromeric heterochromatin. Expl Cell Res 75:304-306 (1972).

Kryštufek B, Nedyalkov N, Astrin J, Hutterer R: News from the Balkan refugium: Thrace has an endemic mole species (Mammalia: Talpidae). Bonn Zool Bull 67:41–57 (2018).

Volleth M, Müller S: Zoo-FISH in the European mole (*Talpa europaea*) detects all ancestral Boreo-Eutherian human homologous chromosome associations. Cytogenet Genome Res 115:154-157 (2006).

Wilson DE, Reeder DM: Mammal Species of the World: A Taxonomic and Geographic Reference, 3rd ed. Johns Hopkins University Press, Baltimore (2005).

Zhdanova NS, Karamishevam TV, Minina J, Astakhova NM, Lansdorp P, Kammori M, Rubtsov NB, Searle JB: Unusual Distribution Pattern of Telomeric Repeats in the Shrews *Sorex araneus* and *Sorex granaries*. Chromosome Res. 13:617-625 (2005).

Zurita F, Sánchez A, Burgos M, Jiménez R, Díaz de la Guardia R: Interchromosomal, intercellular and interindividual variability of NORs studied with silver staining and in situ hybridization. Heredity 77:229-223 (1997).

Zurita F, Jiménez R, Burgos M, Díaz de la Guardia R: Sequential silver staining and in situ hybridization reveal a direct association between rDNA levels and the expression of homologous nucleolar organizing regions: a hypothesis for NOR structure and function. J Cell Sci 111:1433-1439 (1998).

Molecular cytogenetic analysis of karyotype and Y chromosome conservation in species of the genus Talpa (Insectivora)

Authors: Juana Gutierrez, Gaël Aleix-Mata, Juan Alberto Marchal, María Arroyo, Riccardo Castiglia, Antonio Sánchez

Departamento de Biología Experimental, Facultad de Ciencias Experimentales, Universidad de Jaén, Campus Las Lagunillas s/n, 23071 Jaén (Spain).

Department of Biology and Biotechnology "Charles Darwin", University of Rome 'Sapienza', Rome (Italy).

Cytogenet Genome Res (10.1159/000507836)

Cytogenetic and Genome Research

Original Article

Cytogenet Genome Res
DOI: 10.1159/000507836

Received: December 19, 2019
Accepted: March 30, 2020
Published online: May 9, 2020

Molecular Cytogenetic Analysis of Karyotype and Y Chromosome Conservation in Species of the Genus *Talpa* (Insectivora)

Juana Gutierrez[a] Gael Aleix-Mata[a] Juan A. Marchal[a] María Arroyo[a] Riccardo Castiglia[b] Antonio Sánchez[a]

[a]Departamento de Biología Experimental, Facultad de Ciencias Experimentales, Universidad de Jaén, Jaén, Spain;
[b]Department of Biology and Biotechnology "Charles Darwin," University of Rome "Sapienza," Rome, Italy

Abstract:

The Talpidae family has a highly stable karyotype. Most of the chromosome studies of this mammal group, however, employ classical cytogenetic techniques. Molecular cytogenetic analyses are still scarce and, for example, no repeated DNA sequences have been described to date. In this work, we used sequence analysis and the chromosomal location of a LINE1 retro-element sequence, as well as chromosome painting with the whole Y chromosome of *T. occidentalis*, to compare the karyotype of three species of the genus *Talpa (T. occidentalis, T. romana* and *T. aquitania*). Our results demonstrate that in *Talpa* genomes LINE1 sequences are widely distributed on all chromosomes but are enriched on pericentromeric C-band positive regions. In addition, this LINE1 accumulate on the Y chromosomes of the three species of *Talpa* regardless of their euchromatic or heterochromatic condition. The chromosome painting demonstrates that the Y chromosome of these three species is highly conserved. Interestingly, Y chromosomes from the analyzed species share sequences with heterochromatic blocks on chromosome pairs 14 and 16 and, to a lesser degree, with the pericentromeric regions of other autosomes. Together, our analyses demonstrate that the repetitive DNA content of chromosomes from *Talpa* species is highly conserved.

Keywords: heterochromatin, LINE1, repeat sequences, *Talpa*, Y chromosome

Introduction

The genus *Talpa* (family Talpidae) includes eleven species of fossorial mammals belonging to the order Soricomorpha, which is sometimes included as a suborder of the Eulipotyphla [Hutterer 2005; Wilson and Reeder 2005; Nicolas et al. 2015, 2017; Kryštufek et al. 2018].

The species of the Talpidae family – above all moles – are characterized by their highly stable karyotypes [Biltueva and Vorobieva 2012; Ye et al. 2006]. Most of the species of the genus *Talpa* have chromosome number 2n=34, the only exceptions being *T. caeca* (2n=36) and *T. caucasica* (2n=38). Their autosomal fundamental number (FNa) is also conserved and varies between 62 and 66 [Zima 1983; Fedyk and Ivanitskaya 1972; Jiménez et al. 1984; Kawada et al. 2002; Gornung et al. 2008; Gutiérrez et al. 2019]; their sex chromosomes are also well conserved. Their X chromosome is typically medium-sized and biarmed, and, despite appearing submetacentric in some cases, is usually metacentric [Gornung et al. 2008]. The Y chromosome in most species is dot-like with no distinguishable morphology and is classified as metacentric, submetacentric or acrocentric. Recently, in the newly described species *T. aquitania* the Y chromosome has been described as a medium-sized submetacentric biarmed chromosome [Gutiérrez et al. 2019].

As in the rest of the Talpidae family [Yates and Moore 1990], chromosome conservation defines evolutionary dynamics in *Talpa*, although a number of minor differences still account for some karyotypic divergence in this genus [Reumer and Meylan 1986; Kawada et al. 2002; Gornung et al. 2008; Gutiérrez et al. 2019]. In fact, some studies have revealed interspecific differences due to Robertsonian translocations, variation in the morphology of the sex chromosomes, and quantitative changes in heterochromatin. In terms of this latter factor, while *T. europaea* shows only pericentromeric heterochromatin [Zima 1983], *T. altaica, T. occidentalis* and *T. aquitania* all also have heterochromatin in some non-centromeric regions [Jiménez et al. 1984; Kawada et al. 2002; Gutiérrez et al. 2019].

Regarding the analysis of the repeated DNA sequences, only the distribution and location of the major ribosomal genes and telomeric sequences have ever been determined in the genus *Talpa* [Zurita et al. 1997; Gornung et al. 2008; Gutiérrez et al. 2019]. The localization of the 5S ribosomal genes is less well understood and has only been investigated in *T. romana* and *T. europaea*, where differences in the karyotype distribution have been observed [Gornung et al. 2008]. Telomeric sequences have been found at chromosome ends but also arranged as ITSs (interstitial telomeric sequences) in some pericentromeric regions in *Talpa romana*, *Talpa europaea* and *T. aquitania* [Gornung et al. 2008; Gutiérrez et al. 2019]. On the basis of the different distributions of ITSs and 5S ribosomal genes in the chromosomes of *T. romana* and *T. europaea*, it has been suggested that karyotype diversification in *Talpa* may rely on genomic events that are indecipherable using classical cytogenetics [Gornung et al. 2008].

To identify cryptic rearrangements or subtle changes occurring during karyotypic evolution in the genus *Talpa*, a molecular analysis of their chromosomal content is required. Hence, in this work we performed i) a comparative analysis of the karyotype of three species of the genus *Talpa* based on sequence analysis and the chromosomal location of a LINE1 retro-element sequence and ii) chromosome painting using the whole Y chromosome of *T. occidentalis* as a probe.

Material and methods

Specimens analyzed and chromosome preparations

For this study, we used DNA samples from males of four *Talpa* species, namely, *T. occidentalis* (Granada, Spain), *T. europaea* (Pavia, Italy), *T. aquitania* (Torme, Spain) and *T. romana* (Rome, Italy), together with chromosome preparations from one *T. occidentalis*, one *T. romana* and three *T. aquitania*. DNA samples were extracted following the standard phenol-chloroform procedure; chromosomes were prepared from bone marrow cells following the method described by Burgos et al. [1986]. All capture and sacrifice protocols were approved by the Junta de Andalucía Ethics Committee for Animal Experimentation (code: 22/05/2018/094). The chromosomes in this work are numbered according to the karyotypes described for *T. occidentalis* and *T. aquitania* (Jiménez et al. 1984; Gutierrez et al. 2019).

Isolation, cloning and sequencing of repeated DNA

To identify the repeated DNA sequences from *T. occidentalis, T. europaea, T. aquitania* and *T. romana,* genomic DNA was restricted with *Hin*dIII restriction endonuclease, which produces in gel electrophoresis several bands of repeated DNA sequences: 1kb, 0.8kb, 0.6kb and 0.4kb. The 1kb band of *T. europaea* was eluted from the gel and cloned into pGEM-T easy Vector (Promega), as previously described by Sánchez et al. [1996]. Transformed Z-competent bacteria (Zymo Research) were screened using the same digoxigenin-labelled band as a probe.

Southern Blot

Genomic DNAs were digested with *Hin*dIII restriction endonuclease and the resulting fragments were separated onto 1% agarose gels and blotted onto nylon membranes (Amersham), following Bullejos et al. [1997]. Membranes were probed overnight at 55°C using the positive clone TE-1kb-clon 31 (see below) (digoxigenin-labelled by PCR). Alkaline phosphatase detection was carried out according to the supplier's recommendations (Roche).

Sequence Analysis

Plasmids from bacterial clones were Sanger sequenced using universal primers and were analyzed using the Bioedit program (version 7.0.9.0) (http://www.mbio.ncsu.edu/BioEdit/bioedit.html). Pairwise and multiple sequence alignments were carried out with the program CLUSTAL OMEGA [Thompson et al. 1994]. Repeated DNA sequences were screened using the program RepeatMasker version 4.0.6 (http://repeatmasker.org). A GenBank search for sequences was performed with BLASTN [Zhang et al. 2000] and BLASTP [Altschul et al. 1997] (https://blast.ncbi.nlm.nih.gov/).

FISH and Chromosome Painting

Indirect FISH was performed, as previously described by Fernández et al. [2001], using as a probe the clone TE-1kb-clon 31 labelled with biotin-16-dUTP by PCR amplification, with M13 forward and reverse primers.

For chromosome painting, a painting probe was prepared from the whole minute Y chromosome of *T. occidentalis*. For probe preparation, 15 micro-dissected chromosomes were amplified and labelled by DOP-PCR using Spectrum-Orange dUTP (Abbott). We followed the direct FISH protocol for chromosome painting described by Marchal et al. [2004]. The images were captured using a fluorescence microscope (Olympus BX51) equipped with a CCD camera (Olympus DP70).

Results

Cloning and chromosomal localization of the repeated sequences

Electrophoresis of the *Hin*dIII-restricted genomic DNAs of *T. occidentalis*, *T. europaea, T. aquitania* and *T. romana* produced several intense bands of approximately 1kb, 0.8kb, 0.6kb and 0.4kb (Fig. 1a; only *T. europaea* is illustrated). The 1kb band of *T. europaea* was eluted, labelled and used as a probe for a Southern blot on the restricted DNAs of the four species. The Southern blot results were all very similar for *T. occidentalis*, *T. europaea* and *T. aquitania* (the results for *T. romana* are not shown as the quality and amount of the DNA was insufficient). In addition to the corresponding 1kb band, three other bands about 0.8, 1.2 and 1.7 kb in size were clearly visible. These bands could correspond to shorter or larger DNA fragments of the same LINE1 arising as a consequence of sequence modification and variation in restriction pattern, or alternatively correspond to restriction fragments of other similar LINEs. Another additional large intense band was present in *T. europaea*, which was only faintly detectable in the *T. occidentalis* and *T. aquitania* genomes (Fig. 1b).

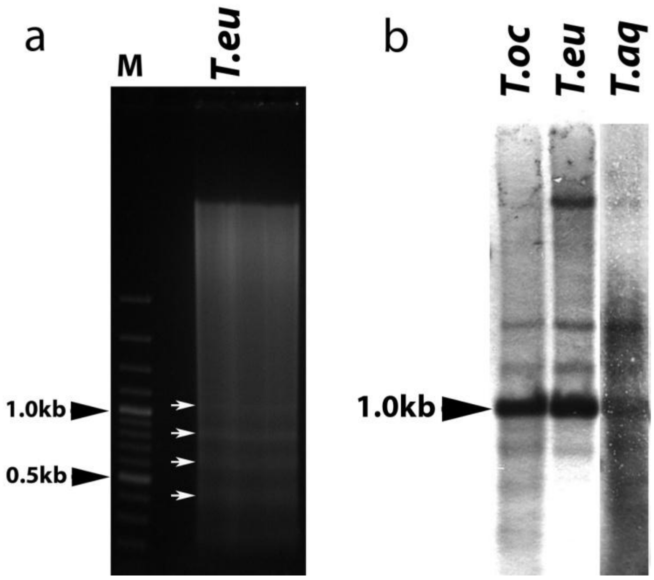


***Figure 1**. a) Gel electrophoresis of genomic DNA from T. europaea (T.eu) digested with HindIII (M: 100-bp DNA molecular weight marker); b) Southern blot on T. occidentalis (T.oc), T. europaea (T.eu) and T. aquitania (T.aq) genomic DNA digested with HindIII and probed with the eluted 1-kb band labelled with digoxigenin.*

The 1-kb-eluted band from *T. europaea* was cloned in pGEM-T Easy Vector to obtain 10 positive clones that were fully sequenced (TE-1kb-clon 31, 32, 38, 51, 135, 139, 169, 182, 194,

196) (GenBank accession numbers: MN212943-MN212952). The clones contained sequences that varied in length between 1033 and 1054 bp. The identity of these sequences was in the range 88.8–99.3% when compared with the consensus; on the other hand, pairwise sequence comparison produced identities of 86.4–99.5%. The AT content of the cloned sequences was on average 64.4%. The nucleotide variation observed in the alignment was reduced to random base changes and a few small insertions of 1–8 bp in length. The RepeatMasker program identified the analyzed sequences as part of the reverse transcriptase (ORF2) from a LINE1 retrotransposon. A BLAST search of the whole genome shotgun databases of the Soricomorpha identified many sequences with 88.0% identity in contigs from *Scalopus aquaticus*, a species of Soricomorpha phylogenetically related to *Talpa* species.

As the sequence translation gave rise to multiple stop codons in most of the sequenced clones, they were probably from non-functional copies (pseudogenes) of LINE1 retroposons. However, clone TE-1kb-clon 139 did not contain premature stop codons in the translated amino acid sequence. A BLASTp search demonstrated 78% of identity with the LINE1 reverse transcriptases available in GenBank, including five of the conserved domains (3–7) that characterize reverse transcriptase proteins [Xiong and Eickbush 1990].

To determine the chromosomal location of the cloned LINE1 sequence we performed FISH on metaphase spreads of *T. occidentalis*, *T. romana* and *T. aquitania* (Fig. 2). The FISH signals in the karyotype of these three species were widely distributed on most autosomes and X chromosomes. Interestingly, LINE1 appeared enriched on the pericentromeric regions of the chromosomes of *T. occidentalis* and *T. romana*, while in *T. aquitania* it was less intense in these regions. LINE1 signals were also detected on the dot-like minute Y chromosomes of *T. occidentalis* and *T. romana*, and on the Y submetacentric chromosome of *T. aquitania* (Fig. 2). In this latter case, the sequences are not homogeneously distributed and seem to be accumulated on the short arm of the Y chromosome (Fig. 2c).

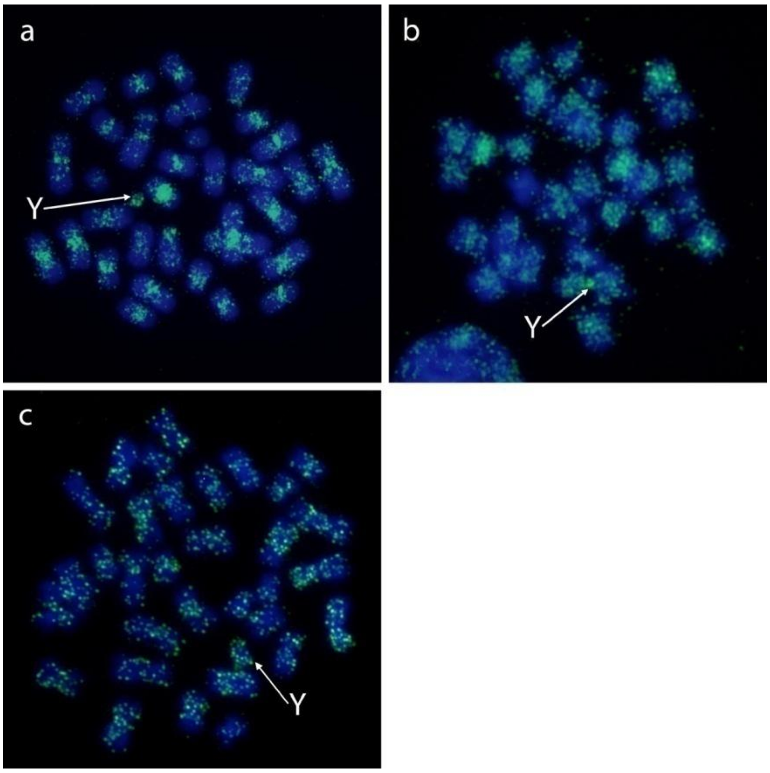


Figure 2.- *In situ hybridisation with the 1-kb LINE1 probe of T. europaea on male metaphases of T. occidentalis (a), T. romana (b) and T. aquitania (c).*

Painting with the T. occidentalis Y chromosome probe

As expected, this probe painted the complete Y chromosome of *T. occidentalis*. Notably, it also strongly hybridized on the small arm of autosomal pair 16 and, somewhat less intensely, on the pericentromeric region and on the small arm of submetacentric pair 14. Faint signals were also observed in pericentromeric regions of certain other chromosomes (Fig. 3a). Identical results were obtained when this probe was hybridized in the *T. romana* karyotype (chromosome pairs 16 and 14 of *T. occidentalis* correspond with *T. romana* autosomal pairs 9 and 1, according to the chromosome nomenclature used by Gornung et al. [2008]); only the Y chromosome stained less intensely than the Y chromosome of *T. occidentalis* (Fig. 3b). In *T. aquitania*, the entire submetacentric Y chromosome, as well as the short arm of submetacentric pair 16, hybridized strongly. The heterochromatin and the short arm of pair 14 of *T. aquitania* were also painted, albeit faintly; no clear signals were observed in the pericentromeric regions of the remaining chromosomes (Fig. 3c).

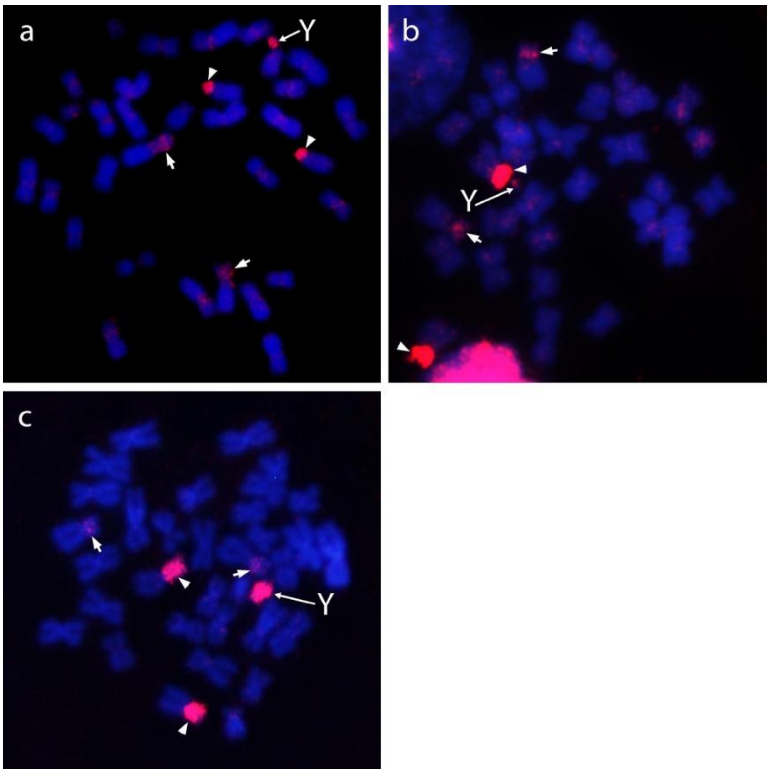


Figure 3.- *Chromosome painting with the probe from the whole Y chromosome of T. occidentalis on male metaphases of T. occidentalis (a), T. romana (b) and T. aquitania (c). The arrows and arrowheads indicate chromosome pairs 14 and 16, respectively.*

Discussion

Eukaryotic genomes contain many interspersed repeated sequences derived from mobile genetic elements [Kirkness et al. 2003; Lander et al. 2001; Venter et al. 2001; Waterston et al. 2002; Gibbs et al. 2004]. LINE1 elements are one of the most important groups of retrotransposons and are abundant in mammal genomes [Hardies et al. 2000; Mears and Hutchison 2001; Adelson et al. 2009]. It is estimated that one third of the mammal genome originated directly or indirectly through LINE1 retrotransposition [Han and Boeke 2005]. A good example is the LINE1 content revealed by the Human Genome Project of around 21% [Lander et al. 2001].

Our results demonstrated that LINE1 sequences are widely distributed in the genome of these three *Talpa* species and form part of their repetitive DNA content. We cloned and analyzed a 1kb fragment of the ORF2 from the LINE1 coding for the reverse transcriptase. Most of the sequences obtained in our study correspond to non-functional copies of LINE1-retroposons. This is not surprising given that LINE1 retroposition generates mostly defective copies that remain on the genome as mutated or rearranged LINE1 sequences [Furano 2000;

Boissinot and Furano 2001]. Complete LINE1 sequences with autonomous mobile capacity are in fact very scarce in mammal genomes [Sassaman et al. 1997; Kazazian 1999]. One of the clones analyzed could derive from a functional *Talpa* LINE1 retro-element since it does not contain stop codons, and the resulting polypeptide includes five conserved domains of the reverse transcriptase, as described by Xiong and Eickbush [1990]. However, we cannot rule out the possibility that this analyzed sequence could also be part of an inactive LINE1 element if it is truncated or harbours mutations from other regions of the protein sequence.

A key question is why the distribution of LINE1 elements in mammal genomes is widely but not randomly spread on chromosomes [Graham and Boissinot 2005; Kvikstad and Makova 2010]. Genome projects are currently quantifying the AT-rich genome distribution bias of LINE1, which could have certain functional implications. At chromosomal level, LINE1 sequences are enriched in heterochromatic regions of autosomes, while in sex chromosomes they are similarly abundant in both euchromatin and heterochromatin [Bailey et al. 2000; Dobigny et al. 2004; Waters et al. 2004; Marchal et al. 2006; Acosta et al. 2008; Kvikstad and Makova 2010]. Our results provide evidence that in *Talpa* karyotypes LINE1 possesses similar dynamics since they are widely distributed on all chromosomes but accumulate at pericentromeric regions, which have previously been described as C-band positive heterochromatin [Jiménez et al. 1984; Gornung et al. 2008; Gutiérrez et al. 2019].

Among the factors proposed to explain the enrichment of retrotransposons in mammal sex chromosomes are the lower rates of recombination, AT-biased composition preferential insertion, local genome landscape features, natural selection and random genetic drift (reviewed in Kvikstad and Makova 2010]. LINE1 accumulation in mammal Y chromosomes could be easily tolerated since this chromosome only contains a few genes and is mainly heterochromatic [Marshall Graves 2000, 2001]. The accumulation of LINE1 on the Y chromosomes of the three species of *Talpa* described here for the first time is not related with if they are mostly euchromatic (C-band negative) or heterochromatic (C-band positive). Indeed, the Y chromosome is heterochromatic in *T. occidentalis* and *T. aquitania* but euchromatic in *T. romana* [Jiménez et al. 1984; Gornung et al. 2008; Gutiérrez et al. 2019] and all of them are LINE1 enriched. These data show that the repetitive DNA content of the Y chromosome of *Talpa* species is well conserved despite differences in chromosome size and morphology.

Several chromosome painting studies in the Soricomorpha (including *Talpa* species) have previously been performed using human [Yang et al. 2006; Ye et al. 2006; Volleth and Müller 2006; Biltueva and Vorobieva 2012] and stone marten (Yang et al. 2006) chromosome painting probes, and, alternatively, using insectivore chromosomes as probes on mammal or human chromosomes, in order to delineate chromosome evolution [Biltueva and Vorobieva, 2012]. Despite the good results for autosomes and X chromosomes, in most of this research no information has been generated for the Y chromosomes. Only Volleth and Müller [2006] indicate that painting with human chromosomes on *T. europaea* yields reproducible results for all probes but not for Y chromosomes. This is customary as painting analyses of the Y chromosome are scarce in the literature [Acosta et al. 2011]. In the Soricomorpha only one previous study of chromosome painting using *Sorex* species exists in which both the X and Y chromosomes are analyzed [Biltueva et al. 2011].

The results of Y-chromosome painting in rodent *Microtus* species and insectivore *Sorex* species demonstrate that a small region is conserved, corresponding to the euchromatic part or the pseudoautosomal region (PAR) with the X and Y chromosomes but not for the rest of Y chromosomes, even in closely related species [Acosta et al. 2011; Biltueva et al. 2011].

By contrast, our painting results using as a probe the Y chromosome of *T. occidentalis* show that the Y chromosome content of the three studied *Talpa* species is well preserved. If the dot-like Y chromosomes of *T. occidentalis* and *T. europaea* are considered to be similar to their ancestral Y chromosomes, the enlargement of the Y chromosome occurring in *T. aquitania* involved the amplification of sequences already present on that minute chromosome. Interestingly, the Y chromosome of the three *Talpa* species contains sequences that are very similar to the sequences that form the heterochromatin of autosomal pair 14 and 16. Two evolutionary contexts could explain this homogenous pattern of heterochromatin in the Y chromosome and autosomal pair 14 and 16. Both contexts are similar and probably only differ in the original source of the repeated heterochromatic sequences – either autosomal or Y chromosome – before amplification and transference to their counterparts. Regardless of the particular evolutionary dynamics that took place, it is highly probable that these repeated sequences were already emplaced in the heterochromatin arranged on the Y chromosome and on the heterochromatin of chromosomes 14 and 16 in the karyotype of the ancestral species of the genus *Talpa*.

It is significant that the scenario described in *Talpa* species for Y chromosome evolution – that is, the conservation of the sequence content – differs significantly from that of other mammals. Many studies have shown that, even between closely related species, the DNA content of the Y chromosome is poorly conserved not only in the heterochromatin but also in the euchromatin regions [Sitnikova et al. 2007; Kirsch et al. 2008; Gifalli-Iughetti and Koiffmann 2009; Acosta et al. 2011]. Such rapid and independent evolution probably reflects the degenerative process that has driven the evolution of this chromosome [Wilson and Makova 2009].

In conclusion, we demonstrate here that the repetitive DNA content of the genome of *Talpa* species is enriched in LINE1 sequences, which are distributed on the karyotypes following a conserved pattern. Furthermore, the Y chromosome of these species is composed of similarly conserved repeated sequences. Finally, the origin of the heterochromatic blocks arranged in some autosomal pairs of these species, in particular pair 16, was parallel to the evolution of the repetitive DNA content of the Y chromosome.

Acknowledgements

This work was funded by the Consejería de Innovación, Ciencia y Empresa de la Junta de Andalucía (research group RNM-924), and by the Universidad de Jaén (via the program *Plan de Apoyo a la Investigación* 2017-2019, Acción 1). G. Aleix-Mata was funded by the Fédération Nationale des Chasseurs, Fédération Départementales des Chasseurs de l'Ariège and Hautes Pyrénées (France) (project FNC-PSN-PR16-2014). The authors wish to express their gratitude to the Servicio Territorial de Medio Ambiente de Burgos (Junta de Castilla y León) and to the Delegación Territorial de la Consejería de Agricultura, Pesca y Medio Ambiente de Granada (Junta de Andalucía), respectively, for permission to capture *T. aquitania* and *T. occidentalis*.

References

Acosta MJ, Marchal JA, Fernández-Espartero CH, Bullejos M, Sánchez A. 2008. Retroelements (LINEs and SINEs) in vole genomes: Differential distribution in the constitutive heterochromatin. Chromosome Res 16:949–959.

Acosta MJ, Romero-Fernández I, Sánchez A, Marchal JA. 2011. Comparative analysis by chromosome painting of the sex chromosomes in arvicolid rodents. Cytogenet Gen Res 132:47-54.

Adelson DL, Raison JM, Edgar RC. 2009. Characterization and distribution of retrotransposons and simple sequence repeats in the bovine genome. PNAS 106:12855–12860.

Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. 1997. Gapped BLAST and PSIBLAST: a new generation of protein database search programs. Nuc. Acids Res. 25:3389-3402.

Bailey JA, Carrel L, Chakravarti A, Eichler EE. 2000. Molecular evidence for a relationship between LINE-1 elements and X chromosome inactivation: the Lyon repeat hypothesis. PNAS 97:6634-6639.

Biltueva L, Vorobieva N, Perelman P, Trifonov V, Volobouev V, Panov V, Ilyashenko V, Onischenko S, O'Brien P, Yang F, Ferguson-Smith M, Graphodatsky A. 2011. Karyotype Evolution of Eulipotyphla (Insectivora): The Genome Homology of Seven *Sorex*Species Revealed by Comparative Chromosome Painting and Banding Data. Cytogenet Genome Res 135:51–64

Biltueva L, Vorobieva N. 2012. Chromosome evolution in Eulipotyphla. Cytogenet Genome Res 137: 154-164.

Boissinot S, Furano AV. 2001. Adaptive evolution in LINE-1 retrotransposons. Mol Biol Evol 18: 2186Y2194.

Bullejos M, Sánchez A, Burgos M, Hera C, Jiménez R, Díaz de la Guardia R. 1997. Multiple, polymorphic copies of SRY in both males and females of the vole *Microtus cabrerae*. Cytogenet Cell Genet 79:167–171.

Burgos M, Jiménez R, Díaz de la Guardia R. 1986. A rapid, simple and reliable combined method for G-banding mammalian and human chromosomes. Stain Technology 61: 257–260.

Dobigny G, Ozouf-Costaz C, Waters PD, Bonillo C, Coutanceau JP, Volobouev V. 2004. LINE-1 amplification accompanies explosive genome repatterning in rodents. Chromosome Res 12: 787Y793.

Fedyk S, Ivanitskaya EY. 1972. Chromosomes of Siberiain shrews. Acta theriol 27:475-492.

Fernández R, Barragán MJ, Bullejos M, Marchal JA, Martínez S, Díaz de La Guardia R, Sánchez A. 2001. Molecular and cytogenetic characterization of highly repeated DNA sequences in the vole *Microtus cabrerae*. Heredity 87: 637–646.

Furano AV. 2000. The biological properties and evolutionary dynamics of mammalian LINE-1 retrotransposons. Prog Nucleic Acid Res Mol Biol 64: 255Y294.

Gifalli-Iughetti C, Koiffmann CP. 2009. The Y chromosome of the Atelidae family (Platyrrhini): study by chromosome microdissection. Cytogenet Genome Res 125: 46-53.

Gibbs RA, Weinstock GM, Metzker ML, Muzny DM, Sodergren EJ, et al. 2004. Genome sequence of the Brown Norway rat yields insights into mammalian evolution. Nature 428:493-521.

Gornung E, Volleth M, Capanna E, Castiglia R. 2008. Comparative cytogenetics of moles (Eulipotyphla, Talpidae): chromosomal differences in *Talpa romana* and *Talpa europaea*. Cytogenet Genome Res 121:249-254

Graham T, Boissinot S. 2005. Genomic Distribution of L1 elements: The role of insertion bias and natural selection. J Biomed Biotechnol 2006:1–5.

Gutiérrez J, Aleix-Mata G, Lamelas L, Arroyo M, Marchal JA, Sánchez A. 2019. Karyotype analysis of the new *Talpa* species *Talpa aquitania* (Talpidae; Insectivora) from northern Spain. Cytogenet Genome Res 159: 26-31.

Han JS, Boeke JD. 2005. LINE-1 retrotransposons: modulators of quantity and quality of mammalian gene expression? Bioessays 27: 775.

Hardies SC, Wang L, Zhou L, Zhao Y, Casavant NC, Huang S. 2000. LINE-1 (L1) lineages in the mouse. Mol. Biol. Evol. 17: 616-628.

Hutterer R. 2005. Order Soricomorpha. In: Wilson, D.E, Reeder, D.M. (Eds.), Mammals Species of the World. A Taxonomic and Geographic Reference, third ed. The Johns Hopkins University Press, Baltimore, pp 220–311.

Jiménez R, Burgos M, Díaz de la Guardia R. 1984. Karyotype and chromosome banding in the mole (*Talpa occidentalis*) from the south-east of the Iberian Peninsula. Implications on its taxonomic position. Caryologia 37:253–258.

Kawada S, Harada M, Grafodatsky AS, Oda S. 2002. Cytogenetic study of the Siberian mole, *Talpa altaica* (Insectivora: Talpidae) and karyological relationships within the genus *Talpa*. Mammalia 66:53-62.

Kazazian HH Jr. 1999. An estimated frequency of endogenous insertional mutations in humans. Nat Genet 22: 130.

Kirkness EF, Bafna V, Halpern AL, Levy S, Remington K, Rusch DB, Delcher AL, Pop M, Wang W, Fraser CM, Venter JC. 2003. The dog genome: survey sequencing and comparative analysis. Science 301:1898-1903.

Kirsch S, Munch C, Jiang Z, Cheng Z, Chen L, Batz C, Eichler EE, Schempp W. 2008. Evolutionary dynamics of segmental duplications from human Y-chromosomal euchromatin/heterochromatin transition regions. Genome Res 18: 1030-1042.

Kryštufek B, Nedyalkov N, Astrin J, Hutterer R. 2018. News from the Balkan refugium: Thrace has an endemic mole species (Mammalia: Talpidae). Bonn Zool Bull 67:41–57.

Kvikstad EM, Makova KD. 2010. The (r)evolution of SINE versus LINE distributions in primate genomes: Sex chromosomes are important. Genome Research 20:600–613.

Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, et al. 2001. Initial sequencing and analysis of the human genome. Nature 409:860-921.

Marchal JA, Acosta MJ, Nietzel H, Sperling K, Bullejos M, Díaz de la Guardia R, Sánchez A. 2004. X chromosome painting in *Microtus*: origin and evolution of the giant sex chromosomes. Chromosome Res 12:767–776.

