

Genómica comparativa en *Triatoma patagonica* (Hemiptera: Reduviidae)

Tesina para optar por el grado de Licenciado en
Ciencias Biológicas

Br. Santiago Pereyra Mello

Tutor: Dr. Sebastián Pita Mimbacas, Sección Genética
Evolutiva.

ÍNDICE

INTRODUCCIÓN	3
OBJETIVOS	8
MATERIALES Y MÉTODOS.....	9
RESULTADOS.....	14
DISCUSIÓN.....	21
ANEXO	26
BIBLIOGRAFÍA	32

INTRODUCCIÓN

La subfamilia Triatominae (Hemiptera: Reduviidae) comprende más de 150 especies descritas de insectos hematófagos agrupadas en 15 géneros, distribuidas a lo largo del continente americano y Asia [1,2]. Dentro de la subfamilia Triatominae, el género más conspicuo es *Triatoma* con más de 80 especies, agrupadas dentro de 8 complejos y 8 subcomplejos según criterios morfológicos, de hábitat, ecológicos y análisis genéticos [3,4].

Estos insectos son vectores del parásito protozooario *Trypanosoma cruzi*, agente causante de la enfermedad de Chagas, reconocida como la enfermedad parasitaria humana más seria de Latinoamérica debido a su impacto social y económico [5]. Si bien existen distintas formas de transmisión, la vectorial representa más del 80% de los casos [6]. El hecho de no existir ninguna vacuna ni tratamiento farmacológico eficaz contra el parásito, conlleva a que la principal estrategia de gestión elegida sea la eliminación de las poblaciones de vectores domésticos, mediante el uso de insecticidas residuales [7]. Por este motivo, es importante el conocimiento en profundidad de las características biológicas de estos insectos, incluyendo su genoma. Con las especies adecuadamente descritas y sus características ecológicas y epidemiológicas bien determinadas, se tendrá una mayor comprensión de los procesos que contribuyen a su diversificación y domesticación, lo cual facilitará mejorar el diseño de las estrategias de control vectorial [1].

La importancia epidemiológica de los vectores de la enfermedad de Chagas depende primordialmente de la habilidad de dispersarse y adaptarse a hábitats domésticos [8]. *Triatoma infestans* es una de las especies de Triatominae más exitosamente adaptada a ambientes domésticos y el vector más importante en el Cono Sur de América Latina. Desde 1991, campañas de fumigación con insecticidas a gran escala fueron implementadas exitosamente en varios países para la eliminación de *T. infestans* doméstica. En algunas regiones, cuando el vector primario es eliminado, otras especies de triatominos se vuelven epidemiológicamente importantes. Incluido en estos vectores secundarios está *Triatoma patagonica*, una especie endémica de Argentina con un amplio rango geográfico desde el norte al sur del país [8]. Esta especie es principalmente

silvestre, pero se ha encontrado colonizando estructuras domiciliarias y peridomiciliarias en varias provincias argentinas [9,10]. Diferentes autores han demostrado la capacidad de la especie como vector de *Tryp. cruzi* [11,12].

Estudios genéticos son esenciales para conocer el estado taxonómico, estructura poblacional y procesos de diferenciación y dispersión. La interacción entre flujo génico, selección y deriva génica a través de regiones biogeográficas a menudo resulta en un mosaico de poblaciones genómicamente diferenciadas. Las especies de Triatominae son consideradas insectos con plasticidad, que responden rápidamente a cambios ambientales, lo que puede representar una ruta evolutiva común a nuevas especies [13]. Métodos basados en análisis de ADN, incluyendo genes mitocondriales, han demostrado ser herramientas efectivas para evaluar la estructura de poblaciones, reconocer especies morfológicamente similares e inferir relaciones filogenéticas en Triatominae [14–17].

Si bien los estudios poblacionales de *T. patagonica* son muy escasos, un análisis preliminar, utilizando un fragmento del gen mitocondrial *citocromo oxidasa I* (*coI*) mostró resultados que merecen mayor atención. El análisis filogenético reveló la existencia de al menos dos linajes mitocondriales con gran distancia genética entre sí, uno perteneciente al norte de Argentina y otro al centro y sur del país (Fig. 1). Además, se realizaron cruzamientos experimentales que apoyan la existencia de barreras post-cigóticas entre los clados del norte y del centro-sur, sugiriendo que las poblaciones de *T. patagonica* están sufriendo un incipiente proceso de especiación [18].

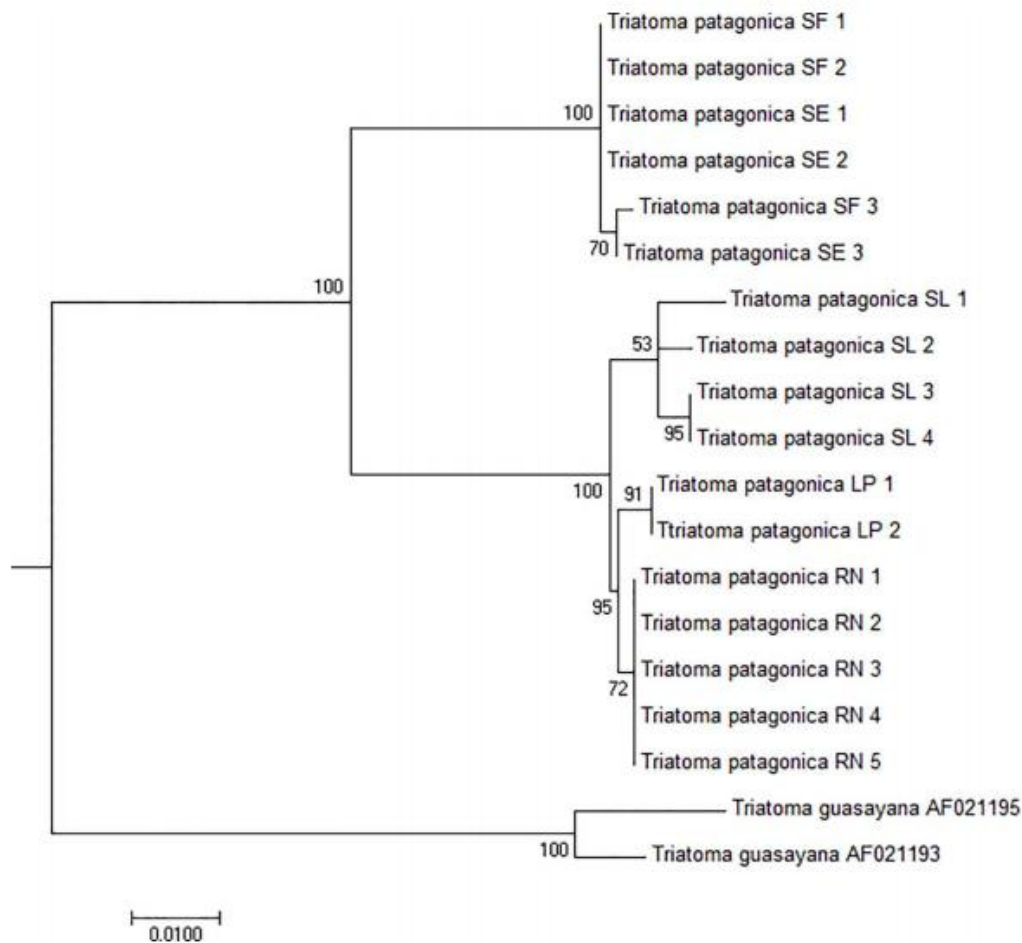


Fig. 1. Árbol filogenético de máxima verosimilitud, obtenido a partir de un fragmento del gen *col* de *T. patagonica*. La topología del árbol revela que *T. patagonica* está compuesta por dos clados soportados, representados por norte (individuos SE y SF), y las regiones centro (individuos SL) y sur (individuos LP y RN). El número sobre los nodos representa el soporte estadístico obtenido mediante 1000 replicaciones bootstrap. (Extraído de [18])

Citogenéticamente, ha sido observada una diferencia en el contenido de heterocromatina constitutiva (heterocromatina-C) entre los linajes mitocondriales norte (35% de contenido heterocromatina C) y sur (0%) de *T. patagonica*. Además, podría describirse un tercer grupo, en el centro de su distribución, con una cantidad intermedia de heterocromatina-C [18]. Aunque frecuentemente se observa un polimorfismo de heterocromatina-C en varias especies de triatominos [19], las pronunciadas diferencias de heterocromatina autosómica entre individuos norte y sur no son usuales en poblaciones conespecíficas. Cambios en cantidad y posición de la heterocromatina-C en los cromosomas pueden afectar el nivel de recombinación y expresión de las regiones de eucromatina adyacentes [20]. En *T. infestans* se ha sugerido que la marcada diferenciación poblacional en heterocromatina-C podría estar reflejando cambios genómicos adaptativos que

contribuyen a la habilidad para sobrevivir y reproducirse en distintos ambientes [20,21]. La variación de heterocromatina en poblaciones de *T. patagonica* también podría estar relacionada a cambios adaptativos asociados a diferentes condiciones ambientales y ser el causante del proceso de especiación incipiente antes mencionado [18].

Las técnicas de secuenciación de próxima generación (*next generation sequencing*: NGS) son herramientas poderosas que nos permiten llevar a cabo un análisis global de las secuencias repetidas que forman parte de un genoma [22–24]. A pesar del progreso reciente en estas tecnologías, la cuantificación precisa y caracterización de secuencias de ADN repetitivo en genomas complejos de animales y plantas sigue siendo una tarea desafiante. La mayoría de las herramientas bioinformáticas disponibles requieren el ensamblado completo del genoma para la identificación de repetidos o dependen de búsquedas de similitud desafiando bases de datos de elementos repetidos ya conocidos. Esto tiene además la contrapartida de que existen elementos repetidos que no tengan similitud con ninguno conocido a la fecha, y por ende no poseen similitud con los utilizados en las bases de datos.

La fracción de ADN repetitivo de un genoma, llamado “repitoma” (“*repeatome*” en inglés) por Maumus y Quesneville [25], ha sido frecuentemente omitido o analizado superficialmente en la mayoría de los análisis de genoma completo [25], probablemente debido a ser secuencias no-codificantes o por la dificultad que representan estas secuencias a la hora de realizar un ensamblaje del genoma. Sin embargo, hay numerosos estudios demostrando el rol fundamental de las secuencias repetitivas en la evolución y conservación del genoma [27]. La investigación sobre la naturaleza de las secuencias repetitivas en los genomas de triatomíneos todavía es escasa, reduciéndose únicamente a dos especies [28,29].

La determinación del repitoma y su posterior análisis puede ser abordado mediante la implementación del programa RepeatExplorer [22], que ofrece un enfoque diferente para el análisis global de repetidos, permitiendo la identificación de repetidos *de novo* (sin conocimiento previo, como por ejemplo una base de datos) y es adecuada para utilizar en *reads* crudos sin necesidad de realizar un ensamblaje. Está basada en encontrar y cuantificar similitudes entre lecturas de NGS o “*reads*”, obtenidas de secuenciaciones de baja cobertura (<1X). Estas

similitudes se utilizan para construir un grafo en el cual las uniones corresponden a lecturas o *reads*. Los *reads* solapantes están conectados por los bordes o uniones y su *score* de similitud se expresa como “peso de unión” (*edge weights* en inglés). La topología de los grafos es entonces analizada para identificar y separar *clusters* (grupos) formados por *reads* que representan familias de elementos repetidos. Esta estrategia ha demostrado ser eficiente para analizar la composición de repetidos en genomas tanto de plantas [30-33] como de animales [29,34-36].

OBJETIVOS

Objetivo general:

Determinar las diferencias a nivel genómico entre individuos de la especie *Triatoma patagonica* pertenecientes a distintos linajes mitocondriales, haciendo énfasis en secuencias de ADN repetidas.

Objetivos específicos:

- 1) Identificar y clasificar el ADN repetitivo contenido en el genoma de cada linaje de *T. patagonica* mediante herramientas bioinformáticas tales como RepeatExplorer, TAREAN, entre otras.
- 2) Determinar las distintas familias de ADN satélite que componen los genomas de cada linaje y establecer para cada una de ellas los monómeros de repetición.
- 3) Reconocer cuáles familias de ADN satélite se comparten, y si existen familias específicas para cada linaje.
- 4) Evaluar una posible marca filogenética en las secuencias de ADN repetitivo entre los linajes de *T. patagonica*.

