Universidad de Buenos Aires

Facultad de Ciencias Exactas y Naturales

Tesis presentada para optar al grado de Licenciada en Ciencias Biológicas

"Análisis de la conservación de la actividad pleiotrópica de *enhancers* a escala genómica entre *Drosophila melanogaster* y *Drosophila virilis*"

Autora: Ailen Altamirano
Director: Dr. Nicolás Frankel
Director Asistente: Lic. Ian Laiker

Lugar de trabajo: Instituto de Fisiologia, Biologia Molecular y Neurociencias (IFIByNE, CONICET-UBA)
Grupo de Evolución y Desarrollo

Marzo de 2024

Resumen

Los genomas animales están compuestos mayormente por ADN no codificante. Una parte del ADN no codificante tiene como función regular la expresión de los genes y, en consecuencia, se conoce como “ADN regulatorio”. El ADN regulatorio contiene elementos llamados *enhancers*, que determinan cuándo, dónde y cuánto se expresa un gen. Históricamente, los *enhancers* fueron caracterizados como elementos que dirigen la expresión génica en un único contexto espacio-temporal, es decir en un único tejido y/o momento del desarrollo. Sin embargo, estudios recientes han demostrado que existen *enhancers* que poseen información regulatoria para generar más de un patrón de expresión y, por ende, son *enhancers* pleiotrópicos. La pleiotropía impone restricciones evolutivas, ya que un cambio en la actividad del *enhancer* puede producir efectos fenotípicos en varios contextos espacio-temporales.

En nuestro grupo se analizó la estructura y función de la región regulatoria del gen *shavenbaby* en *D. virilis* y *D. melanogaster*, dos especies que divergieron hace aproximadamente 40 millones de años. Al examinar la actividad de siete *enhancers* de *svb* en los estadios de embrión, larva y pupa en en *D. virilis* y *D. melanogaster*, observamos que los siete *enhancers* están activos en los tres estadios del desarrollo en ambas especies. Por ende, la actividad pleiotrópica de estos *enhancers* está conservada evolutivamente. Teniendo en cuenta estos resultados, pensamos que la actividad pleiotrópica de numerosos *enhancers* podría estar conservada entre *D. melanogaster* y *D. virilis* (pensamos que una gran cantidad de *enhancers* que eran pleiotrópicos en el ancestro común de estas especies siguen siendo pleiotrópicos en *D. melanogaster* y *D. virilis*).

Para poner a prueba esta hipótesis, identificamos *enhancers* putativos en *D. melanogaster* y los categorizamos como pleiotrópicos o contexto-específicos utilizando información de la estructura de la cromatina en diferentes contextos del desarrollo. Posteriormente, buscamos a los *enhancers* ortólogos en el genoma de *D. virilis* utilizando diferentes métodos. Finalmente, estudiamos la actividad de dichos *enhancers* ortólogos a partir de datos de apertura de la cromatina en distintos contextos del desarrollo de *D. virilis*. De

esta manera, logramos determinar que la proporción de *enhancers* con actividad pleiotrópica conservada en las dos especies es mayor que la proporción de *enhancers* contexto-específicos de *D. melanogaster* que conservan su actividad en *D. virilis.* Estos resultados sugieren que existen presiones selectivas sustanciales sobre la actividad pleiotrópica de los *enhancers.*

Summary

"Analysis of the genomic conservation of the pleiotropic activity of *enhancers* between *Drosophila melanogaster* and *Drosophila virilis*"

Animal genomes are composed mostly of non-coding DNA. A part of non-coding DNA, which is known as “regulatory DNA”, controls the expression of genes. Regulatory DNA contains transcriptional enhancers, the elements that determine when, where and how much a gene is expressed. Historically, enhancers have been characterized as elements that direct gene expression in a single spatio-temporal context (in a single tissue and/or moment of development). However, recent studies have shown that some enhancers have regulatory information to generate more than one expression pattern and, therefore, are pleiotropic enhancers. Pleiotropy imposes evolutionary constraints, since a change in enhancer activity can produce phenotypic effects in various spatio-temporal contexts.

We have studied the regulatory region of the gene *shavenbaby* in *D. virilis* and *D. melanogaster*, two species that diverged approximately 40 million years ago. By examining the activity of seven *svb* enhancers in the embryo, larva and pupa in *D. virilis* and *D. melanogaster*, we observed that the seven enhancers are active in the three stages of development in both species. Thus, the pleiotropic activity of these enhancers is evolutionarily conserved. Given these results, we hypothesized that the pleiotropic activity of numerous enhancers could be conserved between *D. melanogaster* and *D. virilis* (we think that a large number of enhancers that were pleiotropic in the common ancestor of these species remain pleiotropic in *D. melanogaster* and *D. virilis*).

To test this hypothesis, we identified putative enhancers in *D. melanogaster* and categorized them as pleiotropic or context-specific using chromatin structure data in different developmental contexts. Subsequently, we searched for orthologous enhancers in the *D. virilis* genome using two different methods. Finally, we analyzed the activity of orthologous enhancers using open-chromatin in different developmental contexts of *D. virilis.* We determined that the proportion of enhancers with pleiotropic activity that are shared between

the two species is greater than the proportion of context-specific enhancers that are shared between the two species. These results suggest that there are considerable selective pressures on the activity pleiotropic of enhancers.

Agradecimientos

Gracias a la universidad pública, gratuita y de calidad por darme la posibilidad de formarme en una de las mejores instituciones académicas del mundo. Estoy infinitamente agradecida por esta oportunidad de hacer ciencia en mi país, al cual amo profundamente, y espero que sea el inicio de muchas más oportunidades por venir. Estoy orgullosa de ser la primera universitaria de mi familia y estoy segura de que sin educación pública eso no hubiera sido posible. Es por eso que me duele muchísimo ver a nuestras universidades agonizar por el desfinanciamiento que están sufriendo. Deseo que a futuro todos puedan estudiar en una universidad de primer nivel, pública, laica y de calidad, como pude hacerlo yo.

A Nico Frankel por abrirme la puerta de su laboratorio y apoyarme incondicionalmente en este camino. Gracias por aconsejarme frente a cada obstáculo y por compartir mi entusiasmo con cada nuevo avance. Gracias por todas las charlas científicas y las no tan científicas. Es un privilegio investigar bajo la dirección de un gran investigador y, por sobre todas las cosas, de una gran persona.

A Ian por su paciencia inagotable, gracias por las tardes que dedicaste a enseñarme prácticamente todo lo que sé sobre bioinformática. En vos no solo encontré a una persona brillante sino también a un gran amigo. Cuando sea grande quiero ser como vos.

A Nacho, mi segundo director asistente, por creer en mí más que yo misma, por estar siempre para mí y por impulsarme con cada charla a mejorar mi trabajo y volverme una mejor científica. Realmente no encuentro manera de expresar cuánto te quiero.