Marchal JA, Acosta MJ, Bullejos M, Puerma E, Díaz de la Guardia R, Sánchez A. 2006. Distribution of l1-retroposons on the giant sex chromosomes of *Microtus cabrerae* (Arvicolidae, Rodentia): Functional and evolutionary implications. Chromosome Res 14:177-186.

Marshall Graves JA. 2000. Human Y chromosome, sex determination, and spermatogenesis- a feminist view. Biol Reprod 63:667-676.

Marshall Graves JA. 2001. From brain determination to testis determination: evolution of the mammalian sex-determining gene. Reprod Fertil Dev 13:665-672.

Mears ML, Hutchison CA. 2001. III The evolution of modern lineages of mouse L1 elements. J Mol Evol. 52: 51-62.

Nicolas V, Martínez-Vargas J, Hugot JP. 2015. *Talpa aquitania* nov. sp. (Talpidae, Soricomorpha) a new mole species from southwest France and north Spain. Bull Acad Vet Fr. 168:329–334.

Nicolas V, Martínez-Vargas J, Hugot JP. 2017. *Talpa aquitania* sp. nov. (Talpidae, Soricomorpha), a new mole species from SW France and N Spain. Mammalia 81:641–642.

Reumer JWF, Meylan A. 1986. New developments in vertebrate cytotaxonomy IX. Chromosomenumbers in the order Insectivora (Mammalia). Genetica 70: 119-151.

Sánchez A, Bullejos M, Burgos M, Jiménez R, Díaz de la Guardia R. 1996. An alternative to bluntend ligation for cloning DNA fragments with incompatible ends. Trends Genet 12: 44.

Sassaman DM, Dombroski BA, Moran JV, Kimberland ML, Naas TP, DeBerardinis RJ, Gabriel A, Swergold GD, Kazazian HH Jr. 1997. Many human L1 elements are capable of retrotransposition. Nat Genet 16: 37-43.

Sitnikova NA, Romanenko SA, O'Brien PC, Perelman PL, Fu B, Rubtsova NV, Serdukova NA, Golenishchev FN, Trifonov VA, Ferguson-Smith MA, Yang F, Graphodatsky AS. 2007. Chromosomal evolution of Arvicolinae (Cricetidae, Rodentia). I. The genome homology of tundra vole, field vole, mouse and golden hamster revealed by comparative chromosome painting. Chromosome Res 15:447-456.

Thompson JD, Higgins DG, Gibson TJ. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nuc. Acids Res. 22:4673-4680.

Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, et al. 2001. The sequence of the human genome. Science 291:1304-1351.

Volleth M, Müller S. 2006. Zoo-FISH in the European mole (*Talpa europaea*) detects all ancestral Boreo-Eutherian human homologous chromosome associations. Cytogenet Genome Res 115:154-157.

Waters PD, Dobigny G, Pardini AT, Robinson TJ. 2004. LINE-1 distribution in Afrotheria and Xenarthra: implications for understanding the evolution of LINE-1 in eutherian genomes. Chromosoma 113:137-144.

Waterston RH, Lindblad-Toh K, Birney E, Rogers J, Abril JF, et al. 2002. Initial sequencing and comparative analysis of the mouse genome. Nature 420:520-562.

Wilson DE, Reeder DM. 2005. Mammal Species of the World: A Taxonomic and Geographic Reference, 3rd ed. Johns Hopkins University Press, Baltimore.

Wilson MA, Makova KD. 2009. Genomic analyses of sex chromosome evolution. Annu Rev Genomics Hum Genet 10: 333-354.

Xiong Y, Eickbush TH. 1990. Origin and evolution of retroelements based upon their reverse transcriptase sequences. The EMBO Journal 9:3353-3362.

Yates TL, Moore DW. 1990. Speciation and evolution in the family Talpidae (Mammalia: Insectivora). Proc Clin Biol Res 335: 1-22.

Yang F, Graphodatsky AS, Li T, Fu B, Dobigny G, Wang J, Perelman PL, Serdukova NA, Su W, O'Brien PCM, Wang Y, Ferguson-Smith MA, Volobouev V, Nie W. 2006. Comparative genome maps of the pangolin, hedgehog, sloth, anteater and human revealed by cross-species chromosome painting: further insight into the ancestral karyotype and genome evolution of eutherian mammals. Chromosome Res 14:283–296.

Ye J, Biltueva L, Huang L, Nie W, Wang J, Jing M, Su W, Vorobieva NV, Jiang X, Graphodatsky AS, Yang F. 2006. Cross-species chromosome painting unveils cytogenetic signatures for the Eulipotyphla and evidence for the polyphyly of Insectivora. Chromosome Res 14: 151-159.

Zhang Z, Schwartz S, Wagner L, Miller W. 2000. A greedy algorithm for aligning DNA sequences. J. Comput. Biol. 7:203–214.

Zima, J. 1983. Thekaryo type of *Talpa europaea kratochvili* (Talpidae, Insectivora). Folia Zoologica 32: 131-136.

Zurita F, Sánchez A, Burgos M, Jiménez R, Díaz de la Guardia R. 1997. Interchromosomal, intercellular and interindividual variability of NORs studied with silver staining and in situ hybridization. Heredity 77:229-223.

Satellitome analysis on Talpa aquitania genome and inferences about the satDNAs evolution on some Talpidae

Authors: Juana Gutierrez, Gaël Aleix-Mata, Eugenia E. Montiel, Diogo C. Cabral-de-Mello, Juan Alberto Marchal, Antonio Sánchez

Departamento de Biología Experimental, Facultad de Ciencias Experimentales, Universidad de Jaén, Campus Las Lagunillas s/n, 23071 Jaén (Spain).

Departamento de BiologiaGeral e Aplicada, Instituto de Biociências/IB, UNESP—Universidade EstadualPaulista, Rio Claro, São Paulo 13506-900, Brazil.

Genes (10.3390/genes14010117)

genes MDPI

Article

Satellitome analysis on *Talpa aquitania* genome and inferences about the satDNAs evolution on some Talpidae

Juana Gutiérrez[1], Gaël Aleix-Mata[1], Eugenia E. Montiel[1], Diogo C. Cabral-de-Mello[1, 2], Juan Alberto Marchal[1], Antonio Sánchez[1,*]

[1] Departamento de Biología Experimental, Área de Genética, Universidad de Jaén, Paraje de las Lagunillas s/n., 23071 Jaén, Spain.

[2] Departamento de Biologia Geral e Aplicada, UNESP – Universidade Estadual Paulista, Instituto de Biociências/IB, Rio Claro, São Paulo, Brazil.

* Correspondence: abaca@ujaen.es

Abstract:

In the genus *Talpa* a new species, named *Talpa aquitania*, has been recently described. Only cytogenetic data are available for the nuclear genome of this species. In this work, we characterize the satellitome of the *T. aquitania* genome that presents 16 different families, including telomeric sequences, and they represent 1.24% of the genome. The first satellite DNA family (TaquSat1-183) represents 0.558%, and six more abundant families, including TaquSat1-183, comprise 1.13%, while the remaining 11 sat-DNAs represent only 0.11%. The average A + T content of the SatDNA families was 50.43% and the median monomer length was 289.24 bp. The analysis of these SatDNAs indicated that they have different grades of clusterization, homogenization, and degeneration. Most of the satDNA families are present in the genomes of the other *Talpa* species analyzed, while in the genomes of other more distant species of Talpidae, only some of them are present, in accordance with the library hypothesis. Moreover, chromosomal localization by FISH revealed that some satDNAs are localized preferentially on centromeric and non-centromeric heterochromatin in *T. aquitania* and also in the sister species *T. occidentalis* karyotype. The differences observed between *T. aquitania* and the close relative *T. occidentalis* and *T. europaea* suggested that the satellitome is a very dynamic component of the genomes and that the satDNAs could be responsible for chromosomal differences between the species. Finally, in a broad context, these data contribute to the understanding of the evolution of satellitomes on mammals.

Keywords: insectivora; satellite DNAs; satellitome; genus *Talpa*; Talpidae

Introduction

Eukaryotic genomes are divided into two regions, one called euchromatin, where most genes are located [1] and heterochromatin, composed of a large number of repetitive DNA sequences. The repetitive sequences possess high variability in nucleotide base pair composition and number of copies, being involved in the genome size variation and diversity across tree of life [2]. As a whole, the repeated DNA content (repeatome) includes dispersed, represented by all classes of transposable elements (TEs) and tandem satellite DNA (satDNA) repeats [3–6]. The satDNA sequences content of a genome, or satellitome [7], are constituted by non-coding repeated sequences organized in tandem, mainly located in constitutive heterochromatin regions, that in multiple species are observed on pericentromeric and centromeric regions of chromosomes and on a lesser extent on telomeres [2,8,9]. Satellite DNAs could represent a large proportion of the whole eukaryotic genomes, with variation between 0.1% and up to 50% (Revised in [2]). Due to the absence of general selective forces, the rate of evolution of the satDNA sequences is very high, and they show great variation in monomer size, nucleotide composition, chromosomal distribution, and abundance in the genome. Those factors can affect genome composition, structure, or size between species [10]. On many occasions, satDNA can be chromosome- or species-specific and can give rise to large genome variations between closely related species [2,11].

The biological significance of satDNA and heterochromatin itself is a topic that is currently arousing great scientific interest. Traditionally it has been considered as silenced junk DNA lacking any relevant function. However, there are multiple examples in insect, plant, or mammalian species of centromeric or telomeric repeat sequences being transcribed [8,10,12]. Transcription of satDNA could be associated with the heterochromatin assembly mechanism, the regulation of centromere function, the sister chromatid cohesion, chromosome organization, pairing and segregation, control of telomere elongation, gene regulation, karyotypic evolution, and speciation [4,5,13–23].

Massive sequencing techniques combined with computational and molecular cytogenetic analysis have opened a new way to characterize the repetitive DNA present in a genome. The development of powerful graph-based bioinformatics tools such as RepeatExplorer and TAndemREpeatANalyzer (TAREAN) using massive sequencing data have enabled repeatome analysis [24–26]. These programs allowed the characterization of the repeated DNA content of the genomes from Next Generation Sequencing data. In this way, the satellitomes have been studied using these approaches in several animal [7,18,27–32] and plant species [33–36], allowing the testing of satDNA evolution models across distinct eukaryote groups. With the use of these new tools, it is possible to gain knowledge of the different families of satDNAs and to obtain information on the size and variability within a particular genome, as well as its evolution and genetic functions. In addition, this methodology allowed the detection of a large number of satDNA families at once. Illustrative examples are *Triatoma infestans*, *Locusta migratoria*, *Rhynchophorus ferrugineus*, and *Pontastacus leptodactylus* with 42, 62, 112, and 258 families identified, respectively [7,18,31,32].

The genus *Talpa* (family Talpidae) includes 11 species of fossorial mammals [37]. The species of the Talpidae family have very stable karyotypes [38,39]. Most of the species of *Talpa* have chromosome number 2n = 34 (exceptions *T. caeca*: 2n = 36 and *T. caucasica*: 2n = 38), and autosomal fundamental number (FNa) varies between 62 and 66 [39–42]. The sex chromosomes are also well conserved, with the X chromosome typically a medium-sized submetacentric or metacentric and the Y chromosome in most species is a dot-like metacentric, submetacentric, or acrocentric [42,43].

Analysis of repetitive DNAs identified heterochromatin on centromeres of most chromosomes of the karyotype of *T. romana* and *T. europaea* [42], while is less abundant in *T. aquitania* and *T. occidentalis* [37,40]. In addition, in some species, the Y chromosome is completely heterochromatic, and most of the species also present an autosomal pair with non-centromeric heterochromatin [37]. Although, the specific molecular content of the heterochromatin and chromosomal distribution of repeats is poorly known in the genus. Previous studies investigated only the chromosomal distribution of ribosomal genes, telomeric sequences, and a LINE fragment [37,42–44]. Telomeric sequences are located at chromosome ends but also arranged as ITSs (interstitial telomeric sequences) in some pericentromeric regions in *Talpa romana*, *Talpa europaea*, and *T. aquitania* [37,42]. LINE sequences are widely distributed on all chromosomes, as expected, with a noteworthy enrichment on pericentromeric C-band positive regions and the Y chromosome [43]. Interestingly, chromosome painting analyses demonstrate that heterochromatic blocks from two autosomal pairs and Y-heterochromatin share repeat sequences [43]. Recently, the genome of *Talpa occidentalis* has been sequenced, revealing that about 30% of the genome is comprised of repetitive sequences, similar data to other species in this group (15% of LINEs; 7% of SINEs; 4% of LTR; about 1% DNA transposon) [45].

In most of the mammalian genomes sequenced, little information is provided about the repetitive DNA sequence composition, the repeatome, and especially about the satDNA sequence composition, the satellitome. Until now, few satellitome analyses have been performed on mammal species [46], including insectivore species. Moreover, in this family, there is scarce information about repeat in general and satDNAs in particular. Data on satDNAs repeats remain almost completely missing for the *Talpa* species genomes. The aim of this work is to characterize the satDNA families, the satellitome, in the genome of the recently described species of this genus *T. aquitania* by using the Next Generation Sequencing approach. The satellitomes of the *T. aquitania* genome present 16 different satDNA families, including the telomeric sequences. Most of the satDNA families are present in the genomes of the other *Talpa* species analyzed, while in the genomes of other more distant species of the Talpidae family only some of them are present. Moreover, chromosomal localization by Fluorescence in situ Hybridization (FISH) revealed that some satDNAs are localized preferentially on centromeric and non-centromeric heterochromatin in *T. aquitania* and also in the sister species *T. occidentalis* karyotype. The differences observed between *T. aquitania* and the close relative *T. occidentalis* suggested that the satellitome is a very dynamic component of the genomes.

Materials and Methods

Talpa Specimens

For this study, we analyzed samples from three *T. aquitania* (captured in Torme (Burgos); northern Spain) and four *T. occidentalis* (captured in Granada; southern Spain) males. Permission for capture was granted by the Servicio Territorial de Medio Ambiente de Burgos (Junta de Castilla-León) and the Delegación Territorial de la Consejería de Agricultura, Pesca y Medio Ambiente de Granada (Junta de Andalucía). All capture and sacrifice protocols were approved by the Junta de Andalucía Ethics Committee for Animal Experimentation (code: 22/05/2018/094).

DNA Extraction, Genome Sequencing, Sequence Clustering, and Analysis

Liver tissue samples were stored in 100% ethanol at −20 °C and DNA samples for PCR were extracted following the standard phenol-chloroform procedure. For Illumina sequencing, we extracted genomic DNA (gDNA) from one adult male of *T. aquitania* and one adult male of *T. occidentalis* using theGentraPuregene Tissue Kit (Qiagen). The extracted gDNAs were sequenced with Illumina technology. Briefly, for genome sequencing, approximately 3 µg of genomic DNA was used for the construction of a library of 350-bp-length fragments. This library was based on the Illumina® Hiseq™ 2500 platform and paired-end reads with 2 × 150 bp were obtained. A total of 4.5 Gb of sequences were obtained from the *T. aquitania* genome, corresponding to a coverage of about 2 × of the genome, considering the genome of similar size to the *T. occidentalis* genome (2.099 Gb) [45], and 2 Gb were obtained from the *T. occidentalis* genome (approximately a coverage about 1×). Graph-based clustering analysis [24–26,47], a method for similarity-based clustering of sequence reads, was performed with 829.778 reads of the *T. aquitania* genome using RepeatExplorer, implemented within the Galaxy environment (http://repeatexplorer.org/, accessed on 19 April 2021).

The genome proportion for the repetitive DNA clusters was calculated as read percentages. Additionally, all the clusters were analyzed with sequence-similarity searches of the assembled contigs against GenBank using BlastN (http://www.ncbi.nlm.nih.gov/, accessed on 5 September 2021) and Repbase using the program RepeatMasker version 4.0.9 (http://repeatmasker.org; https://www.dfam.org/, accessed on 1 April 2022, [48] and CENSOR (http://www.girinst.org/, accessed on 1 April 2022). In addition, to identify satDNA repeats, contigs were analyzed using Dotmatcher (http://emboss.bioinformatics.nl/cgi-bin/emboss/dotmatcher/, accessed on 5 September 2021). For the analysis of the characteristics of sequences of the clusters, we used the software Bioedit (version 7.2.5) (http://www.mbio.ncsu.edu/BioEdit/bioedit.html, accessed on 5 September 2021) [49] and Clustal Omega (https://www.ebi.ac.uk/Tools/msa/clustalo/, accessed on 5 September 2021) [50].

Estimates of evolutionary divergence between nucleotide sequences were conducted with MEGA 10 using *p*-distance. The satDNA family abundance and divergence were calculated with RepeatMasker with the "-a" option and the RMBlast search engine. Four million reads of *T. aquitania* and *T. occidentalis* samples were selected and aligned to the total collection of satDNA dimers, or monomer concatenations of approximately 200 bp in length in satDNAs with small size monomer. A satellite landscape was generated considering distances from the sequences applying the Kimura 2-parameter model with the Perl script calcDivergenceFromAlign.pl and createRepeatLandscape.pl from the RepeatMasker suite. The tandem structure index (TSI) was calculated from the out file of RepeatMasker output as in Montiel et al. [31]. We calculated the number of reads with at least 89% read length aligned to the dimer (internal/pure satDNA reads) divided by the total number of aligned reads (internal/pure + external/mixed satDNA reads). The satellite landscape created was used to calculate the divergence peak (DivPeak) and the relative abundance size of the peak (RSP) for each satDNA family following the method in [51]. Briefly, the divergence value where the landscape had a maximum (DivPeak) was calculated and the relative peak size as the sum of satDNA abundances at ±2% divergence from the divergence peak (peak size) was divided by the total abundance of that satDNA family. The relative abundance among samples was calculated as the log_2 of the ratio of the compared sample divided by the *T. aquitania* sample in order to facilitate the understanding of the fold-change between samples. Then, to check the presence of *T. aquitania* satDNAs on the closest relative species the consensus sequences were also masked on the chromosome-scale genome assembly of *T. occidentalis*, based on long- (PacBio) and short-read (Illumina) sequencing and scaffolded using Hi-C data, published by Real et al. [45] (GCA_014898055.1). Moreover, aiming to check sharing of satDNAs among representatives of family Talpidae abundance and divergence were calculated from low-coverage sequencing data of other species that were available on NCBI SRA database: *Talpa europaea* (SRX8240408), *Scalopus aquaticus* (SRX4562103), *Condylura cristata* (SRX101046), *Galemys pyrenaicus* (SRX10243621) and *Uropsilus gracilis* (SRX4562112). Calculations and statistical analysis as a correlation between variables using Spearman's rank correlation rho and comparison between paired samples using Wilcoxon signed rank exact test were performed in R base v.4.0.1 [52]. Figures were also plotted in R with ggplot2 and Viridis packages [53,54].

Satellite DNA Probes, Chromosome Preparations, and FISH

The consensus sequence of each satDNA family (Table S1) was used to design one (small consensus sequences) or two oligonucleotides (large consensus sequences) by using the Primer3 software [55] (Table S2). Only three SatDNA families with large monomers (TaquSat4-437-466; TaquSat5-3102 and TaquSat11-71) were amplified and labeled by PCR using the specifically designed primer pairs. PCR was performed in 50 µL of reaction mixture containing 100 ng of *T. aquitania* genomic DNA, 10 pmol of each primer, 1 µL 10 mM dNTPs, 5 µL of 10× NH_4 Reaction Buffer, 2.5 µL of 50 mM $MgCl_2$, and 5 U of BIOTAQ™ DNA Polymerase (Bioline). The PCR program used was 5 min at 95 °C and 30 cycles: 30 s at 95 °C, 30 s at 55 °C, and 60 s at 72 °C, with a final extension of 5 min at 72 °C. PCR amplicons were resolved in ethidium bromide-stained 1% agarose gels; the appropriate bands were isolated from the gel with a QIAquick gel extraction kit (Qiagen), and labeled with biotin-16-dUTP or digoxigenin-11-dUTP by PCR amplification, with the specific primers pairs. For satDNA families with small monomers, the oligonucleotides based on the most conserved regions were directly labeled with biotin-16-dUTP or digoxigenin-11-dUTP using Terminal Transferase (Roche) as described by Pita et al. [18].

Chromosomes were prepared from bone marrow cells following the method described by Burgos et al. [56]. The location of satDNAs sequences by Fluorescence in situ Hybridization (FISH) was performed as previously described by Pita et al. [18], Fernández et al. [57], and

Cabral-de-Mello and Marec [58]. Briefly, the hybridized probes, labeled with biotin-16-dUTP, were detected by avidin-based indirect fluorescence techniques, with two rounds of amplification [18,57]or with Alexa Fluor 488 streptavidin conjugate, while the probes labeled with digoxigenin-11-dUTPwere detected with Anti-Digoxigenin-Rhodamine (Roche) [58].

Slides were mounted with Vectashield (Vector Laboratories, Newark, CA, USA) containing DAPI to counterstain the chromosomes. The images were captured and analyzed using a fluorescence microscope (Olympus BX51) equipped with a CCD camera (Olympus DP70), and processed with Adobe Photoshop software.

Results and Discussion

General Characterization of T. aquitania Satellitome

Sequencing data (4.5 Gb) of the *T. aquitania* genome produced 30,149,458 reads. Nucleotide analyses showed that A + T genome content was 59.06% (GC content 40.94%). A high A + T content is a common feature within the Talpidae genomes (58.09% *T. occidentalis*, 59.5% *T. europaea*, 58.3% *G. pyrenaicus,* 58% *S. aquaticus,* 58.1% *C. cristata*, and 58.9*% U. gracilis*) (GenBanK data; [45,59].

From the total reads, a subset of 8 million pair-end reads was randomly selected and processed in RepeatExplorer. From that subset, 829,778 sequences were selected in the pipeline. Those correspond to about 6% of the genome of *T. aquitania*, considering the genome of similar size as the *T. occidentalis* genome (2.099 Gb) [45].

After RepeatExplorer clustering, 263,881 reads were grouped into 38,295 clusters, with 20% of the *T. aquitania* genome composed of repeated sequences. All the remaining reads (565,897) were classified as singletons. From those clusters, 425 represented at least 0.0024% of the genome and those that featured a star-like or circular graph topology, typically observed in the satDNA families, were deeply analyzed. For each candidate cluster, we examined the contigs assembled by RepeatExplorer to search tandem repeated structures using the Dotmatcher and multiple-sequence alignments and manual inspection to determine the consensus sequences. After the computational analysis, we identified 15 satDNA families (Tables 1 and S1). The satDNA families were named according to Ruiz-Ruano et al. [7], including the species name abbreviation, a number in decreasing order of abundance and the length of the repeat sequence monomer, starting from TaquSat1-183 (the most abundant) to TaquSat15-64 (the least abundant). For the TaquSat4 family, two variants were found with 437 and 466 bp. Telomeric sequence repeat TTAGGG was not found through this analysis.

Nevertheless, the $(TTAGGG)_{50}$ repeat was included in the RepeatMasker analysis showing that the abundance of telomeric repeats was 0.13% (Table 1). In this way, including the telomeric sequences, the *T. aquitania* satellitome is composed of at least 16 satDNA families, corresponding to 1.24% of the genome (Table 1). This percentage of satDNA is in accordance with observed in other Talpidae species as *G. pyrenaicus* (1.08%) [59]. On other species of mammals in which RepeatExplorer analysis was carried out, Valeri et al. [46] found only one satDNA family, corresponding to 0.87% of the genome of *Trichechus manatus*, revealing that the picture of abundance and number of families could be highly divergent on distinct groups, deserving more analysis.

Table 1. *Data of the satDNA families found in T. aquitania: genome proportion (%), the repeat unit length, A + T content and divergence (%), tandem structure index (TSI), divergence peak (DivP) and relative abundance size of the peak (RSP).*

Name	Genome Proportion	Repeat Unit Length (bp)	A + T Percentage	Kimura Divergence	TSI	DivP	RSP
TaquSat1-183	0.55864	183	67.2	12.85	0.85	10	0.36
TaquSat2-107	0.16776	107	55.1	13.63	0.93	13	0.56
TaquSat3-6	0.12463	6	66.7	15.61	0.76	24	0.21
TaquSat4-437-466	0.146 (0.08–0.066)	437–466	30.4–32.6	13.3–18.5	0.83–0.84	12–19	0.44–0.42
TaquSat5-3102	0.04003	3102	53	0.14	0.99	0	1.00
TaquSat6-84	0.04000	84	65.5	20.63	0.53	21	0.28
TaquSat7-60	0.00993	60	48.3	16.72	0.41	17	0.28
TaquSat8-45	0.00532	45	51.1	3.01	0.45	2	0.81
TaquSat9-90	0.00425	90	56.7	16.28	0.40	14	0.34
TaquSat10-24	0.00418	24	50	18.16	0.48	15	0.48
TaquSat11-71	0.00233	71	40.3	2.1	0.09	1	0.98
TaquSat12-101	0.00180	101	50.5	7.73	0.01	6	0.43
TaquSat13-54	0.00175	54	59.3	18.7	0.20	24	0.24
TaquSat14-17	0.00129	17	35.3	21.69	0.36	18	0.33
TaquSat15-64	0.00045	64	45.3	4.89	0.00	8	0.38
Telomeric	0.13160	6	50	18.29	0.91	18	0.57
Total	1.2422						
Mean		289.24	50.43	13.07	0.53	13.06	0.48
SD		737.28	11.19	6.89	0.33	7.63	0.24
Median		71.00	50.50	15.61	0.48	14.00	0.42

The first satDNA family TaquSat1-183 represents 0.558% of the genome. Together with the next five families in the list and the telomeric sequences, they all comprise 1.13% of the genome, while the remaining 11 satDNAs represent only 0.11%. The mean A + T content is 50.43% (with variation between 67.2% and 30.4%), lower than the overall genome A + T content (59.06%). The mean of the monomer length is 289.24, with variation between the largest one with 3102 bp (TaquSat5-3102) and the smallest one with 6 bp (TaquSat3-6 and telomeric sequences), being the monomer length of the most abundant satellite DNA of 183 bp (TaquSat1-183). Most of the satDNA families have repeat units smaller than 183 bp, with the exception of the TaquSat4-437-466 and TaquSat5-3102 in global the mean of monomer length is 289.24 bp (median 71.00).

The relationships between satDNA families were analyzed by comparison of the consensus sequences. Most of the satDNA families did not show similarity with the sequences of other families. However, in the satDNA TaquSat4-437-466, analyzing the contigs alignments, and manually, we identified two different monomer lengths 437 and 466, both monomers shared a fragment with an identity of 78.21%, and the remaining portion of the longer monomer corresponded with the first portion of a new the monomer repetition. On some posterior analysis, they are considered as TaquSat4-437 and TaquSat4-466.

Nucleotide divergence of satDNA families in *T. aquitania* ranged between 0.14% (TaquSat5-3102) and 21.69% (TaquSat14-17). Altogether, the satellitome of *T. aquitania* shows a mean nucleotide divergence value of 13.07% (Table 1). Additionally, we calculated the tandem structure index (TSI) of each satDNA [31]. This index represents the proportion of internal reads of satDNA arrays (reads containing only satDNA sequence) with respect to the total of reads with satDNA sequence. The TSI varies between 0 and 1 and informs about the clusterization level of the satDNA family. Therefore, dispersed families with arrays formed by a low number of repeat units would present a lower value of TSI in comparison to families with their repeat units concentrated in longer arrays, which would present TSI values closer to 1. The satDNAs of *T. aquitania* have a TSI mean of 0.53. The first six most abundant satDNAs (including the telomeric) have the highest TSI values, ranging between 0.99 for TaquSat5-3102 and 0.76 for TaquSat3-6. However, in satDNA families with large monomer sizes, high values of the TSI index could be related more to the size than the structure since short reads used in that calculation could correspond to the same monomers and not to tandem monomers. The remaining ones have the lowest TSI values ranging between 0.53 for TaquSat6-84 and 0.00 for TaquSat15-64.

The satellite landscape is a good representation of both the homogenization or degeneration status of each satDNA family. Those landscapes were obtained by plotting the abundance versus divergence of satDNA sequences against their consensus sequence. SatDNA families with a recent expansion would be formed by nearly identical monomers and almost the total abundance of the family would be contained in the landscape peak. These families would show leptokurtic satellite landscape distributions, with their peaks on low divergence values, as the families TaquSat5-3102, TaquSat8-45, and TaquSat11-71 (Fig. 1). On the contrary, satDNA families with high divergence would contain monomers with a great accumulation of mutations in their sequence and the satDNA family abundance would not be contained in the peak. These divergent families would show platykurtic satellite landscape distribution with the peak on high divergence values. Platykurtic landscapes were shown by the family TaquSat3-6, TaquSat7-60, and TaquSat13-54 (Fig. 1). The rest of the satDNA families displayed mesokurtic satellite landscapes indicating families amplified time ago and in process of degeneration (Fig. 1).

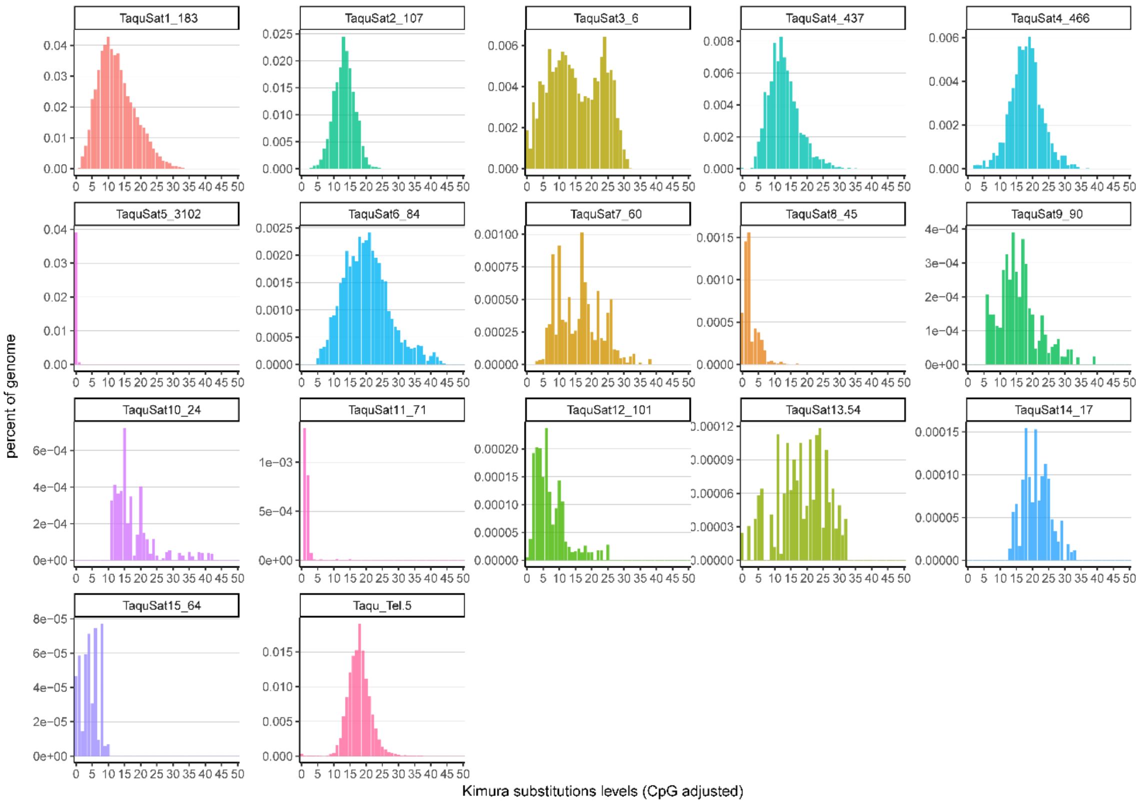


Figure 1. *Satellite landscapes (abundance versus divergence) of T. aquitania satDNA families.*

In addition, we calculated two indices, the divergence value where the landscape had a maximum (DivPeak) (Table 1), as a measure of the family degeneration, and the relative abundance size of the peak (RSP) (Table 1), as a measure of the family homogenization [51]. Both indices were inversely correlated in the *T. aquitania* satellitome (Spearman's rank correlation r_s=−0.75, p <0.001) where the recently expanded families showed low DivPeak and high RSP values, and the highly divergent and degenerated families presented high DivPeak and low RSP values (Fig. 2).

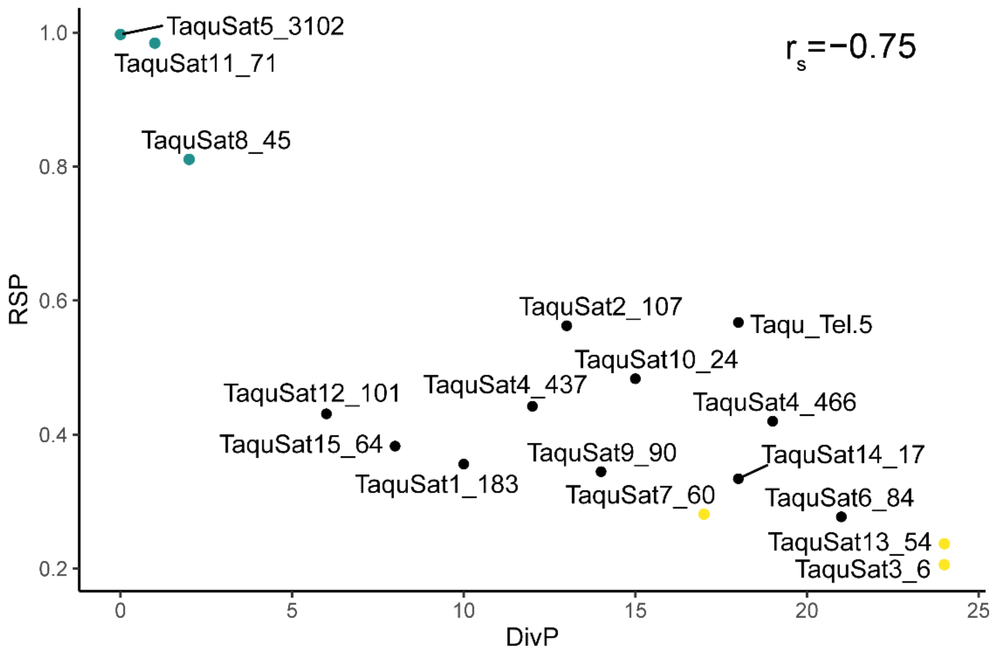


Figure 2. *Correlation between the homogenization (RSP) and degeneration (DivP) indices of T. aquitania satDNA families. Recently expanded families are in blue and highly divergent and degenerated families are in yellow.*

Sister Species Genomic Analysis Reveals Insight about satDNAs Diversification on Talpidae.