MATERIALES Y MÉTODOS

Muestras y extracción de ADN

Se utilizaron 5 individuos machos de la especie *T. patagonica*. Uno de cada una de las siguientes localidades: norte: Mitre, Santiago del Estero (SE) y 9 de Julio, Santa Fé (SF); sur: Utracan, La Pampa (LP) y Avellaneda, Río Negro de Argentina (RN); centro: San Martín, San Luis.

Para la extracción del ADN, se utilizaron patas fijadas en etanol 70% de cada espécimen y el ADN total fue aislado con un kit comercial. Los resultados de las extracciones fueron evaluadas mediante un gel de agarosa al 1%, utilizando como tinción el intercalante GelRed. Las muestras de ADN total extraído fueron enviadas a BGI (Hong Kong) para la secuenciación de ADN de baja cobertura con la tecnología de DNBseq, similar a Illumina, de *paired-end reads* o lecturas pareadas con longitud de 150 pares de bases.

Procesamiento de las secuencias

Para cada una de las bibliotecas se realizó el curado y procesamiento de las secuencias obtenidas, donde se evaluó la calidad y el tamaño de los *reads*, mediante herramientas bioinformáticas implementadas en lenguajes BASH, java, perl y python. Primeramente, se realizó un análisis de la calidad de los *reads* de la secuenciación cruda utilizando el programa fastQC [37].

En segundo lugar, se utilizó la herramienta DeconSeq [38] para poder filtrar los *reads* que correspondían al ADN mitocondrial de las muestras. Para utilizar este programa se emplearon como molde los mitogenomas del género *Triatoma* ya publicados [39,40]. Se obtuvieron dos sets de *reads*, uno conteniendo los filtrados sin el genoma mitocondrial, y el otro conteniendo los similares a los genomas mitocondriales de otras especies. Para los siguientes pasos se utilizó el set de *reads* filtrados sin el genoma mitocondrial.

Se descartaron *reads* utilizando la herramienta Trimmomatic [41], donde se implementaron varias opciones que ofrece el *software*; la eliminación de *reads* que

presentaban adaptadores u otras secuencias específicas de las bibliotecas, o bases que caen por debajo de cierto umbral de calidad ($Q < 30$), recorte mediante enfoque de ventana deslizante (en este caso seteado para que la región de la ventana sea de 4 nucleótidos), que actúa si la calidad promedio de la región queda por debajo del umbral determinado, y descarte de *reads* si son menores a un largo específico (en este caso fue de 150 pb).

Una vez obtenida la salida de Trimmomatic se empleó `fastq_to_fasta`, una herramienta que forma parte de la colección de herramientas de FASTX-Toolkit [42], que nos permite convertir un archivo en formato fastq a uno en formato fasta.

Luego se utilizaron dos *scripts* escritos en lenguaje *python* (datos no publicados), uno para filtrar *reads* con bases indeterminadas (N) y otro para combinar los pares de *reads*. Este último *script* se modificó ya que el original está elaborado para que funcione en archivos fastq, y en este caso se necesitaba utilizar sobre archivos fasta. También se utilizó un *script* escrito en lenguaje perl (datos no publicados), para intercalar las secuencias pareadas, colocando en un archivo multifasta primero el *read* correspondiente a un extremo de la biblioteca (comúnmente llamado R1) y luego el otro (R2). Este *script* también se modificó de manera que pueda funcionar en archivos fasta en lugar de fastq.

Análisis de la fracción repetida

A partir de los archivos resultantes de este proceso, se llevó a cabo el análisis de *clusters* basado en grafos utilizando RepeatExplorer y TAREAN, implementados dentro del entorno Galaxy (<https://repeatexplorer-elixir.cerit-sc.cz/>) [22,43]. Se subieron los datos a la plataforma RepeatExplorer con cada muestra por separado, así como utilizando un dataset con las muestras juntas (análisis comparativo), para luego procesarlas en el software RepeatExplorer. El análisis comparativo facilita la búsqueda de familias de repetidos compartidas entre las muestras, mientras que el análisis del genoma individual facilita la detección de familias de repetidos específicos de cada linaje, así como la determinación de los monómeros de cada familia de ADN satélite para cada muestra [22]. En ambos casos la proporción del genoma para cada *cluster*, o familia de ADN repetitivo, se calculó en base al

porcentaje de *reads* contenido en el *cluster*, en relación al total de *reads* analizados por el pipeline.

En primer lugar, se realizó el análisis comparativo. Una vez obtenidos los resultados de RepeatExplorer, se anotaron algunos de los *clusters* mediante la base de datos implementada dentro del *software*, de manera automática. Adicionalmente se utilizó una base de datos propia del laboratorio, que contiene satélites y elementos transponibles de *Rhodnius prolixus* y *Triatoma infestans*. Algunos *clusters* conteniendo satDNAs que quedaron sin anotar fueron identificados basados en la topología del grafo con forma de esfera o de anillo, y se analizaron utilizando Dotmatcher (disponible en <http://emboss.bioinformatics.nl/cgi-bin/emboss/dotmatcher/>) y/o YASS (<https://bioinfo.lifl.fr/yass/yass.php>). Para anotar otros *clusters* cuyo grafo no tenía forma aparente de ADN repetitivo, la clasificación se realizó por búsquedas de similitud de secuencias utilizando los *contigs* ensamblados por RE contra GenBank utilizando BlastN y BlastX (<http://www.ncbi.nlm.nih.gov/>), Repbase utilizando CENSOR (<http://www.girinst.org/>) y bases de datos propias del laboratorio utilizando blast de forma local.

Cuando se finalizó la anotación de los *clusters* del *dataset* del análisis comparativo, la misma se utilizó para realizar la anotación heredada en los *datasets* de las muestras individuales. Esto se llevó a cabo implementando blastn local de los *contigs* del análisis comparativo como *data base* sobre los *contigs* de las muestras individuales como *query*.

Al finalizar la anotación de todos los *datasets*, se utilizaron las hojas de cálculo de Google (Google Sheets) para calcular la cantidad de cada *cluster* en los genomas de cada individuo de ambos linajes, tanto el porcentaje que ocupa del genoma para cada muestra como la cantidad en Mega pares de bases (Mpb). Luego se agruparon los datos correspondientes a localidades de un mismo linaje para la comparación entre linajes norte y sur. Esto se realizó utilizando el valor C (1,046 pg en linaje norte y 1 pg en el linaje sur) de cada grupo de *T. patagonica* (datos no publicados) y la equivalencia de 1 pg de ADN a 978 Mpb.

Finalmente, con la cantidad de las distintas categorías de ADN repetitivo cuantificado para cada muestra, se procedió a realizar tests estadísticos de chi cuadrado, para determinar si las diferencias observadas de cada categoría de ADN

repetitivo son significativas entre las muestras analizadas. Para ejecutar los tests estadísticos se implementó el programa R [44].

Se realizó el mismo procedimiento de cuantificación de ADN y análisis estadístico para determinar diferencias entre las muestras en las cantidades de cada familia de ADN satélite identificado.

Identificación de las familias de ADN satélite

Para las familias de ADN satélite que fueron identificadas, se extrajo el número de *cluster* de archivo de la salida del RepeatExplorer y mediante la utilización de dos *scripts* desarrollados por nuestro grupo en lenguaje BASH, se extrajeron los *contigs* de mayor cobertura y de mayor largo de los *clusters* ya mencionados. Mediante el *software* Tandem Repeat Finder [45] incluido en uno de los *scripts* mencionados, se determinaron los monómeros para cada familia. Además de determinar la secuencia de los monómeros de los repetidos, se identificaron distintas características de los mismos, como el largo de los monómeros y se calculó el contenido A+T de los satélites mediante el uso de una aplicación web (<http://www.biologicscorp.com/tools/GCContent>).

Para los *clusters* que fueron anotados como ADN satélites pero que Tandem Repeat Finder no pudo generar información, se tomaron varios *contigs* del *cluster* para alinearlos usando MAFFT (<https://mafft.cbrc.jp/alignment/software/>) y generar una secuencia consenso a mano. También se utilizó en el proceso la herramienta YASS (<https://bioinfo.lifl.fr/yass/yass.php>).

Con el objetivo de detectar una marca filogenética y comparar los monómeros de los distintos linajes se empleó el *software* RepeatProfiler [46]. Este *software* es una herramienta que se emplea para el estudio de ADN repetitivo, comparando el perfil y la cobertura de distintas muestras para uno o varios repetidos de interés. Además, a partir de las variantes entre los perfiles de los repetidos en estudio se generan archivos que pueden ser utilizados para análisis filogenéticos.

Para utilizar RepeatProfiler, como secuencias de referencia se utilizaron los dímeros (o las repeticiones del monómero necesarias para alcanzar un mínimo de 200 pb) identificados de las familias de satélites obtenidas a partir del análisis por RepeatExplorer del análisis comparativo. Como base de datos para el análisis se

utilizaron muestras de *T. patagonica* correspondientes a las 4 localidades en estudio. Además, se realizó un segundo análisis donde se utilizaron secuenciaciones de 16 especies de triatomíneos disponibles en el laboratorio. Para este análisis se debió de utilizar una selección al azar de 1 millón de *reads* de cada genoma. Vale destacar que estas 16 especies pertenecen a los 3 principales linajes dentro de la tribu Tiatomini: linaje *T. dispar*, linaje Sudamérica y linaje Norteamérica.

Se utilizaron los archivos de salida en formato PHYLIP generados por RepeatProfiler de las familias de satélites que se pudieron identificar en todos los genomas (CL1 y CL212) para generar árboles filogenéticos (archivos de notación Newick) a través del *software* RAxML [47]. Estos archivos PHYLIP resumen la información hallada en forma de perfiles de variación mediante la identificación de variantes y su posterior codificación como caracteres moleculares-morfométricos de múltiples estados. En base a ellos, utilizando máxima verosimilitud, se calcula el árbol filogenético.

Para visualizar los árboles se subió el archivo Newick a la web iTOL (<https://itol.embl.de/>), una herramienta online para la visualización y anotación de árboles filogenéticos.

RESULTADOS

La extracción y secuenciación de ADN fue exitosa, es decir se obtuvo ADN genómico de alto peso molecular, en 4 individuos de la especie *T. patagonica* pertenecientes a 4 localidades distintas. Originalmente, también se había planeado analizar individuos de San Martín (San Luis) pero en este caso la extracción de ADN no fue satisfactoria, observándose mucha degradación en el gel de agarosa. Al no obtener ADN de buena calidad no fue posible realizar el análisis previsto probablemente debido al inadecuado estado de conservación de las muestras. Se confirmó que hubo degradación, y se detectó una cantidad superior a 1 microgramo de ADN total, en una concentración mayor a 100 ng/ul.

Se realizó el análisis de *clusters* basado en grafos para caracterizar y comparar secuencias de ADN altamente repetitivas, implementando el *software* RepeatExplorer, utilizando 3608114, 2945888, 3360332 y 4280952 reads para SE, SF, RN y LP respectivamente. En las muestras de la región norte, Santiago del Estero (SE) y Santa Fé (SF), el *clustering* de 231.975 y 212.252 *reads* produjo 219 y 206 *top clusters* respectivamente. Mientras que, en las muestras de la región del sur, La Pampa (LP) y Río Negro (RN), el *clustering* de 419.384 y 441.380 *reads* produjo 237 y 284 *top clusters* respectivamente, en todos los casos excluyendo *reads* que corresponden al genoma mitocondrial. Se pudo identificar que en los individuos de las regiones de SE y SF el repitoma ocupa una gran parte del genoma (36,1% en ambos casos), mientras que en los individuos de LP y RN ocupa una proporción menor (26,3% y 28,2% respectivamente). Los *clusters* se clasificaron en 6 categorías: *Long Terminal Repeats* (LTR), *non-Long Terminal Repeats* (NLTR), elementos clase II o transposones de ADN (Clase II), ADN satélite (satélite), ADN ribosomal (ADNr) y repetidos indeterminados (indeterminado). Esta última categoría incluye los *clusters* que no se pudieron asignar a ninguna categoría. Considerando el genoma haploide, la frecuencia y cantidad en Mega pares de bases (Mpb) de cada categoría se muestra en la tabla 1 del anexo. Las secuencias no repetidas de LP Y RN representan el 73,7% y el 71,8% del genoma, equivalentes a 709,78 y 667,26 Mpb por genoma haploide, respectivamente. Mientras que para SE y SF representan el 63,9% del genoma, equivalente a 705,19 Mpb y 705,01 Mpb por genoma haploide, respectivamente (Fig.2). No se encontraron diferencias

significativas en la cantidad de secuencias no repetidas (expresadas en Mpb) entre las muestras de las 4 localidades.