A Niquito y a Caro, mi *dream team*, por todas las charlas, risas, chismes y meriendas. Me hace muy feliz haberme formado al lado de personas tan inteligentes, buenas y capaces como ustedes.

A Juan, Pau, Juli y Dani por las tardes y reuniones de grupo compartidas. Gracias por estar siempre dispuestos a ayudarme y por formar parte de un ambiente de trabajo tan lleno de buena energía.

A Luli y a Mili, mis compañeras de carrera y mis grandes amigas. Su amistad y apoyo son invaluables para mí. Voy a estar por siempre agradecida de haber coincidido en aquel aula de Sociedad y Estado.

Al biogrupo que supo ser mi sostén en los primeros años de carrera.

A Belá, Szajo, Maia y Paula, por aguantar todos mis “este finde no puedo porque el lunes rindo” y por estar siempre ahí para apoyarme en todo. No se que haría sin ustedes, las amo.

A Mati, por ser todo lo que siempre necesité y más. Mis días de estudio fueron mucho más felices gracias a tus cafés y tus besos en la frente. Gracias por ser mi cable a tierra y mi gran amor. Ojalá sigamos creciendo juntos.

A mis abuelos, por cuidarme durante toda mi infancia mientras mis papás salían a trabajar, gracias a ustedes y a los valores que me inculcaron soy la persona que soy hoy en día.

A mis hermanos y a mis papas por absolutamente todo. Gracias por el apoyo, tanto económico como emocional, que me permitió seguir la carrera que elegí y que amo con todo mi corazón. Mamá, gracias por esperarme todas las noches con un plato de comida caliente, y papá, gracias por irme a buscar a la estación cuando salía tarde de cursar para que llegue segura a casa. Gracias a ustedes tuve el privilegio de que mi única preocupación y responsabilidad fuera estudiar y formarme profesionalmente.

Índice

1. Introducción

1.1 El género *Drosophila* como modelo experimental

El género *Drosophila* contiene alrededor de 1600 especies de moscas pequeñas, siendo *Drosophila melanogaster* la especie más conocida del clado. El uso de *D. melanogaster* como modelo experimental ha sido clave para el avance de la Genética del Desarrollo y la Genómica Evolutiva, y ha permitido una mayor comprensión de ciertas enfermedades humanas. El uso de esta especie se popularizó a principios del siglo XX, cuando Thomas H. Morgan, mediante el estudio de la cepa de *D. melanogaster* de ojos blancos (*white*), confirmó la teoría cromosómica de la herencia. Esta teoría, planteada originalmente por Theodor Boveri y Walter Sutton, estableció las bases de la genética moderna, brindando una explicación mecanicista de las Leyes de Mendel (Morgan 1910, Morgan y Bridges 1916, Benson 2001). Los trabajos pioneros de Morgan en este campo le valieron el Premio Nobel de Fisiología y Medicina en 1933.

A continuación se enumeran algunas características valiosas del género *Drosophila*:

- Mantener a las moscas resulta fácil y económico. Un gran número de líneas genéticas puede ser mantenido dentro de un laboratorio, sin requerir demasiada infraestructura ni espacios extensos. Esta característica permite realizar estudios genético-poblacionales a gran escala.
- Presentan un tiempo generacional corto, de aproximadamente 10 días a 25°C para *D. melanogaster* (desde la puesta de huevos hasta que el organismo alcanza la adultez). Además, las hembras producen entre 750 y 1500 huevos durante su tiempo de vida. Por lo tanto, es posible llevar a cabo experimentos que requieran analizar varias generaciones.
- El genoma de la mosca es pequeño y menos redundante que los genomas de mamíferos. Se estima que solamente una o unas pocas copias de genes codifican para

un tipo de proteína. Esto es útil cuando se quiere investigar la función de genes a través de mutantes de pérdida de función.

- En la actualidad existe una gran cantidad de herramientas de genética molecular en *Drosophila*, las cuales posibilitan analizar con detalle la función de los genes.
- Aproximadamente el 75% de los genes identificados como mutados, amplificados o delecionados en enfermedades humanas conocidas tienen genes homólogos en *D. melanogaster* (Mirzoyan et al, 2019). Esto permite estudiar procesos genéticos en *D. melanogaster,* evitando las dificultades relacionadas con la experimentación en humanos.
- Existe una gran comunidad de científicos que trabajan con *Drosophila,* la cual está formada por más de 1600 laboratorios en todo el mundo (Hales et al, 2015). Esto permite la interacción entre grupos de trabajo diversos, favoreciendo el desarrollo de conocimiento y la producción de datos de forma masiva.
- La cantidad y calidad de información genómica de distintas especies del género disponible en bases de datos públicas permite realizar estudios genómicos confiables de forma relativamente sencilla. Gracias a la comunidad de científicos que trabajan con organismos de este género, existen datos análisis genómicos de genes, regiones regulatorias, sitios de unión de proteínas y regiones conservadas evolutivamente disponibles para ser utilizados de forma libre y gratuita.

1.1.1 El ciclo de vida de *Drosophila*

Las especies del género *Drosophila* poseen un ciclo de vida de seis etapas: un estadio embrionario, tres estadios larvales, un estadio pupal, y finalmente el estadio adulto, el cual reinicia el ciclo mediante reproducción sexual (Figura 1). *Drosophila* pertenece al grupo de insectos holometábolos, lo que significa que experimenta una metamorfosis completa durante el estadio pupal. A su vez, la duración de cada uno de los estadios del ciclo de vida de *Drosophila* guarda una relación directa con la temperatura del entorno. A 25 °C, el ciclo de vida de *D. melanogaster* tiene una duración aproximada de 10 días (Demerec y Kaufman, 1996; Ashburner 1989). A pesar de que todas las especies de *Drosophila* atraviesan las mismas etapas, la duración de cada estadío es variable entre especies.

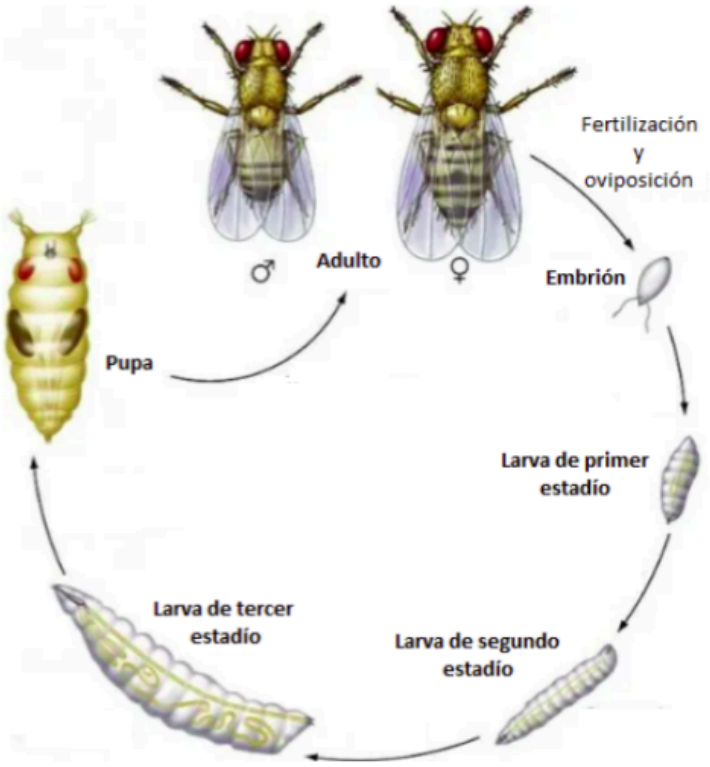


Figura 1. Esquema del ciclo de vida de *Drosophila*.