The satellitome is a very dynamic genome component that could vary between closely related species or even among populations of the same species [15,18,31]. We were interested in studying the putative conservation of satDNAs in *T. occidentalis,* the species considered as the closely related species of *T. aquitania* [37,60,61] and other Talpidae representatives. Therefore, the consensus sequences of *T. aquitania* satDNA families were used to investigate their presence in the species *T. occidentalis* and other Talpidae.

The abundance and divergence of each satDNA family sequence were analyzed with RepeatMasker using a sample of 4 million reads of *T. occidentalis* (aprox. 400 Mb, 0.2× genome coverage). Fifteen of the sixteen satDNAs found on *T. aquitania* genome were shared with *T. occidentalis*, with the unique exception of TaquSat8-45.The satDNA sequences accounted for 0.69% of the *T. occidentalis* genome, corresponding to almost half of the satellitome in *T. aquitania*. The variation of abundance among samples was calculated with the ratio between samples (*T. occidentalis*/*T. aquitania*), which was represented as $\log_2$ (*T. occidentalis*/*T. aquitania*). Most satDNA families (10 out of 15) were less abundant on the *T. occidentalis* genome ($\log_2$ ratio < −0.6). The most abundant family of *T. aquitania*, TaquSat1-183, only accounted for 0.13% of *T. occidentalis*, being the third satDNA family in terms of abundance on

T. occidentalis. In contrast, the two variants of TaquSat4 (TaquSat4-437 and TaquSat4-466) were more abundant on *T. occidentalis* (log_2 ratio > 0.6), with TaquSat4-437 being the most abundant satDNA family on *T. occidentalis* accounting for approximately the 0.20% of the genome (Table S3). As fast-evolving components of the genome, the changes in satDNA abundance among closely related species are not uncommon. The expansion or divergence of these repeated sequences configures specie-specific genetic backgrounds, which could contribute to the speciation process [15,62].

The satellite landscapes for each family were also compared between species, including the degeneration (DivPeak) and homogenization (RSP) indices. The landscape distribution of satDNA families did not show significant differences, besides abundance, between species, and also indices, DivPeak and RSP, were similar among samples (Tables 1 and S3, Fig. S1). It was interesting, however, that satDNA families with a leptokurtic distribution on their satellite landscape on *T. aquitania*, recently expanded families, were low abundance with larger degeneration value (TaquSat11-71) and low (TaquSat5-3102) or even not present (TaquSat8-45) on the homogenization index on *T. occidentalis*. In fact, one would expect that between closely related species, the satellitome would be similar, and between species that are more distant, the satellitome might show significant differences. However, in a comprehensive analysis of the satellitomes of bird species, the observed pattern was not in agreement with this assumption. In a group of birds with more deeply diverged species, satellites were found to be more similar and were less similar among more recently diverged species [63].

Additionally, as the first genome assembly of *T. occidentalis* was published by Real et al. [45], satDNA consensus was mapped to the assembly to test their presence on it. The proportion of satDNA sequences in the assembly was 0.20%, a lower value than the estimation with *T. occidentalis* low-coverage sequencing data (0.69%). Studies identifying satDNA sequences on the assembled genomes are scarce but their underestimation of the assemblies has been a common theme. For instance, the satellitome of *Rhodnius prolixus*, which accounts for 8% of the genome, only makes up 5.6% of the genome assembly [30]. Another example is the satellitome of the red palm weevil, *R. ferrugineus*, which is 10 times underrepresented in the genome assemblies [31]. The abundance variation of each family among *T. occidentalis* sequencing data and the assembly was calculated as above (log_2(*T. occidentalis* assembled genome/*T. occidentalis* genome)) (Table S3). Even though no significant differences were found between the two samples (p = 0.1167 two-sided Wilcoxon signed-rank test), there was a negative correlation between the TSI value and the variation of abundance among samples (Spearman's rank correlation r_S=−0.63 p =0.008). Thereupon, satDNA families with larger TSI values, families with higher clusterization forming large arrays, were less abundant in the assembly, while satDNA families with lower TSI values, dispersed families with arrays made by a low number of monomers, were more abundant (Fig. S2). Repeated DNA sequences tend to be discarded in the assembly process, due to the challenge that their assembly supposes. Therefore, satDNA families dispersed in small clusters scattered in the genome were better assembled than satDNA families composed of large clusters that were likely to be collapsed just in a few monomers. The correlation between TSI and the satDNA estimation in genome assemblies has been reported before. In *R. ferrugineus*, the high-TSI families showed a higher underestimation than low-TIS families, for example, the family RferSat01-169, which accounts for 0.01–0.06% of the assemblies whereas it was estimated at 20% of the genome [31].

The presence of *T. aquitania* satDNA families was tested on the genomes of other species of the family Talpidae, including *T.* europaea, *S. aquaticus*, *C. cristata*, *G. pyrenaicus*, and *U. gracilis*. Raw data of these species were available in the NCBI SRA database and the analysis was performed as above. The number of satDNA families that present those species was variable. On *T.* europaea, the other member of the subfamily Talpinae, it was identified 14 satDNA families, one less than in *T. occidentalis* (Table S3, Fig. S3). The species where fewer families were identified were *U. gracilis* (7 families), the most distant species from *T. aquitania*. Interestingly, those seven families are also presented in the other species and may be conserved in the family Talpidae. The abundance and divergence of satDNA families also seem to be related to the phylogenetic distance to *T. aquitania*. Therefore, for *T. aquitania* satDNA sequences, *T.* europaea displayed a larger abundance (2.27%) and lower Kimura divergence (15.65%), while *U. gracilis* showed low abundance (0.03%) and larger Kimura divergence (27.21%). Analysis of Dugongidae mammals revealed also the conservation of one satDNA family found initially in *T. manatus* in other species of the family with differences in abundance [46]. Those differences would be consistent with the library hypothesis proposed by Fry and Salser [64]. From the same pool of satDNA sequences, certain families could eventually expand and fix in one species, whereas other families could expand in other species. Consequently, at the same time species distance themselves, their satellitomes also do so, contributing to differences among species or even favoring new speciation events [8,65].

SatDNAs Chromosome Localization on T. aquitania and T. occidentalis Reveals Information about Karyotypic Evolution

T. aquitania karyotype has a diploid number of 2n = 34 (NFa = 64) and all chromosomes including the sex chromosomes are biarmed, either metacentric or submetacentric. The X chromosome is medium-sized submetacentric and the Y chromosome is submetacentric and the smallest of the karyotype. As in all the karyotypes of all *Talpa* species studies on *T. aquitania*, one of the largest autosomes (pair 3) harbors a secondary constriction [35,38,40]. In *T. aquitania* karyotype C-banding revealed heterochromatin in centromeric regions of a few pairs of chromosomes, on the proximal part to the centromere of the short arm of pair 14, on the entire short arm of pair 16, and in the entire Y chromosome [35]. The *T. occidentalis* karyotype is very similar to the *T. aquitania,* with the main differences due to the size of the Y chromosome, which is a medium size biarmed submetacentric on *T. aquitania* and a dot-like probably metacentric, in *T. occidentalis*; and in the size of the small arm of the autosome pair 16, that is large on *T. aquitania* and very small in *T. occidentalis* [35,41].

Only seven of the satDNA families of *T. aquitania* were visualized by FISH in the chromosomes of this species and only six on *T. occidentalis* chromosomes. These satDNAs presented very similar distribution on both species, and most of the variations are due to the above-commented differences between the karyotypes (Fig. 3).

Three satDNAs (TaquSat1-183, TaquSat3-6, and TaquSat4-437-466) are located on pericentromeric regions, including the X chromosome, however, only TaquSat4-437-466 is also located on the Y chromosome, and TaquSat3-6 has one interstitial band in the arm of one small autosome in *T. aquitania* not present in *T. occidentalis* (Fig. 3a–f).

SatDNA TaquSat2-107 is located on subtelomeric regions of the long arm of the Y chromosome and on the subtelomeric regions of the heterochromatic short arm or the pair 16 in *T. aquitania* (Fig. 3g), while in *T. occidentalis* it is probably on the small arm of the minute Y chromosome and occupying all the small arm of the pair 16 (Fig. 3h). These localizations suggested that the long arm of the Y chromosome of *T. aquitania* arises by the enlargement of the small arm of the Y chromosomes of *T. occidentalis.* In the same way, the enlargement of the heterochromatic small arm of a pair 16 of *T. occidentalis* arises the big small arm of this pair in *T. aquitania.* Heterochromatin enlargements in both chromosomes amplified the satDNA TaquSat6-84. In fact, TaquSat6-84 is accumulated on the heterochromatin of both, the Y chromosome and the short arm of the pair 16 in *T. aquitania*, but was undetectable in *T. occidentalis* (Fig. 3i,j). The recent amplification of pair 16 heterochromatic small arm in *T. aquitania* could explain the different composition with chromosome 14 heterochromatin.

SatDNA TaquSat5-3102 is distributed along all the chromosomes in both species, with enrichment on heterochromatin regions of pair 14, the centromeric region of some pairs, and on the Y chromosome; however, is not accumulated on the heterochromatin region of pair 16 (Fig. 3k,l). The satDNA TaquSat11-71 is also distributed in both species along all the chromosomes, with enrichment on heterochromatin short arms of pairs 14 and 16, and on the Y chromosome (Fig. 3m,n). Hence, the heterochromatic short arm of pairs 14 and 16 on both species shared satellite sequences, as the TaquSat2-107 and TaquSat5-3102, but also present different compositions on sequences, such as TaquSat6-84 and TaquSat5-3102.

Our data indicate that in *Talpa* genomes satDNA families are preferentially accumulated in the centromeric regions and in the C-band positive heterochromatic regions. Overall, the results demonstrate that the satellitome of *T. occidentalis* and *T. europaea* is very similar to that of *T. aquitania*, but differed significantly in the presence and amount of several satDNA families. As the estimated divergence time between *T. aquitania* and *T. occidentalis* is 2.47 ± 0.12 million of years ago (Mya) and between these species and *T. europaea* is 2.82 ± 0.10 Mya [61], the results demonstrated that the satellitome is a very dynamic component of the genomes due to the differences observed in these closely related species.

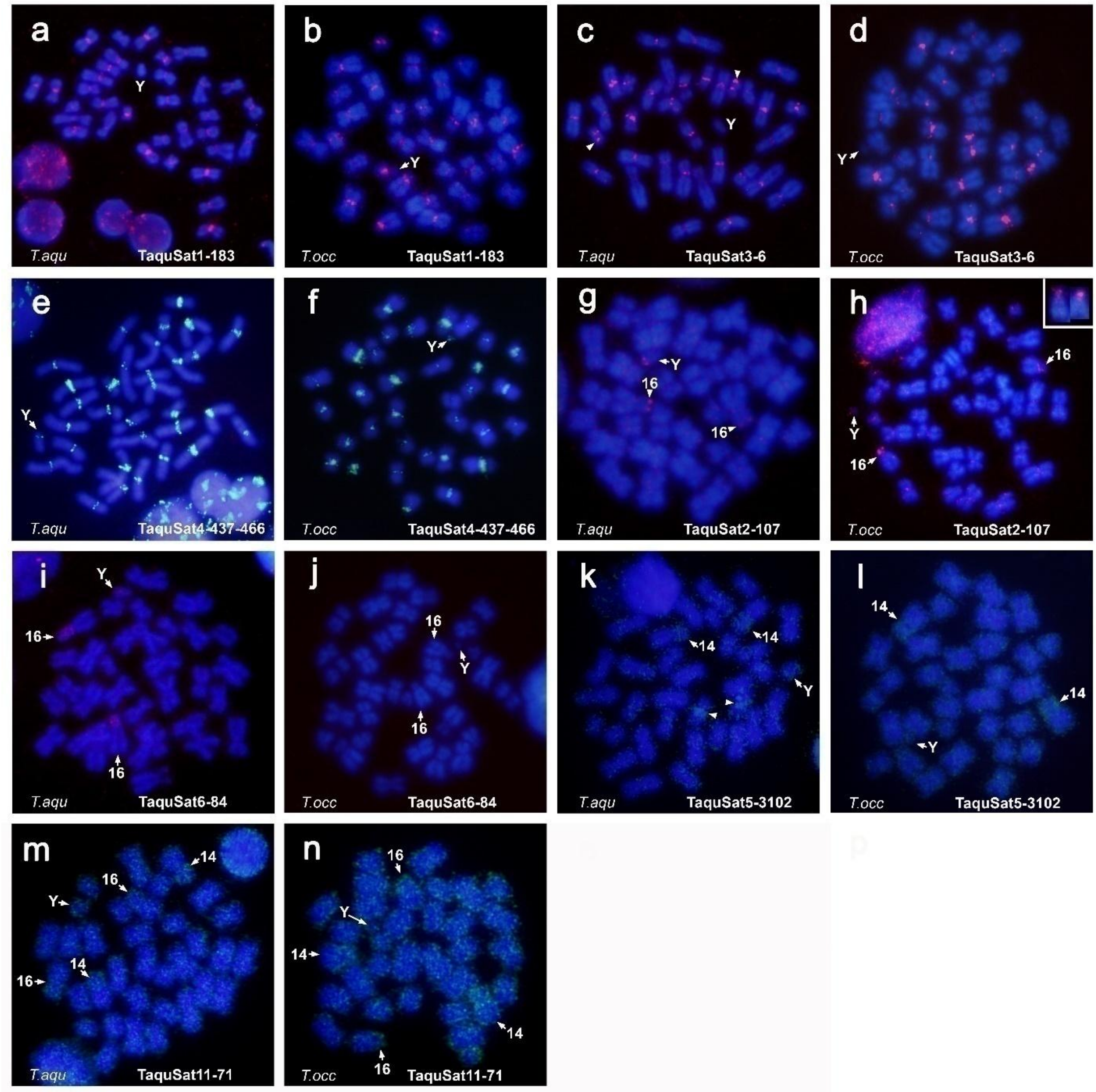


Figure 3. *Male metaphases of T. aquitania (**a**,**c**,**e**,**g**,**i**,**k**,**m**) and T. occidentalis (**b**,**d**,**f**,**h**,**j**,**l**,**n**) hybridized with satDNA probes: (**a**,**b**): TaquSat1-183;(**c**,**d**): TaquSat3-6;(**e**,**f**): TaquSat4-437-466; (**g**,**h**): TaquSat2-107;(**i**,**j**): TaquSat6-84;(**k**,**l**): TaquSat5-3102;(**m**,**n**): TaquSat11-71. Arrowhead denotes the interstitial signal on the small autosome pair in (**c**), accumulation of signal on the centromeric region of one autosome pair in (**k**). Insert in (**h**), pair 16 chromosomes from other metaphase hybridized with the same probe.*

Supplementary Materials: The following supporting information can be downloaded at: www.mdpi.com/xxx/s1, Fig. S1: Satellitome landscapes (abundance versus divergence) of *T. aquitania* satDNA families on *T. aquitania* and *T. occidentalis* genomes; Fig. S2: Correlation between the tandem structure index and the fold-change of satDNA family abundance among *T. occidentalis* assembled genome and *T. occidentalis* genome; Fig. S3: Presence and abundance of *T. aquitania* satDNA families across species of the family Talpaidea; Table S1: *T. aquitania* satDNA families consensus sequences; Table S2: List of oligos used to localize by FISH the repeat DNA sequences of *T. aquitania*; Table S3: Excel file containing data about *T. aquitania* satDNA families abundance in Talpidae species.

Author Contributions: J.G., G.A.-M., E.E.M., D.C.C.-d.-M., J.A.M., and A.S., conceived, designed, and performed the experiments, analyzed the data, and wrote the draft. All authors have read and agreed to the published version of the manuscript.

Funding: This work was funded by the Consejería de Innovación, Ciencia y Empresa of the Junta de Andalucía (Group: RNM924), and by the Universidad de Jaén (as part of the program Plan de Apoyo a la Investigación 2019–2020, Acción 1). DCC-d-M. is supported by the María Zambrano grant (E-23-2022-0219886).

InstitutionalReviewBoardStatement: Permissionfor captures was granted by the Servicio Territorial de Medio Ambiente de Burgos (Junta de Castilla-León) and the Delegación Territorial de la Consejería de Agricultura, Pesca y Medio Ambiente de Granada (Junta de Andalucía). All capture and sacrifice protocols were approved by the Junta de Andalucía Ethics Committee for Animal Experimentation (code: 22/05/2018/094).

Informed Consent Statement: Not applicable.

Data Availability Statement: The datasets generated and analyzed during the current study are available on request from the corresponding author.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Charlesworth, B.; Sniegowski, P.; Stephan, W. The evolutionary dynamics of repetitive DNA in eukaryotes. *Nature* **1994**, *371*, 215–220.
2. Garrido-Ramos, M.A. Satellite DNA: Na Evolving Topic. *Genes* **2017**, *8*, 230.
3. López-Flores, I.; Garrido-Ramos, M.A. The repetitive DNA content of eukaryotic genomes. *Genome Dyn.* **2012**, *7*, 1–28.
4. Biscotti, M.A.; Canapa, A.; Forconi, M.; Olmo, E.; Barucca, M. Transcription of tandemly repetitive DNA: Functional roles. *Chromosome Res.* **2015**, *23*, 463–477.
5. Biscotti, M.A.; Olmo, E.; Heslop-Harrison, J.S. Repetitive DNA in eukaryotic genomes. *Chromosome Res.* **2015**, *23*, 415–420.
6. Choi, I.Y.; Kwon, E.C.; Kim, N.S. The C- and G-value paradox with polyploidy, repeatomes, introns, phenomes and cell economy. *Genes Genom.* **2020**, *42*, 699–714.
7. Ruiz-Ruano, F.J.; López-León, M.D.; Cabrero, J.; Camacho, J.P. High-throughput analysis of the satellitome illuminates satellite DNA evolution. *Sci. Rep.* **2016**, *6*, 28333.
8. Ugarkovic, D.; Plohl, M. Variation in satellite DNA profiles-causes and effects. *EMBO J.* **2002**, *21*, 5955–5959.
9. Palomeque, T.; Lorite, P. Satellite DNA in insects: A review. *Heredity* **2008**, *100*, 564–573.
10. Plohl, M.; Luchetti, A.; Mestrovic, N.; Mantovani, B. Satellite DNAs between selfishness and functionality: Structure, genomics and evolution of tandem repeats in centromeric (hetero)chromatin. *Gene* **2008**, *409*, 72–82.
11. Mehrotra, S.; Goyal, V. Repetitive sequences in plant nuclear DNA: Types, Distribution, Evolution and Function. *Genom. Proteom. Bioinform.* **2014**, *12*, 164–171.
12. Dimitri, P.; Corradini, N.; Rossi, F.; Vernì, F. The paradox of functional heterochromatin. *Bioessays* **2005**, *27*, 29–41.
13. Grewal, S.I.; Elgin, S.C. Transcription and RNA interference in the formation of heterochromatin. *Nature* **2007**, *447*, 399–406.
14. Josse, T.; Teysset, L.; Todeschini, A.-L.; Sidor, C.M.; Anxolabehere, D.; Ronsseray, S. Telomeric trans-silencing: An epigenetic repression combining RNA silencing and heterochromatin formation. *PLoS Genet* **2007**, *3*, 158.
15. Ferree, P.M.; Prasad, S. How can satellite DNA divergence cause reproductive isolation? Let us count the chromosomal ways. *Genet. Res. Int.* **2012**, 2012, 430136. https://doi.org/10.1155/2012/430136.
16. Pezer, Z.; Brajkovic, J.; Feliciello, I.; Ugarkovic, D. Satellite DNA-Mediated Effects on Genome Regulation. *Genome Dyn.* **2012**, *7*, 153–169.

17. Plohl, M.; Mestrovic, N.; Mravinac, B. Centromere identity from the DNA point of view. *Chromosoma* **2014**, *123*, 313–325.
18. Pita, S.; Panzera, F.; Mora, P.; Vela, J.; Cuadrado, A.; Sánchez, A.; Palomeque, T.; Lorite, P. Comparative repeatome analysis on *Triatoma infestans* Andean and Non-Andean lineages, main vector of Chagas disease. *PLoS ONE* **2017**, *12*, e0181635.
19. Pita, S.; Mora, P.; Vela, J.; Palomeque, T.; Sánchez, A.; Panzera, F.; Lorite, P. Comparative analysis of repetitive DNA between the main vectors of Chagas disease: *Triatoma infestans* and *Rhodnius prolixus*. *Int. J. Mol. Sci.* **2018**, *24*, 1277.
20. Fingerhut, J.M.; Moran, J.V.; Yamashita, Y.M. Satellite DNA-containing gigantic introns in a unique gene expression program during *Drosophila* spermatogenesis. *PLoS Genet*. **2019**, *15*, e1008028.
21. Shatskikh, A.S.; Kotov, A.A.; Adashev, V.E.; Bazylev, S.S.; Olenina, L.V. Functional Significance of Satellite DNAs: Insights from *Drosophila*. *Front. Cell Dev. Biol.* **2020**, *8*, 312.
22. Jagannathan, M.; Yamashita, Y.M. Defective satellite DNA clustering into chromocenters underlies hybrid incompatibility in *Drosophila*. *Mol BiolEvol.* **2021**, *38*, 4977–4986.
23. Wei, X.; Eickbush, D.G.; Speece, I.; Larracuente, A.M. Heterochromatin-dependent transcription of satellite DNAs in the *Drosophila melanogaster* female germline. *eLife* **2021**, *10*, e62375.
24. Novák, P.; Neumann, P.; Macas, J. Graph-based clustering and characterization of repetitive sequences in next-generation sequencing data. *BMC Bioinformatics* **2010**, *11*, 378.
25. Novák, P.; Neumann, P.; Pech, J.; Steinhaisl, J.; Macas, J. RepeatExplorer: A Galaxy-based web server for genome-wide characterization of eukaryotic repetitive elements from next-generation sequence reads. *Bioinformatics* **2013**, *29*, 792–793.
26. Novák, P.; Ávila Robledillo, L.; Koblížková, A.; Vrbová, I.; Neumann, P.; Macas, J. TAREAN: A computational tool for identification and characterization of satellite DNA from unassembled short reads. *Nucleic Acids Res.* **2017**, *45*, e111.
27. Utsunomia, R.; Silva, D.M.Z.A.; Ruiz-Ruano, F.J.; Goes, C.A.G.; Melo, S.; Ramos, L.P.; Oliveira, C.; Porto-Foresti, F.; Foresti, F.; Hashimoto, D.T. Satellitome landscape analysis of *Megaleporinus macrocephalus* (Teleostei, Anostomidae) reveals intense accumulation of satellite sequences on the heteromorphic sex chromosome. *Sci. Rep.* **2019**, *9*, 5856.
28. da Silva, M.J.; Fogarin Destro, R.; Gazoni, T.; Narimatsu, H.; Pereira Dos Santos, P.S.; Haddad, C.F.B.; Parise-Maltempi, P.P. Great abundance of satellite DNA in

Proceratophrys (Anura, Odontophrynidae) revealed by genome sequencing. *Cytogenet. Genome Res*. **2020**, *160*, 141–147.

29. Mora, P.; Vela, J.; Ruiz-Ruano, F.J.; Ruiz-Mena, A.; Montiel, E.E.; Palomeque, T.; Lorite, P. Satellitome Analysis in the Ladybird Beetle *Hippodamia variegata* (Coleoptera, Coccinellidae). *Genes* **2020**, *11*, 783.
30. Montiel, E.E.; Mora, P.; Rico-Porras, J.M.; Palomeque, T.; Lorite, P. Satellitome of the red palm weevil, *Rhynchophorus ferrugineus* (Coleoptera: Curculionidae), the most divers among insects. *Front. Ecol. Evol.* **2022**, *10*, 826808.
31. Montiel, E.E.; Panzera, F.; Palomeque, T.; Lorite, P.; Pita, S. Satellitome Analysis of *Rhodnius prolixus*, one of the Main Chagas Disease Vector Species. *Int. J. Mol. Sci.* **2021**, *22*, 6052.
32. Boštjančić, L.L.; Bonassin, L.; Anušić, L.; Lovrenčić, L.; Besendorfer, V.; Maguire, I.; Grandjean, F.; Austin, C.M.; Greve, C.; Hamadou, A.B.; et al. The *Pontastacus leptodactylus* (Astacidae) repeatome provides insight into genome evolution and reveals remarkable diversity of satellite DNA. *Front Genet.* **2021**, *11*, 611745.
33. Weiss-Schneeweiss, H.; Leitch, A.R.; McCann, J.; Jang, T.S.; Macas, J. Employing next generation sequencing to explore the repeat landscape of the plant genome. In *Next Generation Sequencing in Plant Systematics Regnum Vegetabile*; Hörandl, E., Appelhans, M., Eds.; Koeltz Scientific Books: Königstein, Germany, **2015**; pp. 155–179.
34. Ruiz-Ruano, F.J.; Navarro-Domínguez, B.; Camacho, J.P.M.; Garrido-Ramos, M.A. Characterization of the satellitome in lower vascular plants: The case of the endangered fern *Vandenboschia speciose. Ann. Bot.* **2019**, *123*, 578–599.
35. Heitkam, T.; Schulte, L.; Weber, B.; Liedtke, S.; Breitenbach, S.; Kögler, A.; Morgenstern, K.; Brückner, M.; Tröber, U.; Wolf, H.; et al. Comparative Repeat Profiling of Two Closely Related Conifers (*Larix decidua* and *Larix kaempferi*) Reveals High Genome Similarity With Only Few Fast-Evolving Satellite DNAs. *Front. Genet.* **2021**, *12*, 683668.
36. Sader, M.; Vaio, M.;Cauz-Santos, L.A.; Dornelas, M.C.; Vieira, M.L.C.; Melo, N.; Pedrosa-Harand, A. Large vs. small genomes in Passiflora: The influence of the mobilome and the satellitome. *Planta* **2021**, *253*, 86.
37. Gutiérrez, J.; Aleix-Mata, G.; Lamelas, L.; Arroyo, M.; Marchal, J.A.; Sánchez, A. Karyotype analysis of the new *Talpa* species *Talpa aquitania* (Talpidae; Insectivora) from northern Spain. *Cytogenet. Genome Res.* **2019**, *159*, 26–31.
38. Biltueva, L.; Vorobieva, N. Chromosome evolution in Eulipotyphla. *Cytogenet. Genome Res.* **2012**, *137*, 154–164.

39. Ye, J.; Biltueva, L.; Huang, L.; Nie, W.; Wang, J.; Jing, M.; Su, W.; Vorobieva, N.V.; Jiang, X.; Graphodatsky, A.S.; et al. Cross-species chromosome painting unveils cytogenetic signatures for the Eulipotyphla and evidence for the polyphyly of Insectivora. *Chromosome Res.* **2006**, *14*, 151–159.
40. Jimenez, R.; Burgos, M.; Díaz de la Guardia, R. Karyotype and chromosome banding in the mole (*Talpa occidentalis*) from the south-east of the Iberian Peninsula. Implications on its taxonomic position. *Caryologia* **1984**, *37*, 253–258.
41. Kawada, S.; Harada, M.; Grafodatsky, A.S.; Oda, S. Cytogenetic study of the Siberian mole, *Talpa altaica* (Insectivora: Talpidae) and karyological relationships within the genus *Talpa*. *Mammalia* **2002**, *66*, 53–62.
42. Gornung, E.; Volleth, M.; Capanna, E.; Castiglia, R. Comparative cytogenetics of moles (Eulipotyphla, Talpidae): Chromosomal differences in *Talpa romana* and *Talpa europaea*. *Cytogenet. Genome Res.* **2008**, *121*, 249–254.
43. Gutiérrez, J.; Aleix-Mata, G.; Marchal, J.A.; Arroyo, M.; Castiglia, R.; Sánchez, A. Molecular Cytogenetic Analysis of Karyotype and Y Chromosome Conservation in Species of the Genus *Talpa* (Insectivora). *Cytogenet. Genome Res.* **2020**, *160*, 264–271.
44. Zurita, F.; Sánchez, A.; Burgos, M.; Jiménez, R.; Díaz de la Guardia, R. Interchromosomal, intercellular and interindividual variability of NORs studied with silver staining and in situ hybridization. *Heredity* **1997**, *78*, 229–223.
45. Real, F.M.; Haas, S.A.; Franchini, P.; Xiong, P.; Simakov, O.; Kuhl, H.; Schöpflin, R.; Heller, D.; Moeinzadeh, M.H.; Heinrich, V.; et al. The mole genome reveals regulatory rearrangements associated with adaptive intersexuality. *Science* **2020**, *370*, 208–214.
46. Valeri, M.P.; Dias, G.B.; do Espírito Santo, A.A.; Moreira, C.N.;Yonenaga-Yassuda, Y.; Sommer, I.B.; Kuhn, G.C.S.; Svartman, M. First Description of a Satellite DNA in Manatees’ Centromeric Regions. *Front. Genet.* **2021**, *12*, 694866.
47. Novák, P.; Neumann, P.; Macas, J. Global analysis of repetitive DNA from unassembled sequence reads using RepeatExplorer2. *Nat. Protoc.* **2020**, *15*, 3745–3776.
48. Storer, J.; Hubley, R.; Rosen, J.; Wheeler, T.J.; Smit, A.F. The Dfam community resource of transposable element families, sequence models, and genome annotations. *Mob. DNA* **2021**, *12*, 2.
49. Hall, T.A. BioEdit: A user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucl. Acids. Symp. Ser.* **1999**, *41*, 95–98.
50. Madeira, F.; Park, Y.M.; Lee, J.; Buso, N.; Gur, T.; Madhusoodanan, N.; Basutkar, P.; Tivey, A.R.N.; Potter, S.C.; Finn, R.D.; et al. The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Res.* **2019**, *47*, W636–W641.

51. Camacho, J.P.M.; Cabrero, J.; López-León, M.D.; Martín-Peciña, M.; Perfectti, F.; Garrido-Ramos, M.A.; Ruiz-Ruano, F.J. Satellitome comparison of two oedipodine grasshoppers highlights the contingent nature of satellite DNA evolution. *BMC Biol.* **2022**, *20*, 36. https://doi.org/10.1186/s12915-021-01216-9.
52. R Core Team. *A Language and Environment for Statistical Computing; R Foundation for Statistical Computing*: Vienna, Austria, 2019.
53. Warnes, G.R.; Bolker, B.; Bonebakker, L.; Gentleman, R.; Huber, W.;Liaw, A.; Lumley, T.;Maechler, M.; Magnusson, A.; Moeller, S.; et al. Gplots: Various R Programming Tools for Plotting Data. R Package Version 3.1.1. 2020. Available online: https://CRAN.R-project.org/package=gplots (accessed on 15 June 2022).
54. Garnier, S.; Ross, N.; Rudis, R.; Camargo, A.P.; Sciaini, M.; Scherer, C. Rvision—Colorblind-Friendly Color Maps for R. R Package Version 0.6.1. **2021**. Available online: https://cran.r-project.org/package=viridisLite (accessed on 15 June 2022).
55. Untergasser, A.; Cutcutache, I.; Koressaar, T.; Ye, J.; Faircloth, B.C.; Remm, M.; Rozen, S.G. Primer3—New capabilities and interfaces. *NucleicAcids Res.* **2012**, *40*, e115.
56. Burgos, M.; Jiménez, R.; Díaz de la Guardia, R. A rapid, simple and reliable combined method for G-banding mammalian and human chromosomes. *Stain Technol.* **1986**, *61*, 257–260.
57. Fernández, R.; Barragán, M.J.; Bullejos, M.; Marchal, J.A.; Martínez, S.; Díaz de la Guardia, R.; Sánchez, A. Molecular and cytogenetic characterization of highly repeated DNA sequences in the vole *Microtus cabrerae*. *Heredity* **2001**, *87*, 637–646.
58. Cabral-de-Mello, D.C.; Marec, F. Universal fluorescence in situ hybridization (FISH) protocol for mapping repetitive DNAs in insects and other arthropods. *Mol. Genet. Genomics* **2021**, *296*, 513–526.
59. Escoda, L.; Castresana, J. The genome of the Pyrenean desman and the effects of bottlenecks and inbreeding on the genomic landscape of an endangered species. *Evol. Appl.* **2021**, *14*, 1898–1913.
60. Nicolas, V.; Martínez-Vargas, J.; Hugot, J.P. Preliminary note: *Talpa aquitania* nov. sp. (Talpidae, Soricomorpha) a new mole species from southwest France and north Spain. *Bul. Acad. Vet. Fr.* **2015**, *168*, 329–334.
61. Nicolas, V.; Martínez-Vargas, J.; Hugot, J.P. Molecular data and ecological niche modelling reveal the evolutionary history of the common and Iberian blind moles (Talpidae) in *Europe. Zool. Scr.* **2017**, *46*, 12–26.
62. Hughes, S.E.; Hawley, R.S. Heterochromatin: A rapidly evolving species barrier. *PLoS Biol.* **2009**, *7*, e1000233. https://doi.org/10.1371/journal.pbio.1000233.

63. Peona, V.; Kutschera, V.E.; Blom, M.P.K.; Irestedt, M.; Suh, A. Satellite DNA evolution in Corvoidea inferred from short and long reads. *Mol. Ecol.* **2022,** 10.1111/mec.16484. https://doi.org/10.1111/mec.16484.
64. Fry, K.; Salser, W. Nucleotide sequences of HS-α satellite DNA from kangaroo rat *Dipodomys ordii* and characterization of similar sequences in other rodents. *Cell* **1977**, *12*, 1069–1084. https://doi.org/10.1016/0092-8674(77)90170-2.
65. Mestrović, N.; Plohl, M.; Mravinac, B.; Ugarković, D. Evolution of satellite DNAs from the genus *Palorus*—Experimental evidence for the" library" hypothesis. *Mol. Biol. Evol.* **1998**, *15*, 1062–1068. https://doi.org/10.1093/oxfordjournals.molbev.a026005.