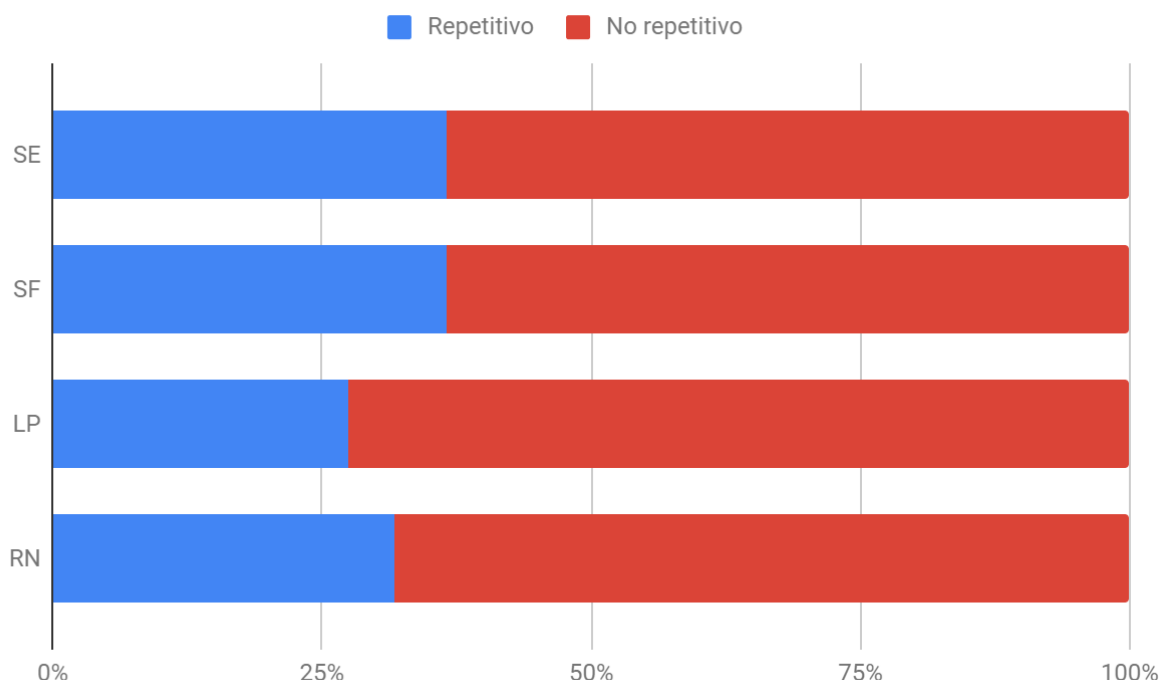


Fig. 2. Porcentaje que ocupa el ADN repetitivo y no repetitivo del genoma de *T. patagonica* en cada una de las muestras de las 4 localidades en estudio: Santiago del Estero (SE), Santa Fé (SF), La Pampa (LP) y Río Negro (RN).

En la muestra de SE se encontró que el ADN satélite ocupa un 12,7% del total del genoma, el ADN ribosomal un 0,3%, los transposones clase I un 11,4% (NLTR un 9,5% y LTR un 1,9%) y transposones clase II un 5,3%, mientras que hubo un 6,4% de elementos repetidos que no se pudo clasificar ([Fig. 3, a](#)). En SF el ADN satélite ocupa un 13,1% del total del genoma, el ADN ribosomal un 0,3%, los transposones clase I un 11% (NLTR un 9,2% y LTR un 1,8%) y transposones clase II un 5,4%, mientras que hubo un 6,3% de elementos repetidos que no se pudo clasificar ([Fig. 3, b](#)). En LP el ADN satélite ocupa un 8,5% del total del genoma, el ADN ribosomal un 0,3%, los transposones clase I un 7,7% (NLTR un 6,4% y LTR un 1,3%) y transposones clase II un 4,2%, mientras que hubo un 5,6% de elementos repetidos que no se pudo clasificar ([Fig. 3, c](#)). En RN el ADN satélite ocupa un 7,7% del total del genoma, el ADN ribosomal un 0,5 %, los transposones clase I un 8% (NLTR un 6,5 % y LTR un 1,5%) y transposones clase II un 3,8%, mientras que hubo un 8,3% de elementos repetidos que no se pudo clasificar ([Fig. 3, d](#)).

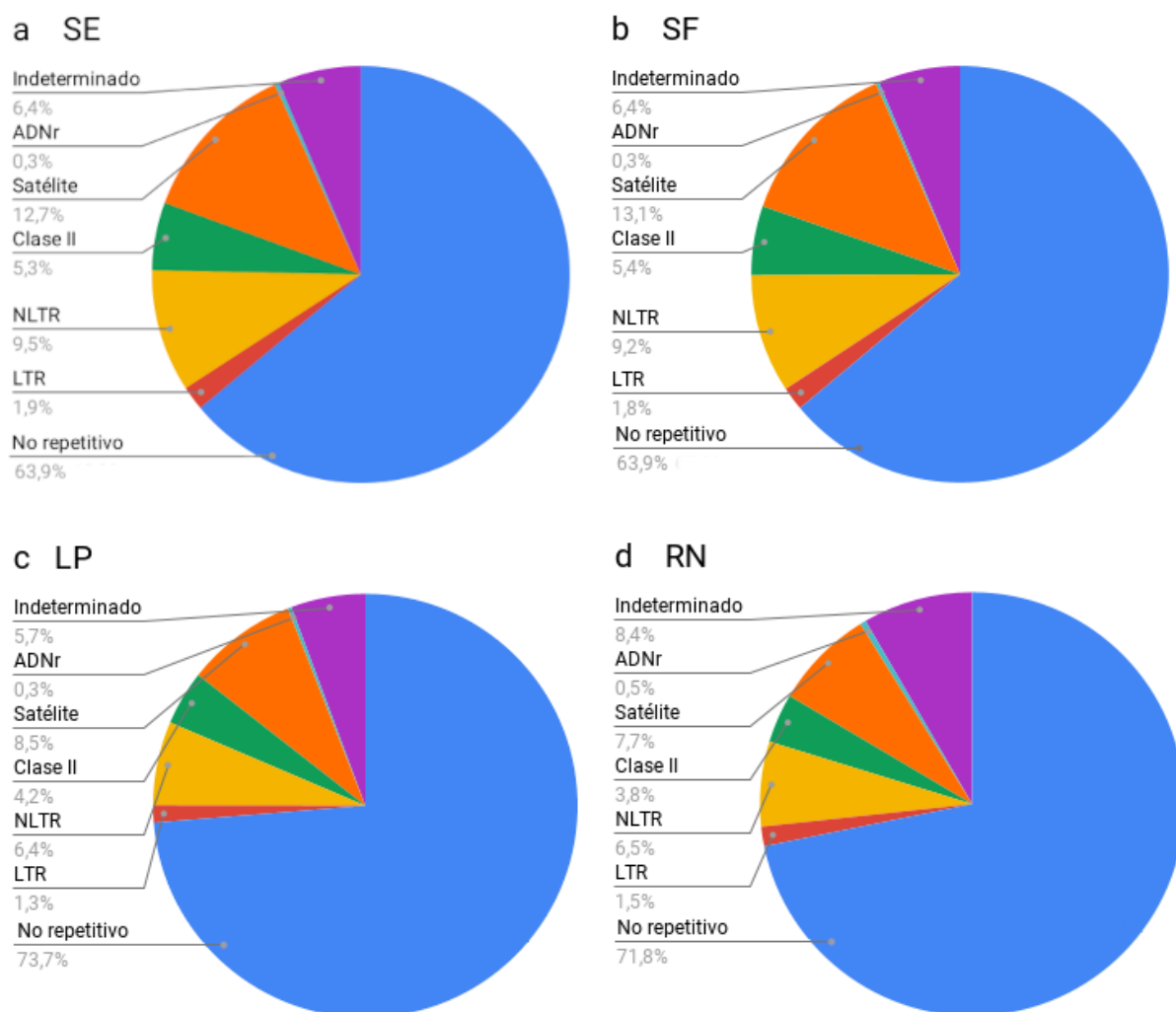


Fig. 3. Gráficos circulares muestran los porcentajes totales de cada categoría de ADN repetitivo en el genoma de los individuos pertenecientes a las cuatro localidades en estudio: **(a)** Santiago del Estero (SE) **(b)** Santa Fé (SF) **(c)** La Pampa (LP) **(d)** Río Negro (RN). En azul se representa el ADN no repetitivo. Long Terminal Repeats (LTR), non-Long Terminal Repeats (NLTR), elementos clase II o transposones de ADN (Clase II), ADN satélite (satélite), ADN ribosomal (ADNr) y repetidos indeterminados (indeterminado).

Mediante tests estadísticos de chi cuadrado se pudo observar que entre las muestras de los linajes norte y sur hay diferencias estadísticamente significativas en las cantidades en Mpb de ADN satélite ($p < 0,001$) y ADN correspondiente a NLTRs ($p < 0,01$). En el caso de los transposones clase II, cuando se compararon las muestras del norte (SE y SF) con RN se encontró una diferencia significativa ($p < 0,05$), pero no fue así cuando se compararon con LP. La cantidad de ADN correspondiente a LTRs y ADNr entre los individuos de las cuatro localidades no

fueron estadísticamente significativas (Fig. 4). En conclusión, la diferencia en la cantidad de ADN del repitoma entre los individuos del linaje de la región norte con los individuos de la región sur está dada por ADN satélite, NLTRs y en menor medida de transposones clase II.

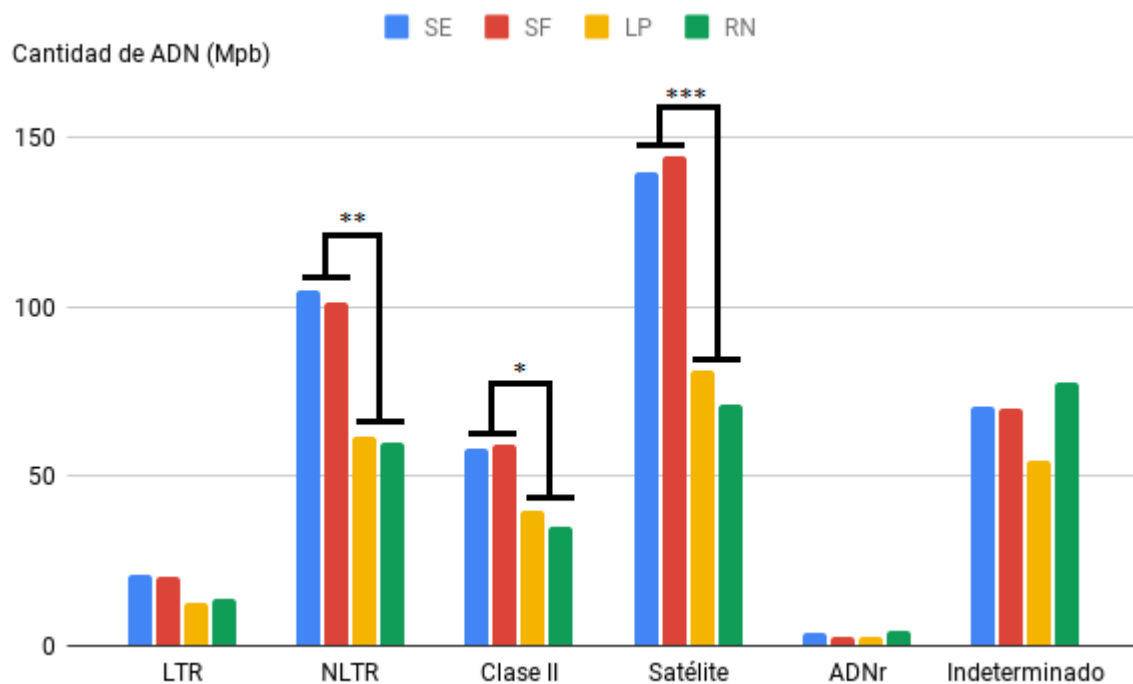


Fig. 4. Gráficas comparativas mostrando la cantidad de ADN para cada categoría de ADN repetido expresadas en mega pares de bases (Mpb) por genoma haploide de *T. patagonica* para cada uno de las 4 localidades estudiadas. Diferencias significativas por test de Chi-cuadrado están marcadas con asteriscos, se explica en la sección resultados. Santiago del Estero (SE), Santa Fé (SF), La Pampa (LP) y Río Negro (RN). Long Terminal Repeats (LTR), non-Long Terminal Repeats (NLTR), elementos clase II o transposones de ADN (Clase II), ADN satélite (satélite), ADN ribosomal (ADNr) y repetidos indeterminados (indeterminado).