La embriogénesis comienza con la fertilización del oocito y la formación del cigoto. En *D. melanogaster* el desarrollo embrionario dura 24 horas a 25°C y se encuentra dividido en 17 estadios (Campos-Ortega y Hartenstein, 1985) (Figura 2). El desarrollo comienza con una etapa sincicial, en la que los núcleos se dividen sin que haya división del citoplasma (estadio 2). Después de 10 rondas de divisiones sincronizadas, los núcleos migran a la periferia y se produce la celularización (estadios 3 a 5). La transcripción del genoma cigótico (fenómeno denominado transición materno-cigótica o MZT, por sus siglas en inglés) comienza durante el ciclo mitótico 14, justo antes de la celularización (Langley et al, 2014). Este momento marca el inicio de las divisiones celulares asincrónicas. Posteriormente se forma la gástrula (estadio 8), donde quedan especificadas las tres capas germinales. Luego se produce la migración celular desde la zona anterior hacia la zona posterior del embrión durante la elongación de la banda germinal (estadio 9), seguido de una retracción de la misma (estadio 12). Posteriormente, las células epiteliales migran hacia la línea media dorsal en el cierre dorsal (estadio 13), y las estructuras de la cabeza comienzan a madurar (involución de la cabeza; estadio 15).

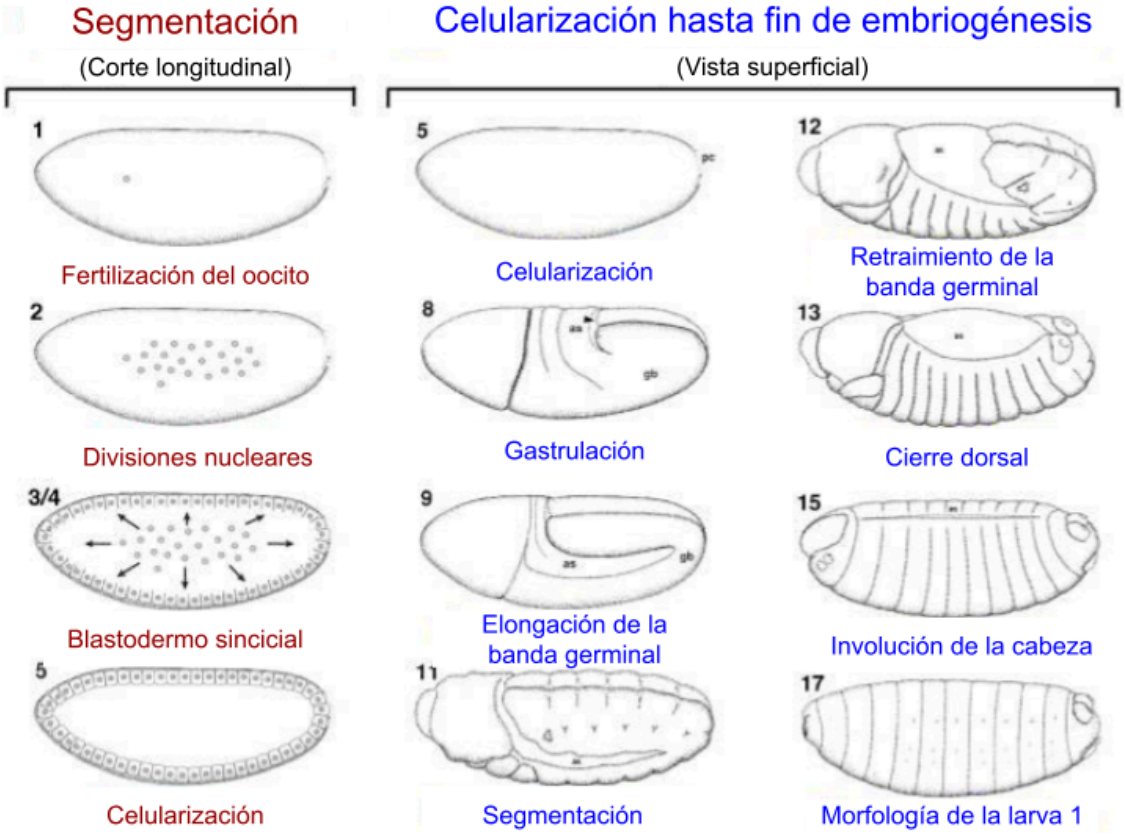


Figura 2. Estadios del desarrollo embrionario. El número correspondiente a cada estadio se indica en la esquina superior izquierda de cada esquema.

Una vez concluido el desarrollo del embrión, *Drosophila* atraviesa tres estadios larvales consecutivos (larva 1, larva 2 y larva 3) separados por mudas de cutícula. En *D. melanogaster* los estadios de larva 1 y larva 2 tienen una duración de un día cada uno, mientras que el estadio de larva 3 abarca entre 2 y 3 días. Durante los tres estadios larvales, la larva se alimenta y aumenta considerablemente su tamaño, incrementando su peso alrededor de 200 veces. El crecimiento larval se logra mediante un aumento en el volumen celular y la replicación de ADN, sin que medie la división celular en el proceso. Al alcanzar un peso crítico, la larva 3 abandona el sustrato y busca una zona limpia y seca a la cual adherirse (fase conocida como "larva *wandering*") y allí se inmoviliza y forma el pupario. En algunos tejidos de la larva, como por ejemplo las glándulas salivales, es posible observar células con cromosomas politénicos. A su vez, la larva posee grupos de células llamados histoblastos y órganos denominados discos

imaginales, los cuales darán origen a diversas estructuras en el adulto. Los discos imaginales son precursores de la epidermis (excepto el abdomen) y los apéndices del adulto: las alas, las patas, los ojos, la genitalia y las partes bucales (Figura 3). Los histoblastos son pequeños grupos celulares que formarán la epidermis abdominal y varios órganos internos del adulto (Tyler 2000).