Supplementary Materials:

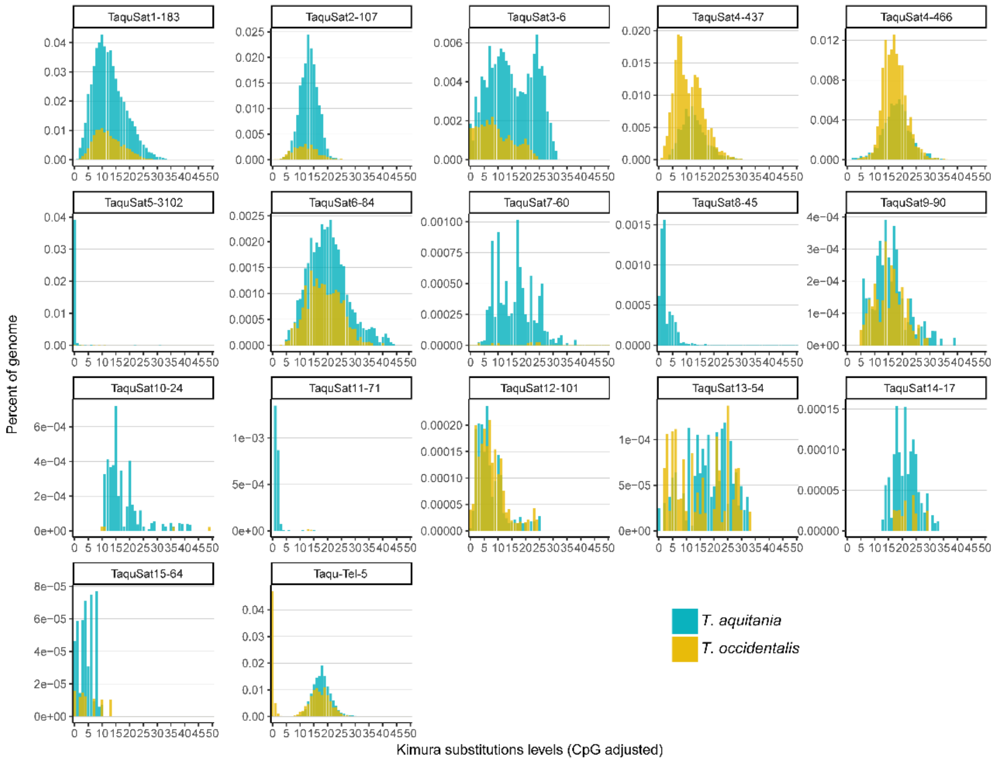


Figure S1. *Satellitome landscapes (abundance versus divergence) of Talpa aquitania satDNA families on Talpa aquitania and Talpa occidentalis genomes.*

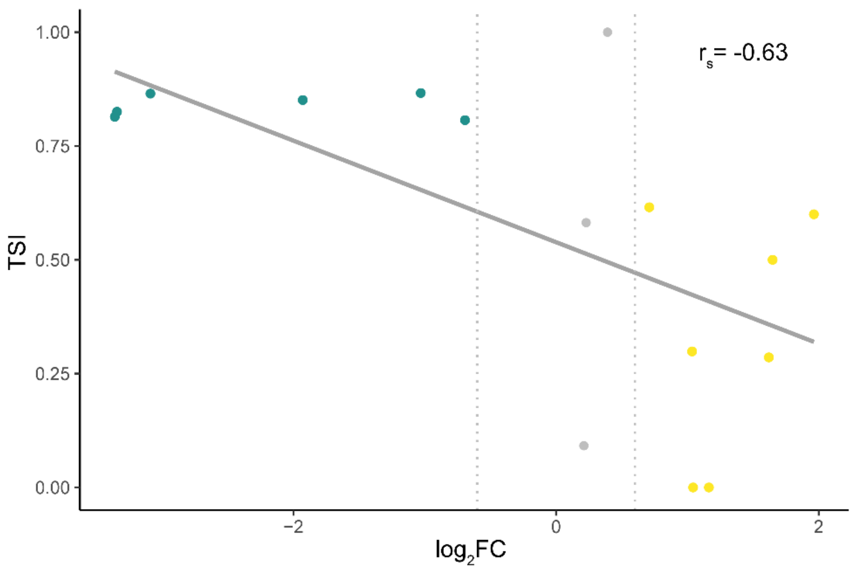


Figure S2. *Correlation between the tandem structure index and the fold-change of satDNA family abundance among Talpa occidentalis assembled genome and T. occidentalis genome. The fold-change is expressed as log2 of genome percentage among samples. SatDNA families with less of 1.5 fold-change between samples (-0.6 < log2 ratio < 0.6) are in grey, most abundant satDNA families on T. occidentalis assembly in yellow and most abundant on T. occidentalis unassembled reads in blue. Regression line is depicted in dark grey.*

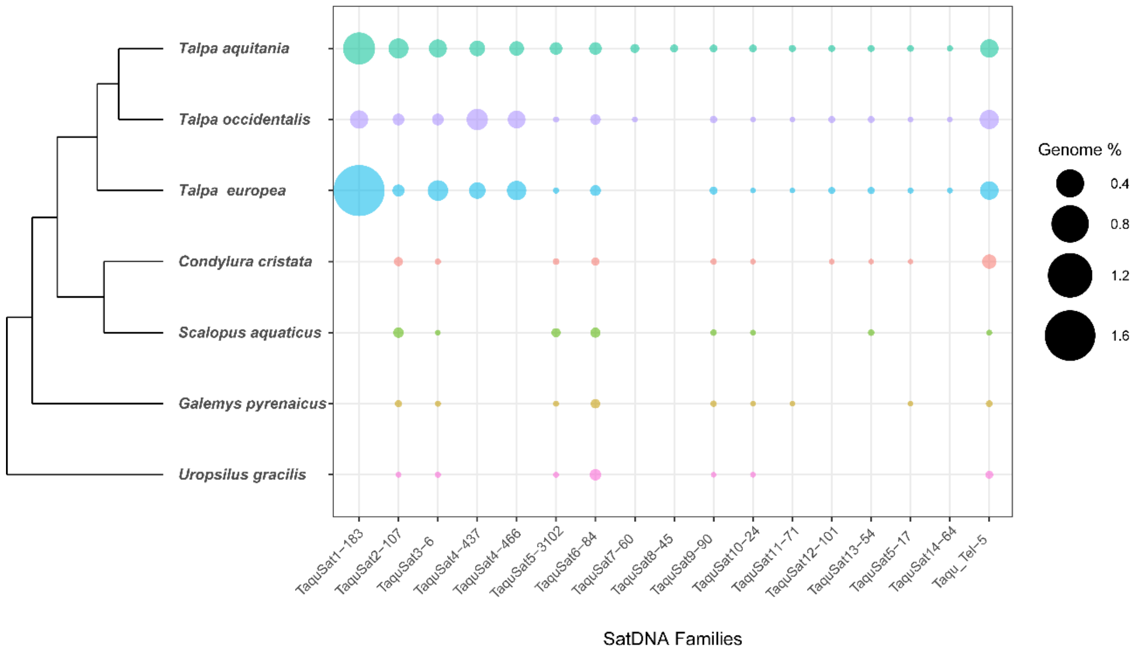


Figure S3*: Presence and abundance of Talpa aquitania satDNA families across species of the family Talpaidea. The bubble size indicates the abundance of satDNA family.*

***Table S1**: T. aquitania satDNA families consensus sequences*

satDNAname	consensussequence
TaquSat1-183	GAGATGTAGAACTACAAGAACAACAACAAGAACTAGAAGAACAAGAAGAACAAGAAGAAGAAACAGAAGAATAAGTCA AAAAAGCAGAAATAACAGAAGAATAAGCTGAGGAAGAAGCAGAAAAAGAAGTAATTGAAAAAGCAGAAAATGCAGAAG AAACAGAAGAAGATGAAGAAGAAGAAG
TaquSat2-107	TTTCTCCTCAGAGGCCTGTCTTCTCCTCAGAATCTGCTTCCCAGGGTGTCATTTCTCCTCAGATATTTCTCCTCAACA CTCCTTACTACAGTATTTCTCCTCAGTTT
TaquSat3-6	CTTCTT
TaquSat4-437-470	Consensuspattern 437 bp GGTGTGCCCCGAGGGGAAGATTTACCGTTCCCCCAAGGGGGAAGAGGGTCGCGTTTCCCCCTGGGGTAGAGGGTCCCC TACCCCCGTGGGGGCTGAGGGTCCCCATCCCCCTCGCCGGCGCCCCCTCGCGGCGGCTCTGTGAAGTGCACCTGTGCT CAGGGAGAGTCCGGTGGGCGGGGTCAGGCTCCGGCCGGGGGCGGAGCCTGGGGCTCATCAGCGCTGCCTGGCCAGGTG AGGACTCTTCCCCTTTGAGTGGTGCTGGGTTGCTCGGGCCTTTCCCCGCCGGCCTGTCTTCCTGCAGTTTTGGAGGAG GCCGCGGTCGCCGTCTGTCTGTGGCTTGCAGGGGGTGCCCTCCCTGACGGTGACCCCGCCCCTTCGTCTATGGCGTCT TGGGCGCAATGTCCGCCCCTCGGGTTGTGAGTGCGGAGGAGGTTGCA Consensuspattern 466 bp GGCTTCCCCCAAGGGGAAGATAGCCCCGTTCCCCCATGGGGGAGGAAAATCCCGTTCCCCCATGGGTGCAGAGGTCCC CCAAACCCCATGAAGGCTGAGGTCCCCGAACCCCCTCACCAGCGCCCCCTAGCGGCAGCCCTGAGAAGTGCACCTTCG AGCATGCAGAGTCCAGTGGGCGGGGTCAGGCTACGGCCAGGGGCGGAGCCTGGGGCTCATCAGCGCTGCGGGCCAGGT GAGGACTCTTACCCTTTCAGTGGGGCTGGCTTGCCCGGGGCTTTCCGCGCCACCCTGTCTGCCTGCAGTTTTGGAAGA GGCTCAGGTCGCCGCTTGTCCCTGGCGTACAGGGGGTGCCCTCCCCGGACAGGGACCCCACCCCTCCGTCTAGGGCGC CTGAGGGCGCATCGGCATGCCCCTAGGGCTGTGAGTGCGAGAGGAGGGTGCAGGTGTTCCCCGGGGGAAAGATTTG
TaquSat5-3102	TGATCATAGATGGGGAAATACCAATCCACATCAACAAGAAGAGGGGGTTCCTCTTCCTGAGGACACAAGTGGATTATT CCACCTTTCAGGCTCTACATGAACTGCCTTCCATAGGAGAGCTGGCGTAATATAAATATTTATGAAAAGAAATGCACT AAAACTAAGATGAAAGTGGGCGCAGACCCAGAGAGCCACAGGGTGGGGGGTACCTAGAAGTCAGTGACCTGGCACCCC TTTGAAACAACCATTTGCTTAACAGGGACTTGGGATTTGCCAGACAAGATATCGATTGCACAGAAGCAAGAAGGACTT GGTTTGGTTTGGAGGCAGGAATGAAGCATGACAATGAGCATCACATTGTGCTCCTTACCTCTCCCTGGAGGGTTCTGC CTCCCAGGTTCTGCAGCGGCTGCTTTCTCCTCCACATTTTCCACGATCTCTAATCTTGAAGATAGAACAAACTGGTCA TCCATACAAATGACCAAACCAAGTGCTTTACTAGACCTGAAACTGTAGTTGAGGACCTCAACTCCACCCTGAAGCAGA ATCACCACAGCCTGCTACCTGCTTGGCTAGTGGGCGGCTTCCCTGCGCTGCTCCTGCATCTCCACTCCATCTGACACG CCAGGACCTGTGGGGAAACACAGCCACTGGTCAGGCCCAGTACAGTGCAGAGAAGGGGGGAAGGAGAGTCAACTTACC CATTTCTCCTGAGCCTTCCTTCCAGCCTGATCCACCGAAACTGATCCTTGGAGCCTGAGCCTTTGAGCCTTCAAGCCC TGGAGCCTTCTTCTCCTGCTGGGCAATCTGAAAAGAGCAACACAGCCAAGCCCTGGGGTTAGGTCCCTGGGGCTTTGC AAACATTTGTTTGAGGAGCAGGAGGATTGTCTGTGGGGAGAGGCAGGTTACCTGGAGTTGATGAACAAAGTCAGCTTG CCTCAGTTTTTCTGCAGTTTCTTCAGTTTCTTGCCTATTATGTTTGATTTCATTAGAGAAAGCAGAATATTAAGAATC ATTTAAAGTCATTTCCCAAGTGCCTATTACCTTCACATCTATTGTTCCCTAATACAACTGACCTATACACAGAAAACA ATTTTTCAAGGTTATAAACACACTGTGGGGTCAAAGCTTTTTGATGCATCATTTCCTTATTTACTGTTTGATGCCCAC CAGATTCTCAGCCTGTGTAACTTACTTACCCATCTCAACATTTCGCTAATTGATCATGTCCTCAATAGCTTGTCTCAT AAAGCTCTAATCTCCTTTGATTTATATCCCTATCATATTTGCCAGAATTTCGAAAGTCTAGGTCCATCAAATGACAAA AGCATATTTACGGATATGATTCACCCTACAGTGAGCTAGTTGTTCGTGATTATAAAAATTACACTGTCCTCAGAGAC ATTTGTATGAACTAAAAGGCTAAGAGCATTGGGGGATCAAGAAATATTTATGAACTTATTTTTTGTTGTTCTTTGTAG AACAACTTTTGTTCCTTGTGTTGTGTCTCTATATGACTTCTGAAACTTTGCTGATATCACTAAACATCCTGAGTCTAA ACATTACAAGTTTAAAAACAAAAAGCCGGGATTTCTTGACCTGCTGGTATCTACGAACTCACTTGTATTTTTCAATTT CTTGAAGAATTTCTTTCAGTTTCATCTCCAGAGCTACAAGCTCCTTCTCCTTTTTCTCCCTTTCATGCTGTTTCCATT TCAGCATCCTGAATTGGGAAAGACAGAGATGATGCCCAGCACCCCTCACGAAATCCTGCCTTTCCCAGCTCAGGGCCT GGATTGCAAAGCCCCCTCCTGCGCTGGCCTCTGCTGTGCCCTCTGCAAACTCGGCTTTTCCTCACAGTCTCCACCAGC TCATTGCTCTTCAGTCATCAGGACAAGGGCGTGGGGTGCCTCCCTGTGCTGGGGCTGCTGTTTGAGAAGGTGAAGGCT TTTTCTTTCAAACTGGAAACGTACACGTGAAATTTCCACCTTGAAATAAAGGGGATGCCTGTCCAACTCACGACAACC ACATACATTAGGGCAAAGAAGGGGATCCTACCAGTCCGCATCTCCGAGGGGAGCCTGATTATCAAGCCCAGGCTGTGC TTATGCAGGACAAGTAGCTATGAAGACAGTCTTACAGGCTGGGACTTCCCGAGCCCTCTGCCATCAGCCCCACATAGT CTGAAGGCTCGGGAGCAGAGCTTTGCCTCCTACTGCAAACACCTCGTGGGAGAAAAGGTCCATCTTATCGCTGCTAGG ATGTACTGAAATATAAAATGCTGGAAACTGTCCCTTAGTCAGGGTATTATCTCTAGGAAGCAATTTCAAATGGACAAC AAACAGCTTCAGGGCGTCAATTTCCTTTACCAAGACCCAGATCACAGCTAGTGTGGGCAGCAATGAGCTGTGGCTGGG AATCCTTCCTTTGCTCTGCCGCCCTCAGACTATTTTTCTCATGCTGCTGACTCCAAGCTGACACTCTATTAACTGTAT AGATTGACACACAAAATAATAAGGAACCGGGACAAAGTAAGCACTGAAGAGAGAAGAGGACACACACTTCTTTCTAGC CTCCCTACTGTTCCTGCAGCAACCGCTTCCTATAGTGACTGCAGTACTTATGGCTTTGCTGTTTACATGCCCACATTA CATGCCCACAGAAAAAGATCACAGAGGAAGGAAAGCTGGTGGGGACATCATCCCCCTACCCTTCTCTGCGTCCTCAGC TGGCACAGAGTTGCCATCACTCACGGGAGCTGCACCAGGTCCTGAGTCTGACAAAGACCACAGAAGTGACAGAGGTGT ATTGGGACTTGCAGAGATGGTGAGGAAACCCTCCTTGCCTTCCCCCATTACCCTCCATTGTCCCCACATTCAGTCCCT TCTCAGGATATGGGGCTCTCCAGGAGCCTCACTGGCCCCAGAGCCCAAGGGGGACAGGAATCACCCCAGTCCCCTGGG TCAGACGCTTTCTCTCAACTTCATCACCCAAGTGGGAACCTGTCAGGGCTCACCTAGTTGGAGTCTGGCACAAAAACA GAGGCTCCCCAATGCTTGGCAAGACAAATTGCCATTTCAATGTAGGTTAAAGAAAAAGTA
TaquSat6-84	GAAACCYTAYRAATGTGAAGARTGTGGGAAAGCYTTTWMCCRKAAGTCAHMCCTCACTRAACATCAGAGAATTCAYAC TGGAGA
TaquSat7-60	GATCTGTCAGCAGCTAGGTATGCCTTGCGATCACAAGCCTCAACTAGTGGTACCTGGGTT
TaquSat8-45	GGCTGAAGACCATCCACCAGTCTCATCAACATCTCGAGACGTAAA
TaquSat9-90	GAACCAAGTACAGAGGANAGGAGACAAATGATAGCACAGGGTGGAGGCTCAGATGAAGANAGGGAATACAGACAGRST AGAAGGAATCTA
TaquSat10-24	GAGCCTTGATCATGGTCACAGACT
TaquSat11-71	TCAGCCTACCCCCCAAGGGGTCTACAGCCACTCACAGTTTAACAGGCAGTGCCTCAGCGCCTGCCACTGAAG
TaquSat12-101	CGGAGGAAAACCACTGAATGACAGGACTTGGTGTTGAGAGCCTGGAAAGTGACCTGCGTGGAGTCGGGCTTCTACATC TCCACAGTCTGAGAATAACAAAA
TaquSat13-54	CATGCAGGATCAAGTGCAACGACAAAATCAGACTTTTGAGTCAGCTTTTCAAAG
TaquSat14-17	GACCCCAGAACCCCAGT
TaquSat15-64	AGCCTGCCGCAGAGCAGGCTGGGAAGGATGAGACCAAGCAATAGAATAAAGAGGGCTGCCTTCA

Table S2: *List of oligos used to localize by FISH the repeat DNA sequences of T.aquitania*

Oligo	Sequentia (5' a 3')
TaquSat1-183-F	AATAAGTCAAAAAAGCAGAAATAA
TaquSat1-183-R	TTGTTCTTCTAGTTCTTGTTGTTG
TaquSat2-107	TCWCCTCAGATWTTTCTCCTC
TaquSat3-6	TTCTTCTTCTTCTTCTTCTTCTTC
TaquSat4-437-466-F	TGCCTGCAGTTTTGGAGG
TaquSat4-437-466-R	GACTCTGCCTGCGCACAG
TaquSat5-3102-F1	CACATCAACAAGAAGAGGGGGTTC
TaquSat5-3102-R1	TGAGTTCGTAGATACCAGCAGGTC
TaquSat5-3102-F2	AGCATCCTGAATTGGGAAAGACAG
TaquSat5-3102-R2	GAGCCTCTGTTTTTGTGCCAGAC
TaquSat6-84-F	CATCAGAGAATTCAYACTGGAG
TaquSat6-84-R	AAARGCTTTCCCACAYTCTTCAC
TaquSat7-60	CGATCACAAGCCTCAACTAGTGGT
TaquSat8-45	AGACCATCCACCAGTCTCATCAAC
TaquSat9-90-F	CAGATGAAGANAGGGAATACAGAC
TaquSat9-90-R	CACCCTGTGCTATCATTTGTCTC
TaquSat10-24	AGTCTGTGACCATGATCAAGGCTC
TaquSat11-71-F	CCTCAGCGCCTGCCACTGAAG
TaquSat11-71-R	GTGAGTGGCTGTAGACCCCTTG
TaquSat12-101-F	GGAAAACCACTGAATGACAGGAC
TaquSat12-101-R	GTTATTCTCAGACTGTGGAGATGT
TaquSat13-54	GCAGGATCAAGTGCAACGACAAAA
TaquSat14-17	CAGAACCCCAGTGACCCCAGAAC
TaquSat15-64	GGCTGGGAAGGATGAGACCAAGCA

***Table S3-1**: Data about T. aquitania satDNA families abundance in read of sequencing of T. occidentalis.*

SatDNAFamily	RUL	A-T%	Genome %	Kimura Divergence %	TSI	DivPeak	RSP	log2(T. occidentalis/T. aquitaniaGenome %)
TaquSat1-183	183	67,2	0,13105418	12,77	0,81	10	0,39	-2,09
TaquSat2-107	107	55,1	0,0312906	12,41	0,87	12	0,43	0,18
TaquSat3-6	6	66,7	0,03105831	9,62	0,83	8	0,31	-2,00
TaquSat4-437	437	30,4	0,19805295	11,34	0,85	7	0,39	1,27
TaquSat4-470	466	32,6	0,12019506	17,67	0,81	17	0,46	0,87
TaquSat5-3102	3102	53	0,00025284	3,13	1,00	0	0,62	-7,31
TaquSat6-84	84	65,5	0,02065225	19,12	0,58	14	0,28	-0,95
TaquSat7-60	60	48,3	0,00014409	17,32	0,29	22	0,45	-6,11
TaquSat9-90	90	56,7	0,00293611	15,75	0,62	14	0,31	-0,53
TaquSat10-24	24	50	9,30E-05	26,36	0,50	10	0,52	-5,49
TaquSat11-71	71	40,3	2,88E-05	13,71	0,00	13	1,00	-6,34
TaquSat12-101	101	50,5	0,0020102	8,08	0,09	7	0,42	0,16
TaquSat13-54	54	59,3	0,0014128	15,5	0,30	25	0,14	-0,31
TaquSat14-17	17	35,3	0,00022302	22,17	0,60	24	0,30	-2,53
TaquSat15-64	64	45,3	8,62E-05	5,39	0,00	0	0,32	-2,40
Taqu_Tel-5	6	50	0,14870516	11,44	0,87	0	0,36	-2,42
Total	16		0,69					
Media	304,50	50,39	0,04	13,86	0,56	11,44	0,42	-2,25
Mediana	77,5	50,25	0,00247315	13,24	0,61	11	0,39	-2,05
SD	758,68	11,55	0,07	6,00	0,33	7,95	0,19	2,71
Min	6	30,4	2,8817E-05	3,13	0	0	0,14	-7,31
Max	3102	67,2	0,19805295	26,36	1	25	1	1,27

Table S3-2: *Data about T. aquitania satDNA families abundance in the assembly of T. occidentalis (Real et al., 2020) (GCA_014898055.1).*

SatDNAFamily	RUL	A-T%	Genome %	Kimura Divergence %	DivPeak	RSP	log2(T. occidentalis assembly/T. occidentalis Genome %)
TaquSat1-183	183	67,2	0,01275579	11,14	12	0,43	-3,36
TaquSat2-107	107	55,1	0,01530967	11,09	12	0,46	-1,03
TaquSat3-6	6	66,7	0,00305826	13,98	22	0,31	-3,34
TaquSat4-437	437	30,4	0,05197857	12,81	11	0,71	-1,93
TaquSat4-470	466	32,6	0,07428941	19,26	18	0,70	-0,69
TaquSat5-3102	3102	53	0,0003316	7,08	0	0,46	0,39
TaquSat6-84	84	65,5	0,024196	24,35	24	0,25	0,23
TaquSat7-60	60	48,3	0,00044285	25,33	17	0,24	1,62
TaquSat8-45	45	51,1	0,0000258	4,46	6	0,66	
TaquSat9-90	90	56,7	0,00479986	19,51	13	0,31	0,71
TaquSat10-24	24	50	0,00029137	38,96	43	0,30	1,65
TaquSat11-71	71	40,3	0,0000645	13,01	1	0,32	1,16
TaquSat12-101	101	50,5	0,00232686	8,57	3	0,32	0,21
TaquSat13-54	54	59,3	0,0028954	20,45	22	0,32	1,04
TaquSat14-17	17	35,3	0,0008692	22,06	21	0,51	1,96
TaquSat15-64	64	45,3	0,00017764	6,19	6	0,54	1,04
Taqu_Tel-5	6	50	0,01746542	5,32	0	0,74	-3,09
Total	17		0,21				
Media	289,24	50,43	0,01	15,50	13,59	0,45	-0,21
Mediana	71	50,5	0,0028954	13,01	12	0,43	0,31
SD	737,28	11,19	0,02	9,04	11,09	0,17	1,83
Min	6	30,4	2,5834E-05	4,46	0	0,24	-3,36
Max	3102	67,2	0,07428941	38,96	43	0,74	1,96

Table S3-3: *Data about T. aquitania satDNA families abundance in Talpa europaea (SRX8240408) available in NCBI SRA database.*

SatDNA Family	RUL	A-T%	Genome %	Kimura Divergence %	TSI	DivPeak	RSP
TaquSat1-183	183	67,2	1,63558163	15,18	0,89	13	0,40
TaquSat2-107	107	55,1	0,03436573	13,7	0,86	15	0,48
TaquSat3-6	6	66,7	0,17862143	8,85	0,77	0	0,30
TaquSat4-437	437	30,4	0,09674406	15,27	0,83	13	0,46
TaquSat4-470	466	32,6	0,15377198	18,77	0,88	18	0,53
TaquSat5-3102	3102	53	0,0003582	4,89	0,75	0	0,40
TaquSat6-84	84	65,5	0,02245188	22,1	0,50	24	0,28
TaquSat9-90	90	56,7	0,00518514	15,55	0,45	15	0,39
TaquSat10-24	24	50	4,70E-05	34,55	0	35	0,78
TaquSat11-71	71	40,3	3,54E-05	8,59	0	5	0,35
TaquSat12-101	101	50,5	0,00187013	8,98	0	2	0,30
TaquSat13-54	54	59,3	0,00226377	18,52	0,17	13	0,30
TaquSat14-17	17	35,3	0,00040854	23,77	0,21	26	0,38
TaquSat15-64	64	45,3	0,00015699	7,22	0	4	0,63
Taqu_Tel-5	6	50	0,13334408	18,88	0,68	0	0,14
Total	15		2,27				
Media	320,80	50,53	0,15	15,65	0,47	12,20	0,41
Mediana	84	50,5	0,00518514	15,27	0,50	13	0,39
SD	782,40	11,94	0,42	7,68	0,37	10,59	0,16
Min	6	30,4	3,5439E-05	4,89	0	0	0,14
Max	3102	67,2	1,63558163	34,55	0,89	35	0,78

Table S3-4. *Data about T. aquitania satDNA families abundance in Condylura cristata (SRX101046) available in NCBI SRA database.*

SatDNAFamily	RUL	A-T%	Genome %	Kimura Divergence %	TSI	DivPeak	RSP
TaquSat2-107	107	55,1	0,01070824	29,32	0,19	31	0,38
TaquSat3-6	6	66,7	0,00040774	3,27	0,67	2	0,76
TaquSat5-3102	3102	53	0,00111034	23,15	0,39	24	0,40
TaquSat6-84	84	65,5	0,00633523	21,02	0,31	22	0,30
TaquSat9-90	90	56,7	0,00074344	27,15	0,37	25	0,37
TaquSat10-24	24	50	7,33E-05	23,12	0	25	0,51
TaquSat12-101	101	50,5	8,57E-05	30,46	0	26	0,40
TaquSat13-54	54	59,3	2,33E-05	18,89	0	19	1,00
TaquSat14-17	17	35,3	1,81E-05	16,89	0	16	1,00
Taqu_Tel-5	6	50	0,0620409	3,04	0,85	0	0,52
Total	10		0,08				
Media	359,10	54,2	0,0082	19,63	0,28	19,00	0,56
Mediana	69	54,1	0,0006	22,07	0,25	23	0,46
SD	921,45	18,4	0,02	10,93	0,30	11,34	0,30
Min	6	35,3	0,00002	3,04	0,00	0	0,30
Max	3102	66,7	0,06	30,46	0,85	31	1,00

Table S3-5*. Data about T. aquitania satDNA families abundance in Scalopus aquaticus (SRX4562103) available in NCBI SRA database.*

SatDNAFamily	RUL	A-T%	Genome %	Kimura Divergence %	TSI	DivPeak	RSP
TaquSat2-107	107	55,1	0,0193296	34,32	0,19	33	0,50
TaquSat3-6	6	66,7	5,49E-05	9,64	0,00	4	0,61
TaquSat5-3102	3102	53	0,01108782	21,89	0,75	22	0,50
TaquSat6-84	84	65,5	0,01684471	25,14	0,46	25	0,33
TaquSat9-90	90	56,7	0,00088241	23,81	0,07	19	0,43
TaquSat10-24	24	50	0,00016664	35,15	0,20	31	0,29
TaquSat13-54	54	59,3	0,00084433	26,65	0,00	27	0,57
Taqu_Tel-5	6	50	0,00020652	31,61	0,06	42	0,23
Total	8		0,05				
Media	434,13	57,04	0,01	26,03	0,21	25,38	0,43
Mediana	69	55,9	0,00086337	25,90	0,13	26	0,47
SD	1078,68	6,43	0,01	8,24	0,26	11,20	0,14
Min	6	50	5,4913E-05	9,64	0	4	0,23
Max	3102	66,7	0,0193296	35,15	0,75	42	0,61

Table S3-6. *Data about T. aquitania satDNA families abundance in Galemys pyrenaicus (SRX10243621) available in NCBI SRA database.*

SatDNAFamily	RUL	A-T%	Genome %	Kimura Divergence %	TSI	DivPeak	RSP
TaquSat2-107	107	55,1	0,00249114	30,24	0,24	31	0,37
TaquSat3-6	6	66,7	0,00034445	4,3	0,73	0	0,54
TaquSat5-3102	3102	53	0,00018663	16,19	0,5	15	0,31
TaquSat6-84	84	65,5	0,01263538	25,12	0,50	23	0,28
TaquSat9-90	90	56,7	0,00076707	26,63	0,31	24	0,59
TaquSat10-24	24	50	8,10E-05	36,97	0	30	0,44
TaquSat11-71	71	40,3	3,38E-05	33,34	0	37	0,65
TaquSat14-17	17	35,3	1,97E-05	20,83	0	20	1,00
Taqu_Tel-5	6	50	0,00146127	20,27	0,49	0	0,18
Total	9		0,018				
Media	389,67	52,5	0,0020	23,77	0,31	20,00	0,48
Mediana	71	53,0	0,0003	25,12	0,31	23	0,44
SD	967,53	19,2	0,00	11,94	0,27	13,82	0,28
Min	6	35,3	0,00002	4,30	0,00	0	0,18
Max	3102	66,7	0,01	36,97	0,73	37	1,00

Table S3-7*. Data about T. aquitania satDNA families abundance in Uropsilus gracilis (SRX4562112) available in NCBI SRA database.*

SatDNAFamily	RUL	A-T%	Genome %	Kimura Divergence %	TSI	DivPeak	RSP
TaquSat2-107	107	55,1	0,00014379	24,73	0,43	24	0,38
TaquSat3-6	6	66,7	0,00033618	7,02	0,03	3	0,43
TaquSat5-3102	3102	53	0,00018507	28,8	0,67	32	0,43
TaquSat6-84	84	65,5	0,02775217	27,09	0,41	24	0,30
TaquSat9-90	90	56,7	2,79E-05	34,27	0,00	34	1,00
TaquSat10-24	24	50	7,11E-05	42,01	0,00	44	0,36
Taqu_Tel-5	6	50	0,00477876	26,55	0,65	26	0,38
Total	7		0,033				
Media	488,43	56,71	0,00	27,21	0,31	26,71	0,47
Mediana	84	55,1	0,00018507	27,09	0,41	26	0,38
SD	1153,24	6,87	0,01	10,69	0,30	12,63	0,24
Min	6	50	2,7856E-05	7,02	0	3	0,30
Max	3102	66,7	0,02775217	42,01	0,67	44	1

Complete mitocondrial genome of the Iberian Mole Talpa occidentalis (Talpidae, Insectivora) and comparison with Talpa europaea

Authors: Juana Gutiérrez, Luz Lamelas, Gaël Aleix-Mata, María Arroyo, Juan Alberto Marchal, Teresa Palomeque, Pedro Lorite, Antonio Sánchez

Departamento de Biología Experimental, Área de Genética, Universidad de Jaén, Paraje de las Lagunillas s/n., 23071 Jaén, Spain.

Genetica (10.1007/s10709-018-0033-z)

Genetica (2018) 146:415–423
https://doi.org/10.1007/s10709-018-0033-z

ORIGINAL PAPER

Complete mitochondrial genome of the Iberian Mole *Talpa occidentalis* (Talpidae, Insectivora) and comparison with *Talpa europaea*

Juana Gutiérrez[1] · Luz Lamelas[1] · Gaël Aleix-Mata[1] · María Arroyo[1] · Juan Alberto Marchal[1] · Teresa Palomeque[1] · Pedro Lorite[1] · Antonio Sánchez[1]

Abstract

The complete mitogenome of *Talpa occidentalis*, the Iberian mole, was sequenced using a combination of the Illumina and Sanger methods. The 16,962 bp genome obtained contains 13 protein-coding genes, 22 transfer RNAs, 2 ribosomal RNAs, and a control region. Thirty-seven identical repetitions of a 10-nucleotide (CACACGTACG) repeat element were identified in the non-coding control region (D-loop). The number, order, and orientation of the mitochondrial genes are the same as in *T. europaea*, the only mitogenome published so far for this genus. These two mitogenomes differ only at the repeat element included in the control region. The phylogeny obtained for the Talpidae species using the protein-coding genes of these mitogenomes agrees with the current classification of this family.

Keywords: Iberian mole *(Talpa occidentalis);* complete mitogenome; mitochondrial, phylogeny

Introduction

The family Talpidae (Insectivora) includes moles, shrew moles, and desmans. Despite certain controversy, three Talpidae subfamilies are commonly distinguished, the Scalopinae, Talpinae, and Uropsilinae (Hutterer 1993, 2005; Shinohara et al. 2003). This family is thought to have originated in Europe during the Eocene and to have subsequently undergone explosive radiation in Eurasia and North America during the Miocene (Yates and Moore 1990). The tribe Talpini (Talpinae) encompasses the genera *Euroscaptor, Mogera*, *Parascaptor,* and *Scaptochirus,* as well as *Talpa.* The genus *Talpa* includes 10 extant species of fossorial insectivores (Wilson and Reeder 2005; Hutterer 2005; Nicolas et al. 2015, 2017).

The species of this genus are fossorial and well adapted to a subterranean existence, having minute eyes and no external ears. Two of their most remarkable characteristics include the exaggerated modification of the bones of their forelimbs (Campbell 1939; Bickelmann et al. 2017) and their gonadal development, with in females bilateral ovotestes instead of normal ovaries (Jiménez et al. 1993; Sánchez et al. 1996).

Due to maternal inheritance, a fast rate of evolution, a relative lack of recombination, conserved gene content, and the high number of copies in extracted DNAs, mitochondrial DNA (mtDNA) gene sequences are considered to be excellent markers system for phylogenetic analyses (Boore et al. 2005; Triant and DeWoody 2006; Wang et al. 2015), population genetics (Kivisild 2015; Lu et al. 2015), phylogeography (Jaarola and Searle 2002; Wang et al. 2013; Feuda et al. 2015), and species identification (Boonseub et al. 2009; Jun et al. 2011). The use of the complete mitochondrial genome (mitogenome) sequence enables these analyses to be performed with several markers; concatenated sequences provide the most reliable results (Ding et al. 2016).