Se pudieron caracterizar 11 familias de satélite en el análisis comparativo de RepeatExplorer, en la tabla 1 se detallan las mismas, número de bases del monómero, contenido A+T y abundancia en cada muestra. Los monómeros que se pudieron identificar se detallan en la tabla 2 del anexo. Se pudo apreciar que todas las familias de ADN satélite están presentes en ambos linajes, pero en distintas proporciones. Las 6 familias de satélites más frecuentes representan el 98,28% y 97,03% del total de ADN satélite en el linaje norte y sur respectivamente. Sin

embargo, la cantidad de cada familia de ADN satélite varía entre los dos linajes. Por ejemplo, el satélite CL11 es la segunda familia principal en el linaje del norte (1,64% del repitoma), pero no en el linaje del sur (solo un 0,19%). De los 11 satélites, se observó que sólo en los satélites CL1 y CL11 hay una diferencia estadísticamente significativa ($p < 0,01$) en la cantidad de genoma que ocupan entre el linaje norte y el linaje sur. Hay gran variación en relación a la longitud de los monómeros de los repetidos (de 4 a 1000 pb), sin embargo, se pudo identificar que solo 3 de las 11 familias de ADN satélite tienen un monómero mayor a 125 pb. El contenido A+T de las familias de ADN satélite varía entre 60% y 78%.

Tabla 1. Número de bases del monómero, contenido A + T y cuantificación en linajes norte y sur de las familias de ADN satélite identificadas en *Triatoma patagonica*. Santiago del Estero (SE), Santa Fé (SF), La Pampa (LP) y Río Negro (RN).

Satélite	N° de bases del monómero	Contenido A+T	Mpb (SE)	Mpb (SF)	Mpb (LP)	Mpb (RN)
CL1	1000	66%	102,76	103,88	64,18	55,53
CL11	79	64%	11,27	15,53	0,96	1,18
CL16	4	75%	6,31	7,44	3,46	2,57
CL18	88	78%	6,52	5,85	3,34	2,88
CL23	59	60%	4,43	4,48	4,01	3,81
CL24	124	72%	5,89	4,96	3,25	2,95
CL100	232	69%	0,76	0,73	0,33	0,38
CL134	531	61%	0,27	0,24	0,34	0,28
CL160	119	69%	0,13	0,17	0,22	0,17
CL212	84	65%	0,08	0,13	0,11	0,14
CL214	53	78%	0,15	0,11	0,07	0,13

Mediante RepeatProfiler se pudo comprobar que las familias de satélites CL1 y CL212 se encuentran presentes en los 20 genomas utilizados en el análisis para detectar una marca filogenética (Figs. 1 y 2, anexo). En ambos casos no se halló una marca filogenética distintiva de la especie *T. patagonica*, pero en el árbol

filogenético del satélite CL1 las 16 especies se agruparon en 3 linajes como era esperado (Norteamérica, Sudamérica y Dispar) ([Fig. 5](#)).

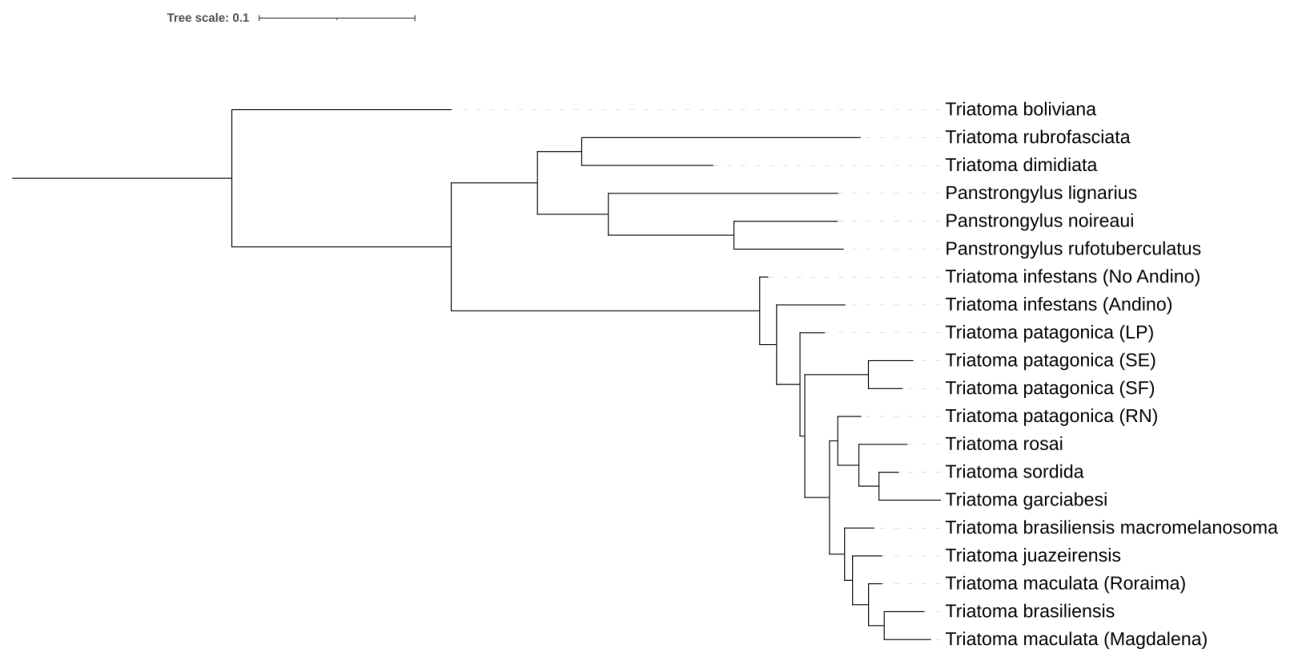


Fig. 5. Árbol filogenético inferido por máxima verosimilitud, utilizando caracteres moleculares-morfométricos, obtenido a partir del consenso del satélite CL1 de *T. patagonica*, y un subconjunto de un millón de *reads* crudos de 20 genomas de triatomini pertenecientes a los tres principales clados de la tribu Triatomini.

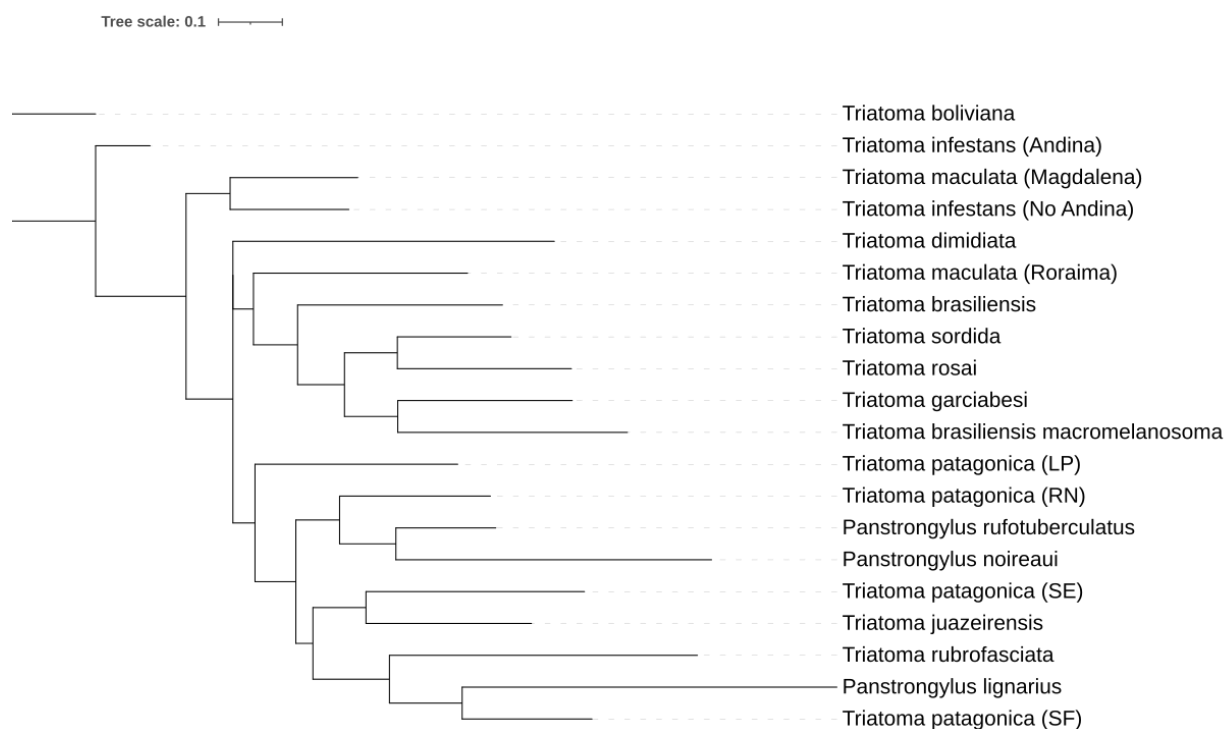


Fig. 6. Árbol filogenético inferido por máxima verosimilitud, utilizando caracteres moleculares-morfométricos, obtenido a partir del consenso satélite CL212 de *T. patagonica*, y un subconjunto de un millón de *reads* crudos de 20 genomas de triatomini pertenecientes a los tres principales clados de la tribu Triatomini.

DISCUSIÓN

El contenido de ADN repetitivo total (36,1% en el linaje norte y 27,3% en el linaje sur) está en el mismo rango que otras especies del orden Hemiptera, como el pulgón verde *Acyrtosiphon pisum* (33.3%) [48]. Sin embargo, los repitomas de *T. patagonica* de los 2 linajes analizados tienen un porcentaje similar de ADN satélite y de elementos transponibles (ETs). Esto se diferencia de otras especies como *N. lugens* y el pulgón verde, donde el mayor porcentaje del repitoma está ocupado por ETs (39% y 38% respectivamente). Otros insectos no-hemípteros con cromosomas holocéntricos como el lepidóptero *Bombyx mori* tiene en su genoma un 43.6% de secuencias altamente repetitivas, también compuesta principalmente por ETs [49].

Comparando con especies más cercanas, la proporción que ocupa el repitoma en el genoma de *T. patagonica* se encuentra en un valor intermedio entre los de especies como *T. infestans* (39%) [4] y *R. prolixus* (24%) [29]. A pesar de esto, hay varios aspectos en los que el genoma de *T. patagonica* presenta una mayor similitud con el de *T. infestans* que con *R. prolixus*. Cuando observamos el repitoma de los *T. infestans* y ambos linajes de *T. patagonica*, encontramos que el ADN satélite ocupa la mayor proporción del mismo [4]. Esto también ha sido observado en otras especies de la subfamilia Triatomini (datos no publicados). Sin embargo, este no es el caso en *R. prolixus* (tribu Rhodniini), donde la mayor proporción del repitoma está ocupado por elementos transponibles clase II. Otra semejanza que *T. patagonica* representa en relación a *T. infestans* es que el satelitoma (conjunto de familias de ADN satélite presentes en un genoma) está conformado por pocas familias de ADN satélite. En el caso particular de *T. patagonica*, la familia principal de satélite (CL1) representa el 72,70% en el linaje norte y 78,42 % en el linaje sur del total del satelitoma. A partir de otros estudios realizados en triatominos [29], donde se observa que esta misma familia de ADN está distribuida en la eucromatina, podríamos esperar que la localización se mantenga en *T. patagonica*. Todos estos datos evidencian una diferencia en la composición del genoma entre los géneros *Triatoma* y *Rhodnius*, lo que se ve reflejado en su distancia evolutiva y diferencias observadas en otros aspectos, incluso a nivel de citogenética [50]. A partir de los resultados podríamos apoyar la

hipótesis planteada que postula que *T. patagonica* se divide principalmente en dos linajes: norte y sur. Estos linajes muestran una diferencia de alrededor del 10% en el tamaño de sus genomas (1113 Mpb y 978 Mpb por genoma haploide respectivamente), que se correlaciona con una cantidad diferencial de heterocromatina-C [18]. Si bien ambos linajes tienen una cantidad similar de secuencias no repetitivas (sin diferencias estadísticamente significativas), los resultados obtenidos con RepeatExplorer evidenciaron que en los repitomas de *T. patagonica* se puede apreciar una diferencia significativa ($p < 0,01$) entre las cantidades de ADN repetitivo entre los linajes norte y sur (407,90 Mpb y 289,48 Mpb por genoma haploide, respectivamente) (Tabla 1, anexo), siendo éste el principal factor de la diferencia de tamaño de los genomas entre ambos linajes.