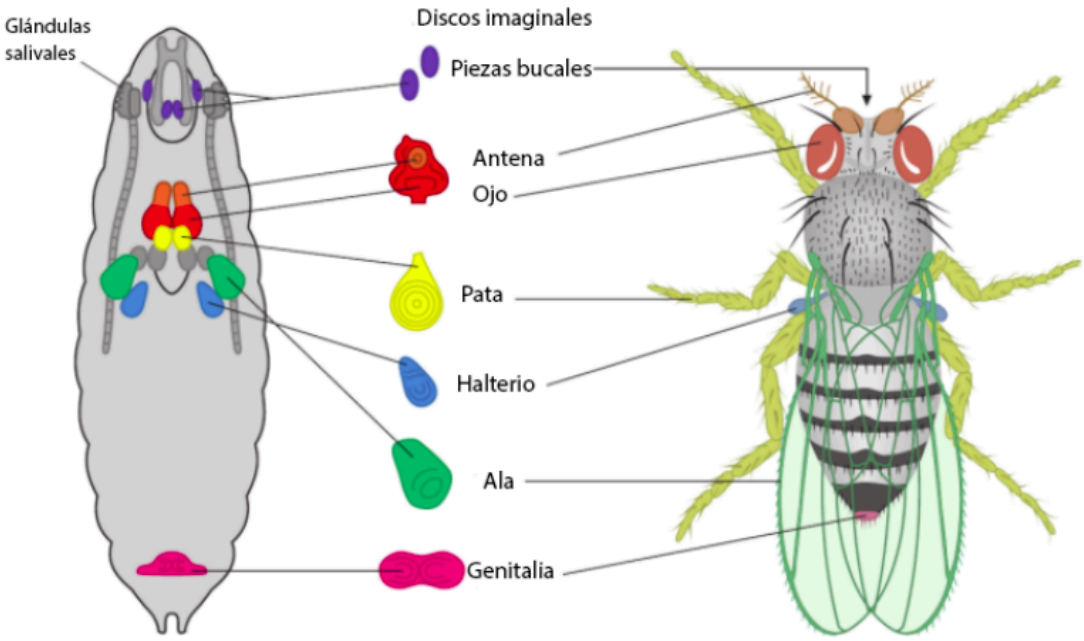


Figura 3. Localización de los discos imaginales en larva y estructuras que generarán en el adulto.

La pupación, al igual que las sucesivas mudas de la larva, está finamente regulada por la liberación de la hormona esteroidea ecdisona (Thummel 1995). El estadio de pupa de *D. melanogaster* dura aproximadamente 100 horas a 25 °C (Bainbridge y Bownes, 1981). El final del estadio pupal y comienzo de la etapa adulta ocurre a los 10-12 días de iniciado el ciclo a 25°C, cuando el adulto completamente formado (adulto farado) eclosiona del pupario. La esperanza de vida de un individuo adulto es de alrededor de 2 meses (Mohr, 2018).

1.1.2 *Drosophila melanogaster* y *Drosophila virilis*, dos especies que divergieron hace aproximadamente 40 millones de años

Las especies del género *Drosophila* presentan una amplia distribución geográfica. *Drosophila* incluye tanto especies cosmopolitas, que colonizaron todo el planeta, como *D. melanogaster* y *D. simulans,* así como especies con distribución geográfica acotada, como *D. sechellia,* que se encuentra en una única isla. *Drosophila* incluye especies muy diversas en cuanto a morfología, ecología y comportamiento. A pesar de estas diferencias, las especies de *Drosophila* comparten un plan corporal y presentan un ciclo de vida muy similar.

Drosophila es parte de la familia *Drosophilidae,* que contiene cerca de 4000 especies descriptas distribuidas en más de 70 géneros (O'Grady y DeSalle 2018). El género *Drosophila* incluye al ~50% de las especies de *Drosophilidae* (alrededor de 1600 especies). La gran cantidad de especies de *Drosophila* es el resultado de una importante radiación adaptativa, la cual permitió a las distintas especies ocupar una amplia variedad de nichos (Markow y O'Grady 2005, 2006). Tanto *D. melanogaster* como *D. virilis* forman parte de las 12 especies cuyos genomas fueron secuenciados por el *Drosophila 12 Genomes Consortium* (Clark et al, 2007). Estas dos especies divergieron hace aproximadamente 40 millones de años (Figura 4).

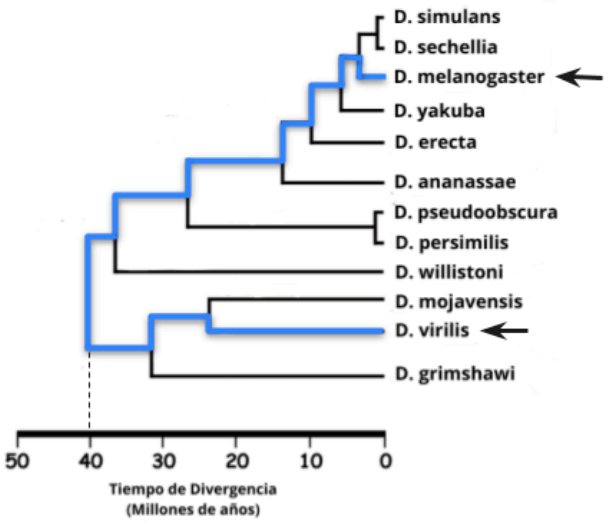


Figura 4. Relaciones filogenéticas entre las primeras 12 especies del género *Drosophila* en contar con un genoma de referencia. En la imágen se incluye una escala de tiempos evolutivos (en MA). Las flechas negras indican la ubicación de *D. virilis* y *D. melanogaster* en el árbol.

Históricamente, se considera que *D. virilis* es parte del grupo conocido como *virilis-repleta*. Este linaje fue propuesto por Throckmorton (1975), quien agrupó especies en base a carácteres morfológicos y hábitos ecológicos, como el hecho de ocupar nichos

cactófilos. *D. virilis* pertenece al subgénero *Drosophila,* mientras que *D. melanogaster* forma parte del subgénero *Sophophora* (O'Grady y DeSalle 2018*).*

Actualmente *D. virilis*, al igual que *D. melanogaster,* es considerada una especie cosmopolita. Ambas especies ampliaron sus distribuciones geográficas ancestrales, probablemente como consecuencia de su comensalismo con los humanos y su capacidad para reproducirse en una gran variedad de frutas en descomposición (Mirol 2008).

El ciclo de vida de *D. virilis* es más largo que el de *D. melanogaster,* con una duración aproximada de 18 días a 25°C (Markow y O´Grady 2005, 2007). Los machos de *D. melanogaster* alcanzan la madurez sexual 2 horas antes que las hembras, mientras que en *D. virilis* son las hembras quienes alcanzan la madurez sexual 6 horas antes que los machos (Markow y O´Grady 2007). Una posible explicación a estas diferencias reside en la complejidad relativa de la gametogenesis o de la maduración del tracto reproductivo de un sexo respecto del otro. A su vez, se ha observado que las especies en las que los machos maduran antes que las hembras, como *D. melanogaster,* tienden a producir espermatozoides más cortos que aquellas en las que las hembras maduran antes (Markow y O´Grady 2007).