Typically, complete mitogenomes have been sequenced using primer pairs PCR amplifications of overlapping fragments (Simon et al. 1994, 2006; Kim and Park 2015; Cabria et al. 2006). However, a new generation of sequencing methods now allows us to assemble mitogenomes from whole genome sequence data, even in cases of low coverage due to the high-copy portion of the mitogenome in DNA samples (Straub et al. 2012; Ye et al. 2014; Pita et al. 2017; Fernández-Pérez et al. 2017)

Of the genus *Talpa*, only the mitogenome of *T. europaea* has ever been completely sequenced and analyzed. Complete mitochondrial genome sequences are available in GenBank for nine other species of the family Talpidae. In this study, we sequenced and characterized the complete mitogenome of *Talpa occidentalis,* a species found in Spain and Portugal that is included on the IUCN Red List of threatened species (Cassola 2016).

Materials and methods

DNA extraction, sequencing and mitogenome assembly

A single adult of *T. occidentalis* from southern Spain (Granada province) was used. Tissue samples were stored in 100% ethanol at -20°C. The total DNA was extracted from the liver using the GentraPuregene Tissue Kit (Qiagen). The extracted Genomic DNA was sequenced with the Illumina technology in Macrogen Europe. Briefly, for genome sequencing, approximately 3 µg of genomic DNA were used for the construction of a library of 750-bp-length fragments. This library was used in Illumina® Hiseq™ 2000 paired-end sequencing with 100 bp reads. Two Gbp of sequences were obtained (coverage about 1X of the genome). Graph-based clustering analysis (Novák et al. 2010 2013), a method for similarity-based clustering of sequence reads, was performed using RepeatExplorer, implemented within the Galaxy environment (http://repeatexplorer.org/). Sequence clusters corresponding to mtDNA were selected, assembled, and revised manually using the *T. europaea* complete mitogenome as a template (Mouchaty et al. 2000; accession number: Y19192).

Polymerase chain reaction (PCR) amplification and sequencing

Only the control region (D-loop) of the mitochondrial genome was not completely recovered with the genome sequencing and clustering as consequence of a gap caused by the failure in the assembly of a repetitive DNA sequence. This gap was filled by amplification of the complete control region by PCR using the same *T. occidentalis* DNAsample as used for the genome sequencing. For PCR amplification we used the primer pair Pro+ (5'-ACCATCAGCACCCAAAGCTG-3') and Phe- (5'-AAGCATTTTCAGTGCTTTGCTT-3'), as described previously by Haring et al. 2000. PCR was performed in 50 µl of reaction mixture containing 100 ng of genomic DNA, 10 pmol of each primer, 1 µl 10 mM dNTPs, 1 µl DMSO, 5 µl of 10x NH_4 Reaction Buffer, 2.5 µl of 50 mM $MgCl_2$, and 5 U of BIOTAQ™ DNA Polymerase (Bioline). The PCR program used was 5 min at 95°C and 30 cycles: 30s at 95°C, 30s at 55°C, and 90s at 72°C, with a final extension of 5 min at 72°C. PCR amplicons were resolved in ethidium bromide-stained 1% agarose gels; the appropriate band of 1,592 bp was isolated from the gel with QIAquick gel extraction kit (Qiagen) and cloned in JM109 bacteria using PGEMT-easy vector (Promega). Two positive clones were sequenced in both directions. Sequences were analyzed using the Bioedit program (version 7.0.9.0) (http://www.mbio.ncsu.edu/BioEdit/bioedit.html).

Annotation and sequence analysis

The final circularized version was confirmed by manual inspection. The annotation of the *T. occidentalis* mitogenome was fulfilled using web-based services MITOS (http://mitos.bioinf.uni-leipzig.de/help.py) (Bernt et al. 2013) and tRNA scan-SE (http://lowelab.ucsc.edu/tRNAscan-SE/) (Lowe and Eddy 1997). The annotations of the protein coding genes (PCGs), transfer RNAs (tRNAs), and rRNA genes were refined by comparing manually with the *T. europaea* mitogenome. The base composition and codon usage were analyzed using MEGA version 7 (Kumar et al. 2016). The circularized drawing of the mitogenome was carried out using the OrganellarGenomeDRAW tools (https://chlorobox.mpimp-golm.mpg.de/OGDraw.html) (Lohse et al. 2013).

Phylogenetic analysis

For the phylogenetic analysis we employed complete mitochondrial genomes from other Talpidae species available in GenBank (*Condylura cristata* KU144678, *Galemys pyrenaicus* AY833419, *Mogera robusta* KT934322, *Mogera wogura* AB099482, *Scapanulus oweni* KM506754, *Talpa europaea* Y19192, *Uropsilus soricipes* JQ658979, *Uropsilus gracilis* KM379136, and *Urotrichus talpoides* AB099483). As an outgroup we chose two species of the family Soricidae (*Crocidura russula* AY769263 and *Sorex araneus* KT210896).

The sequences from the 13 PCGs were aligned using ClustalW and then concatenated for phylogenetic analysis. The phylogenetic relationships were reconstructed using the Maximum-Likelihood (Saitou and Nei 1987) and Neighbor-Joining methods, both implemented in the program MEGA version 7 (Kumar et al. 2016), and Bayesian inference (BI) implemented in MrBayes version 3.1 (Ronquist and Huelsenbeck 2003). Node supports in Maximum-Likelihood and Neighbor-Joining analysis were assessed with 2000 bootstrap replicates. In Bayesian inference, runs of two million generations were conducted. Trees were sampled every 1000 generations with a burn-in of 25%. The best-fit nucleotide substitution model with the lowest BIC (Bayesian Information Criterion) value was chosen (GTR+G+I) using MEGA version 7.

Results and discussion

Gene organization

The complete mitogenome (MF958963) of *T. occidentalis* contains 16,962 bp, which is slightly larger in size than that of *T. europaea* (16,884 bp) (Mouchaty et al. 2000) and that of the mitogenomes of the other analyzed species of the family Talpidae, which vary between 16,510 in *Galemys pyrenaicus* (Cabria et al. 2006) and 16,904 in *Urotrichus talpoides* (Nikaido et al. 2003). Its mitogenome consists of a control region and a conserved set of 37 vertebrate mitochondrial genes, including 13 protein coding genes (PCGs), 22 tRNA genes, and two rRNA genes (*12S rRNA* and *16S rRNA*) (Table 1).

The order and orientation of the *T. occidentalis* mitogenome is identical to that of *T. europaea* (Mouchaty et al. 2000) (Fig. 1). *Nd6* and eight *tRNAs* are found on the light strand, while the other 12 PCGs, 14 *tRNAs*, and two *rRNAs* are located on the heavy strand. The control region is located between *tRNA-Pro* and *tRNA-Phe*, as in the *T. europaea* mitogenome (Mouchaty et al. 2000). The identity percentage between the two complete mitogenomes is 90.85% but is higher still (92.29%) if we exclude the control region (Table 2).

***Table 1**. Gene organization of the Talpa occidentalis mitogenome*

Gene	Start position	Stop position	Length (bp)	Anticodon	Strand
tRNA-Phe	1	70	70	GAA	H
12S rRNA	73	1,041	969		H
tRNA-Val	1,042	1,109	68	TAC	H
16S rRNA	1,110	2,682	1,573		H
tRNA-Leu(UUR)	2,683	2,757	74	TAA	H
Nd1	2,760	3,713	954		H
tRNA-Ile	3,716	3,784	69	GAT	H
tRNA-Gln	3,782	3,854	73	TTG	L
tRNA-Met	3,856	3,924	69	CAT	H
Nd2	3,925	4,968	1,044		H
tRNA-Trp	4,967	5,034	68	TCA	H
tRNA-Ala	5,041	5,109	69	TGC	L
tRNA-Asn	5,111	5,183	73	GTT	L
OR	5,184	5,219	36		H
tRNA-Cys	5,217	5,283	67	GCA	L
tRNA-Tyr	5,284	5,350	67	GTA	L
CoI	5,352	6,896	1,545		H
tRNA-Ser(UCN)	6,898	6,966	69	TGA	L
tRNA-Asp	6,974	7,042	69	GTC	H
CoII	7,043	7,726	684		H
tRNA-Lys	7,730	7,797	68	TTT	H
Atp8	7,799	8,002	204		H
Atp6	7,960	8,640	681		H
CoIII	8,640	9,423	784		H
tRNA-Gly	9,424	9,492	69	TCC	H
Nd3	9,493	9,838	346		H
tRNA-Arg	9,839	9,906	68	TCG	H
Nd4L	9,907	10,203	297		H
Nd4	10,197	11,574	1,378		H
tRNA-His	11,575	11,642	68	GTG	H
tRNA-Ser(AGY)	11,643	11,703	61	GCT	H
tRNA-Leu(CUN)	11,706	11,775	70	TAG	H
Nd5	11,776	13,596	1,821		H
Nd6	13,580	14,107	528		L
tRNA-Glu	14,108	14,176	69	TTC	L
Cyt B	14,181	15,320	1,140		H
tRNA-Thr	15,321	15,389	69	TGT	H
tRNA-Pro	15,389	15,458	69	TGG	L
Control region	15,459	16,962	1,504		H

***Table 2.** Sequence similarity between the mitogenomes of T. occidentalis and T. europaea*

Gene region	Gene length		AT content		Sequence similarity (%)
	T. occ	*T. eur*	*T. occ*	*T. eur*	
Protein coding genes (PCGs)	11,406	11,406	60.87	61.08	90.86
12S rRNA and 16S rRNA	2,542	2,545	61.29	61.34	95.68
22 tRNAs	1,518	1,517	63.37	63.28	97.96
Control region	1,504	1,422	56.45	58.51	76.00
Total length excluding the control region	15,458	15,462	61.13	61.32	92.29
Total	16,962	16,884	60.71	61.08	90.85

Table 3. *Sequence similarity between the PCGs of T. occidentalis and T. europaea mitogenomes*

Gene	Length (bp)		AT content		% Identity	Start/Stop codons		Protein length	
	T.occ	*T.eur*	*T.occ*	*T.eur*		*T.occ*	*T.eur*	*T.occ*	*T.eur*
Nd1	954	954	60.38	59.75	91.51	ATG/TAA	ATG/TAA	317	317
Nd2	1,044	1,044	64.37	65.23	89.08	ATA/TAG	ATA/TAG	347	347
CoI	1,545	1,545	58.90	58.77	91.52	ATG/TAA	ATG/TAA	514	514
CoII	684	684	61.84	63.01	93.27	ATG/TAA	ATG/TAA	227	227
Atp8	204	204	68.14	67.65	91.18	ATG/TAA	ATG/TAA	67	67
Atp6	681	681	58.44	60.79	91.19	ATG/TAA	ATG/TAA	226	226
CoIII	784	784	58.29	58.29	91.71	ATG/T--	ATG/T--	261	261
Nd3	346	346	59.54	59.54	88.15	ATT/T--	ATT/T--	115	115
Nd4L	297	297	61.95	63.64	87.88	ATG/TAA	ATG/TAA	98	98
Nd4	1,378	1,378	61.83	61.39	89.84	ATG/T--	ATG/T--	459	459
Nd5	1,821	1,821	61.23	61.50	90.72	ATT/TAA	ATT/TAA	606	606
Nd6	528	528	63.07	62.69	91.48	ATG/TAA	ATG/TAA	175	175
Cytb	1,140	1,140	59.47	59.30	91.49	ATG/AGA	ATG/AGA	379	379
Total	11,406	11,406	60.87	61.08	90.86			3,791	3,791

Nucleotide composition

As is typical in mitogenomes, there is a bias towards A and T nucleotides. The *T. occidentalis* overall nucleotide composition of the H-strand was 34.08% A, 26.62% T, 14.36% G, and 24.93% C, with an A+T content of 60.71%, which is very similar to the data for *T. europaea* (34.10% A, 26.98% T, 14.30% G, and 24.62% C, with an A+T content of 61.08%). The 13 mitochondrial PCGs are AT-biased, with a total AT content of 60.87%, ranging from 58.29% in *CoIII*to 68.14% in *Atp8* (Table 3). The control region, the two rRNA genes, and the 22 tRNAs are also AT-biased in both species (Table 2).

Protein-coding genes and codon usage

The 13 mitochondrial PCGs consisted of 11,373 bp, excluding the stop codons (33 bp), which encode 3791 amino acids (Table 3). The mitogenome of *T. occidentalis* contains some overlapping nucleotides between PCGs or between PCGs and tRNAs (Table 1). We detected a 43-bp overlap between *Atp8* and *Atp6*, a 7-bp overlap between *Nd4L* and *Nd4*, a 17-bp overlap between *Nd5* and *Nd6*, a 1-bp overlap between *Atp6* and *CoIII*, a 3-bp overlap between *tRNA-*

Ile and *tRNA-Gln*, a 2-bp overlap between *Nd2* and *tRNA-Trp*, a 3-bp overlap between OR and *tRNA-Cys*, and a 1-bp overlap between *tRNA-Thr* and *tRNA-Pro*.

The 13 mitochondrial PCGs of *T. occidentalis* for translation initiation use three start codons (ATA for *Nd2*, ATT for *Nd3* and *Nd5,* and ATG for the other 10 PCGs). On the other hand, PCGs genes used four stop codons for translation termination, one incomplete (T-- for *CoIII*, *Nd3,* and *Nd4*), and three complete (TAG for *Nd2*, AGA for *CytB*, and TAA for the other eight PCGs) (Table 3). The mitogenome sequences from *T. occidentalis* and *T. europaea* (Mouchaty et al. 2000) show high sequence similarity (Table 3). Nevertheless, these authors describe ATT as the start codon for the Nd6 gene and ATG for Nd3 and Nd5 genes in *T. europaea*. By contrast, we found that ATG is the start codon for Nd6 and ATT for Nd3 and Nd5. Our results agree with the Genbank annotation for these genes in the *T. europaea* mitogenome. Mouchaty et al. (2000) also propose that Nd2 has a TA- incomplete stop codon in *T. europaea*, while we consider that it is more likely that the stop codon is TAG if there is a bp overlap of this gene with the following *tRNA-Gln.*

The most abundant start and stop codons were ATG and TAA, respectively. The frequent pattern of ATG and TAA usage as start and stop codons has also been observed in *T. europaea* and other mammalian mitogenomes (Mouchaty et al. 2000; Kim and Park 2015; Cabria et al. 2006; Xu et al. 2016; Kim et al. 2017). Incomplete stop codons (T-- or TA), like those used in three PCGs (*CoIII*, *Nd3,* and *Nd4*) in the *T. occidentalis* mitogenome, are not unusual in metazoans, and may be completed by poly-adenylation of the 3'-end of the mRNA after transcription, giving rise to the complete stop codon TAA (Ojala 1981, Mouchaty et al. 2000).

The A+T bias in the composition of the *T. occidentalis* mtDNA is reflected by the codon usage of the PCGs, with a strong bias toward AT-rich codons (Table 4). The most frequently used codons were CTA (Leu, 7.81%), ATA (Met, 4.87%), ACA (Thr, 4.55%), ATT (Ile, 4.34%), and ATC (Ile, 4.00%). Accordingly, leucine (16.14%), isoleucine (8.36%), threonine (8.10%), serine (7.73%), alanine (6.83%), and methionine (6.41%) are the most common amino acids in the mitochondria protein sequences. The high proportion in A+T in PCGs appears to be a shared feature in mammalian species (Kim et al. 2017). For an amino acid, the relative synonymous codon usage value (RSCU) is the number of times that a codon appears in a gene in relation to the number of expected occurrences under an assumption of equal codon usage. The six codons with highest RSCU values (CTA(L): 2.91; CGA(R): 2.88; CCA(P): 2.79; TCA(S): 2.74; GTA(V): 2.45; ACA(T): 2.25) are AT-rich as all of them have two positions with T and/or A bases, with the exception of codons CGA and CCA (Table 4).

Table 4. Codon usage of T. occidentalis mitochondrial genome protein coding genes

Codon	n	%	RSCU	Codon	n	%	RSCU
UUU(F)	113	2.97	0.97	UAU(Y)	61	1.60	0.94
UUC(F)	121	3.18	1.03	UAC(Y)	69	1.82	1.06
UUA(L)	127	3.34	1.25	UAA(*)	8	0.21	3.2
UUG(L)	23	0.61	0.23	UAG(*)	1	0.03	0.4
CUU(L)	47	1.24	0.46	CAU(H)	27	0.71	0.55
CUC(L)	57	1.50	0.56	CAC(H)	72	1.89	1.45
CUA(L)	297	7.81	2.91	CAA(Q)	73	1.92	1.64
CUG(L)	61	1.60	0.6	CAG(Q)	16	0.42	0.36
AUU(I)	165	4.34	1.04	AAU(N)	51	1.34	0.66
AUC(I)	152	4.00	0.96	AAC(N)	103	2.71	1.34
AUA(M)	185	4.87	1.52	AAA(K)	88	2.32	1.81
AUG(M)	58	1.53	0.48	AAG(K)	9	0.24	0.19
GUU(V)	34	0.89	0.7	GAU(D)	27	0.71	0.81
GUC(V)	21	0.55	0.43	GAC(D)	40	1.05	1.19
GUA(V)	119	3.13	2.45	GAA(E)	83	2.18	1.68
GUG(V)	20	0.53	0.41	GAG(E)	16	0.42	0.32
UCU(S)	41	1.08	0.84	UGU(C)	13	0.34	1.08
UCC(S)	45	1.18	0.92	UGC(C)	11	0.29	0.92
UCA(S)	134	3.53	2.74	UGA(W)	95	2.50	1.84
UCG(S)	12	0.32	0.25	UGG(W)	8	0.21	0.16
CCU(P)	27	0.71	0.57	CGU(R)	8	0.21	0.5
CCC(P)	20	0.53	0.42	CGC(R)	4	0.11	0.25
CCA(P)	132	3.47	2.79	CGA(R)	46	1.21	2.88
CCG(P)	10	0.26	0.21	CGG(R)	6	0.16	0.38
ACU(T)	55	1.45	0.72	AGU(S)	24	0.63	0.49
ACC(T)	58	1.53	0.76	AGC(S)	37	0.97	0.76
ACA(T)	173	4.55	2.25	AGA(*)	1	0.03	0.4
ACG(T)	21	0.55	0.27	AGG(*)	0	0.00	0
GCU(A)	39	1.03	0.6	GGU(G)	33	0.87	0.61
GCC(A)	95	2.50	1.47	GGC(G)	46	1.21	0.85
GCA(A)	108	2.84	1.67	GGA(G)	100	2.63	1.84
GCG(A)	17	0.45	0.26	GGG(G)	38	1.00	0.7

rRNA and tRNA genes

The two rRNA genes (*12S rRNA* and *16S rRNA*) are located between *tRNA-Phe*and *tRNA-Leu(UUR)*, and are separated by *tRNA-Val* (Table 1, Fig. 1). The combined size of the two genes is 2,542 bp (Table 2). The combined size of the 22 tRNA genes is 1518 bp (Table 2), ranging in size from 61 bp (*tRNA-Ser(AGY)*) to 74 bp (*tRNA-Leu(UUR)*) (Table 1). The tRNA genes include two leucine-tRNA genes (*tRNA-Leu(UUR)* and *tRNA-Leu(CUN)*), and two serine-tRNA genes (*tRNA-Ser(UCN)* and *tRNA-Ser(AGY)*) (Table 1).

Non-coding regions

Some non-coding regions such as the origin of replication (*OR*), intergenic spacers, and the control region are important during replication and maintenance of the mitogenomes

(Fernández-Silva et al. 2003). The non-coding regions were also found in the mitogenome of *T. occidentalis* (Table 1). The mitochondrial *OR* of *T. occidentalis* consists of36 bp and is located between *tRNA-Asn*and *tRNA-Cys*in the WANCY region, which consists of a cluster of five tRNA genes (*tRNA-Trp*, *tRNA-Ala*, *tRNA-Asn*, *tRNA-Cys*, and *tRNA-Tyr*), as in *T. europaea* and other mammals (Mouchaty et al. 2000, Kim and Park 2015). Intergenic spacers were found in 12 regions of the mitogenome, with sizes ranging between 1 to 7 bp (Table 1).

The control region of the *T. occidentalis* mitogenome is 1,504 bp in length and 1,422 bp in *T. europaea*, and in both species is located between *tRNA-Pro* and *tRNA-Phe* (Tables 1, 2) (Fig. 1). The control regions are the most clearly differentiated regions of the two mitogenomes, with an identity percentage of just 76.0%. The main difference is due to the presence of repeat units: thirty-seven identical repetitions of a 10-nucleotide (CACACGTACG) repeat in *T. occidentalis* and nineteen identical repetitions of a 16-nucleotide (ACAGGCGTATACACCC) repeat in *T. europaea*.

The conserved sequences block CSB-1, CSB-2, and CSB-3 (Walberg and Clayton 1981) and were identified in the 3′ ends of the control region. CSB-1 has a 100% and an 88% of identity with the same block identified in *T. europaea* and *Crocidura russula*, respectively. CSB-2 and CSB-3 were identical to those of *T. europaea* and *C. russula* (Sbisa et al. 1997). As well, the extended termination-associated sequence ETAS-1 was identified in the 5′ ends of the control region, which has an identity percentage of 87% and 73% with the same sequences of *T. europaea* and *C. russula*, respectively (Sbisa et al. 1997). The ETAS-2 was not identified in the *T. occidentalis* mitochondrial control region (Sbisa et al. 1997).

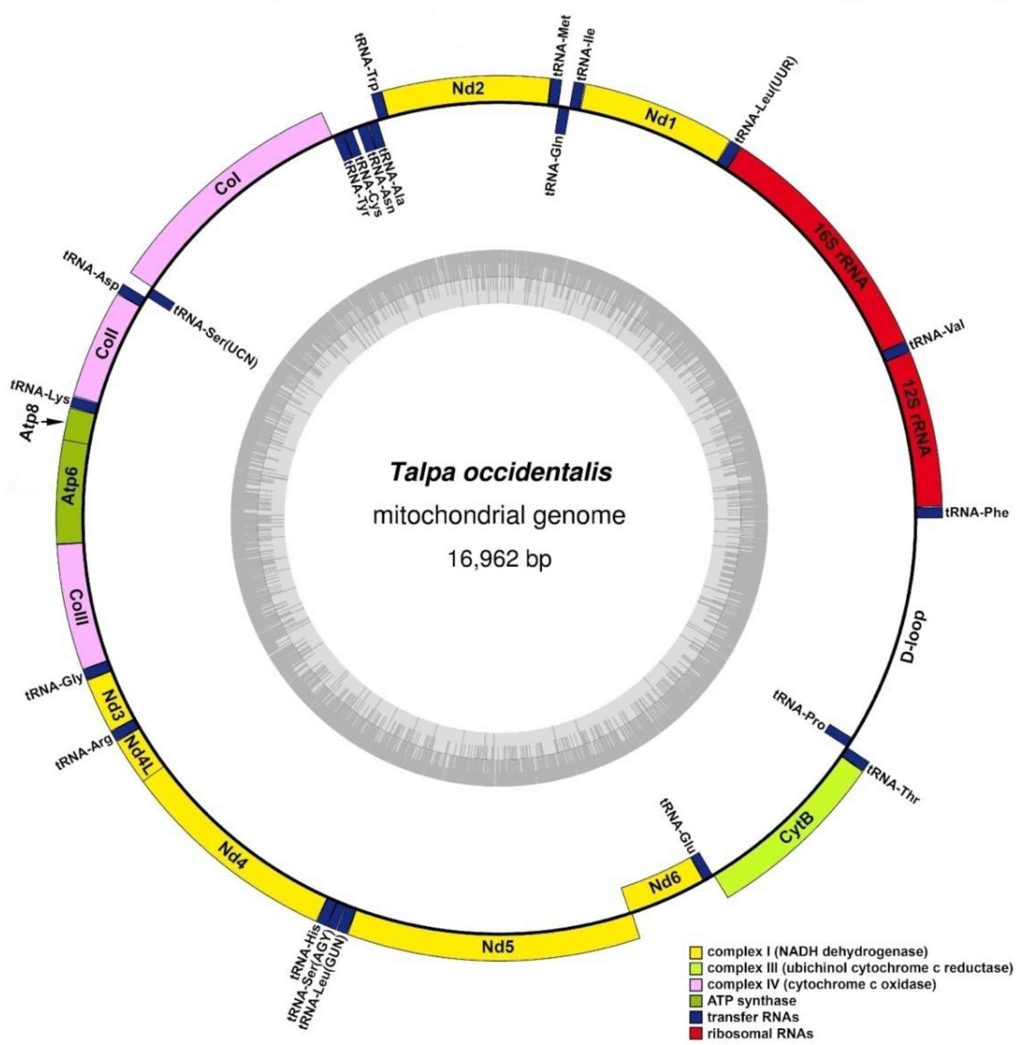


Figure 1*. Map of the mitochondrial genome of T. occidentalis. Genes encoded by the heavy strand are shown outside the circle and those encoded by the light strand are shown inside the circle. The inner ring shows the GC content of this genome.*

Phylogenetic analysis

The phylogenetic position of *T. occidentalis* within the family Talpidae was assessed using the Maximum-Likelihood, Neighbor-Joining, and Bayesian inference techniques based on 13 PCGs from the available mitochondrial genome of 10 taxa representing the main lineages within this family. Two members of the family Soricidae were used as an outgroup. Our analysis did not take into account biases in phylogenic relationships due to mitochondrial introgression. Nevertheless, although ancient genetic introgressions between *T. europaea* and *T. romana* in a parapatric contact zone have been hypothesized (Loy et al. 2001), recent introgression events between *T. occidentalis* and *T. europaea* mitogenomes were excluded due to the high mitochondrial cytochrome b divergence (Feuda et al. 2015).

The Maximum-Likelihood and Neighbor-Joining analyses yielded identical phylogenetic trees, with only small differences in the support for two nodes (2a). The Bayesian inference tree is very similar, the main difference being that it groups *U. talpoides*, *S. oweni*, and *C. cristata* in one branch, while with the other methods *U. talpoides*is grouped with *Talpa* and *Mogera*, and *S. oweni* and *C. cristata* are grouped in an independent branch. Our results agree with other studies that have also failed to clarify the phylogenetic relationships between these three genera (Shinohara et al. 2003; Zhu et al. 2017).

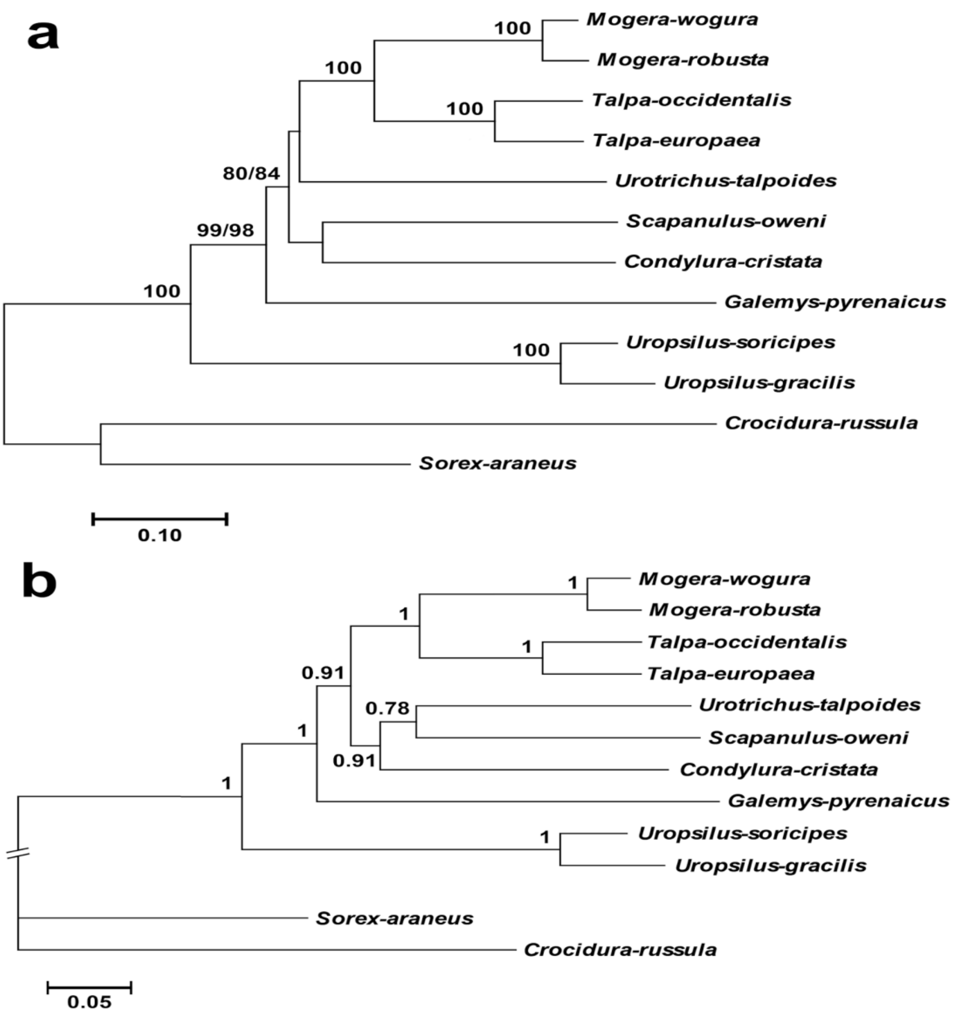


Figure 2. *Maximum likelihood and Neighbor Joining (a) and Bayesian Inference (b) trees using the concatenated sequences from the 13 protein coding genes (PCGs). The analysis includes the 10 sequenced Talpidae species, with the two Soricidae species, Crocidura russula and Sorex araneus, as the outgroup. Bootstrap supports greater than 75 and the Bayesian posterior probability are indicated at each node. In nodes with two different Bootstrap supports, the Maximum likelihood values are on the left and Neighbor-Joining values are on the right.*

Nevertheless, two well-supported clades are present in all trees, one of which includes the species of the sister genera *Talpa* and *Mogera* spliced into two branches, and the other the species from the genus *Uropsilus* (subfamily Uropsilinae) in a basal position. A recent phylogenetic study using complete mitogenomes has shown that species from the subfamilies Uropsilinae and Talpinae can be grouped into two well-supported clades (Tu et al. 2012, 2014). Additionally, the Pyrenean desman (*Galemys pyrenaicus*) in all trees was placed basally to the Talpinae species, as has been previously stated (Shinohara et al. 2003; Cabria et al. 2006).

As indicated above, the subfamily classification of Talpidae is regarded as controversial. Although for some authors the family Talpidae includes only three subfamilies (Scalopinae, Talpinae, and Uropsilinae), others defend the existence of a fourth subfamily, Desmaninae (McKenna and Bell 1997). This subfamily includes species of the genera *Desmana* and *Galemys*. Our results agree with the maintenance of the subfamily Desmaninae. Further studies are still necessary, nevertheless, to clarify the phylogenetic relationships between species of the family Talpidae.

Acknowledgments

This work was funded by the *Consejería de Innovación, Ciencia y Empresa* of the Junta de Andalucía, and by the Universidad de Jaén (as part of the program *Plan de Apoyo a la Investigación 2017-2019, Acción 1*). The authors would like to express their gratitude to Francisco J. Ruiz-Ruano Campaña and Juan Pedro M. Camacho for their assistance in the RepeatExplorer program, and to the Consejería de Medio Ambiente y Ordenación del Territorio de la Junta de Andalucía (Delegación Territorial de Granada) for the capture permits for *T. occidentalis.*

References

Bernt M, Donath A, Jühling F, Externbrink F, Florentz C, Fritzsch G, Pütz J, Middendorf M, Stadler PF, (2013) MITOS: improved de novo metazoan mitochondrial genome annotation. Mol Phylogenet Evol 69:313-319

Bickelmann C, van der Vos W, de Bakker MA, Jiménez R, Maas S, Sánchez-Villagra MR, (2017) Hox gene expression in the specialized limbs of the Iberian mole (*Talpa occidentalis*). Evol Dev 19:3-8

Boonseub S, Tobe SS, Linacre AMT, (2009) The use of mitochondrial DNA genes to identify closely related avian species. Forensic Sci Int: Genet Suppl Series vol 2, no. 1, pp. 275-277

Boore JL, Macey JR, Medina M, (2005) Sequencing and comparing whole mitochondrial genomes of animals. Methods Enzymol 395:311-348

Cabria MT, Rubines J, Gómez-Moliner B, Zardoya R, (2006) On the phylogenetic position of a rare Iberian endemic mammal, the Pyrenean desman (*Galemys pyrenaicus*). Gene 375:1-13

Campbell B, (1939) The shoulder anatomy of the moles.A study in phylogeny and adaptation. Am J Anat 64:1-39

Cassola F, (2016) *Talpa occidentalis*. The IUCN Red List of Threatened Species 2016: e.T41483A2953593.http://dx.doi.org/10.2305/IUCN.UK.2016-.RLTS.T41483A2953593.en

Ding L, Li W, Liao J, (2016) Mitochondrial genome of *Cricetulus migratorius* (Rodentia: Cricetidae): Insights into the characteristics of the mitochondrial genome and the phylogenetic relationships of *Cricetulus* species. Gene 595:121-129

Fernández-Pérez J, Nantón A, Ruiz-Ruano FJ, Camacho JPM, Méndez J, (2017) First complete female mitochondrial genome in four bivalve species genus *Donax* and their phylogenetic relationships within the Veneroida order. PLoS One 8:12(9)

Fernández-Silva P, Enriquez JA, Montoya J, (2003) Replication and transcription of mammalian mitochondrial DNA. Exp Physiol 88:41-56

Feuda R, Bannikova AA, Zemlemerova ED, Di Febbraro M, Loy A, Hutterer R, Aloise G, Zykov AE, Annesi F, Colangelo P, (2015) Tracing the evolutionary history of the mole, *Talpa europaea*, through mitochondrial DNA phylogeography and species distribution modelling. Biol J Linnean Soc 114:495-512

Haring E, Herzig-Straschil B, Spitzenberger F, (2000) Phylogenetic analysis of Alpine voles of the *Microtus* multiplex complex using the mitochondrial control region. J Zool Evol Research 38:231-238

Hutterer R, (1993) Order Insectivora. In Mammal species of the world: 69-130. Wilson, D. E, Reeder, D. M. (Eds). Washington: Smithsonian Institution Press.