El análisis de las familias de ADN satélite muestra que están conservadas entre los dos linajes, ya que todas las familias de ADN satélite identificadas se encuentran en ambos. A pesar de esto, la razón de la diferencia en contenido de ADN repetitivo entre ambos linajes parece ser debida a la cantidad diferencial de estas familias en cada linaje (Tabla 1). Esto coincide con la hipótesis de la "biblioteca genómica", la cual predice que especies relacionadas comparten un conjunto ancestral de familias de ADN satélite, las cuales podrían estar amplificadas diferencialmente en cada especie debido a mecanismos estocásticos de evolución [51]. De hecho, la mayor parte de la diferencia de los genomas entre linajes es debida a solo 2 familias de ADN satélite (CL1 y CL11). Estas familias de ADN satélite constituyen 116,73 Mpb en el genoma del linaje norte y solo 60,93 Mpb en el genoma del linaje sur (Tabla 1). El CL11, segundo en abundancia, emerge como un posible candidato para ser el responsable de la heterocromatina autosómica presente únicamente en el linaje norte de *T. patagonica*. Esta hipótesis se apoya en dos motivos: i) que el CL1 es un satélite distribuido en la eucromatina en otras especies de *Triatoma*. ii) que el CL11 presenta mayor diferencia entre los linajes (15 veces más aproximadamente, mientras que CL1 es el doble aproximadamente). Cabe destacar que además de las dos familias de ADN satélite mencionadas se encontró que para la cantidad transposones NLTR hay una diferencia significativa entre los linajes, aunque afecta en menor medida al tamaño genómico total. A diferencia de los satélites, no se realizó un análisis particular para cada uno de los *clusters* correspondientes a transposones NLTR. A pesar de que la posible función de ADN repetitivo todavía es controvertida, el ADN satélite ha sido relacionado con

el aislamiento reproductivo y por lo tanto con la aparición de nuevas especies [52]. Su importancia en la integridad del genoma y la evolución cariotípica también ha sido resaltada [53]. El número de cromosomas en el género *Triatoma* está altamente conservado, indicando que la diferenciación entre especies parecería no estar acompañada por rearrreglos cromosómicos que alteren en número de cromosomas. Tampoco existe evidencia de la ocurrencia de otros reordenamientos cromosómicos estructurales, como inversiones o translocaciones, que no alteran el número de cromosomas. Los resultados aquí obtenidos muestran que el cambio principal entre ambos linajes (norte y sur) es la variación en la cantidad de ADN satélite. Estos cambios en la composición u organización del genoma podrían estar relacionados con la diversificación de especies en Triatominae u otros grupos de especies animales con baja variación cariotípica que altere el número cromosómico.

La definición de especie más utilizada es el llamado concepto biológico de especie, concebido por Ernst Mayr en 1942. Define a las especies como grupos de poblaciones naturales que se entrecruzan real o potencialmente, y están reproductivamente aisladas de otros grupos similares [54,55]. Se ha ampliado sobre esta definición, determinando que se pueden diferenciar dos especies cuando el cruzamiento entre individuos de los dos grupos no produce descendencia o produce descendencia infértil. En el caso de *T. patagonica* presenta una mayor complejidad, ya que al realizar cruzamientos entre individuos de poblaciones de diferentes regiones se obtuvo que un gran porcentaje no producía descendencia (40%) [18]. A partir de los resultados del cruzamiento, nos podríamos cuestionar si la definición de especie en cuestión es adecuada, o si habría que considerar otros elementos para poder determinar si dos individuos pertenecen o no a la misma especie, en casos como estos. Además, como se mencionó anteriormente, se pudo determinar que existe una diferencia citogenética entre individuos de las regiones norte y sur, específicamente en contenido de heterocromatina-C [18], que demostramos que está relacionada al ADN repetitivo [56,57]. Ha sido demostrado que los satélites son uno de los elementos del genoma que evoluciona más rápidamente, exhibiendo una considerable diferencia de contenido y distribución genómica entre especies cercanamente relacionadas [58,59]. La rápida expansión y divergencia del ADN satélite podría contribuir con procesos de aislación reproductiva y especiación, creando un contexto genético específico en híbridos interespecíficos y causando incompatibilidades entre híbridos [60]. En *T. patagonica*, podría suponerse que esta

divergencia genética contribuye a que la descendencia híbrida no sea completamente fértil, ya que uno de los mecanismos subyacentes de este fenómeno es la disrupción del apareamiento de cromosomas en la mitosis o en la meiosis. [61]. Con el paso del tiempo este proceso puede derivar en incompatibilidades al momento de copular, aumentando a su vez las diferencias genéticas y de esta manera consumando el proceso de especiación.

Los resultados obtenidos respecto a las diferencias en cantidad de ADN repetitivo entre los linajes en estudio se complementan con las diferencias ya reportadas anteriormente sobre aspectos citogenéticos, moleculares y a los experimentos de cruzamiento realizados en estas poblaciones, por lo que podríamos apoyar la hipótesis de que estamos frente a un proceso incipiente de especiación. El hecho de que estas poblaciones se encuentren en un proceso de especiación es particularmente relevante, debido a que se trata de un vector secundario de la enfermedad de Chagas en Argentina [18] que podría ser candidato a ser controlado con el fin de disminuir las incidencias de la enfermedad en la población. También es relevante considerando que recientemente ha aumentado la utilización de métodos para colapsar poblaciones de especies plagas a través de la tecnología de genética dirigida [62]. Siendo el flujo genético entre los individuos un componente fundamental de las campañas de control vectorial basadas fundamentalmente en el cruzamiento de individuos estériles con la población objetivo, es importante conocer la estructura de las poblaciones en cuestión para determinar cómo se relacionan entre sí a modo de que el empleo de esta tecnología sea efectiva y se pueda predecir acertadamente el efecto que tendrá sobre las mismas. [63]

A futuro, sería interesante llevar a cabo el mismo análisis con individuos de la región de San Luis (San Martín) para poder determinar si en el aspecto genómico se asemeja a las muestras del linaje sur, coincidiendo con lo ya hallado a nivel genético y morfológico [18]. Además, en el análisis genómico se podría evaluar si la diferencia en la cantidad de ADN del genoma de esta localidad también está dada por el ADN repetitivo y si los responsables de este aumento son las mismas categorías que ya observamos, o si hay variación en este aspecto. Por otro lado, para complementar los resultados hallados, se podrían mapear físicamente los satélites encontrados en ambos linajes mediante técnica FISH (hibridación

fluorescente *in situ*) para poder determinar cuál o cuáles son los responsables de la variación de heterocromatina-C entre los linajes.

Si bien no se pudo hallar una marca filogenética de la especie *T. patagonica* en los satélites CL1 y CL212, se podrían emplear variaciones en la metodología para obtener resultados más precisos. A pesar de esto, el hecho de que dos familias de ADN satélite se mantengan presentes en todas las especies analizadas es un hallazgo interesante, ya que podría sugerir que estas secuencias cumplen una función en los genomas. Se podría profundizar en este punto mediante el empleo de otro enfoque metodológico, como la evaluación de la transcripción de estas familias de ADN satélite, ya que en otros estudios se ha hallado transcripción de este tipo de ADN en especies relacionadas [29]. En caso de que efectivamente se transcriban, se podría hipotetizar que cumplen una función de regulación de genes o modulación de la cromatina, como ya ha sido descrito en *Drosophila* [61].

Conforme se caracterice el repitoma de más especies que no se hayan estudiado en profundidad genómicamente, se podrán establecer nuevas conclusiones sobre su importancia y el rol que juega en procesos evolutivos. A través de metodologías como la ofrecida por RepeatExplorer se facilita el estudio de estas regiones del genoma, previamente ignoradas o dejadas de lado por la dificultad de su estudio.

ANEXO

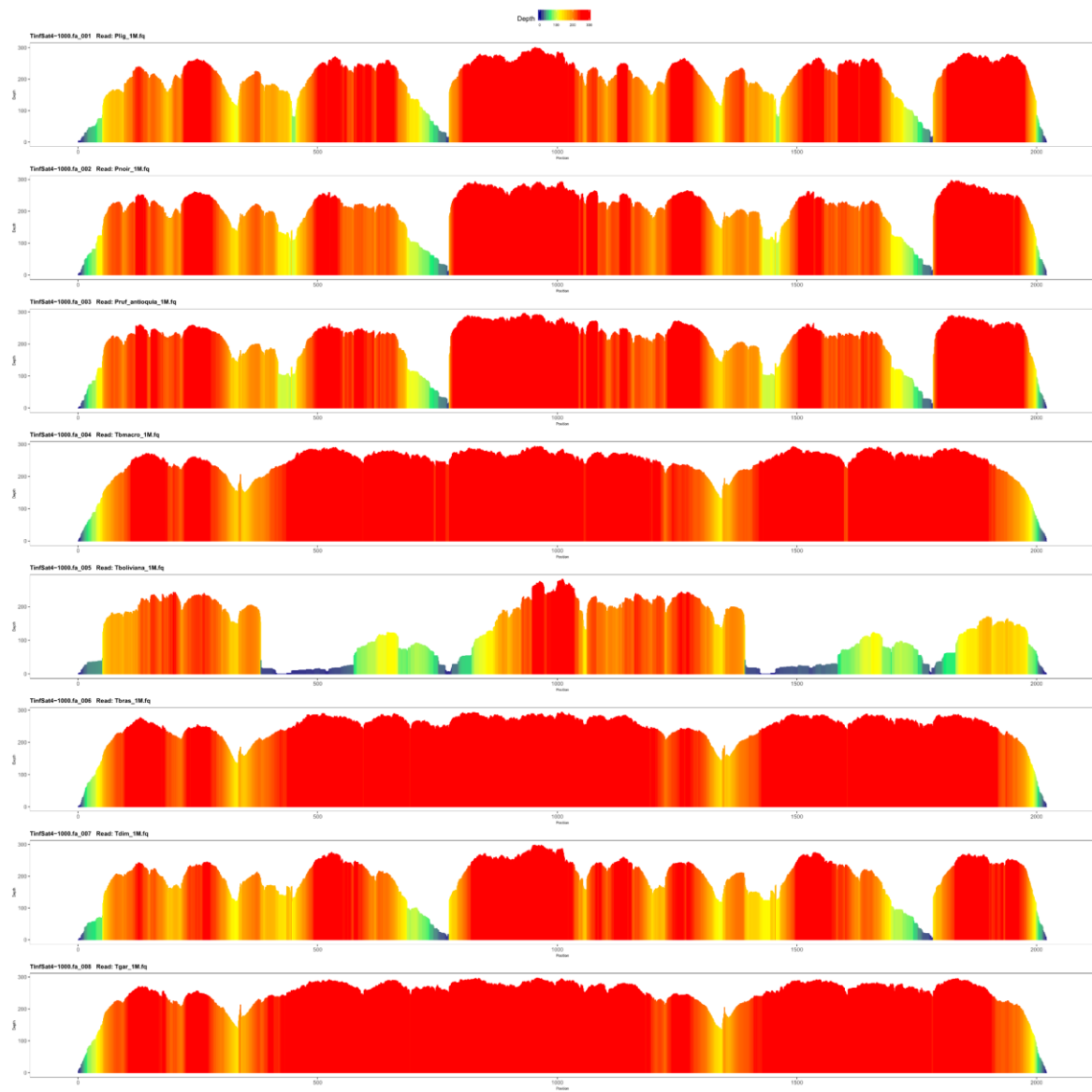
Tabla 1, anexo. Cuantificación de la cantidad de ADN en las categorías del repitoma de *T. patagonica* de las muestras de las 4 localidades estudiadas: Santiago del Estero (SE), Santa Fé (SF), La Pampa (LP) y Río Negro (RN). Long Terminal Repeats (LTR), non-Long Terminal Repeats (NLTR), elementos clase II o transposones de ADN (Clase II), ADN satélite (satélite), ADN ribosomal (ADNr) y repetidos indeterminados (indeterminado).