Los tres estadios larvales, así como la pupa y el adulto presentan tamaño mayor en *D. virilis.* Las moscas adultas de *D. virilis* alcanzan casi el doble de tamaño que los adultos de *D. melanogaster* (Figura 5A). A la vez, tanto el adulto como la pupa tardía de *D. virilis* presenta una pigmentación oscura, la cual no está presente en *D. melanogaster* (Figura 5B).

En *Drosophila* el sistema de determinación del sexo está dado por un mecanismo de "dosaje" de la cantidad de cromosomas X en células diploides, cuyo factor blanco río abajo es el gen Sex-lethal (*Sxl*). Este sistema está conservado entre *D. melanogaster* y *D. virilis* (Bopp et al 1996). En individuos XX, *Sxl* comienza a expresarse pocas horas después de la fertilización del oocito, mientras que, en individuos XY, *Sxl* no es expresado durante los primeros estadios embrionarios (Salz et al, 1987). Posteriormente, un promotor tardío permite la transcripción de *Sxl* tanto en individuos XX como XY. Sin embargo, el ARNm del gen *Sxl* difiere entre machos y hembras; el ARNm de machos da lugar a una proteína no funcional. En individuos XX se produce una variante específica de splicing gracias a la unión de la proteína SXL, la cual se expresó tempranamente en el desarrollo. Esto lleva a la activación de SXL, iniciando la cascada de señalización que resulta en un individuo de sexo femenino y previeniendo el mecanismo de compensación de dosis. En el genotipo XY, SXL permanece inactivo, dando

como resultado un individuo macho y generando la activación del mecanismo de compensación de dosis del cromosoma X.

Figura 5. Características morfológicas de *D. melanogaster* y *D. virilis* en distintos estadios de su ciclo de vida. (A) Comparación entre individuos adultos. A la izquierda se ubica *D. melanogaster* y a la derecha *D. virilis.* **(B)** Comparación entre distintas etapas del estadio pupal. De izquierda a derecha: prepupa, pupa temprana y pupa intermedia/tardía. Fila inferior *D.melanogaster* y fila superior *D. virilis.* Las imágenes de (B) no están en escala.

1.1.3 Genoma de *D. melanogaster* y de *D. virilis*

En el año 2000, el genoma de *Drosophila melanogaster* fue uno de los primeros en ser secuenciado por completo. Utilizando estrategias de *whole-genome shotgun sequencing*, se logró determinar la secuencia de la porción de eucromatina del genoma de *D. melanogaster* (Adams et al, 2000). Siete años después, el *Drosophila 12 Genomes Consortium* publicó los genomas completos de 12 especies del género *Drosophila* (Clark et al, 2007), entre ellos el de

D. virilis, proporcionando a la comunidad científica un recurso invaluable para realizar estudios comparativos.

En la actualidad, gracias a los avances tecnológicos y a la disminución de costos en materia de secuenciación, contamos con numerosos genomas secuenciados de especies de *Drosophila*. Un claro ejemplo de estos avances es un trabajo reciente en el que se ensamblaron los genomas de 101 líneas de 93 especies de drosofílidos (Kim et al, 2021). La gran cantidad de información disponible posibilita la ejecución de estudios comparativos a gran escala, que exploren aspectos como la expresión génica, la estructura de proteínas, los mecanismos del desarrollo, y las adaptaciones ecológicas.

A nivel citológico, el genoma de *D. virilis* está organizado en 5 pares de autosomas y un par de cromosomas sexuales (2n = 2x = 12) (Figura 6A). En cambio, *D. melanogaster* posee 3 pares de autosomas y un par de cromosomas sexuales (2n = 2x = 8) (Figura 6A). En su histórica publicación "*Bearings of the 'Drosophila' work on systematics*", Muller definió 6 elementos cromosómicos fundamentales, los llamados elementos de Muller, a los cuales designó con letras de la A a la F.

Para *D. melanogaster*, el cromosoma sexual X se corresponde con el elemento de Muller A, mientras que el cromosoma Y no posee elemento designado. Los cromosomas 2 y 3 de *D. melanogaster* son metacéntricos, correspondiendo los brazos 2L, 2R, 3L y 3R a los elementos de Muller B, C, D y E respectivamente (Figura 6B). El cromosoma 4 de *D. melanogaster* es conocido como cromosoma punto, y corresponde al elemento de Muller F.

En cambio, para *D. virilis* los 5 pares de autosomas corresponden a los elementos de Muller de la B a la F, mientras que el cromosoma X corresponde al elemento A, al igual que en *D. melanogaster* (Figura 6B). En ambas especies, el elemento de Muller F presenta características particulares: posee una alta densidad de elementos repetitivos, genes de mayor tamaño, menor *codon bias*, y una alta tasa de rearreglos de genes (Leung et al, 2010).

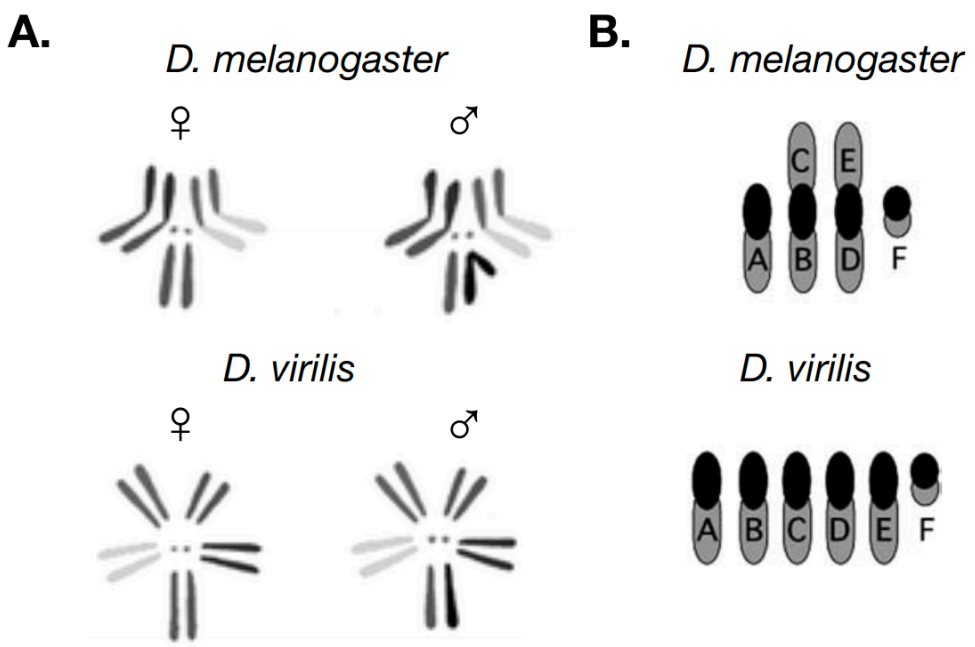


Figura 6. Cariotipos de *D. melanogaster* y de *D. virilis.* (A) Representación del cariotipo de *D. melanogaster* (arriba) y de *D. virilis* (abajo). Para ambas especies, se representa el cariotipo de machos (derecha) y hembras (izquierda). Los elementos de Muller están representados en distintos tonos de grises. Modificado de Adryan, B., y Russell, S. 2012. **(B)** Organización de los seis elementos de Muller en *D. melanogaster* (arriba) y de *D. virilis* (abajo).