Hutterer R, (2005) Order Soricomorpha. In: Wilson, D.E, Reeder, D.M. (Eds.), Mammals Species of the World. A Taxonomic and Geographic Reference, third ed. The Johns Hopkins University Press, Baltimore, pp 220-311

Jaarola M, Searle JB, (2002) Phylogeography of field voles (*Microtus agrestis*) in Eurasia inferred from mitochondrial DNA sequences. Mol Ecol 11:2613-2621

Jiménez R, Burgos M, Sánchez A, Sinclair AH, Alarcón FJ, Marín JJ, Ortega E, Díaz de la Guardia R, (1993) Fertile females of the mole *Talpa occidentalis* are phenotypic intersexes with ovotestes. Development 118:1303-1311

Jun J, Han SH, Jeong TJ, Park HC, Lee B, Kwak M, (2011) Wildlife forensics using mitochondrial DNA sequences: Species identification based on hairs collected in the field and confiscated tanned Felidae leathers. Gene Genom 33:721-726

Kim NH, Lim SJ, Chae HM, Park YC, (2017) Complete mitochondrial genome of the Amur hedgehog *Erinaceus amurensis* (Erinaceidae) and higher phylogeny of the family Erinaceidae. Genet Mol Res 16: 10.4238/gmr16019300

Kim JY, Park YC, (2015) Gene organization and characterization of the complete mitogenome of *Hypsugo alaschanicus* (Chiroptera: Vespertilionidae). Genet Mol Res 14:16325-16331

Kivisild T, (2015) Maternal ancestry and population history from whole mitochondrial genomes. Investig Genet 6:3

Kumar S, Stecher G, Tamura K, (2016) MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. Mol Biol Evol 33:1870-1874

Lohse M, Drechsel O, Kahlau S, Bock R, (2013) OrganellarGenomeDRAW-a suite of tools for generating physical maps of plastid and mitochondrial genomes and visualizing expression data sets. Nucleic Acids Res 41 (Web Server issue):W575-81

Lowe TM, Eddy SR, (1997) tRNAscan-SE: A program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res 25:955-964

Loy A, Capula M, Palombi A, Capanna E, (2001) Genetic and morphometric evidence of introgression between two species of moles (Insectivora: *Talpa europaea* and *Talpa romana*) in central Italy. J Zool 254:229-238

Lu L, Ge D, Chesters D, Ho SYW, Ma Y, Li G, Wen Z, Wu Y, Wang J, Xia L, Liu J, Guo T, Zhang X, Zhu C, Yang Q, Liu Q, (2015) Molecular phylogeny and the underestimated species diversity of the endemic white-bellied rat (Rodentia: Muridae: Niviventer) in Southeast Asia and China. ZoolScr 44:475-494

McKenna MC, Bell SK, (1997) Classification of mammals above the species level. Columbia University Press, New York, 631 pp.

Mouchaty SK, Gullberg A, Janke A, Arnason U, (2000) The phylogenetic position of the Talpidae within Eutheria based on analysis of complete mitochondrial sequences. Mol Biol Evol 17:60-67

Nicolas V, Martínez-Vargas J, Hugot JP, (2015) *Talpa aquitania* nov. sp. (Talpidae, Soricomorpha) a new mole species from southwest France and north Spain. Bull Acad Vét Fr 168: 329-334

Nicolas V, Martínez-Vargas J, Hugot JP, (2017) *Talpa aquitania* sp. nov. (Talpidae, Soricomorpha), a new mole species from SW France and Spain. Mammalia 81:641–642

Nikaido M, Cao Y, Harada M, Okada N, Hasegawa M, (2003) Mitochondrial phylogeny of hedgehogs and monophyly of Eulipotyphla. Mol Phylogenet Evol 28:276-284

Novák P, Neumann P, Macas J, (2010) Graph-based clustering and characterization of repetitive sequences in next-generation sequencing data. BMC Bioinformatics 11:378

Novák P, Neumann P, Pech J, Steinhaisl J, Macas J, (2013) RepeatExplorer: A Galaxy-based web server for genome-wide characterization of eukaryotic repetitive elements from next-generation sequence reads. Bioinformatics 29:792-793

Ojala D, Montoya J, Attardi G, (1981) tRNA punctuation model of RNA processing in human mitochondria. Nature 290:470-474

Pita S, Panzera F, Vela J, Mora P, Palomeque T, Lorite P, (2017) Complete mitochondrial genome of *Triatoma infestans* (Hemiptera, Reduviidae, Triatominae), main vector of Chagas disease. Infect Genet Evol 54:158-163

Ronquist F, Huelsenbeck JP, (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19:1572-1574

Saitou N, Nei M, (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4:406-425

Sánchez A, Bullejos M, Burgos M, Hera C, Stamatopoulos C, Diaz De la Guardia R, Jiménez R, (1996) Females of four mole species of genus *Talpa* (Insectivora, Mammalia) are true hermaphrodites with ovotestes. Mol Reprod Dev 44:289-294

Sbisà E, Tanzariello F, Reyes A, Pesole G, Saccone C, (1997) Mammalian mitochondrial D-loop region structural analysis: identification of new conserved sequences and their functional and evolutionary implications. Gene 205:125-140

Shinohara A, Campbell KL, Suzuki H, (2003) Molecular phylogenetic relationships of moles, shrew moles, and desmans from the new and old worlds. Mol Phylogenet Evol 27: 247-258

Simon C, Buckley TR, Frati F, Stewart JB, Beckenbach AT, (2006) Incorporating molecular evolution into phylogenetic analysis, and a new compilation of conserved polymerase chain reaction primers for animal mitochondrial DNA. Annu Rev Ecol Evol Syst 37: 545-579

Simon C, Frati F, Beckenbach A, Crespi B, Liu H, Flook P. (1994) Evolution, weighting, and phylogenetic utility of mitochondrial gene sequences and a compilation of conserved polymerase chain reaction primers. Ann Entomol Soc Am 87: 651-701

Straub SC, Parks M, Weitemier K, Fishbein M, Cronn RC, Liston A, (2012) Navigating the tip of the genomic iceberg: Next-generation sequencing for plant systematics. Am J Bot 99:349-364

Triant DA, DeWoody JA, (2006) Accelerated molecular evolution in *Microtus* (Rodentia) as assessed via complete mitochondrial genome sequences. Genetica 128:95-108

Tu F, Fan Z, Chen S, Yin Y, Li P, Zhang X, Liu S, Yue B, (2012) The complete mitochondrial genome sequence of the Gracile shrew mole, *Uropsilus gracilis* (Soricomorpha: Talpidae). Mitochondrial DNA 23:382-384

Tu F, Fan Z, Murphy R, Chen S, Zhang X, Yan C, Liu Y, Sun Z, Fu J, Liu S, Yue B, (2014) Molecular phylogenetic relationships among Asiatic shrewlike moles inferred from the complete mitogenomes. J Zool Syst Evol Res 53. 10.1111/jzs.12081

Walberg MW, Clayton DA, (1981) Sequence and properties of the human KB cell and mouse L cell D-loop regions of mitochondrial DNA. Nucleic Acids Res 9:5411-5421

Wang Y, Chen J, Jiang LY, Qiao GX, (2015) Hemipteran mitochondrial genomes: features, structures and implications for phylogeny. Int J Mol Sci 16:12382-12404

Wang Y, Zhao LM, Fang FJ, Liao JC, Liu NF, (2013) Intraspecific molecular phylogeny and phylogeography of the *Meriones meridianus* (Rodentia: Cricetidae) complex in northern China reflect the processes of desertification and the Tianshan Mountains uplift. Biol J Linn Soc 110:362-383

Wilson DE, Reeder DM (2005) Mammals species of the world. A taxonomic and geographic reference, 3rd edn. The Johns Hopkins University Press, Baltimore, MD

Xu Y, Huang X, Hu Y, Tu F, (2016) Description of the mitogenome of Gansu mole (*Scapanulus oweni*). Mitochondrial DNA A DNA Mapp Seq Anal 27:2083-2084

Yang L, Tan Z, Wang D, Xue L, Guan MX, Huang T, Li R, (2014) Species identification through mitochondrial rRNA genetic analysis. Sci Rep 4:4089

Yates TL, Moore DW, (1990) Speciation and evolution in the family Talpidae (Mammalia: Insectivora). In: Nevo, E., Reig, O.A. (Eds.), Evolution of Subterranean Mammals at the Organismal and Molecular Levels. Alan R. Liss, New York, pp 1–22

Ye F, Samuels DC, Clark T, Guo Y, (2014) High-throughput sequencing in mitochondrial DNA research. Mitochondrion 17:157-163

Zhu L, Ao-Nan W, Jun-Sheng Z, Xi Y, Huan L, Zhi-Min J, (2017) Sequencing and analysis of the complete mitochondrial genome of flat-skulled shrew (*Sorex roboratus*) from China. Mitochondrial DNA Part B: Resources 2:369-371

The complete mitochondrial genome of Talpa aquitania (Talpidae; Insectivora), a mole species endemic to northern Spain and southern France.

Authors: Gaël Aleix-Mata, Juana Gutiérrez, Francisco J. Ruiz-Ruano, Pedro Lorite, Juan Alberto Marchal, Antonio Sánchez

Departamento de Biología Experimental, Área de Genética, Universidad de Jaén, Campus de las Lagunillas s/n., 23071 Jaén, Spain.

Departamento de Biología Animal, Biología Vegetal y Ecología; Universidad de Jaén; Campus Las Lagunillas, s/n., 23071, Jaén, Spain.

Department of Ecology and Genetics, Evolutionary Biology Centre, Uppsala University, SE-752 36 Uppsala, Sweden.

Molecular Biology Reports (10.1007/s11033-020-05296-8)

Molecular Biology Reports (2020) 47:2397–2403
https://doi.org/10.1007/s11033-020-05296-8

SHORT COMMUNICATION

Check for updates

The complete mitochondrial genome of *Talpa aquitania* (Talpidae; Insectivora), a mole species endemic to northern Spain and southern France

Gaël Aleix-Mata[1,2] · Juana Gutiérrez[1] · Francisco J. Ruiz-Ruano[3] · Pedro Lorite[1] · Juan A. Marchal[1] · Antonio Sánchez[1]

Abstract

The complete mitogenome sequence of *Talpa aquitania*, a recently described *Talpa* species, was assembled using whole-genome sequencing data. It varies in length from 16,776 to 16,846 bp, contains 13 protein-coding genes, two ribosomal RNA genes, 22 transfer RNA genes, one origin of L-strand replication, and a control region. In the control region, which varied from 1,320 to 1,390 bp, we identified the extended termination-associated sequence (ETAS-1 and ETAS-2) and the conserved sequence blocks (CSB-1, 2, 3, B, C, D, E, F). In addition, this region includes a 10 bp tandem repeat DNA sequence, with a variable number of repeats that suggest the existence of heteroplasmy. Phylogeny reconstructions based on Maximum Likelihood, Neighbor-joining and Bayesian inference analyses yielded phylogenies with similar topologies demonstrating that *T. aquitania* and *T. occidentalis* are sister species.

Keywords: Control region; mitogenome; phylogenetic; *Talpa aquitania*

Introduction

Recently, a combination of morphological, molecular and cytogenetic methods has led to the description of a new*Talpa* species*, Talpa aquitania* [1,2,3,4]. This species is found west and south of the river Loire and in a broad area of northern Spain [1]. Its exact distribution and population status have not yet been well characterized and so it is not currently included in IUCN *Red Lists; the related species T. europaea* and *T. occidentalis*, however, are both *classified as of Least Concern*.

This new separation means that there are now 11 recognized species in the genus *Talpa*, which can be divided into seven western (*T. europaea* (Linnaeus, 1758), *T. caeca* (Savi, 1882), *T. occidentalis* (Cabrera, 1907), *T. romana* (Thomas, 1902), *T. stankovici* (Martino and Martino, 1931), *T. levantis* (Thomas, 1906) and *T. aquitania* (Nicolas, Martínez-Vargas et Hugot, 2015)) and four eastern (*T. caucasica* (Satunin, 1908), *T. davidiana* (Milne-Edwards, 1884), *T. altaica* (Nikolsky, 1883) and *T. martinorum* (Kryštufek, Nedyalkov, Astrin et Hutterer, 2018)) species [1,2,3,4,5,6,7]. Species identification and taxonomic studies of *Talpa* species have traditionally been based on morphometric analysis. Molecular phylogenies using either just the mitochondrial *cytb* gene [8] or combined with nuclear genes [9] have been applied but do not cover all known species as they were performed before the recent descriptions of new species. A common finding derived from the previous molecular phylogenetic analyses was the inclusion of all six western *Talpa* species in a single well-supported clade separated from the eastern species; however, these eastern species have not been grouped into another independent clade. It is worth noting that intra- and interspecific variations are not clearly defined in the genus *Talpa* and several cryptic species have been described. Therefore, species delimitation in this genus is a relevant but complex task that requires further investigations [9].

In the recently described *T. aquitania*, the cytochrome b analysis suggests that this species is more closely related to *T. occidentalis* than to *T. europaea* [1,2]). To confirm this assumption and establish the precise phylogenetic relationship of *T. aquitania* to its congeneric species, further data analyses including the establishing of new molecular markers are still required.

To date, the complete mitogenomes of two species of the genus *Talpa* (*T. europaea* and *T. occidentalis*) and that of nine other species of the family Talpidae have been completely sequenced. We present here the complete mitochondrial genome of a male adult *T. aquitania* (Torme, Spain) identified by mesostyle, cytochrome b gene and cytogenetic analyses [4].

Materials and Methods

A male *T. aquitania* mole (TA-3) captured at Torme (42°59'36.7"N 3°33'47.2"W; Burgos, Spain) was analyzed. Permission for capture was granted by the Servicio Territorial de Medio Ambiente de Burgos (Junta de Castilla-León). All capture and sacrifice protocols were approved by the Ethics Committee for Animal Experimentation of the Junta de Andalucía (code: 22/05/2018/094). The total DNA was extracted from ethanol-fixed tissues with the Quick-DNA™ Tissue/Insect Miniprep Kit (Zymo Research) and 5 Gbp of sequences were obtained using the Illumina® Hiseq™ 2000 platform in paired-end reads with length 2x100 nt. We used the MITObim v1.8 program to assemble the mitogenome [10], with a random selection of one million read pairs. We used as a reference the mitogenomes of *T. occidentalis* and *T. europaea* [11,12] in independent runs and generated two similar partial sequences. Then, we selected homologous reads for both sequences from the library with BLAT [13] using a custom script (https://github.com/fjruizruano/ngs-protocols/blob/master/mapping_blat_gs.py). Finally, we clusterized and assembled the selected reads using RepeatExplorer (http://repeatexplorer.org/) [14] to reach a single consensus sequence of the mitogenome. Since the control region (D-loop) was incomplete, we finished the assembly by PCR, cloning and Sanger sequencing five clones following the protocols described in Gutiérrez et al. [11]. The resulting *T. aquitania* mitogenome was annotated using MITOS (http://mitos.bioinf.uni-leipzig.de) [15] and tRNA scan-SE (http://lowelab.ucsc.edu/tRNAscan-SE) [16].

For the phylogenetic analysis we employed both the complete mitochondrial genomes (including the tRNAs genes and the control region) and the concatenated sequences of 13 PCGs from 29 other insectivore species (11 Talpidae and 18 representative species from genera belonging to the remaining insectivore families; available at GenBank). Sequences were aligned using ClustalW. Poorly aligned positions and divergent regions were removed using Gblocksprogram v 0.91b (http://molevol.cmima.csic.es/castresana/Gblocks_server.html) [17]; the nucleotide substitution models were evaluated using MEGA version X [18]. The models with low BIC scores (Bayesian Information Criterion) were considered to best describe the substitution pattern. In all the alignments the best nucleotide substitution model was GTR+G+I. The phylogenetic relationships were reconstructed with both complete genomes and PCGs concatenated sequences. Maximum-Likelihood and Neighbor-Joining methods, both implemented in the program MEGA version X [18], and Bayesian inference (BI) implemented in MrBayes version 3.1 [19], were used. Node supports in Maximum-Likelihood and Neighbor-Joining analysis were assessed with 1000 bootstrap replicates. In Bayesian inference, two independent runs were performed with four MCMC (Markov Chain Monte Carlo) chains and run for 2 million generations. Trees were sampled each 100th generation and the burn-in was set to 25% of samples. Convergence was considered to be reached when the average standard deviation of split frequencies was less than 0.001. Finally, a 50% majority-rule consensus tree was calculated from the obtained trees and the posterior proportions were calculated using the command "sumt" in MrBayes.

Results and discussion

The mitochondrial genome (GenBank: MN443911) in the male *T. aquitania* varied in length from 16,776 to 16,846 bp due to variations in the control region (see below) and has a base composition of A (34.2%), C (24.8%), T (26.8%) and G (14.2%). The gene organization is identical to the mitogenomes of *Talpa* [11, 12], other Talpidae [20,21,22,23,24] and most mammal species having 13 protein coding genes, two rRNA genes, 22 tRNA genes, L-strand replication origin (OR) and a control region (Table 1) (Fig. 1).

The control region contains three domains: domain I includes the extended termination-associated sequence domain (ETAS-1 and ETAS-2), the central conserved domain II includes five conserved sequence blocks (CSB-B, C, D, E, F), and domain III includes three conserved sequence blocks (CSB-1, 2, 3), as has previously been described in other mammal species [25, 26] (Fig. 2). Between CSB-1 and CSB-2 there is a variable number of repetitions (19–26) of a tandem repeat composed of 10 bp monomers (consensus: CACACRTACG). Consequently, the control region length varied between 1,320 and 1,390 bp. Of the five clones of the control region that were sequenced (GenBank: MN450194-MN450198), three have 24 repetitions and the other two 19 and 26. In each clone, the repetitive fragments started with one variant of the monomer (CAGACGTACG) and finished with another (CACACGTGCG) (only one clone finished with CACACATCCG). Approximately half of the remaining monomers are identical to those present in the *T. occidentalis* mitogenome (CACACGTACG) ([11]), while the other half have one variant with a base substitution (CACACATACG) (Fig. 2). This variation denotes the existence of heteroplasmy in the mitogenome for the numbers and organization of the different variants of the tandem repeat monomers.

The PCGs of the male *T. aquitania* were identical in length and in their start and stop codons to those of *T. occidentalis* and *T. europaea*. For translation initiation, 10 mitochondrial PCGs of *T. aquitania*use the start codons ATG, while *nd2* uses ATA and *nd3* and *nd5* useATT. For translation termination, eight genes use TAA, *nd2* uses TAG and *cytB*usesAGA, while threegenes *(coIII*, *nd3* and *nd4)* use one incomplete stop codon (T--). The length of the tRNA genes varied from 61 to 75 bp (Table 1).

Overlapping occurred in some genes that varied from one nucleotide (between tRNA-Thr and tRNA-Pro) to 43 nucleotides (between ATP8 and ATP6). The WANCY region that contained five tRNA genes (tRNA-Trp, tRNA-Ala, tRNA-Asn, tRNA-Cys, tRNA-Tyr) also included the 37-bp long L-strand replication origin (Table 1; Fig. 1).

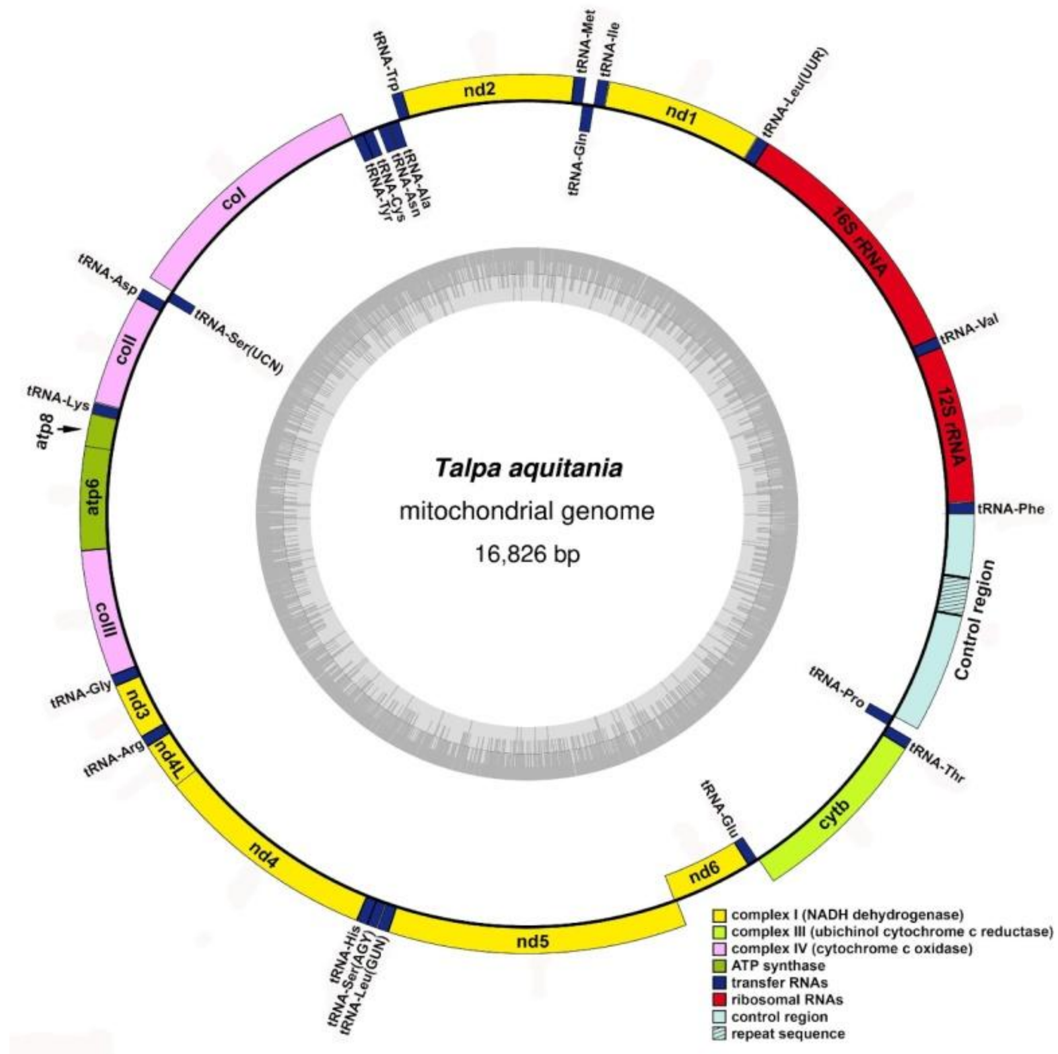


Figure 1. *Map of the T. aquitania mitochondrial genome. Genes encoded by the heavy strand are shown outside the circle and those encoded by the light strand are shown inside the circle. The inner ring shows the GC content of this genome.*

Table 1. *Gene organization of the Talpa aquitania mitochondrial genome*

Gene	Start position	Stop position	Length (bp)	Intergenic Nucleotides (bp)	Anticodon	Start codon	Stop codon	Strand
tRNAPhe	1	70	70	2	GAA			H
12S rRNA	73	1040	968	0				H
tRNAVal	1041	1108	68	0	TAC			H
16S rRNA	1109	2680	1572	0				H
tRNALeu(UUR)	2681	2755	75	2	TAA			H
nd1	2758	3711	954	2		ATG	TAA	H
tRNAIle	3714	3782	69	-3	GAT			H
tRNAGln	3780	3852	73	1	TTG			L
tRNAMet	3854	3922	69	0	CAT			H
nd2	3923	4966	1044	-2		ATA	TAG	H
tRNATrp	4965	5032	68	4	TCA			H
tRNAAla	5038	5106	69	1	TGC			L
tRNAAsn	5108	5180	73	0	GTT			L
OR	5181	5217	37	-3				H
tRNACys	5215	5281	67	0	GCA			L
tRNATyr	5282	5348	67	1	GTA			L
coI	5350	6894	1545	1		ATG	TAA	H
tRNASer(UCN)	6896	6964	69	7	TGA			L
tRNAAsp	6972	7040	69	0	GTC			H
coII	7041	7724	684	3		ATG	TAA	H
tRNALys	7728	7795	68	1	TTT			H
atp8	7797	8000	204	-43		ATG	TAA	H
atp6	7958	8638	681	-1		ATG	TAA	H
coIII	8638	9421	784	0		ATG	T--	H
tRNAGly	9422	9490	69	0	TCC			H
nd3	9491	9836	346	0		ATT	T--	H
tRNAArg	9837	9904	68	0	TCG			H
nd4L	9905	10201	297	-7		ATG	TAA	H
nd4	10195	11572	1378	0		ATG	T--	H
tRNAHis	11573	11640	68	0	GTG			H
tRNASer(AGY)	11641	11701	61	2	GCT			H
tRNALeu(CUN)	11704	11773	70	0	TAG			H
nd5	11774	13594	1821	-17		ATT	TAA	H
nd6	13578	14105	528	0		ATG	TAA	L
tRNAGlu	14106	14174	69	4	TTC			L
cytb	14179	15318	1140	0		ATG	AGA	H
tRNAThr	15319	15387	69	-1	TGT			H
tRNAPro	15387	15456	69	0	TGG			L
control region	15457	16826	1370	0				H

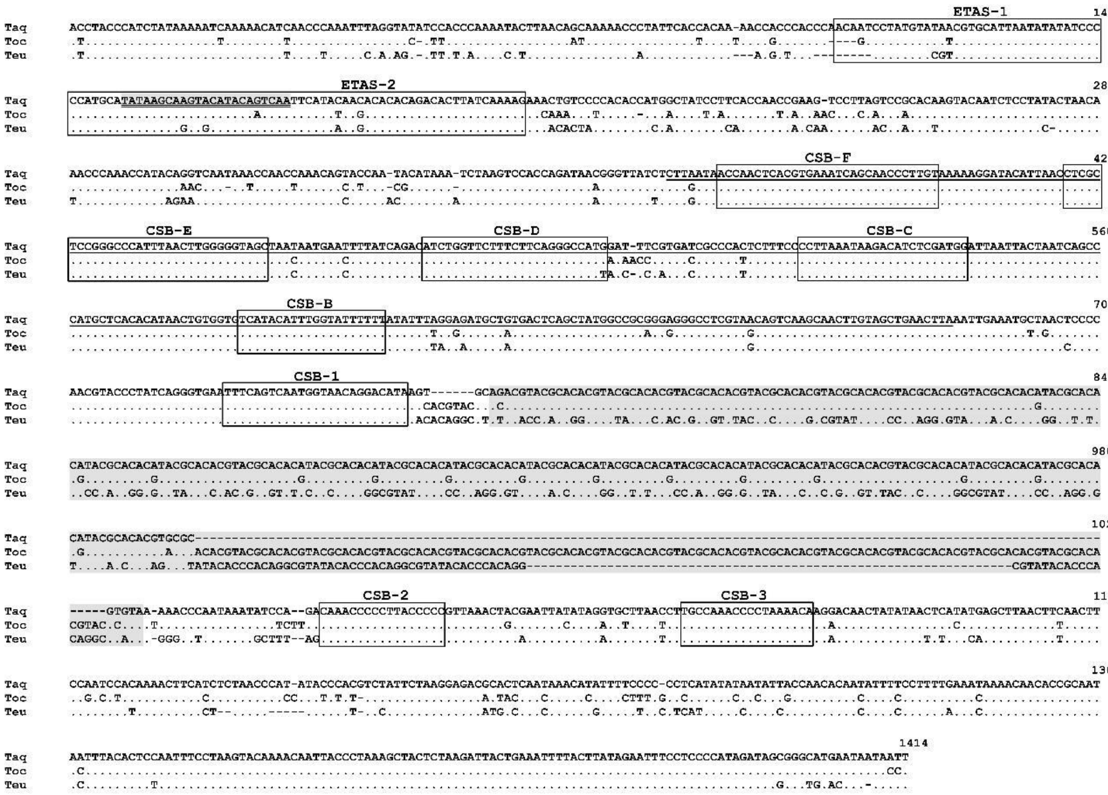


Figure 2*. Aligned nucleotide sequences of the control region of T. aquitania (Taq), T. occidentalis (Toc) and T. europaea (Teu). Dots indicate the identity of the nucleotides and dashes the indels. All the conserved sequences identified are included in boxes. In domain I the two conserved extended termination-associated sequences (ETAS-1 and ETAS-2) are in a single box; the overlapping nucleotides are in grey and double underlined. Central conserved domain II (underlined) includes five conserved blocks (CSB-F, E, D, C and B) and in domain III three conserved blocks (CSB-1, CSB-2 and CSB-3). The repeated DNA sequences between CSB-1 and CSB-2 are marked in grey.*

The phylogenetic position of *T. aquitania* in relation to *T. occidentalis* and *T. europaea* and the rest of the Talpidae family was assessed using both the available mitochondrial genomes and concatenated PCGs of 11 taxa representing the main lineages of this family and 18 additional representative species covering all genera from the other insectivore families available in GenBank (Fig. 3). Analyses based on Maximum Likelihood, Neighbor-Joining and Bayesian inference models yielded very similar phylogenetic trees for both PCGs and complete mitogenomes. All of them discriminate the included species according to the established families and subfamilies of insectivores. The three *Talpa* species grouped together, with *T. aquitanica* always included on the same branch as *T. occidentalis* (1 BI and 100% BP). Our analyses also indicate that *Talpa* and *Mogera* could be regarded as sister groups as the species of both genera group together in one single branch (1 BI and 100% BP). Another interesting feature was the inclusion of the three species of the genus *Uropsilus* (subfamily Uropsilinae) in one early diverging branch. *Galemys pyrenaicus* was always located on another early diverging branch that separated from the other Talpidae species, as previous analysis has found

[11,20,27]. Exceptionally, the Bayesian tree obtained using the complete genome sequences grouped *Galemys pyrenaicus* with *Urotrichus talpoides* on the same branch. The results for *Condylura cristata*, *Scapanulus oweni* and *Urotrichus talpoides* were not conclusive since their associations on branches on different trees were highly variable.

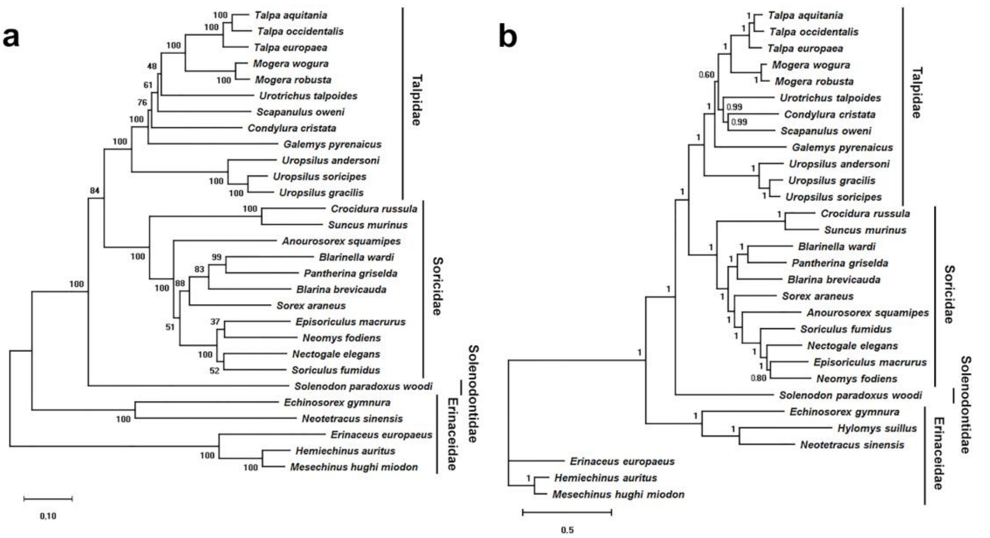


***Figure 3**. Maximum likelihood tree using (a) the complete mitochondrial genome and (b) a Bayesian inference phylogenetic tree with concatenated PCGs sequences of T. aquitania and related Talpidae species, and species of different insectivore genera (GTR+G+I model). Bootstrap supports and the Bayesian posterior probabilities are indicated at each node. The scale bar represents the number of nucleotide substitutions per site. GenBank Accession numbers: Anourosorex squamipes, NC_024563; Blarina brevicauda, NC_042734; Blarinella wardi, NC_041145; Condylura cristata, KU144678; Crocidura russula, AY769263; Echinosorex gymnura, NC_002808; Episoriculus macrurus, NC_029840; Erinaceus europaeus, NC_002080; Galemys pyrenaicus, AY833419; Hemiechinus auritus, NC_005033; Hylomys suillus, NC_010298; Mesechinus hughimiodon, KT824773; Mogera robusta, KT934322; Mogera wogura, AB099482; Nectogale elegans, NC_023351; Neomys fodiens, NC_025559; Neotetracus sinensis, NC_019626; Pantherina griselda, MH801935; Scapanulus oweni, KM506754; Solenodon paradoxus woodi, KU697363; Sorex araneus, KT210896; Soriculus fumidus, AF348081; Suncus murinus, NC_024604; Talpa aquitania, MN443911; Talpa europaea, Y19192; Talpa occidentalis, MF958963; Uropsilus andersoni, MF280389; Uropsilus gracilis, KM379136; Uropsilus soricipes, JQ658979; Urotrichus talpoides, AB099483. Note: Hylomys suillus is not in the complete genome tree as only the mitochondrial PCGs were available.*

In the family Soricidae two clades were observed in all the trees, one including the two species of the subfamily Crocidurinae and the other the nine Soricinae species analyzed. The position of the single species of Solenodontidae analyzed was not conserved on the different trees: in one half of the analysis it is located on an early diverging branch of the clade of Tapidae, while in the other trees it is contained in the clade that includes Talpidae and Soricidae. The species of the family Erinaceidae are also divided into two clades corresponding to the subfamilies Galericinae and Erinaceinae.

In conclusion, as indicated by their shared repeated DNA sequences and phylogenetic analyses, our findings show that *T. aquitania* and *T. occidentalis* are sister species.

Acknowledgments

This work was funded by the Consejería de Innovación, Ciencia y Empresa de la Junta de Andalucía (researchgroup RNM-924), and by Universidad de Jaén (via the program Plan de Apoyo a la Investigación 2017-2019, Acción 1). FJ Ruiz-Ruano was funded by a Lawski scholarship (Sweden) and G. Aleix-Mata by Fédération Nationale des Chasseurs, Fédération Départementales des Chasseurs de l'Ariège and Hautes Pyrénées (France) (project FNC-PSN-PR16-2014). The authors would like to express their gratitude to the Servicio Territorial de Medio Ambiente de Burgos (Junta de Castilla y León) for permission to capture *T. aquitania*.