Muestra	SE (Mpb)	SF (Mpb)	LP (Mpb)	RN (Mpb)
Repetitivo	407,81	407,99	268,22	310,74
No repetitivo	705,19	705,01	709,78	667,26
LTR	20,71	20,08	12,59	13,83
NLTR	104,84	101,44	61,59	60,16
Clase II	57,96	59,23	39,96	34,89
Satélite	139,64	144,6	81,57	71,1
ADNr	3,71	2,8	2,76	4,21
Indeterminado	70,82	70,11	54,59	77,85
Genoma (pg)	1,046	1,046	1	1
Genoma (Mpb)	1113	1113	978	978

Tabla 2, anexo. Monómeros identificados de las familias de ADN satélite de *T. patagonica*.

Satélite	Monómero
CL1	ATCGTCGGAGCCATTTTTGAGAAAATCGCAAAAAAGTAAAAAAAAACAACTAGTAAAGTGGTA CTTCCGGTTGAGGAATTTTGACAGATAACGTACTGTCAGATCCCATGGTGATACCTAAACGGA ATATCAAGTTTCAACTTTCTACGGTTTTTCGTTTTTGAGCTATGCTGTTACATACATACACAC ACACACACACACACACACAGACCATTGCTAAAAACCACTTTTTTGGACTCAGGGGACCTC AAAACGGATATTTCCGGTGAAAACCTCGATATCGAAAATTTGACACGATTACAATACTTCTCTT ACTAGGAGTAAGAGAAAGTAAAAAAAAAAATCTGGTGTGAAACACTCACACAACCTTCTCTT ACTCCAGTTCTCAAAATTATAATTGACAAATTTTATAGTTTGTGTACCAAATTTAATGAAATTCG TCGAAAACCTGAAAAAACTGTTTCAGTAAAAGCGCACTTCCGGTTTACTAATTTTCTTGAAACT CGGAATTTTGCCATCTATTTATTTCTAATTAATTTTGTGGATGATAGAATTGTATAATCAAC TATGATAGCACTTTAGAGAAAAACATAACCCACCCCTCCCCTAAAAGTGCCCTTAAATGAAT TTTCCCAGAGAAATGTTTTAAATAAAAGTTGTAGATCTTTGTATGTTTAGTTTATATAAGAAGTT TCAAGAAAATCGTCGGAGCCATTTTTGAGAAAATCGCGAAAAAGTGAAAAA- TCATTTAGTAAAGTGTTAAAGTAAACGCACTTCCGGTTATCCGATTTTTTTCAAACCTCGGAATT TCAACAACATATTTAATTCTAATTAACATTATTTAGATGACAGAATTGTCTAATAAACTGAAATA GCACTTTAGATGAAAACCTAACCCACCCCTCCCCAAAAGTGCCCTTAAATGAATTTTCCCA GAGAAATGTTTCAAATAAAAGTTGTAGCTCTTTGTATGTGTAGTTGATATACCAAGTTTGAAGA AA
CL11	GTGAGTGTTGTACGAAAATTTACTCTTCCGGTTTAAACGCCGTATTTAGTGATTTTCTTGAACA TTACACCCTACTGTG
CL16	GATA
CL18	AAGTCGAGGGCTAAGTTTTTATAATATTTAGCTACAAACATTAAATATATTATCATCCTATACTA TGAATAAAGACTTTGTTTTTTTT
CL23	AGAACCGCCGCAAAACCCTTTTTGTATTCTCTTCAAACAAGGGTCAATATACACTCAAG
CL24	ATTTAGCTAAACACTTTAATATTTTTTTGTTTCATGTTTACACTTATAATATTTTAGATAAAATACAT TTTAAGCACTTTAACATTGTTTTTGAGCCGGAAGTGGCATCCAAGTAGAGGGCCAAT
CL100	TTTCGAAAAGTTAGAGGTGATTTTCAGCCAGAGAAGAAATTAGGTAAGCATTATACCAGAAAC TGGATTATTTAAAAAAGTGTTATACTTGTAATATAACAAGAATTGTTTTCTCAAAACATTTAG GTTGTAGAGAAAAGCCCAAGGTCCACATTACTTGCACGAAAGCAACTAATGTGTAGATAAAT TAAGGTCCACGTTACATTGCTCGAAATCGATATTTTCGATA
CL134	ATTAACAAAGATATAAAGAAAAAAAAGTGGCCAGTACAACCTTGACGTCAAGTTAGACTGGCCA CCAATTTATATTCACCTTTTTTAACTTTTCTCTTGCCAATCGTTAGTGCTGCGTTTGTGCCGG CGATTCACCAATTTTCATCGTTTGTGCCGTAAAATTGACAAAGATATTGACGAAAACTGTTT GTCGTAATTGAAAACCGAAGGCCATGGCTGCGGGTAATATTTGAAAGTGCGCGTGCTCCACT GACCACCAACGGCCCTTATCAGGGCCAGCAATTAATGTACTGGTCCTCGATTTCTAATCAAA TTTCTGGAATTTTTCTTCGCTAATCTTAATGGATTCAATTTGTGCCAGCGACCCACCAATTTT AGCGTTTGTGCTGCAAAATTGACAAAGTTATTGAGGCGTCAAGTTAGACTGGCCACCAATTTA TATTCACTTTTTTAACTTTTCTCTTTACCAATCGTTAGTGCTGCGTTTGTGCCGGCGATTAC CAAATTTTCATCGTTTGTGCCATAAA
CL160	TAAACACCCACTACCAAGGTTAGAAATTTTACTTTTACTTCCAACCCAATCCAACCTAATG AAGTTTTGCGGAATGAGATTCCATTTTCTTGATTTAAAAAGTATTATCCTTTAAA
CL212	ATATGTCTTTTCAAATGTCCAGATTGTGTAAAACCTGAAATCACATTCACATTATACGGCT

	TCTCACCTGTATGTGTTTCGC
CL214	AAAGTTCTACATTCTAATACTAATAAGATGGAAATATGCAATAATGATGAA



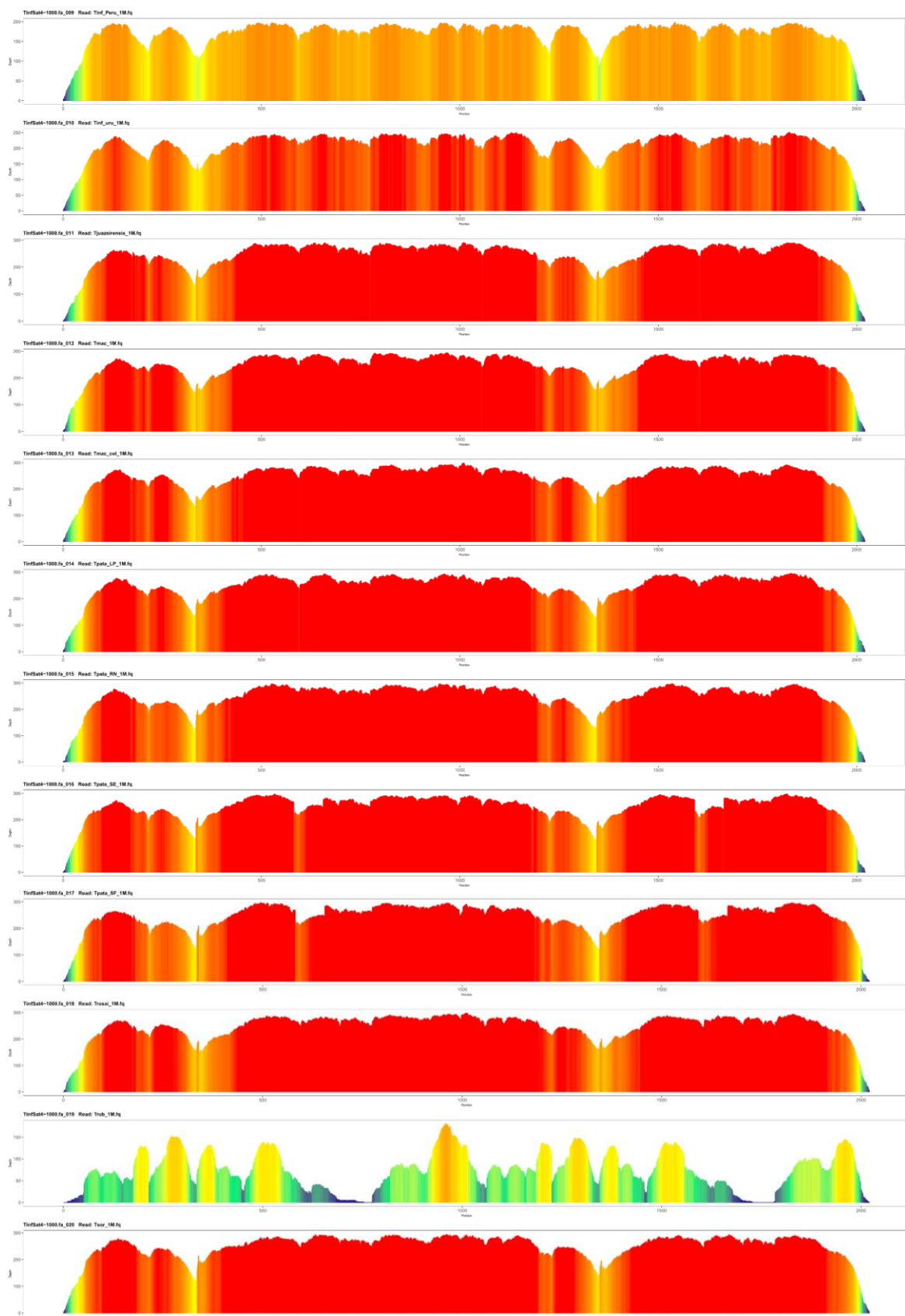
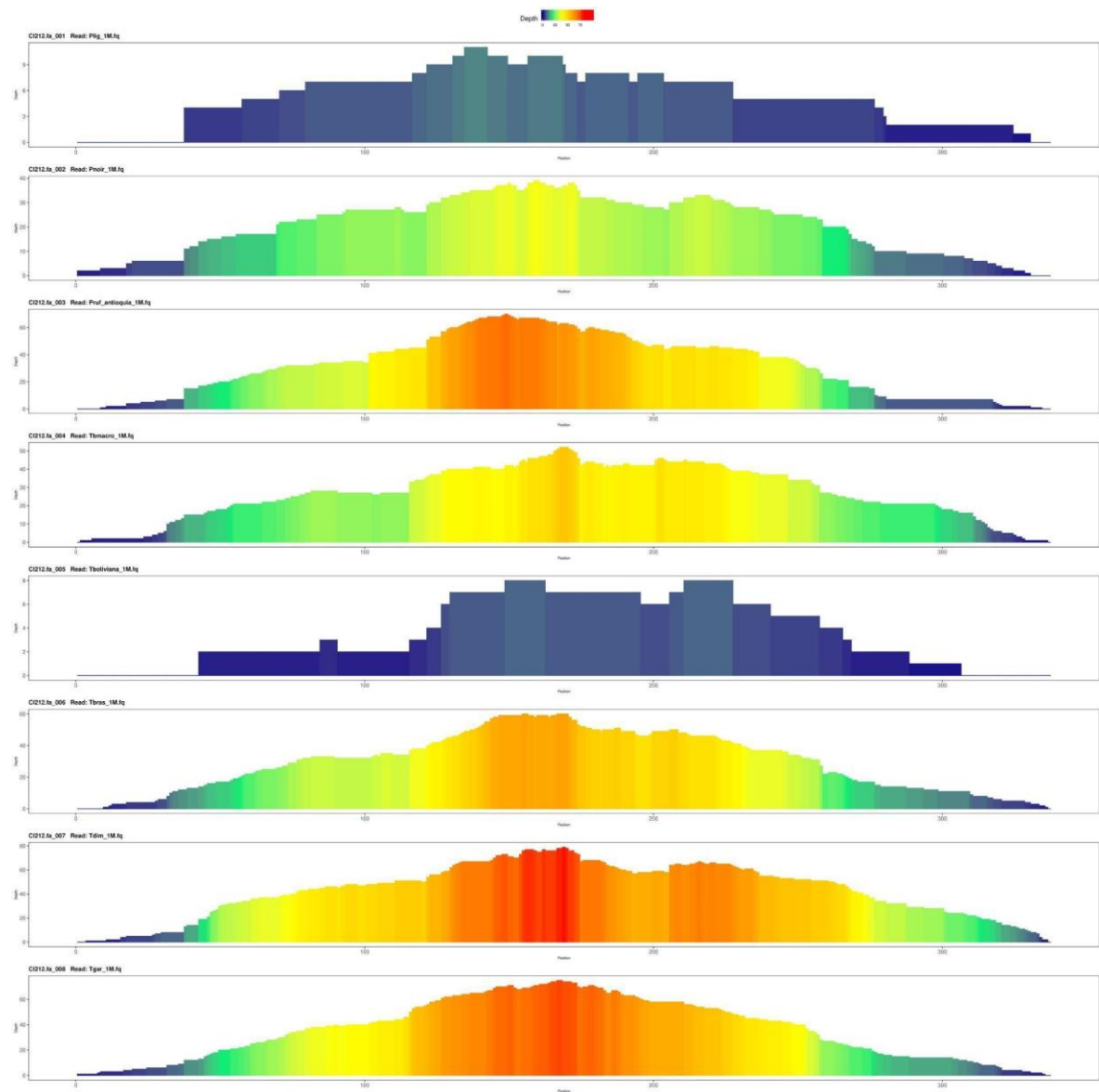