El tamaño estimado por citometría de flujo del genoma de *D. melanogaster* es de ~180 Mb, de los cuales ~120 Mb corresponden a eucromatina (Adams et al, 2000). El genoma de *D. virilis* es aproximadamente un 50% más grande que el de D. melanogaster (Drosophila 12 Genomes Consortium 2007). Aproximadamente el 24% del genoma de *D. melanogaster* está constituido por elementos repetitivos, como microsatélites, secuencias de baja complejidad, transposones de DNA y retrotransposones. En *D. virilis* se registra un porcentaje de contenido repetitivo similar (Kim et al, 2021).

El número estimado de genes de *D. melanogaster* en base a la última versión del genoma es de 13.986 (https://www.ensembl.org/Drosophila_melanogaster/Info/Annotation, último acceso: 9 de diciembre de 2023). En el genoma de *D. virilis*, se estima que existen 13.685 genes que codifican proteínas (www.ncbi.nlm.nih.gov/datasets/genome/GCF_000001215.4/, último acceso: 9 de diciembre de 2023).

1.2 Elementos cis-regulatorios: *enhancers* transcripcionales

Un objetivo central de la biología evolutiva es entender la base genética de la gran diversidad morfológica observada en los metazoos. A lo largo de los años, los investigadores han intentado entender qué tipo de cambios en los genomas son responsables de dicha diversidad morfológica. En 1975, Mary-Claire King y Allan Wilson describieron la gran similitud existente entre proteínas de chimpancé y de humanos, y se preguntaron cómo dos especies con genes tan similares pueden diferir tanto en su anatomía y forma de vida. Concluyeron que los cambios evolutivos en la anatomía y el estilo de vida se basan mayormente en cambios en los mecanismos que controlan a los genes, más que en las secuencias de las proteínas (King y Wilson, 1975). Hoy en día, sabemos que la mayor parte de la variabilidad morfológica puede ser explicada por cambios en los patrones de expresión de proteínas altamente conservadas. Esos cambios son el resultado de mutaciones en regiones no codificantes, más precisamente en las regiones regulatorias de genes que intervienen en complejas redes de regulación durante el desarrollo (Carroll, 2008).

Las regiones regulatorias de los genes contienen distintos tipos de elementos regulatorios, entre ellos los *enhancers,* que activan la transcripción de un gen, los *silencers,* que inhiben la transcripción, y los *insulators,* que son elementos que aíslan un gen de las influencias del ADN vecino (Figura 7) (Hardison y Taylor, 2012).

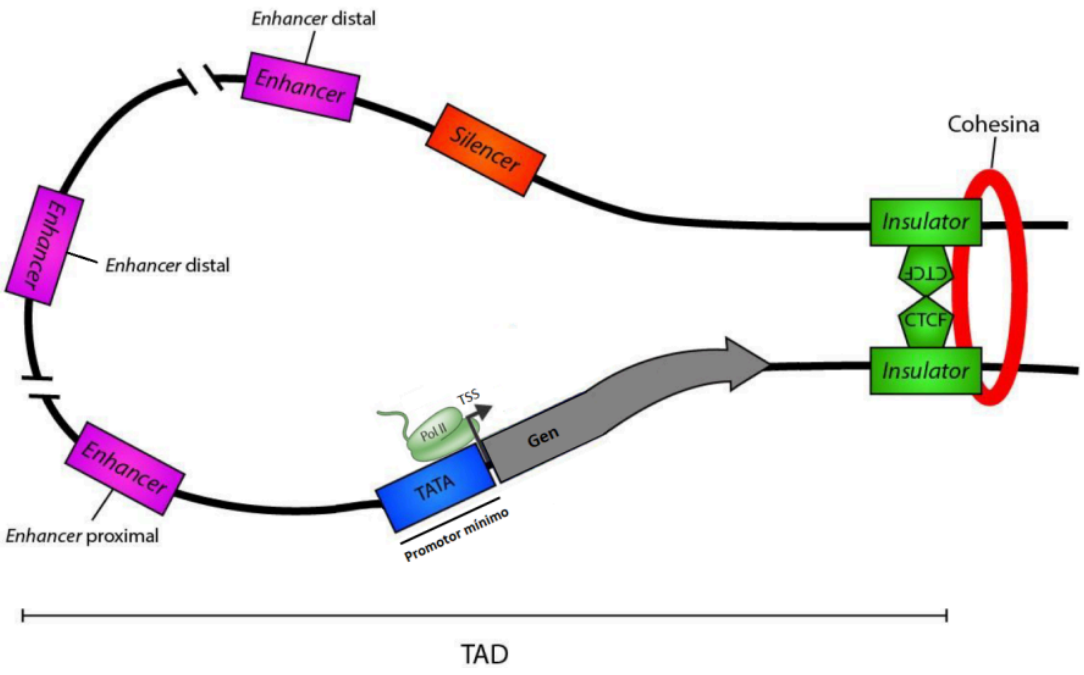


Figura 7. Elementos regulatorios de la transcripción en metazoos. Representación de un gen con su promotor (que incluye la secuencia *TATA-box*), en el que se sitúa la ARN polimerasa II para comenzar la transcripción. Se esquematizan tres *enhancers* a diferentes distancias del promotor y un *silencer*. También se muestra un *insulator* que delimita un TAD (dominio topológicamente asociado) a través de su unión a Cohesina y a la proteína CTCF. Modificado de Ong y Corces, 2011.

Los *enhancers* son secuencias de ADN no codificante que determinan cuándo, dónde y cuánto se expresa un gen (Wittkopp y Kalay, 2011; Frankel, 2012). Los *enhancers* interaccionan físicamente con factores de transcripción (FTs), que reconocen secuencias cortas y específicas en el ADN (Spitz y Furlong, 2012). Las secuencias reconocidas por los FTs se denominan sitios de unión de factores de transcripción (TFBSs) y, típicamente, tienen una longitud inferior a 20 pares de bases (Davidson 2010). La unión de múltiples FTs y cofactores a *enhancers* hace posible la interacción entre los elementos regulatorios y la maquinaria basal de transcripción en el sitio de inicio de la transcripción (TSS) de sus genes blanco (Levine, 2010; Rebeiz y Tsiantis, 2017). Al establecer este contacto físico, los *enhancers* son capaces de aumentar la tasa de transcripción de la ARN polimerasa II.