References

1. Nicolas V, Martínez-Vargas J, Hugot JP (2015) *Talpa aquitania* nov. sp. (Talpidae, Soricomorpha) a new mole species from southwest France and north Spain. Bull Acad Vet Fr 168:329–334
2. Nicolas V, Martínez-Vargas J, Hugot JP (2017a) Molecular data and ecological niche modelling reveal the evolutionary history of the common and Iberian blind moles (Talpidae) in Europe. Zoologica Scripta 46:12–26
3. Nicolas V, Martínez-Vargas J, Hugot JP (2017b) *Talpa aquitania* sp. nov. (Talpidae, Soricomorpha), a new mole species from SW France and N Spain. Mammalia 81:641–642
4. Gutiérrez J, Aleix-Mata G, Lamelas L, Arroyo M, Marchal JA, Sánchez A (2019) Karyotype analysis of the new Talpa species *Talpa aquitania* (Talpidae; Insectivora) from northern Spain. Cytogenet Genome Res 159:26-31
5. Hutterer R (2005) Order Soricomorpha. In: Wilson, D.E, Reeder, D.M. (Eds.), Mammals Species of the World. A Taxonomic and Geographic Reference, third ed. The Johns Hopkins University Press, Baltimore, pp 220–311
6. Wilson DE, Reeder DM (2005) Mammal Species of the World: A Taxonomic and Geographic Reference, 3rd ed. Johns Hopkins University Press, Baltimore
7. Kryštufek B, Nedyalkov N, Astrin J, Hutterer R (2018) News from the Balkan refugium: Thrace has an endemic mole species (Mammalia: Talpidae). Bonn Zool Bull 67:41–57
8. Colangelo P, Bannikova AA, Krystufek B, Lebedev VS, Annesi F, Capanna E, Loy A (2010) Molecular systematics and evolutionary biogeography of the genus *Talpa* (Soricomorpha: Talpidae). Mol Phylogenet Evol 55:372-380
9. Bannikova AA, Zemlemerova ED, Colangelo P, Sözen M, Sevindik M, Kidov AA, Dzuev RI, Kryštufek B, Lebedev VS (2015) An underground burst of diversity - a new look at the phylogeny and taxonomy of the genus *Talpa* Linnaeus, 1758 (Mammalia: Talpidae) as revealed by nuclear and mitochondrial genes. Zool J Linnean Soc 175:930–948
10. Hahn C, Bachmann L, Chevreux B (2013) Reconstructing mitochondrial genomes directly from genomic next-generation sequencing reads-a baiting and iterative mapping approach. Nucleic Acids Res 41:e129
11. Gutiérrez J, Lamelas L, Aleix-Mata G, Arroyo M, Marchal JA, Palomeque T, Lorite P, Sánchez A (2018) Complete mitochondrial genome of the Iberian Mole *Talpa occidentalis* (Talpidae, Insectivora) and comparison with *Talpa europaea*. Genetica 146:415–423

12. Mouchaty SK, Gullberg A, Janke A, Arnason U (2000) The phylogenetic position of the Talpidae within Eutheria based on analysis of complete mitochondrial sequences. Mol Biol Evol 17:60-67
13. Kent WJ (2002) BLAT-the BLAST-like alignment tool. Genome Res 12:656–664
14. Novák P, Neumann P, Pech J, Steinhaisl J, Macas J (2013) Repeat Explorer: A Galaxy-based web server for genome-wide characterization of eukaryotic repetitive elements from next-generation sequence reads. Bioinformatics 29:792-793
15. Bernt M, Donath A, Jühling F, Externbrink F, Florentz C, Fritzsch G, Pütz J, Middendorf M, Stadler PF (2013) MITOS: improved de novo metazoan mitochondrial genome annotation. Mol Phylogenet Evol 69:313-319
16. Lowe TM, Eddy SR (1997) tRNAscan-SE: A program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res 25:955-964
17. Talavera G, Castresana J (2007) Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. Syst Biol 56:564-577
18. Kumar S, Stecher G, Li M, Knyaz C, Tamura K (2018) MEGA X: Molecular Evolutionary Genetics Analysis across computing platforms. Mol Biol Evol 35:1547-1549
19. Ronquist F, Huelsenbeck JP (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19:1572-1574
20. Cabria MT, Rubines J, Gómez-Moliner B, Zardoya R (2006) On the phylogenetic position of a rare Iberian endemic mammal, the Pyrenean desman (*Galemys pyrenaicus*). Gene 375:1-13
21. Hou Q, Tu F, Liu Y, Liu S (2016) Characterization of the mitogenome of *Uropsilus gracilis* and species delimitation. Mitochondrial DNA Part A 27:1836-1837
22. Xu Y, Huang X, Hu Y, Tu F (2016) Description of the mitogenome of Gansu mole (*Scapanulus oweni*). Mitochondrial DNA A DNA Mapp Seq Anal 27:2083–2084
23. Jia X, Yang L, Shi H (2018) The complete mitochondrial genome of Anderson's shrew mole, *Uropsilus andersoni* (Talpidae). Conserv Genet Resour. 10:583–585
24. Liu Z, Wang A, Li JJ (2019) Sequencing and analysis of the complete mitochondrial genome of the large mole (*Mogera robusta*) from China. Mitochondrial DNA Part B 4:1528-1530
25. Sbisà E, Tanzariello F, Reyes A, Pesole G, Saccone C. (1997) Mammalian mitochondrial D-loop region structural analysis: identification of new conserved sequences and their functional and evolutionary implications. Gene 205:125-140

26. Ketmaier V, Bernardini C (2005) Structure of the mitochondrial control region of the Eurasian otter (*Lutra lutra*; Carnivora, Mustelidae): patterns of genetic heterogeneity and implications for conservation of the species in Italy. J Hered 96:318-328

27. Shinohara A, Campbell KL, Suzuki H (2003) Molecular phylogenetic relationships of moles, shrew moles, and desmans from the new and old worlds. Mol Phylogenet Evol 27:247-258

Capítulo 5: Discusión general

DISCUSIÓN GENERAL

Diversidad cariotípica y ADN repetitivo

Las especies de la Familia Talpidae, y especialmente los topos, se caracterizan por tener unos cariotipos altamente conservados (Ye et al. 2006; Biltueva y Vorobieva 2012). La mayoría de especies del género *Talpa* presentan un número cromosómico diploide de 2n = 34. Sólo *T. caeca* y *T. caucasica* difieren, teniendo 2n = 36 y 2n = 38, respectivamente (Gornung et al. 2008). El NFa también está muy conservado y varía entre 62 (*T. romana* y *T. davidiana*) y 66 (*T. caeca*) (Fedyk e Ivanitskaya 1972; Gornung et al. 2008; Arslan y Zima 2014). Asimismo, los cromosomas sexuales también están bien conservados: el cromosoma X suele ser de medio tamaño submetacéntrico o metacéntrico (Gornung et al. 2008) y el cromosoma Y es, en la mayoría de las especies, un elemento pequeño y puntiforme de morfología variable. A pesar de esta conservación cromosómica, existen diferencias interespecíficas observables en el cariotipo, que se deben fundamentalmente a reordenamientos cromosómicos, cambios en la morfología de los cromosomas sexuales y a la cantidad y localización de la heterocromatina (Zima 1983; Jiménez et al. 1984; Kawada et al. 2002).

Hasta la fecha, los topos del norte de la Península Ibérica se identificaban como pertenecientes a la especie *T. europaea* (Feuda et al. 2015), aunque nunca se habían realizado análisis cromosómicos. Tras la reciente descripción de *T. aquitania* en el norte de España como una nueva especie muy próxima a *T. occidentalis* (Nicolas et al. 2015, 2017b), hemos analizado su cariotipo, para determinar las similitudes y características diferenciales entre estas tres especies.

Los cariotipos de *T. aquitania*, *T. europaea* y *T. occidentalis* son muy similares, con un número cromosómico 2n = 34 y NFa = 64, aunque con algunas diferencias como la cantidad de heterocromatina centromérica, el tamaño del brazo corto del cromosoma 16, la cantidad y localización de la heterocromatina en el cromosoma 14 y el tamaño, morfología y cantidad de heterocomatina del cromosoma Y (Jiménez et al. 1984; Volleth y Müller 2006; Gornung et al. 2008; Arslan y Zima 2014).

Uno de los pares cromosómicos más variables dentro del género *Talpa* es el par 16, debido a las variaciones en la longitud del brazo corto heterocromático. La amplificación/eliminación de la heterocromatina en el brazo corto de este cromosoma es una de las características diferenciales encontradas en los cariotipos de *Talpa*. Así, en algunas especies, los cromosomas de este par son telocéntricos (*T. davidina*) (Sözen et al. 2012), acrocéntricos (*T. europaea* y *T. romana*) (Gornung et al. 2008; Arslan y Zima 2014), submetacéntricos (*T. aquitania*), o incluso la mayor pareja de metacéntricos del cariotipo, gracias a la amplificación de la heterocromatina en dicho brazo, como ocurre en *T. altaica* (Kawada et al. 2002). Independientemente de la morfología de este cromosoma, el patrón de bandas G del brazo largo es idéntico en todos los cariotipos de *Talpa* (Jiménez et al. 1984; Volleth y Müller 2006; Gornung et al. 2008; Selçuk y Kefelioğlu 2017).

También se observan diferencias en la pareja de cromosomas número 14. Mientras que en *T. aquitania* sólo la parte proximal al centrómero del brazo corto es heterocromática, en *T. occidentalis* la heterocromatina se localiza en la región proximal centromérica de ambos brazos, y además el brazo corto es casi completamente heterocromático. En *T. romana* y *T. europaea* la heterocromatina de este cromosoma se extiende a parte del brazo corto (Jiménez et al. 1984; Gornung et al. 2008).

En cuanto al cromosoma Y, en la mayoría de las especies del género *Talpa* (incluidas *T. europaea y T. occidentalis*) es un elemento puntiforme y de morfología variable, aunque en *T. levantis* se han descrito tanto cromosomas Y puntiformes como metacéntricos de tamaño medio (Jiménez et al. 1984; Arslan y Zima 2014). Nuestros resultados muestran que en *T. aquitania* es claramente submetacéntrico y de tamaño medio. También hemos podido observar que el cromosoma Y de *T. aquitania* es completamente heterocromático como en *T. occidentalis*, mientras que en otras especies como *T. europaea*, *T. romana* y *T. altaica* es eucromático (Kawada et al. 2002; Gornung et al. 2008).

Hemos podido demostrar en el cariotipo de *T. aquitania,* la presencia de una constricción secundaria que porta los genes del ADNr en un único par cromosómico, lo cual es una característica compartida en las especies de *Talpa* analizadas (Jiménez et al. 1984; Zurita et al. 1997, 1998; Gornung et al. 2008; Selçuk y Kefelioğlu 2017). En este mismo par cromosómico de *T. aquitania*, se localizaron también secuencias teloméricas intersticiales (ITSs). Esta colocalización de los genes de ADNr y de las secuencias teloméricas se ha descrito también en *T. romana* y *T. europaea* (Gornung et al. 2008), lo que sugiere que esto podría ser una característica compartida en los cariotipos de *Talpa*.

Además de en los telómeros con intensidad de señal variable, en *T. aquitania* las secuencias teloméricas son claramente visibles como ITSs en las regiones pericentroméricas de algunos cromosomas, lo que también ocurre en los cariotipos de *T. romana* y *T. europaea* (Gornung et al. 2008). Es interesante observar que el par 16 en *T. aquitania* presenta ITSs centroméricas, al igual que sucede en el par correspondiente en *T. romana* y *T. europaea* (Gornung et al. 2008). La distribución no telomérica de las repeticiones de los telómeros se ha reportado en otros insectívoros (Zhdanova et al. 2005) y también es común en varias especies y grupos de mamíferos (Rovatsos et al. 2011). Las ITSs también están presentes en otros grupos de vertebrados y podrían haber surgido tras reordenamientos cromosómicos durante la evolución del cariotipo. Sin embargo, hay muchos ejemplos en los que su origen no está relacionado con este proceso (para una revisión, ver Ruíz-Herrera et al. 2008).

En relación a las secuencias de ADN repetitivo interdisperso, los elementos LINE1 son uno de los grupos más importantes de retrotransposones y son abundantes en los genomas de los mamíferos (Hardies et al. 2000; Mears y Hutchison 2001; Adelson et al. 2009). Se estima que un tercio del genoma de los mamíferos se originó directa o indirectamente a través de la retrotransposición LINE1 (Han y Boeke 2005). En este trabajo hemos clonado y estudiado la distribución de un fragmento de 1 kb del ORF2 de una LINE 1 en el genoma de *T. occidentalis*, *T. romana* y *T. aquitania.* Nuestros resultados demostraron que la mayoría de las secuencias obtenidas en nuestro estudio corresponden a copias no funcionales de LINE1. Esto no es sorprendente, dado que la retroposición de LINE1 genera en su mayoría copias defectuosas que permanecen en el genoma como secuencias mutadas o reordenadas (Furano 2000; Boissinot y Furano 2001). Además, estas secuencias LINE1 están ampliamente representadas en el genoma de estas tres especies de *Talpa* siguiendo un patrón de distribución cromosómica bastante conservado. Así, las secuencias LINE1 están ampliamente distribuidos en todos los cromosomas, pero se acumulan especialmente en las regiones heterocromáticas pericentroméricas de los autosomas. En los cromosomas sexuales son igualmente abundantes, tanto en la eucromatina como en la heterocromatina, como ya se ha descrito en otros genomas de mamíferos (Bailey et al. 2000; Dobigny et al. 2004; Waters et al. 2004; Marchal et al. 2006; Acosta et al. 2008; Kvikstad y Makova 2010).

La distribución de los retrotransposones LINE 1 en el cromosoma Y de las tres especies de *Talpa* está igualmente conservada y no parece estar relacionada con su mayor o menor contenido heterocromático, como se ha sugerido en otros estudios (Marshall Graves 2000, 2001). De hecho, las secuencias LINE1 se acumulan en los cromosomas Y de las tres especies, tanto si son heterocromáticos (*T. occidentalis* y *T. aquitania*) como eucromáticos (*T. romana*) (Jiménez et al. 1984; Gornung et al. 2008) y a pesar de las diferencias en el tamaño y la morfología de dicho cromosoma.

En las especies de *Talpa,* las investigaciones mediante “chromosome painting” son escasas. Se han realizado algunos estudios, utilizando sondas de cromosomas humanos o de insectívoros, con el fin de obtener datos acerca de la evolución cromosómica (Volleth y Müller 2006; Yang et al. 2006; Ye et al. 2006; Biltueva y Vorobieva 2012). Nuestros resultados de “painting” utilizando como sonda el cromosoma Y de *T. occidentalis,* muestran que el contenido del cromosoma Y de las tres especies de *Talpa* estudiadas está bien conservado. Si consideramos que los cromosomas Y puntiformes de *T. occidentalis* y *T. europaea* son similares al cromosoma Y ancestral, el aumento de tamaño del cromosoma Y que se produjo en *T. aquitania,* implicó la amplificación de secuencias ya presentes en ese cromosoma Y ancestral, que son compartidas actualmente por todos los cromosomas Y de las especies del género *Talpa*. Además, el "chromosome painting" ha demostrado que el cromosoma Y contiene secuencias muy similares a las que forman el bloque heterocromático de los pares autosómicos 14 y 16. Independientemente de la dinámica evolutiva particular que haya tenido lugar, es muy probable que estas secuencias repetidas ya estuvieran localizadas en la heterocromatina del cromosoma Y y de los cromosomas 14 y 16 en el cariotipo de la especie ancestral del género *Talpa*.

Es significativo que el escenario descrito en las especies de *Talpa* para la evolución del cromosoma Y, es decir la conservación del contenido de las secuencias, difiere significativamente del de otros mamíferos. Muchos estudios han demostrado que, incluso entre especies estrechamente relacionadas, el contenido de ADN del cromosoma Y está poco preservado, no sólo en la heterocromatina sino también en las regiones de la eucromatina, reflejo del proceso degenerativo que sigue la evolución de este cromosoma (Sitnikova et al. 2007; Kirsch et al. 2008; Gifalli-Iughetti y Koiffmann 2009; Wilson y Makova 2009; Acosta et al. 2011).

Satelitoma de *Talpa aquitania*

El análisis con RepeatExplorer ha revelado que un 20% del genoma de *T. aquitania* está compuesto por secuencias repetidas. Incluyendo las secuencias teloméricas, el satelitoma de *T. aquitania* está compuesto por al menos 16 familias de ADNsat, que constituyen el 1,24% del genoma. Este porcentaje de ADNsat concuerda con el observado en otras especies de Talpidae como *G. pyrenaicus* (1,08%) (Escoda y Castresana 2021). Sin embargo, parece que tanto la abundancia como el número de familias en otras especies podrían ser muy diferentes, como han demostrado algunos estudios en mamíferos (Valeri et al. 2021).

Las cuatro primeras familias en abundancia de ADNsat y las secuencias teloméricas comprenden el 1,13% del genoma (la familia más abundante representa por sí sola el 0,558%), mientras que los 11 ADNsats restantes representan sólo el 0,11%. El contenido medio de A + T de las familias de ADNsat es del 50,43%, inferior al contenido global de A + T del genoma (59,06%). Un alto contenido de A + T es una característica común en los genomas de Talpidae (datos GenBanK; Real et al. 2020; Escoda y Castresana 2021). La longitud media de los monómeros es de 289,24 pb, presentando gran variación entre 6 pb y 3102 pb. La longitud monomérica del ADN satélite más abundante es de 183 pb y la mayoría de las familias restantes tienen unidades de repetición menores.

Las relaciones entre las familias de ADNsat se han analizado mediante la comparación de todas las secuencias consenso, demostrando que no tenían similitud entre ellas. En una de las familias analizadas de ADNsat (TaquSat4-437-466) hemos identificado dos monómeros de longitudes diferentes (437pb y 466 pb), el monómero más largo alineaba con el monómero completo de menor tamaño presentando una identidad de 78,21%, y la porción no compartida se correspondía con el inicio de otro monómero.

El análisis de estos ADNsat indicó que presentan diferentes grados de clusterización, homogeneización y diversificación. En conjunto, el satelitoma de *T. aquitania* muestra un valor medio de divergencia nucleotídica del 13,07% y varía entre el 0,14% (muy poca variación entre monómeros) y el 21,69% (monómeros muy diferentes). Además, hemos calculado el índice de estructuración en tándem (TSI) de cada ADNsat (Montiel et al. 2021). Este índice varía entre 0 y 1 e informa sobre el nivel de clusterización de las familias de ADNsat. Representa la proporción de lecturas que contienen sólo secuencias de ADNsat en tandem con respecto al total de lecturas con secuencias de ADNsat. Los ADNsat de *T. aquitania* tienen un valor medio de TSI de 0,53, siendo generalmente más alto en las familias más abundantes, incluyendo el telomérico.

También hemos calculado el valor del pico de divergencia (divergence peak, DivP) que es una buena representación del estado de degeneración de cada familia de ADNsat. Las familias de ADNsat con una expansión reciente estarían formadas por monómeros casi idénticos y muestran valores bajos de divergencia mientras que las familias que se expandieron hace tiempo presentarán más variación en sus monómeros y por tanto valores altos de divergencia. Esto se debe a que, a lo largo del proceso evolutivo, han acumulado más mutaciones en su secuencia, y sus monómeros se han diversificado y hecho más diferentes. El valor de DivP en los ADNsat de *T. aquitania* varía entre 0 (TaquSat5-3102) y 24 (TaquSat3-6 y TaquSat13-54), lo que implica que tenemos familias de expansión muy reciente y otras más antiguas.

Como medida de la homogeneización de la familia (Camacho et al. 2022) hemos medido el valor de la abundancia relativa del pico (relative abundance size of the peak, RSP) que varía entre 0 (familias con monómeros muy diversificados en las que el satélite lleva mucho tiempo mutando y evolucionando) y 1 (familias muy homogéneas con monómeros casi idénticos). Este valor varia ente 0,21 (TaquSat3-6) y 1 (TaquSat5-3102).

Ambos índices (DivPeak y RSP) estaban inversamente correlacionados en el satelitoma de *T. aquitania* de manera que las familias recientemente expandidas (p.ej. TaquSat5-3102) mostraban valores bajos de DivPeak y altos de RSP, y las familias más antiguas altamente divergentes y degeneradas (p.ej. TaquSat3-6 y TaquSat13-54) presentaban valores altos de DivPeak y bajos de RSP.

El satelitoma es un componente genómico muy dinámico que podría variar entre especies estrechamente emparentadas o incluso entre poblaciones de la misma especie (Ferree y Prasad 2012; Pita et al. 2017; Montiel et al. 2021). Utilizando las secuencias consenso de las familias de ADNsat de *T. aquitania* hemos comparado el satelitoma de esta especie con el de *T. occidentalis*, considerada como la especie con la que está más estrechamente emparentada (Nicolas et al. 2015, 2017a y b) así como con otros representantes de la familia Talpidae.

La mayoría de los ADNsat encontrados en el genoma de *T. aquitania* fueron compartidos con *T. occidentalis* (15 de 16), aunque en esta última specie la abundancia del ADNsat resultó ser más baja (representaba el 0,69% del genoma en *T. occidentalis*, lo que corresponde a casi la mitad del satelitoma de *T. aquitania).* Como componente del genoma de rápida evolución, los cambios en la abundancia de ADNsat entre especies estrechamente relacionadas no son infrecuentes. La expansión o divergencia de estas secuencias repetidas configura efectos de fondo genético específicos de cada especie, que podrían contribuir al proceso de especiación (Hughes et al. 2009; Ferree y Prasad 2012). Es interesante, que las familias de ADNsat recientemente expandidas en el satelitoma de *T. aquitania*, no estén presentes en el satelitoma de *T. occidentalis*, como ocurre con TaquSat8-45, o presenten muy baja abundancia, aunque su valor de degeneración sea mayor que en el satelitoma de *T. aquitania*.

Aunque en número variable, la mayoría de las familias de ADNsat de *T. aquitania* están presentes también en los genomas de las otras especies de Talpidae analizadas, como *Talpa europaea*, *Scalopus aquaticus, Condylura cristata, Galemys pyrenaicus* y *Uropsilus gracilis*. En *T. europaea,* más próxima evolutivamente a *T. aquitania*, se identificaron 14 familias de ADNsat, mientras que en *U. gracilis*, la especie más distante, se presentaron solo 7 familias. Además, se da la circunstancia de que esas 7 familias también se presentan en las otras especies de *Talpa* analizadas, por lo que puede ser que estén conservadas en la familia Talpidae. La abundancia y divergencia de las familias de ADNsat también parecen estar relacionadas con la distancia filogenética entre especies, de acuerdo con la "hipótesis de la librería", propuesta por Fry y Salser (1977). A partir del mismo pool de secuencias de ADNsat, ciertas familias podrían eventualmente expandirse y fijarse en una especie, mientras que otras familias podrían hacerlo en otras, amplificándose o reemplazándose de manera diferente. En consecuencia, al mismo tiempo que las especies se distancian, sus satelitomas también lo hacen, contribuyendo a las diferencias entre especies o incluso favoreciendo nuevos eventos de especiación (Mestrović et al. 1998; Ugarkovic y Plohl 2002).

Sólo siete de las familias de ADNsat de *T. aquitania* fueron visualizadas por FISH en los cromosomas de esta especie y sólo seis en los cromosomas de *T. occidentalis*. La distribución cromosómica en ambas especies es muy similar, y las variaciones observadas están más relacionadas con las diferencias cromosómicas. Nuestros datos indican que en los genomas de *Talpa,* las familias de ADNsat se acumulan preferentemente en las regiones centroméricas y en las regiones heterocromáticas con bandas C positivas. Los resultados demuestran que el satelitoma de *T. occidentalis* y *T. europaea* es muy similar al de *T. aquitania*, aunque difiere significativamente en la presencia/ausencia y en la cantidad de varias familias de ADNsat. Las diferencias observadas entre *T. aquitania* y sus parientes cercanos *T. occidentalis* y *T. europaea* sugieren que el satelitoma es un componente muy dinámico del genoma y que el ADNsat podría ser responsable de las diferencias cariotípicas entre las especies. Por último, en un contexto amplio, estos datos contribuyen a la comprensión de la evolución de los satelitomas en los mamíferos.

Análisis del mitogenoma

La identificación y los estudios taxonómicos de las especies de *Talpa* se han basado tradicionalmente en el análisis morfométrico, análisis citogenético y análisis de secuencias como del citocromo b (Jiménez et al. 1984; Gornung et al. 2008; Arslan y Zima 2014; Nicolas et al. 2015, 2017a). El análisis del genoma mitocondrial de estas especies ha recibido hasta el momento poca atención. De hecho, hasta ahora sólo se había secuenciado completamente el mitogenoma de una especie del género *Talpa, T. europaea*, y el de otras nueve especies de la familia Talpidae (Mouchaty et al. 2000; Cabria et al. 2006; Hou et al. 2016; Xu et al. 2016; Jia et al. 2018; Liu et al. 2019). En este trabajo, hemos secuenciado y caracterizado el genoma mitocondrial completo de *Talpa occidentalis* y de *Talpa aquitania* mediante una combinación de los métodos de secuenciación masiva Illumina y Sanger, y por ensamblaje mediante métodos bioinformáticos.

Nuestros resultados indican que, como para la mayoría de los genomas mitocondriales (Cabria et al. 2006; Hou et al. 2016; Xu et al. 2016; Jia et al. 2018; Liu et al. 2019), los genomas de las especies del género *Talpa* contienen 13 genes codificadores de proteínas, 22 ARNs de transferencia, 2 ARNs ribosómicos (12 S y 16 S), el origen de replicación de la cadena L y una región de control (D-loop). Estos genomas no mostraron diferencias en los codones de inicio o de terminación ni en los tamaños de las proteínas codificadas. La diferencia más destacable entre estos mitogenomas afecta a la secuencia, tamaño y número de repeticiones de un monómero de ADN repetido en tándem localizado en la región control. Mientras que *T. aquitania* tiene entre 19 y 26 repeticiones de tres monómeros diferentes de 10 pb, que presentan mucha identidad con las 37 repeticiones de 10 pb de *T. occidentalis*, *T. europaea* tiene 19 repeticiones de 16 pb con muy poca identidad con las presentes en las otras dos especies (Mouchaty et al. 2000).

En nuestro análisis de los genomas mitocondriales, identificamos y describimos por primera vez regiones conservadas dentro de la región control, que han sido previamente descritas para otras especies de mamíferos (Sbisà et al. 1997; Ketmaier y Bernardini 2005). Así, hemos identificado en el dominio I los dominios conocidos como ETAS-1 y 2 (extended termination-associated sequence), y en los dominios II y III cinco (CSB-B, C, D, E, F) y tres (CSB-1, 2, 3) bloques de secuencias conservados.

La reconstrucción filogenética con Maximum-Likelihood, Neighbor-Joining y Bayesian inference, utilizando genomas mitocondriales completos y las secuencias de los 13 genes codificadores de proteínas, reproducen las relaciones filogenéticas aceptadas actualmente para las familias y subfamilias de insectívoros analizadas. Las tres especies de *Talpa* se agruparon, con *T. aquitanica* siempre incluida en la misma rama que *T. occidentalis* y separadas de *T. europaea,* lo que sugiere que las dos primeras son especies hermanas. Nuestros análisis también indican que *Talpa* y *Mogera* podrían considerarse grupos muy relacionados, ya que las especies de ambos géneros se agrupan en una sola rama. Otra característica interesante es la inclusión de las tres especies del género *Uropsilus* (subfamilia Uropsilinae) en una rama de divergencia temprana. *Galemys pyrenaicus* muestra una separación clara de las demás especies de Talpidae como se ha observado en estudios previos (Shinohara et al. 2003; Cabria et al. 2006). Basándose en resultados similares, algunos autores defienden la existencia en la familia Talpidae, además de las subfamilias Scalopinae, Talpinae y Uropsilinae, de la subfamilia Desmaninae (McKenna y Bell 1997), que incluiría especies de los géneros *Desmana* y *Galemys*. Nuestros resultados podrían apoyar el mantenimiento de esta subfamilia. Los resultados para *Condylura cristata*, *Scapanulu soweni* y *Urotrichus talpoides* no fueron concluyentes, ya que sus asociaciones en las ramas de diferentes árboles fueron muy variables. En la familia Soricidae se observaron dos clados en todos los árboles, uno que incluía las dos especies de la subfamilia Crocidurinae y el otro las nueve especies de Soricinae analizadas. La posición de la única especie de Solenodontidae analizada no se conservó en los diferentes árboles: en la mitad de los análisis se encuentra en una rama divergente temprana del clado de Tapidae, mientras que en los otros árboles está contenida en el clado que incluye a Talpidae y Soricidae. Las especies de la familia Erinaceidae también se dividen en dos clados correspondientes a las subfamilias Galericinae y Erinaceinae.

Capítulo 6: Conclusiones

CONCLUSIONES

1. El análisis mediante técnicas de citogenética clásica y molecular del cariotipo de *Talpa aquitania,* muestra que posee las características básicas distintivas de los cariotipos de las especies del género *Talpa*, diferenciándose en el tamaño y organización del cromosoma Y y en la longitud de los brazos cortos de algunas parejas de autosomas.

2. El análisis de las secuencias repetidas mediante técnicas clásicas y "chromosome painting" demuestran que éstas se acumulan principalmente en las regiones centroméricas y en los bloques heterocromáticos de los cromosomas, con apenas diferencias en las distintas especies de *Talpa* investigadas.

3. El satelitoma de *Talpa aquitania* representa el 1,24% del genoma y presenta al menos 16 familias, muchas de ellas compartidas con *Talpa occidentalis* y *Talpa europaea,* aunque con ciertas variaciones en la abundancia y el número de familias presentes en cada especie.

4. Los mitogenomas de *Talpa occidentalis* y *Talpa aquitania* son prácticamente idénticos, diferenciándose únicamente en la repetición de la región D-Loop, y presentan las características básicas de los mitogenomas de la mayoría de especies de mamíferos.

Bibliografía

BIBLIOGRAFÍA

Abad JP, Agudo M, Molina I, Losada A, Ripoll P (2000). Pericentromeric regions containing 1.688 satellite DNA sequences show anti-kinetochore antibody staining in prometaphase chromosomes of *Drosophila melanogaster*. Mol Gen Genet 264: 371-377.

Acosta MJ, Marchal JA, Fernández-Espartero CH, Bullejos M, Sánchez A (2008). Retroelements (LINEs and SINEs) in vole genomes: Differential distribution in the constitutive heterochromatin. Chromosome Res 16: 949-959.

Acosta MJ, Romero-Fernández I, Sánchez A, Marchal JA (2011). Comparative analysis by chromosome painting of the sex chromosomes in arvicolid rodents. Cytogenet Gen Res 132: 47-54.

Adelson DL, Raison JM, Edgar RC (2009). Characterization and distribution of retrotransposons and simple sequence repeats in the bovine genome. PNAS 106: 12855-12860.

Altschul SF, Stephen F, Thomas L, Madden L, Schaffer AA, Zhang Z, Miller W, Lipman DJ (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25: 3389-3402.

Ando K, Tagawa T, Uchida TA (1980). C-banding pattern on the chromosomes of the Japanese house shrew, *Suncus murinus riukiuanus*, and its implication. Experientia 36: 1040-1041.

Arslan A, Zima J (2014). Karyotypes of the mammals of Turkey and neighbouring regions: review. Folia Zool 63: 1-62.

Bailey JA, Carrel L, Chakravarti A, Eichler EE (2000). Molecular evidence for a relationship between LINE-1 elements and X chromosome inactivation: the Lyon repeat hypothesis. PNAS 97: 6634-6639.

Balakirev ES, Ayala FJ (2003). Pseudogenes: are they ‘‘junk’’ or functional DNA? Annu Rev Genet 37: 123-151.

Balloux F, Ecoffey E, Fumagalli L, Goudet J, Wyttenbach A, Hausser J (1998). Microsatellite conservation, polymorphism, and GC content in shrews of the genus *Sorex* (Insectivora, Mammalia). Mol Biol Evol 15: 473-475.

Bannikova AA, Lavrenchenkob LA, Kramerovc DA (2005). Phylogenetic relationships between Afrotropical and Palaearctic *Crocidura* species inferred from Inter-SINE-PCR. Biochemical Syst Ecol 33: 45-59.

Benson G (1999). "Tandem repeats finder: a program to analyze DNA sequences." Nucleic Acids Res 27: 573-580.

Bernard P, Maure, JF, Partridge JF, Genier S, Javerzat JP, Allshire RC (2001). Requirement of heterochromatin for cohesion at centromeres. Science 294: 2539-2542.

Bernt M, Donath A, Jühling F, Externbrink F, Florentz C, Fritzsch G, Pütz J, Middendorf M, Stadler PF (2013). MITOS: improved de novo metazoan mitochondrial genome annotation. Mol Phylogenet Evol 69: 313-319.

Biamonti G (2005). Structural and functional characterization of noncoding repetitive RNAs transcribed in stressed human cells. Mol Biol Cell 16: 2597-2604.

Bickelmann C, van der Vos W, de Bakker MA, Jiménez R, Maas S, Sánchez-Villagra MR (2017). Hox gene expression in the specialized limbs of the Iberian mole (*Talpa occidentalis*). Evol Dev 19: 3-8.

Biltueva L, Vorobieva N (2012). Chromosome evolution in Eulipotyphla. Cytogenet Genome Res 137: 154-164.

Biscotti MA, Canapa A, Forconi M, Olmo E, Barucca M (2015a). Transcription of tandemly repetitive DNA: functional roles. Chromosome Res 23: 463-477.

Biscotti MA, Olmo E, Heslop-Harrison JS (2015b). Repetitive DNA in eukaryotic genomes. Chromosome Res 23: 415-420.

Boissinot S, Furano AV (2001). Adaptive evolution in LINE-1 retrotransposons. Mol Biol Evol 18: 2186-2194.

Borodulina, OR, Kramerov, DA (2001). Short interspersed elements (SINEs) from insectivores. Two classes of mammalian SINEs distinguished by A-rich tail structure. Mamm Genome 12: 779-786.

Bullejos M, Sánchez A, Burgos M, Hera C, Jiménez R, Díaz de la Guardia R (1997). Multiple, polymorphic copies of SRY in both males and females of the vole *Microtus cabrerae*. Cytogenet Cell Genet 79: 167-171.

Burgos M, Jiménez R, Díaz de la Guardia R (1986). A rapid, simple and reliable combined method for G-banding mammalian and human chromosomes. Stain Technol 61: 257-260.

Butler PM (1988). Phylogeny of the insectivores. In The Phylo-geny and Classification of the Tetrapods, Vol. 2: Mammals (ed. M. J. Benton), pp. 117-141. Systematics Association Special Volume 35b. Clarendon Press, Oxford.