Fig. 1, anexo. Perfiles del satélite CL1 obtenidos a partir del análisis con RepeatProfiler, evidenciando la presencia de esta familia de satélite en todos los genomas analizados.



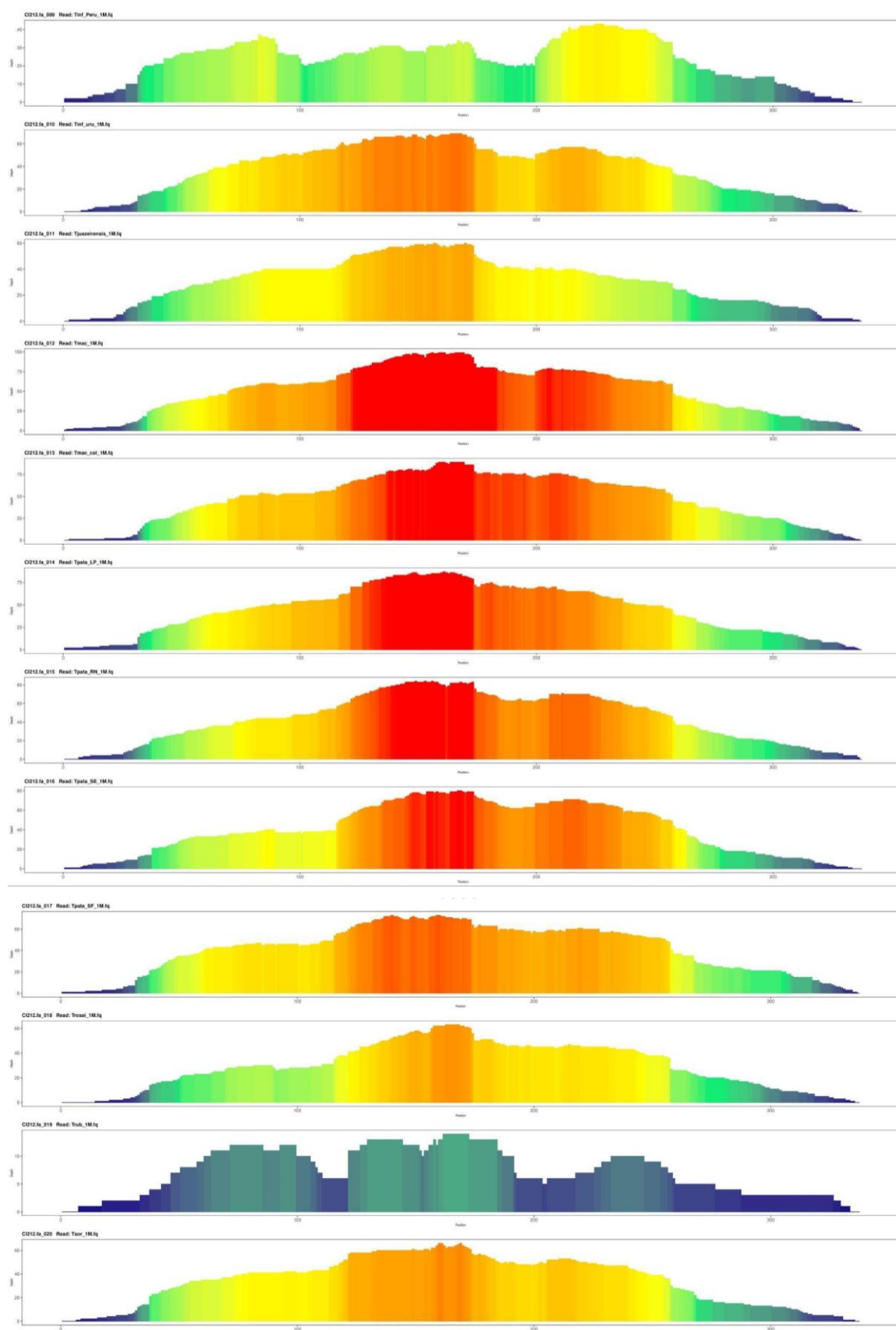


Fig. 2, anexo. Perfiles del satélite CL212 obtenido a partir del análisis con RepeatProfiler, evidenciando la presencia de esta familia de satélite en todos los genomas analizados.

BIBLIOGRAFÍA

1. Justi, S. A., & Galvão, C. (2017). The Evolutionary Origin of Diversity in Chagas Disease Vectors. *Trends in parasitology*, 33(1), 42–52. <https://doi.org/10.1016/j.pt.2016.11.002>
2. Monteiro, F. A., Weirauch, C., Felix, M., Lazoski, C., & Abad-Franch, F. (2018). Evolution, Systematics, and Biogeography of the Triatominae, Vectors of Chagas Disease. *Advances in parasitology*, 99, 265–344. <https://doi.org/10.1016/bs.apar.2017.12.002>
3. Schofield, C. J., & Galvão, C. (2009). Classification, evolution, and species groups within the Triatominae. *Acta tropica*, 110(2-3), 88–100. <https://doi.org/10.1016/j.actatropica.2009.01.010>
4. Pita, S., Lorite, P., Nattero, J., Galvão, C., Alevi, K. C., Teves, S. C., Azeredo-Oliveira, M. T., & Panzera, F. (2016). New arrangements on several species subcomplexes of *Triatoma* genus based on the chromosomal position of ribosomal genes (Hemiptera - Triatominae). *Infection, genetics and evolution : journal of molecular epidemiology and evolutionary genetics in infectious diseases*, 43, 225–231. <https://doi.org/10.1016/j.meegid.2016.05.028>
5. Schofield, C. J., Jannin, J., & Salvatella, R. (2006). The future of Chagas disease control. *Trends in parasitology*, 22(12), 583–588. <https://doi.org/10.1016/j.pt.2006.09.011>
6. Rassi, A., Jr, Rassi, A., & Marin-Neto, J. A. (2010). Chagas disease. *Lancet* (London, England), 375(9723), 1388–1402. [https://doi.org/10.1016/S0140-6736\(10\)60061-X](https://doi.org/10.1016/S0140-6736(10)60061-X)
7. Gourbière, S., Dorn, P., Tripet, F., & Dumonteil, E. (2012). Genetics and evolution of triatomines: from phylogeny to vector control. *Heredity*, 108(3), 190–202. <https://doi.org/10.1038/hdy.2011.71>
8. Lent, H., Wygodzinsky, P. (1979). Revision of the Triatominae (Hemiptera, Reduviidae) and their significance as vector of Chagas disease. *Bull Am Mus Nat Hist* 163,: 123–520.
9. Ferrero, A. A., Visciarelli, E. C., Torno, O., & Costamagna, S. R. (1999). Presencia de *Triatoma patagonica* en viviendas humanas en la ciudad de Río Colorado, provincia de Río Negro. *Revista de la Sociedad Entomológica Argentina*, 58(3-4).
10. Giraldez, E., López Ureta, M. P., Nepote, M. (2009). Colonización domiciliaria por *Triatoma patagonica* en el departamento San Cristóbal, provincia de

11. Nattero, J., Crocco, L. B., & Rodriguez, C. S. (2002). Feeding and defaecation behaviour of *Triatoma patagonica* (Del Ponte, 1929) (Hemiptera: Reduviidae). *Memorias do Instituto Oswaldo Cruz*, 97(7), 1063–1065. <https://doi.org/10.1590/s0074-02762002000700022>
12. Rodríguez, C. S., Carrizo, S. A., & Crocco, L. B. (2008). Comparison of feeding and defecation patterns between fifth-instar nymphs of *Triatoma patagonica* (Del Ponte, 1929) and *Triatoma infestans* (Klug, 1934) under laboratory conditions. *Revista da Sociedade Brasileira de Medicina Tropical*, 41(4), 330–333. <https://doi.org/10.1590/s0037-86822008000400002>
13. Dujardin, J. P., Costa, J., Bustamante, D., Jaramillo, N., & Catalá, S. (2009). Deciphering morphology in Triatominae: the evolutionary signals. *Acta tropica*, 110(2-3), 101–111. <https://doi.org/10.1016/j.actatropica.2008.09.026>
14. Hypsa, V., Tietz, D. F., Zrzavý, J., Rego, R. O., Galvao, C., & Jurberg, J. (2002). Phylogeny and biogeography of Triatominae (Hemiptera: Reduviidae): molecular evidence of a New World origin of the Asiatic clade. *Molecular phylogenetics and evolution*, 23(3), 447–457. [https://doi.org/10.1016/s1055-7903\(02\)00023-4](https://doi.org/10.1016/s1055-7903(02)00023-4)
15. Monteiro, F. A., Barrett, T. V., Fitzpatrick, S., Cordon-Rosales, C., Feliciangeli, D., & Beard, C. B. (2003). Molecular phylogeography of the Amazonian Chagas disease vectors *Rhodnius prolixus* and *R. robustus*. *Molecular ecology*, 12(4), 997–1006. <https://doi.org/10.1046/j.1365-294x.2003.01802.x>
16. Calleros, L., Panzera, F., Bargues, M. D., Monteiro, F. A., Klisiowicz, D. R., Zuriaga, M. A., Mas-Coma, S., & Pérez, R. (2010). Systematics of *Mepraia* (Hemiptera-Reduviidae): cytogenetic and molecular variation. *Infection, genetics and evolution : journal of molecular epidemiology and evolutionary genetics in infectious diseases*, 10(2), 221–228. <https://doi.org/10.1016/j.meegid.2009.12.002>
17. Justi, S. A., Russo, C. A., Mallet, J. R., Obara, M. T., & Galvão, C. (2014). Molecular phylogeny of Triatomini (Hemiptera: Reduviidae: Triatominae). *Parasites & vectors*, 7, 149. <https://doi.org/10.1186/1756-3305-7-149>
18. Nattero, J., Pita, S., Calleros, L., Crocco, L., Panzera, Y., Rodríguez, C. S., & Panzera, F. (2016). Morphological and Genetic Differentiation within the Southernmost Vector of Chagas Disease: *Triatoma patagonica* (Hemiptera - Reduviidae). *PloS one*, 11(12), e0168853. <https://doi.org/10.1371/journal.pone.0168853>
19. Panzera, F., Pérez, R., Panzera, Y., Ferrandis, I., Ferreira, M. J., & Calleros, L. (2010). Cytogenetics and genome evolution in the subfamily Triatominae (Hemiptera, Reduviidae). *Cytogenetic and genome research*, 128(1-3), 77–