La actividad de los *enhancers* puede ser controlada en varios niveles. La actividad de un *enhancer* puede ser regulada por la compactación de la cromatina. Los *enhancers* activos suelen tener la cromatina “abierta”. Los FTs pioneros tienen la capacidad de unirse al ADN en

un contexto nucleosomal (ADN unido a histonas), a diferencia de la mayoría de los FTs que sólo se unen a regiones libres de nucleosomas. Al unirse a *enhancers*, los FTs pioneros generalmente interactúan con remodeladores de la cromatina, generando la apertura de la cromatina en esa región (Iwafuchi-Doi y Zaret, 2014). En segundo lugar, luego de la unión de FTs pioneros, la unión de otros FTs es fundamental para la actividad de los *enhancers.* En muchos casos, los *enhancers* poseen motivos de unión para FTs que actúan como represores, inhibiendo la expresión génica en contextos inadecuados (Mellerick y Nirenberg, 1995). Incluso, se sabe que los represores son capaces de refinar el patrón de expresión de los genes (Crocker et al, 2017). Por último, algunos *enhancers* requieren la transducción de señales adicionales para permitir la unión de cofactores, los cuales posibilitan la interacción con la maquinaria de transcripción (Barolo y Posakony 2002).

Se postula que los *enhancers* tienen la capacidad de regular la expresión génica independientemente de su orientación y distancia relativa al promotor de su gen blanco (Long et al, 2016). En *Drosophila*, los *enhancers* suelen activar al gen más cercano (Massouras et al, 2012; Kvon et al, 2014). Sin embargo, un *enhancer* y el promotor de su gen blanco pueden encontrarse a decenas de kilobases (kb) de distancia. En estos casos, la estructura tridimensional de la cromatina en TADs (dominios topológicamente asociados, Szabo et al, 2018) facilita la interacción *enhancer*-promotor (Figura 7).

Existe evidencia que demuestra que los *enhancers* establecen contacto con la región promotora del gen que regulan mediante la formación de un bucle en la cromatina (Bulger y Groudine, 1999). La cohesina es una proteína que actúa como mediadora en la formación de estos bucles, que pueden establecerse a larga distancia. Dichos bucles pueden ser estudiados con técnicas que estudian la estructura 3D de la cromatina (Berkum y Dekker, 2014). Una vez que establecido el bucle entre el *enhancer* y su promotor blanco, se cree que los *enhancers* estimulan la transcripción al reclutar al complejo mediador y a factores de transcripción basales (Kornberg 2005) (Figura 8).

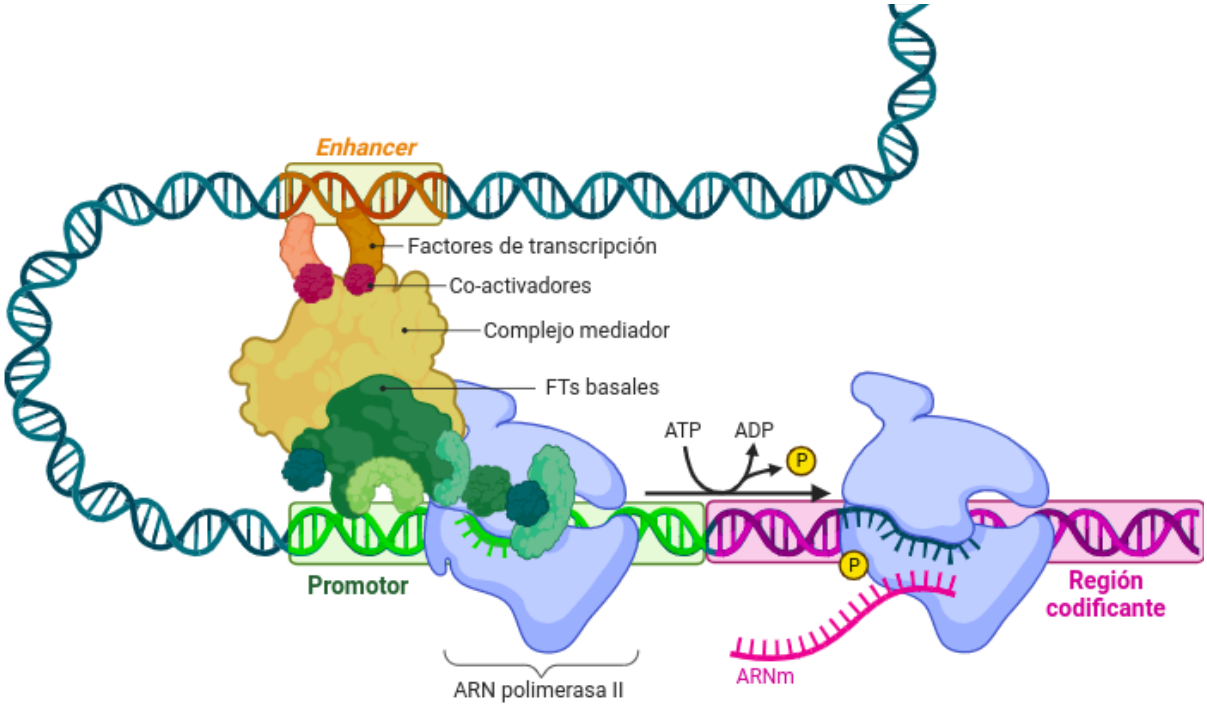


Figura 8. Activación de la transcripción en metazoos. El acercamiento espacial del *enhancer* al promotor de su gen blanco se produce mediante la formación de un bucle en el ADN. Los FT "activadores" (naranja) se unen a los *enhancers* y, mediante factores co-activadores (rojo), se acercan al complejo mediador (amarillo). Este complejo establece un aumento en la concentración local de factores de transcripción basales (representados en distintos tonos de verde) en la región promotora, lo que lleva al aumento en la tasa de transcripción.

El funcionamiento incorrecto de *enhancers,* ya sea por alteraciones genéticas, estructurales o epigenéticas, altera la expresión génica en distintos tipos de cáncer y en numerosas enfermedades humanas (Claringbould y Zaugg, 2021). Se ha calculado que el 90% de los SNPs (polimorfismos de nucleótido único) asociados a enfermedades humanas por GWAS (estudio de asociación entre la variación nucleotídica a nivel genómico con la variación observada en relación al fenotipo estudiado) están en regiones no codificantes del genoma (Edwards et al, 2013). Es muy probable que la mayoría de estos SNPs afecten la función de *enhancers* (Corradin y Scacheri, 2014). De hecho, la gran mayoría de los SNPs asociados con cáncer de mama y cáncer de próstata están en regiones del genoma que tienen una estructura cromatínica propia de *enhancers* (Corradin y Scacheri, 2014). A su vez, el mal funcionamiento de *enhancers* es la causa de las enfermedades mendelianas llamadas "enhanceropatias*".* Un ejemplo de enhanceropatia es la polidactilia, causada por mutaciones en un *enhancer* del gen

Shh, el cual se encuentra altamente conservado en vertebrados (Lettice et al 2003, Dahn et al 2007). Comprender cómo funcionan los *enhancers* resulta entonces fundamental para entender las bases moleculares de numerosas enfermedades en humanos.