Cabral-de-Mello DC, Marec F (2021). Universal fluorescence in situ hybridization (FISH) protocol for mapping repetitive DNAs in insects and other arthropods. Mol Genet Genomics 296: 513-526.

Cabria MT, Rubines J, Gómez-Moliner B, Zardoya R (2006). On the phylogenetic position of a rare Iberian endemic mammal, the Pyrenean desman (*Galemys pyrenaicus*). Gene 375: 1-13.

Cam HP, Sugiyama T, Chen ES, Chen X, FitzGerald PC, Grewal SI (2005). Comprehensive analysis of heterochromatin- and RNAi-mediated epigenetic control of the fission yeast genome. Nat Genet 37: 809-819.

Camacho JPM, Cabrero J, López-León MD, Martín-Peciña M, Perfectti F, Garrido-Ramos MA, Ruiz-Ruano FJ (2022). Satellitome comparison of two oedipodine grasshoppers highlights the contingent nature of satellite DNA evolution. BMC Biol *20*: 36.

Campbell B (1939). The shoulder anatomy of the moles. A study in phylogeny and adaptation. Am J Anat 64: 1-39.

Carone DM, Longo MS, Ferreri GC, Hall L, Harris M, Shook N, Bulazel KV, Carone BR, Obergfell C, O'Neill MJ, O'Neill RJ (2009). A new class of retroviral and satellite encoded small RNAs emanates from mammalian centromeres. Chromosoma 118: 113-125.

Charlesworth B, Sniegowski P, Stephan W (1994). The evolutionary dynamics of repetitive DNA in eukaryotes. Nature 371: 215-220.

Colangelo P, Bannikova AA, Krystufek B, Lebedev VS, Annesi F, Capanna E, Loy A (2010). Molecular systematics and evolutionary biogeography of the genus *Talpa* (Soricomorpha: Talpidae). Mol Phylogenet Evol 55: 372-380.

Craig JM, Earnshaw WC, Vagnarelli P (1999). Mammalian centromeres: DNA sequence, protein composition, and role in cell cycle progression. Exp Cell Res 246: 249-262.

Descorps-Declère S, Richard GF (2022). Megasatellite formation and evolution in vertebrate genes. Cell Rep. 40:111347.

Dimitri P, Corradini N, Rossi F, Vernì F (2005). The paradox of functional heterochromatin. Bioessays 27: 29-41.

Djupedal I, Kos-Braun IC, Mosher RA, Söderholm N, Simmer F, Hardcastle TJ, Fender A, Heidrich N, Kagansky A, Bayne E, Wagner EG, Baulcombe DC, Allshire RC, Ekwall K (2009). Analysis of small RNA in fission yeast; centromeric siRNAs are potentially generated through a structured RNA. EMBO J 28: 3832-3844.

Dobigny G, Ozouf-Costaz C, Waters PD, Bonillo C, Coutanceau JP, Volobouev V (2004). LINE-1 amplification accompanies explosive genome repatterning in rodents. Chromosome Res 12: 787Y793.

Douady CJ, Catzeflis F, Kao DJ, Springer MS, Stanhope MJ (2002a). Molecular evidence for the monophyly of Tenrecidae (Mammalia) and the timing of the colonization of Madagascar by Malagasy tenrecs. Mol Phylogenet Evol 22: 357-363.

Douady CJ, Chatelier PI, Madsen O, De Jong WW, Catzeflis F, Springer MS, Stanhope MJ (2002b). 124 Matthew R. E. Symonds Molecular phylogenetic evidence confirming the

Eulipotyphla concept and in support of hedgehogs as the sister group to shrews. Mol Phylogenet Evol 25: 200-209.

Ducroz JF, Volobouev V, Granjon L (2001). An assessment of the systematics of arvicanthine rodents using mitochondrial DNA sequences: evolutionary and biogeographical implications. J Mamm Evol 8: 173-206.

Ekwall K, Javerzat JP, Lorentz A, Schmidt H, Cranston G, Allshire R (1995). The chromodomain protein Swi6: a key component at fission yeast centromeres. Science 269: 1429-1431.

Elrod DA, Beck ML, Kennedy ML (1996). Chromosomal variation in the southern short-tailed shrew (*Blarina carolinensis*). Genetica 98: 199-203.

ENCODE Project Consortium, Birney E, Stamatoyannopoulos JA, Dutta A, Guigó R, Gingeras TR, Margulies EH, Weng Z, Snyder M (2007). Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. Nature 447: 799-816.

Escoda L, Castresana J (2021). The genome of the Pyrenean desman and the effects of bottlenecks and inbreeding on the genomic landscape of an endangered species. Evol App 14: 1898-1913.

Fedyk S, Ivanitskaya EY (1972). Chromosomes of Siberian shrews. Acta theriol 27: 475-492.

Fernández R, Barragán MJ, Bullejos M, Marchal JA, Martínez S, Díaz de la Guardia R, Sánchez A (2001) Molecular and cytogenetic characterization of highly repeated DNA sequences in the vole *Microtus cabrerae*. Heredity 87: 637-646.

Fernández-Pérez J, Nantón A, Ruiz-Ruano FJ, Camacho JPM, Méndez J (2017). First complete female mitochondrial genome in four bivalve species genus *Donax* and their phylogenetic relationships within the Veneroida order. PLoS One 8: 12(9).

Ferree PM, Prasad S (2012). How can satellite DNA divergence cause reproductive isolation? Let us count the chromosomal ways. *Genet Res Int* 2012: 430136.

Feuda R, Bannikova AA, Zemlemerova ED, Febbraro M, Loy A, Hutterer R, Aloise G, Zykov AE, Annesi F, Colangelo P (2015). Tracing the evolutionary history of the mole, *Talpa europaea*, through mitochondrial DNA phylogeography and species distribution modelling. Biol J Linn Soc 114: 495-512.

Fredga K (1973). A new chromosome race of the common shrew (*Sorex araneus*) in Sweden. Hereditas 73: 153-157.

Fry K, Salser W (1977). Nucleotide sequences of HS-alpha satellite DNA from kangaroo rat *Dipodomysordii* and characterization of similar sequences in other rodents. Cell 12: 1069-1084.

Furano AV (2000). The biological properties and evolutionary dynamics of mammalian LINE-1 retrotransposons. Prog Nucleic Acid Res Mol Biol. 64: 255-294.

Galián J, Vogler AP (2003). Evolutionary dynamics of a satellite DNA in the tiger beetle species pair *Cicindela campestris* and *C. maroccana*. Genome 46: 213-223.

Gambogi CW, Dawicki-McKenna JM, Logsdon GA, Black BE (2020). The unique kind of human artificial chromosome: Bypassing the requirement for repetitive centromere DNA. Exp Cell Res 1:111978.

Garrido-Ramos MA (2017). Satellite DNA: An Evolving Topic. Genes 8:230.

Gifalli-Iughetti C, Koiffmann CP (2009). The Y chromosome of the Atelidae family (Platyrrhini): study by chromosome microdissection. Cytogenet Genome Res 125: 46-53.

Gornung E, Volleth M, Capanna E, Castiglia R (2008). Comparative cytogenetics of moles (Eulipotyphla, Talpidae): chromosomal differences in *Talpa romana* and *Talpa europaea*. Cytogenet Genome Res 121: 249-254.

Grady DL, Ratlif RL, Robinson DL, McCanlies EC, Meyene J, Moyzis RK (1992). Highly conserved repetitive DNA sequences are present at human centromeres. Proc. Natl. Acad. Sci.89: 1695-1699.

Graphodatsky A, Radjabli S, Zaitsev M, Sharshov A (1991). The levels of chromosome conservatism in the different groups of insectivores (Mammalia, Insectivora), in Zaitsev EV (ed): Questions of Systematic, Faunistics and Paleontology of Small Mammals, pp 47-57 (Zoological Institute USSR, St. Petersburg).

Grewal SI, Elgin SC (2007). Transcription and RNA interference in the formation of heterochromatin. Nature 447: 399-406.

Guenatri M, Bailly D, Maison C, Almouzni G (2004). Mouse centric and pericentric satellite repeats form distinct functional heterochromatin. J Cell Biol 166: 493-505.

Gutiérrez J, Lamelas L, Aleix-Mata G, Arroyo M, Marchal JA, Palomeque T, Lorite P, Sánchez A (2018). Complete mitochondrial genome of the Iberian Mole *Talpa occidentalis* (Talpidae, Insectivora) and comparison with *Talpa europaea*. Genetica 146: 415-423.

Hahn C, Bachmann L, Chevreux B (2013). Reconstructing mitochondrial genomes directly from genomic next-generation sequencing reads-a baiting and iterative mapping approach. Nucleic Acids Res 41:e129.

Hall TA (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symp Ser 41: 95-98.

Han JS, Boeke JD (2005). LINE-1 retrotransposons: modulators of quantity and quality of mammalian gene expression? Bioessays 27: 775.

Harada M, Ando A, Tsuchiya K, Koyasu K (2001). Geographical Variations in Chromosomes of the Greater Japanese Shrew-Mole, *Urotrichus talpoides* (Mammalia: Insectivora). Zool Sci 18: 433-442.

Hardies SC, Wang L, Zhou L, Zhao Y, Casavant NC, Huang S (2000). LINE-1 (L1) lineages in the mouse. Mol. Biol. Evol. 17: 616-628.

Haring E, Herzig-Straschil B, Spitzenberger F (2000). Phylogenetic analysis of Alpine voles of the *Microtus* multiplex complex using the mitochondrial control region. J Zool Evol Res 38: 231-238.

Hartley SE, Davidson WS (1994). Characterization and distribution of genomic repeat sequences from arctic char (*Salvelinus alpinus*). En "Genetics and Evolution of Aquatic Organisms." Beamont AR (Ed). Chapman and Hall, London pp. 271-279.

He K, Shinohara A, Helgen KM, Springer MS, Jiang XL, Campbell KL (2017). Talpid Mole Phylogeny Unites Shrew Moles and Illuminates Overlooked Cryptic Species Diversity. Mol Biol Evol 34: 78-87.

Hou Q, Tu F, Liu Y, Liu S (2016). Characterization of the mitogenome of *Uropsilus gracilis* and species delimitation. Mitochondrial DNA A DNA Mapp Seq Anal 27: 1836-1837.

Hughes SE, Hawley RS (2009). Heterochromatin: a rapidly evolving species barrier. PLoS Biol7: e1000233.

Hugot JP, Gu SH, Feliu C, Ventur J, Ribas A, Dormion J, Yanagihara R, Nicolas V (2014). Genetic diversity of *Talpa europaea* and Nova Hanta Virus (NVAV) in France. Bull Acad Vet Fr 167.

Hutterer R (2005). Order Soricomorpha. In: Wilson, D.E, Reeder, D.M. (Eds.). Mammals Species of the World. A Taxonomic and Geographic Reference, third ed. The Johns Hopkins University Press, Baltimore, pp 220-311.

Jia X, Yang L, Shi H (2018). The complete mitochondrial genome of Anderson’s shrew mole, *Uropsilus andersoni* (Talpidae). Conserv Genet Resour 10: 583-585.

Jiménez R, Burgos M, Díaz de la Guardia R (1984). Karyotype and chromosome banding in the mole (*Talpa occidentalis*) from the south-east of the Iberian Peninsula. Implications on its taxonomic position. Caryologia 37: 253-258.

Jiménez R, Burgos M, Sánchez A, Díaz de la Guardia R (1990). Synaptonemal complex analysis in *Talpa occidentalis* spermatocytes (Insectivora, Mammalia). II. Evolution of the X-chromosome self-pairing process. Cytogenet Cell Genet 54:35-37.

Jiménez R, Burgos M, Sánchez A, Sinclair AH, Alarcón FJ, Marín JJ, Ortega E, Díaz de la Guardia R (1993). Fertile females of the mole *Talpa occidentalis* are phenotypic intersexes with ovotestes. Development 118: 1303-1311.

John B, Milkos GLG (1988). The eukaryote genome in development and evolution (Allen & Unwin Eds) *London*.

Josse T, Teysset L, Todeschini A-L, Sidor CM, Anxolabehere D, Ronsseray S (2007). Telomeric trans-silencing: an epigenetic repression combining RNA silencing and heterochromatin formation. PLOS Genet 3: 158.

Kawada S, Harada M, Grafodatsky AS, Oda S (2002). Cytogenetic study of the Siberian mole, *Talpa altaica* (Insectivora: Talpidae) and karyological relationships within the genus *Talpa*. Mammalia 66: 53-62.

Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C, Thierer T, Ashton B, Meintjes P, Drummond A (2012). Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28: 1647-1649.

Ketmaier V, Bernardini C (2005). Structure of the mitochondrial control region of the Eurasian otter (*Lutra lutra*; Carnivora, Mustelidae): patterns of genetic heterogeneity and implications for conservation of the species in Italy. J Hered 96: 318-328.

Kim JY, Park YC (2015). Gene organization and characterization of the complete mitogenome of *Hypsugo alaschanicus* (Chiroptera: Vespertilionidae). Genet Mol Res 14: 16325-16331.

Kirsch S, Munch C, Jiang Z, Cheng Z, Chen L, Batz C, Eichler EE, Schempp W (2008). Evolutionary dynamics of segmental duplications from human Y-chromosomal euchromatin/heterochromatin transition regions. Genome Res 18: 1030-1042.

Kryštufek B, Nedyalkov N, Astrin J, Hutterer R (2018). News from the Balkan refugium: Thrace has an endemic mole species (Mammalia: Talpidae). Bonn Zool Bull 67: 41-57.

Kumar S, Stecher G, Li M, Knyaz C, Tamura K (2018). MEGA X: molecular evolutionary genetics analysis across computing platforms. Mol Biol Evol 35: 1547-1549.

Kumar S, Stecher G, Tamura K (2016). MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. Mol Biol Evol 33: 1870-1874.

Kvikstad EM, Makova KD (2010). The (r)evolution of SINE versus LINE distributions in primate genomes: Sex chromosomes are important. Genome Res 20: 600-613.

Lamelas L, Arroyo M, Fernández FJ, Marchal JA, Sánchez A (2018). Structural and Evolutionary Relationships in the Giant Sex Chromosomes of Three *Microtus* Species. Genes (Basel). 9: 27.

Liu Z, Wang A, Li JJ (2019). Sequencing and analysis of the complete mitochondrial genome of the large mole (*Mogera robusta*) from China. Mitochondrial DNA Part B 4: 1528-1530.

Lohse M, Drechsel O, Kahlau S, Bock R (2013). Organellar Genome- DRAW- a suite of tools for generating physical maps of plastid and mitochondrial genomes and visualizing expression data sets. Nucleic Acids Res 41: W575-W581 (Web Server issue).

Lopatin AV (2006). Early Paleogene insectivore mammals of Asia and establishment of the major groups of Insectivora. Paleontological Journal. 40: S205-S405.

Louzada S, Lopes M, Ferreira D, Adega F, Escudeiro A, Gama-Carvalho M, Chaves R (2020). Decoding the role of satellite DNA in genome architecture and plasticity-an evolutionary and clinical affair. Genes (Basel) 11:72.

Lowe TM, Eddy SR (1997). tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res 25: 955-964.

Lu J, Gilbert DM (2007). Proliferation-dependent and cell cycle regulated transcription of mouse pericentric heterochromatin. Cell Biol 179: 411-421.

Lu J, Gilbert DM (2008). Cell cycle regulated transcription of heterochromatin in mammals vs. fission yeast: functional conservation or coincidence. Cell Cycle 7: 1907-1910.

Madeira F, Park YM, Lee J, Buso N, Gur T, Madhusoodanan N, Basutkar P, Tivey ARN, Potter SC, Finn RD, Lopez R (2019). The EMBL-EBI search and sequence analysis tools APIs in 2019. Nucleic Acids Res 47(W1): W636-W641.

Marchal JA, Acosta MJ, Bullejos M, Díaz de la Guardia R, Sánchez A (2003). Sex chromosomes, sex determination, and sex-linked sequences in Microtidae. Cytogenet. Genome Res. 101: 266-273.

Marchal JA, Acosta MJ, Bullejos M, Puerma E, Díaz de la Guardia R, Sánchez A (2006). Distribution of l1-retroposons on the giant sex chromosomes of *Microtus cabrerae* (Arvicolidae, Rodentia): Functional and evolutionary implications. Chromosome Res 14: 177-186.

Marchal JA, Acosta MJ, Nietzel H, Sperling K, Bullejos M, Díaz de la Guardia R, Sánchez A (2004). X chromosome painting in *Microtus*: origin and evolution of the giant sex chromosomes. Chromosome Res 12: 767-776.

Marshall Graves JA (2000). Human Y chromosome, sex determination, and spermatogenesis, a feminist view. Biol Reprod 63: 667-676.

Marshall Graves JA (2001). From brain determination to testis determination: evolution of the mammalian sex-determining gene. Reprod Fertil Dev 13: 665-672.

Martínez-Guitarte JL, Díez JL, Morcillo G (2008). Transcription and activation under environmental stress of the complex telomeric repeats of *Chironomus thummi*. Chromosome Res 16: 1085-1096.

Maumus F, Quesneville H (2014). Deep investigation of *Arabidopsis thaliana* junk DNA reveals a continuum between repetitive elements and genomic dark matter. *PloS ONE*, 9: e94101.

McKenna MC, Bell SK (1997). Classification of mammals above the species level. Columbia University Press, New York, 631 pp.

McNulty SM, Sullivan BA (2018). Alpha satellite DNA biology: finding function in the recesses of the genome. Chromosome Res 26:115-138.

Mears ML, Hutchison CA (2001). The evolution of modern lineages of mouse L1 elements. J Mol Evol. 52: 51-62.

Mestrović N, Plohl M, Mravinac B, Ugarković D (1998). Evolution of satellite DNAs from the genus *Palorus*--experimental evidence for the "library" hypothesis. Mol Biol Evol 15: 1062-1068.

Mighell AJ, Smith NR, Robinson PA, Markham AF (2000). Vertebrate pseudogenes. FEBS Lett 468: 109-114.

Montiel EE, Panzera F, Palomeque T, Lorite P, Pita S (2021). Satellitome Analysis of *Rhodnius prolixus*, one of the Main Chagas Disease Vector Species. Int J Mol Sci *22*: 6052.

Mouchaty SK, Gullberg A, Janke A, Arnason U (2000). The phylogenetic position of the Talpidae within Eutheria based on analysis of complete mitochondrial sequences. Mol Biol Evol 17: 60-67.

Myers P (2001). "Animalia" (On-line), Animal Diversity Web. Accessed June 08, 2021 at https://animaldiversity.org/accounts/Animalia/.

Nakata A, Yoshimura A, Kuro-o M, Obara Y (2005). Karyotypic evolution and organization of the highly repetitive DNA sequences in the Japanese shrew-moles, *Dymecodon pilirostris* and *Urotrichus talpoides*. Cytogenet Genome Res 111: 152-158.

Nicolas V, Martínez-Vargas J, Hugot JP (2015). *Talpa aquitania* nov. sp. (Talpidae, Soricomorpha) a new mole species from southwest France and north Spain. Bull Acad Vet Fr 168: 329-334.

Nicolas V, Martínez-Vargas J, Hugot JP (2017a). Molecular data and ecological niche modelling reveal the evolutionary history of the common and Iberian blind moles (Talpidae) in Europe. Zool Scr 46: 12-26.

Nicolas V, Martínez-Vargas J, Hugot JP (2017b). *Talpa aquitania* sp. nov. (Talpidae, Soricomorpha), a new mole species from SW France and N Spain. Mammalia 81: 641-642.

Novacek MJ, Wyss AR. (1986). Higher-level relation-ships of the recent eutherian orders: morphological evidence. Cladistics 2: 257-287.

Novák P, Ávila-Robledillo L, Koblížková A, Vrbová I, Neumann P, Macas J (2017). TAREAN: a computational tool for identification and characterization of satellite DNA from unassembled short reads. NucleicAcids Res. 45: e111.

Novák P, Neumann P, Macas J (2010). Graph-based clustering and characterization of repetitive sequences in next-generation sequencing data. BMC Bioinformatics 11: 378.

Novák P, Neumann P, Macas J (2020). Global analysis of repetitive DNA from unassembled sequence reads using RepeatExplorer2. Nat Protoc. 15: 3745-3776.

Novák P, Neumann P, Pech J, Steinhaisl J, Macas J (2013). RepeatExporer: a Galaxy-based web server for genome-wide characterization of eukaryotic repetitive elements from next-generation sequence reads. Bioinformatics 29: 792-793.

Nurk et al. (2022). The complete sequence of a human genome. Science 376: 44-53.

Obara Y, Miyai T (1981). A preliminary study on the sex chromosome variation of the Ryukyu house shrew, *Suncus murinus riukiuanus*. Jpn J Genet 56: 365-371.

Ohzeki JI, Otake K, Masumoto H (2020). Human artificial chromosome: Chromatin assembly mechanisms and CENP-B. Exp Cell Res 389:111900.

Palomeque T, Lorite P (2008). Satellite DNA in insects: a review. Heredity 100: 564-573.

Perea-Resa C, Blower MD (2017). Satellite transcripts locally promote centromere formation. Dev Cell 42: 201-202.

Pita S, Panzera F, Mora P, Vela J, Cuadrado Á, Sánchez A, Palomeque T, Lorite P (2017). Comparative repeatome analysis on *Triatoma infestans* Andean and Non-Andean lineages, main vector of Chagas disease. PLoS One 12:e0181635.

Plohl M, Luchetti A, Mestrovic N, Mantovani B (2008). Satellite DNAs between selfishness and functionality: structure, genomics and evolution of tandem repeats in centromeric (hetero)chromatin. Gene 409: 72-82.

Plohl M, Meštrović N, Mravinac B (2012). Satellite DNA evolution. Genome dyn 7: 126-152.

Pons J, Petitpierre E, Juan C (1993). Characterization of the heterochromatin of the darkling beetle *Misolampus goudoti*: cloning of two satellite DNA families and digestion of chromosomes with restriction enzymes. Hereditas 119: 179-185.

Prasanth KV, Sacco-Bubulya PA, Prasanth SG, Spector DL (2003). Sequential entry of components of the gene expression machinery into daughter nuclei. Mol Biol Cell 14: 1043-1057.

Real FM, Haas SA, Franchini P, Xiong P, Simakov O, Kuhl H, Schöpflin R, Heller D, Moeinzadeh MH, Heinrich V, Krannich T, Bressin A, Hartmann MF, Wudy SA, Dechmann DKN, Hurtado A, Barrionuevo FJ, Schindler M, Harabula I, Osterwalder M, Hiller M, Wittler L, Visel A, Timmermann B, Meyer A, Vingron M, Jiménez R, Mundlos S, Lupiáñez DG (2020). The mole genome reveals regulatory rearrangements associated with adaptive intersexuality. Science 370: 208-214.

Reed KM, Phillips RB (1995). Molecular characterization and cytogenetics analysis of highly repeated DNAs of lake trout, *Salvelinus namaycush*. Chromosoma 104: 242-251.

Reumer JWF, Meylan A (1986). New developments in vertebrate cytotaxonomy IX. Chromosome numbers in the order Insectivora (Mammalia). Genetica 70: 119-151.

Ronquist F, Huelsenbeck JP (2003). MrBayes 3: bayesian phylogenetic inference under mixed models. Bioinformatics 19: 1572-1574.

Rosic S, Kohler F, Erhardt S (2014). Repetitive centromeric satellite RNA is essential for kinetochore formation and cell division. J Cell Biol 207: 334-349.

Rovatsos MT, Marchal JA, Romero-Fernández I, Fernández FJ, Giagia-Athanosopoulou EB, Sánchez A (2011). Rapid, independent, and extensive amplification of telomeric repeats in pericentromeric regions in karyotypes of arvicoline rodents. Chromosome Res 19: 869-882.

Rudert F, Bronner S, Garnier JM, Dolle P (1995). Transcripts from opposite strands of gamma satellite DNA are differentially expressed during mouse development. Mamm Genome 6: 76-83.

Ruedi M, Vogel P (1995). Chromosomal evolution and zoogeographic origin of southeast Asian shrews (genus *Crocidura*). Experientia 51: 174-178.

Ruíz-Herrera A, Nergadze SG, Santagostino M, Giulotto E (2008). Telomeric repeats far from the ends: mechanisms of origin and role in evolution. Cytogenet Genome Res 122: 219-228.

Ruiz-Ruano FJ, López-León MD, Cabrero J, Camacho J (2016). High-throughput analysis of the satellitome illuminates satellite DNA evolution. SciRep 6: 28333.

Sánchez A, Bullejos M, Burgos M, Jiménez R, Díaz de la Guardia R (1996). An alternative to blunt-end ligation for cloning DNA fragments with incompatible ends. Trends Genet 12: 44.

Sbisà E, Tanzariello F, Reyes A, Pesole G, Saccone C (1997). Mammalian mitochondrial D-loop region structural analysis: identification of new conserved sequences and their functional and evolutionary implications. Gene 205: 125-140.

Searle JB (1984). Hybridization between Robertsonian karyotypic races of the common shrew *Sorex araneus*. Experientia 40: 876-878.

Selçuk AY, Kefelioğlu H (2017). Cytogenetic characteristic of the Caucasian pygmy shrew (*Sorex volnuchini*) and Levant mole (*Talpa levantis*) (Mammalia: Eulipotyphla) in northern Anatolia, Turkey. Turk J Zool 41: 963-969.

Shatskikh AS, Kotov AA, Adashev VE, Bazylev SS, Olenina LV (2020). Functional Significance of Satellite DNAs: Insights From *Drosophila*. Front Cell Dev Biol 8:312.

Shi X, Seluanov A, Gorbunova V (2007). Cell divisions are required for L1 retrotransposition. Mol Cell Biol 27: 1264-1270.

Shinohara A, Campbell KL, Suzuki H (2003). Molecular phylogenetic relationships of moles, shrew moles, and desmans from the new and old worlds. Mol Phylogenet Evol 27: 247-258.

Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Söding J, Thompson JD, Higgins DG (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol Syst Biol 7: 539.

Simon C, Buckley TR, Frati F, Stewart JB, Beckenbach AT (2006). Incorporating molecular evolution into phylogenetic analysis, and a new compilation of conserved polymerase chain reaction primers for animal mitochondrial DNA. Annu Rev Ecol Evol Syst 37: 545-579.

Simon C, Frati F, Beckenbach A, Crespi B, Liu H, Flook P (1994). Evolution, weighting, and phylogenetic utility of mitochondrial gene sequences and a compilation of conserved polymerase chain reaction primers. Ann Entomol Soc Am 87: 651-701.

Simpson GG (1945). The principles of classification and a classification of mammals. Bulletin of the American Museum of Natural History 85: 1-350.

Singer ML (1982). Highly repeated sequences in mammalian genomes. Int Rev Cytol 76: 67-112.

Sitnikova NA, Romanenko SA, O'Brien PC, Perelman PL, Fu B, Rubtsova NV, Serdukova NA, Golenishchev FN, Trifonov VA, Ferguson-Smith MA, Yang F, Graphodatsky AS (2007). Chromosomal evolution of Arvicolinae (Cricetidae, Rodentia). I. The genome homology of tundra vole, field vole, mouse and golden hamster revealed by comparative chromosome painting. Chromosome Res 15: 447-456.

Sözen M, Matur F, Çolak F, Irmak S (2012). Karyological characteristics, morphological peculiarities, and a new distribution locality for *Talpa davidiana* (Mammalia: Soricomorpha) in Turkey. Turk J Zool 36: 806-813.

Stanhope MJ, Waddell VG, Madsen O, de Jong W, Hedges SB, Cleven GC, Kao D, Springer MS (1998). Molecular evidence for multiple origins of Insectivora and for a new order of endemic African insectivore mammals. PNAS 95: 9967-9972.

Storer J, Hubley R, Rosen J, Wheeler TJ, Smit AF (2021). The Dfam community resource of transposable element families, sequence models, and genome annotations. Mob DNA 12:2.

Straub SC, Parks M, Weitemier K, Fishbein M, Cronn RC, Liston A (2012). Navigating the tip of the genomic iceberg: Next-generation sequencing for plant systematics. Am J Bot 99: 349-364.

Sumner AT (1972). A simple technique for demonstrating centromeric heterochromatin. Expl Cell Res 75: 304-306.

Talavera G, Castresana J (2007). Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. Syst Biol 56: 564–577.

Ugarković D (2005). Functional elements residing within satellite DNAs. *EMBO* Rep 6: 1035-1039.

Ugarković D, Plohl M (2002). Variation in satellite DNA profiles-causes and effects. EMBO J 21: 5955-5959.

Untergasser A, Cutcutache I, Koressaar T, Ye J, Faircloth BC, Remm M, Rozen SG (2012). Primer3 - new capabilities and interfaces. Nucleic Acids Res 40: e115.

Usakin L, Abad J, Vagin VV, de Pablos B, Villasante A, Gvozdev VA (2007). Transcription of the 1.688 satellite DNA family is under the control of RNA interference machinery in *Drosophila melanogaster* ovaries. Genetics 176: 1343-1349.

Valeri MP, Dias GB, do Espírito Santo AA, Moreira CN, Yonenaga-Yassuda Y, Sommer IB, Kuhn GCS, Svartman M (2021). First Description of a Satellite DNA in Manatees' Centromeric Regions. Front Genet 12: 694866.

Vicari MR, Nogaroto V, Noleto RB, Cestari MM, Cioffi MB, Almeida MC, Moreira-Filho O, Bertollo LAC, Artoni RF (2010). Satellite DNA and chromosomes in Neotropical fishes: methods, applications and perspectives. J Fish Biol 76: 1094-1116.

Volleth M, Müller S (2006). Zoo-FISH in the European mole (*Talpa europaea*) detects all ancestral Boreo-Eutherian human homologous chromosome associations. Cytogenet Genome Res 115: 154-157.

Waddell PJ, Cao Y, Hauf J, Hasegawa M (1999). Using novel phylogenetic methods to evaluate mammalian mtDNA, including amino acid-invariant sites-LogDet plus site stripping, to detect internal conflicts in the data, with special reference to the positions of hedgehog, armadillo, and elephant. Syst Biol 48: 31-53.

Waters PD, Dobigny G, Pardini AT, Robinson TJ (2004). LINE-1 distribution in Afrotheria and Xenarthra: implications for understanding the evolution of LINE-1 in eutherian genomes. Chromosoma 113: 137-144.

Wilson MA, Makova KD (2009). Genomic analyses of sex chromosome evolution. Annu Rev Genomics Hum Genet 10: 333-354.

Wilson DE, Reeder DM (2005). Mammal Species of the World: A Taxonomic and Geographic Reference, 3rd ed. Johns Hopkins University Press, Baltimore.

Wong LH, Brettingham-Moore KH, Chan L, Quach JM, Anderson MA, Northrop EL, Hannan R, Saffery R, Shaw ML, Williams E, Choo KH (2007). Centromere RNA is a key component for the assembly of nucleoproteins at the nucleolus and centromere. Genome Res 17: 1146-1160.

Wyttenbach A, Goudet J, Cornuet JM, Hausser J (1999). Microsatellite variation reveals low genetic subdivision in a chromosome race of *Sorex araneus* (Mammalia, Insectivora). J Hered 90: 323-327.

Xu Y, Huang X, Hu Y, Tu F (2016). Description of the mitogenome of Gansu mole (*Scapanulus oweni*). Mitochondrial DNA A DNA Mapp Seq Anal 27: 2083-2084.

Yang F, Graphodatsky AS, Li T, Fu B, Dobigny G, Wang J, Perelman PL, Serdukova NA, Su W, O'Brien PCM, Wang Y, Ferguson-Smith MA, Volobouev V, Nie W (2006). Comparative genome maps of the pangolin, hedgehog, sloth, anteater and human revealed by cross-species chromosome painting: further insight into the ancestral karyotype and genome evolution of eutherian mammals. Chromosome Res 14: 283-296.

Yates TL, Moore DW (1990). Speciation and evolution in the family Talpidae (Mammalia: Insectivora). In: Nevo, E., Reig, O.A. (Eds.), Evolution of Subterranean Mammals at the Organismal and Molecular Levels. Alan R. Liss, New York, pp 1-22.

Ye J, Biltueva L, Huang L, Nie W, Wang J, Jing M, Su W, Vorobieva NV, Jiang X, Graphodatsky AS, Yang F (2006). Cross-species chromosome painting unveils cytogenetic signatures for the Eulipotyphla and evidence for the polyphyly of Insectivora. Chromosome Res 14: 151-159.

Ye F, Samuels DC, Clark T, Guo Y (2014). High-throughput sequencing in mitochondrial DNA research. Mitochondrion 17: 157-163.

Yong HS (1971). Chromosome polymorphism in the Malayan house shrew, *Suncus murinus* (Insectivora, Soricidae). Experientia 27: 589-591.

Yosida TH (1982). Cytogenetical studies on Insectivora. II. Geographical variation of chromosomes in the house shrew, *Suncus murinus* (Soricidae), in East, Southeast and Southwest Asia, with a note on the karyotype evolution and distribution. Jpn J Genet 57: 101-111.

Zhang J, Xu F, Hashimshony T, Keshet I, Cedar H (2002). Establishment of transcriptional competence in early and late S phase. Nature 420: 198-202.

Zhang Z, Schwartz S, Wagner L, Miller W (2000). A greedy algorithm for aligning DNA sequences. J Comput Biol 7: 203-214.

Zheng D, Zhang Z, Harrison PM, Karro J, Carriero N, Gerstein M (2005). Integrated pseudogene annotation for human chromosome 22: evidence for transcription.J. Mol. Biol. 349: 27-45.

Zhdanova NS, Karamishevam TV, Minina J, Astakhova NM, Lansdorp P, Kammori M, Rubtsov NB, Searle JB (2005). Unusual Distribution Pattern of Telomeric Repeats in the Shrews *Sorex araneus* and *Sorex granarius*. Chromosome Res. 13: 617-625.

Zima J (1983). The karyotype of *Talpa europaea kratochvili* (Talpidae, Insectivora). Folia Zool 32: 131-136.

Zurita F, Jiménez R, Burgos M, Díaz de la Guardia R (1998). Sequential silver staining and in situ hybridization reveal a direct association between rDNA levels and the expression of homologous nucleolar organizing regions: a hypothesis for NOR structure and function. J Cell Sci 111:1433-1439.

Zurita F, Sánchez A, Burgos M, Jiménez R, Díaz de la Guardia R (1997). Interchromosomal, intercellular and interindividual variability of NORs studied with silver staining and in situ hybridization. Heredity 77: 229-223.

Zwick MS, Hanson RE, Islam-Faridi MN, Stelly DM, Wing RA, Price HJ, McKnight TD (1997). A rapid procedure for the isolation of C0t-1 DNA from plants. Genome 40: 138-142.