20. Panzera, F., Dujardin, J. P., Nicolini, P., Caraccio, M. N., Rose, V., Tellez, T., Bermúdez, H., Bargues, M. D., Mas-Coma, S., O'Connor, J. E., & Pérez, R. (2004). Genomic changes of Chagas disease vector, South America. *Emerging infectious diseases*, 10(3), 438–446. <https://doi.org/10.3201/eid1003.020812>
21. Panzera, F., Ferreiro, M. J., Pita, S., Calleros, L., Pérez, R., Basmadjian, Y., Guevara, Y., Brenière, S. F., & Panzera, Y. (2014). Evolutionary and dispersal history of *Triatoma infestans*, main vector of Chagas disease, by chromosomal markers. *Infection, genetics and evolution : journal of molecular epidemiology and evolutionary genetics in infectious diseases*, 27, 105–113. <https://doi.org/10.1016/j.meegid.2014.07.006>
22. Novák, P., Neumann, P., & Macas, J. (2010). Graph-based clustering and characterization of repetitive sequences in next-generation sequencing data. *BMC bioinformatics*, 11, 378. <https://doi.org/10.1186/1471-2105-11-378>
23. Heckmann, S., Macas, J., Kumke, K., Fuchs, J., Schubert, V., Ma, L., Novák, P., Neumann, P., Taudien, S., Platzer, M., & Houben, A. (2013). The holocentric species *Luzula elegans* shows interplay between centromere and large-scale genome organization. *The Plant journal : for cell and molecular biology*, 73(4), 555–565. <https://doi.org/10.1111/tpj.12054>
24. Ruiz-Ruano, F. J., López-León, M. D., Cabrero, J., & Camacho, J. P. M. (2016). High-throughput analysis of the satellitome illuminates satellite DNA evolution. *Scientific reports*, 6, 28333. <https://doi.org/10.1038/srep28333>
25. 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(4), e94101. <https://doi.org/10.1371/journal.pone.0094101>
26. Altemose, N., Miga, K. H., Maggioni, M., & Willard, H. F. (2014). Genomic characterization of large heterochromatic gaps in the human genome assembly. *PLoS computational biology*, 10(5), e1003628. <https://doi.org/10.1371/journal.pcbi.1003628>
27. Biscotti, M. A., Olmo, E., & Heslop-Harrison, J. S. (2015). Repetitive DNA in eukaryotic genomes. *Chromosome research : an international journal on the molecular, supramolecular and evolutionary aspects of chromosome biology*, 23(3), 415–420. <https://doi.org/10.1007/s10577-015-9499-z>
28. 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(7), e0181635. <https://doi.org/10.1371/journal.pone.0181635>

29. Montiel, E. E., Panzera, F., Palomeque, T., Lorite, P., & Pita, S. (2021). Satellitome Analysis of *Rhodnius prolixus*, One of the Main Chagas Disease Vector Species. *International journal of molecular sciences*, 22(11), 6052. <https://doi.org/10.3390/ijms22116052>
30. Ruiz-Ruano, F. J., Navarro-Domínguez, B., Camacho, J. P. M., Garrido-Ramos, M. A. (2019) Characterization of the satellitome in lower vascular plants: the case of the endangered fern *Vandenboschia speciosa*. *Annals of Botany*. *Annals of Botany*. 123. 587–599. 10.1093/aob/mcy192.
31. Belyayev, A., Jandová, M., Josefiová, J., Kalendar, R., Mahelka, V., Mandák, B., & Krak, K. (2020). The major satellite DNA families of the diploid *Chenopodium album* aggregate species: Arguments for and against the "library hypothesis". *PloS one*, 15(10), e0241206. <https://doi.org/10.1371/journal.pone.0241206>
32. Sader, M., Vaio, M., Cauz-Santos, L. A., Dornelas, M. C., Vieira, M. L. C., Melo, N., & Pedrosa-Harand, A. (2021). Large vs small genomes in *Passiflora*: the influence of the mobilome and the satellitome. *Planta*, 253(4), 86. <https://doi.org/10.1007/s00425-021-03598-0>
33. Heitkam, T., Schulte, L., Weber, B., Liedtke, S., Breitenbach, S., Kögler, A., Morgenstern, K., Brückner, M., Tröber, U., Wolf, H., Krabel, D., & Schmidt, T. (2021). Comparative Repeat Profiling of Two Closely Related Conifers (*Larix decidua* and *Larix kaempferi*) Reveals High Genome Similarity With Only Few Fast-Evolving Satellite DNAs. *Frontiers in genetics*, 12, 683668. <https://doi.org/10.3389/fgene.2021.683668>
34. Utsunomia, R., Ruiz-Ruano, F. J., Silva, D. M. Z. A., Serrano, É. A., Rosa, I. F., Scudeler, P. E. S., Hashimoto, D. T., Oliveira, C., Camacho, J. P. M., & Foresti, F. (2017). A Glimpse into the Satellite DNA Library in Characidae Fish (Teleostei, Characiformes). *Frontiers in genetics*, 8, 103. <https://doi.org/10.3389/fgene.2017.00103>
35. 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. (2019). Satellitome landscape analysis of *Megaleporinus macrocephalus* (Teleostei, Anostomidae) reveals intense accumulation of satellite sequences on the heteromorphic sex chromosome. *Scientific reports*, 9(1), 5856. <https://doi.org/10.1038/s41598-019-42383-8>
36. Dos Santos, R. Z., Calegari, R. M., Silva, D. M. Z. A., Ruiz-Ruano, F. J., Melo, S., Oliveira, C., Foresti, F., Uliano-Silva, M., Porto-Foresti, F., & Utsunomia, R. (2021). A Long-Term Conserved Satellite DNA That Remains Unexpanded

in Several Genomes of Characiformes Fish Is Actively Transcribed. *Genome biology and evolution*, 13(2), evab002. <https://doi.org/10.1093/gbe/evab002>

37. Andrews, S. (2010). FastQC: A Quality Control Tool for High Throughput Sequence Data [Online]. Available online at: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>
38. Schmieder, R., & Edwards, R. (2011). Fast identification and removal of sequence contamination from genomic and metagenomic datasets. *PloS one*, 6(3), e17288. <https://doi.org/10.1371/journal.pone.0017288>
39. 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. *Infection, genetics and evolution : journal of molecular epidemiology and evolutionary genetics in infectious diseases*, 54, 158–163. <https://doi.org/10.1016/j.meegid.2017.06.003>
40. Dotson, E. M., & Beard, C. B. (2001). Sequence and organization of the mitochondrial genome of the Chagas disease vector, *Triatoma dimidiata*. *Insect molecular biology*, 10(3), 205–215. <https://doi.org/10.1046/j.1365-2583.2001.00258.x>
41. Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics (Oxford, England)*, 30(15), 2114–2120. <https://doi.org/10.1093/bioinformatics/btu170>
42. Hannon, G.J. (2010) FASTX-Toolkit. http://hannonlab.cshl.edu/fastx_toolkit.
43. 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 (Oxford, England)*, 29(6), 792–793. <https://doi.org/10.1093/bioinformatics/btt054>
44. R Core Team (2020) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.r-project.org/>
45. Benson G. (1999). Tandem repeats finder: a program to analyze DNA sequences. *Nucleic acids research*, 27(2), 573–580. <https://doi.org/10.1093/nar/27.2.573>

46. Negm, S., Greenberg, A., Larracuenta, A. M., & Sproul, J. S. (2021). RepeatProfiler: A pipeline for visualization and comparative analysis of repetitive DNA profiles. *Molecular ecology resources*, 21(3), 969–981. <https://doi.org/10.1111/1755-0998.13305>
47. Stamatakis A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* (Oxford, England), 30(9), 1312–1313. <https://doi.org/10.1093/bioinformatics/btu033>
48. International Aphid Genomics Consortium (2010). Genome sequence of the pea aphid *Acyrtosiphon pisum*. *PLoS biology*, 8(2), e1000313. <https://doi.org/10.1371/journal.pbio.1000313>
49. Xia, Q., Zhou, Z., Lu, C., Cheng, D., Dai, F., Li, B., Zhao, P., Zha, X., Cheng, T., Chai, C., Pan, G., Xu, J., Liu, C., Lin, Y., Qian, J., Hou, Y., Wu, Z., Li, G., Pan, M., Li, C., ... Biology Analysis Group (2004). A draft sequence for the genome of the domesticated silkworm (*Bombyx mori*). *Science* (New York, N.Y.), 306(5703), 1937–1940. <https://doi.org/10.1126/science.1102210>
50. Pita, S., Panzera, F., Sánchez, A., Panzera, Y., Palomeque, T., & Lorite, P. (2014). Distribution and evolution of repeated sequences in genomes of Triatominae (Hemiptera-Reduviidae) inferred from genomic in situ hybridization. *PloS one*, 9(12), e114298. <https://doi.org/10.1371/journal.pone.0114298>
51. Fry, K., & Salser, W. (1977). Nucleotide sequences of HS-alpha satellite DNA from kangaroo rat *Dipodomys ordii* and characterization of similar sequences in other rodents. *Cell*, 12(4), 1069–1084. [https://doi.org/10.1016/0092-8674\(77\)90170-2](https://doi.org/10.1016/0092-8674(77)90170-2)
52. Hughes, S. E., & Hawley, R. S. (2009). Heterochromatin: a rapidly evolving species barrier. *PLoS biology*, 7(10), e1000233. <https://doi.org/10.1371/journal.pbio.1000233>
53. Biscotti, M. A., Olmo, E., & Heslop-Harrison, J. S. (2015). Repetitive DNA in eukaryotic genomes. *Chromosome research : an international journal on the molecular, supramolecular and evolutionary aspects of chromosome biology*, 23(3), 415–420. <https://doi.org/10.1007/s10577-015-9499-z>
54. Mayr, Ernst. (1942). *Systematics and the origin of species from the viewpoint of a zoologist*. New York : Columbia University Press
55. de Queiroz K. (2005). Ernst Mayr and the modern concept of species. *Proceedings of the National Academy of Sciences of the United States of*

56. Panzera, F., Ferrandis, I., Ramsey, J., Salazar-Schettino, P. M., Cabrera, M., Monroy, C., Bagues, M. D., Mas-Coma, S., O'Connor, J. E., Angulo, V. M., Jaramillo, N., & Pérez, R. (2007). Genome size determination in chagas disease transmitting bugs (hemiptera-triatominae) by flow cytometry. *The American journal of tropical medicine and hygiene*, 76(3), 516–521.
57. Panzera, F., Pérez, R., Panzera, Y., Ferrandis, I., Ferreiro, M. J., & Calleros, L. (2010). Cytogenetics and genome evolution in the subfamily Triatominae (Hemiptera, Reduviidae). *Cytogenetic and genome research*, 128(1-3), 77–87. <https://doi.org/10.1159/000298824>
58. Garrido-Ramos M. A. (2015). Satellite DNA in Plants: More than Just Rubbish. *Cytogenetic and genome research*, 146(2), 153–170. <https://doi.org/10.1159/000437008>
59. Garrido-Ramos M. A. (2017). Satellite DNA: An Evolving Topic. *Genes*, 8(9), 230. <https://doi.org/10.3390/genes8090230>
60. Jagannathan, M., & Yamashita, Y. M. (2021). Defective Satellite DNA Clustering into Chromocenters Underlies Hybrid Incompatibility in *Drosophila*. *Molecular biology and evolution*, 38(11), 4977–4986. <https://doi.org/10.1093/molbev/msab221>
61. Shatskikh, A. S., Kotov, A. A., Adashev, V. E., Bazylev, S. S., & Olenina, L. V. (2020). Functional Significance of Satellite DNAs: Insights From *Drosophila*. *Frontiers in cell and developmental biology*, 8, 312. <https://doi.org/10.3389/fcell.2020.00312>
62. Bier E. (2022). Gene drives gaining speed. *Nature reviews. Genetics*, 23(1), 5–22. <https://doi.org/10.1038/s41576-021-00386-0>
63. Dhole, S., Lloyd, A. L., & Gould, F. (2020). Gene Drive Dynamics in Natural Populations: The Importance of Density Dependence, Space, and Sex. *Annual review of ecology, evolution, and systematics*, 51(1), 505–531. <https://doi.org/10.1146/annurev-ecolsys-031120-101013>