1.3 Predicción de *enhancers* activos a nivel genómico

A pesar de los grandes avances en tecnología, secuenciación y técnicas experimentales, la identificación de *enhancers* activos hoy en día continúa siendo un problema desafiante.

Históricamente, los *enhancers* han sido validados y estudiados con construcciones reporteras. Esta técnica consiste en clonar la secuencia de un *enhancer* putativo río arriba de un promotor mínimo y un gen reportero. De esta manera, estudiando la expresión del gen reportero, se evalúa si el *enhancer* putativo es capaz de recapitular, total o parcialmente, la expresión del gen. (Andersson y Sandelin, 2019). A pesar de ser una técnica muy útil, posee algunas desventajas. Por un lado, los resultados de ensayos con genes reporteros realizados de manera episomal (sin integración al genoma) pueden ser sustancialmente diferentes a los resultados de evaluar la misma secuencia con inserción de la construcción al genoma (Inoue et al, 2016). A la vez, estos ensayos evalúan a los *enhancers* fuera de su contexto nativo y se sabe que, en estas condiciones, los patrones de expresión generados pueden diferir de los del gen nativo (Kvon, 2015).

Con el gran avance de la secuenciación masiva (NGS) se generaron métodos para la identificación de *enhancers* a nivel genómico*.* Estos métodos identifican características de los *enhancers*: la unión de factores de transcripción, la presencia de determinadas marcas epigenéticas, las regiones de cromatina abierta y la transcripción bidireccional de eARNs (ver más adelante) (modENCODE Consortium et al. 2011; ENCODE Project Consortium 2013).

Los nucleosomas flanqueantes a *enhancers* activos suelen poseer acetilada la lisina 27 de la histona 3 (H3K27ac) (Creyghton et al, 2010). Esta marca, al igual que la presencia de factores de transcripción, puede ser detectada mediante la técnica ChIP-seq. El nombre ChIP-seq hace referencia a la inmunoprecipitación de la cromatina seguida de secuenciación masiva.

Como se mencionó anteriormente, los *enhancers* activos se encuentran en regiones de cromatina abierta, las cuales pueden ser identificadas utilizando DNase-seq ó ATAC-seq. La técnica DNAse-seq permite identificar regiones accesibles gracias a la enzima DNase I, la cual produce cortes doble cadena de manera preferencial en cromatina accesible (Hesselberth et al, 2009). Por otro lado, ATAC-seq (*Assay for Transposase-Accessible Chromatin using sequencing*) utiliza una versión hiperactiva de la transposasa bacteriana Tn5 para insertar adaptadores en regiones accesibles de la cromatina (Buenrostro et al, 2015). Los protocolos de DNAse-seq y ATAC-seq poseen un último paso en común: la secuenciación masiva en paralelo. Con el uso de programas bioinformáticos, las lecturas obtenidas son mapeadas a un genoma de referencia, generando picos de mayor señal en regiones del genoma con cromatina accesible.

En la actualidad, sabemos que muchos *enhancers* se transcriben, produciendo lo que se conoce como eARNs. Los eARNs típicamente son moléculas cortas, poseen CAP en su extremo 5', se transcriben de manera bidireccional, son abundantes y no se encuentran poliadenilados (Mikhaylichenko et al, 2018). Los ARNs que poseen 5' CAP pueden ser detectados mediante CAGE-seq (*Cap Analysis of Gene Expression using sequencing*), técnica basada en la captura del 5' CAP mediante biotinilación. De esta forma, es posible inferir la presencia de *enhancers* activos en regiones alejadas de TSSs con expresión bidireccional (Andersson et al, 2015; The FANTOM Consortium, RIKEN PMI y CLST, 2014).

Los análisis de la cromatina o de unión de proteínas permiten predecir la presencia de *enhancers* activos en un determinado tejido o momento del desarrollo. Los resultados de las técnicas descriptas anteriormente suelen analizarse en conjunto para inferir la presencia de *enhancers* activos. Por ejemplo, en nuestro laboratorio hemos realizado predicciones de *enhancers* humanos identificando regiones de cromatina abierta que tengan la marca H3K27ac (datos de ChIP-seq y DNAse-seq) o regiones que tengan expresión bidireccional (datos de CAGE-seq) (Laiker y Frankel 2022).

1.4 Métodos de predicción de *enhancers* ortólogos

La predicción de *enhancers* a nivel genómico es una tarea desafiante, incluso cuando se cuenta con evidencia experimental proveniente de experimentos de NGS. La predicción de

enhancers se vuelve aún más complicada cuando la disponibilidad de información sobre la estructura de la cromatina es limitada. En las últimas décadas se desarrollaron métodos para identificar *enhancers* ortólogos. Estos métodos utilizan la secuencia de *enhancers* previamente identificados en una especie (típicamente de la especie con mayor información experimental disponible) y buscan localizar esos *enhancers* en otro/s genoma/s.

Un método simple es utilizar herramientas como BLAST (Altschul et al, 1990) para buscar regiones con alta identidad de secuencia. Encontrar *enhancers* en especies evolutivamente distantes es posible en casos donde los elementos están sujetos a selección purificadora, lo que resulta en la conservación de su secuencia a lo largo de la evolución (Siepel et al, 2005). Hoy en día, uno de los métodos más populares para encontrar *enhancers* ortólogos es *liftOver* (Hinrichs et al 2006). El programa *liftOver* permite convertir coordenadas genómicas entre especies, empleando un alineamiento entre los genomas en cuestión. Numerosos trabajos han tenido éxito identificando *enhancers* con *liftOver* (Carelli et al, 2018; Chen et al, 2018; Aurizio et al, 2022*;* Ramirez et al, 2022; Li et al, 2023*).*

Los *enhancers* suelen tener menos restricciones evolutivas que las secuencias codificantes (Markstein y Levine, 2002). A lo largo de su evolución, los *enhancers* pueden sufrir inserciones, deleciones y rearreglos a escala de TFBSs y, a pesar de estos cambios, conservar su función (Weirauch y Hughes 2010; Wong et al 2020). En otras palabras, los *enhancers* pueden conservar su función ancestral, a pesar de que sus secuencias sean muy diferentes. En consecuencia, identificar *enhancers* ortólogos pero con baja conservación de secuencia no siempre es posible. Una estrategia alternativa para detectar *enhancers* con secuencias divergentes son los métodos llamados "*Alignment-free*". Los métodos *alignment-free* permiten inferir ortología a partir de la similitud en la composición de palabras (*k-mers)* entre secuencias, independientemente del orden y orientación de los *k-mers* (Kantorovitz et al, 2007; Arunachalam et al, 2010; Göke et al, 2012).

Actualmente existe un auge por las técnicas de aprendizaje automático . El aprendizaje automático es una rama de la inteligencia artificial centrada en el desarrollo de algoritmos que le permiten a una computadora identificar patrones y realizar una determinada tarea sin ser programada explícitamente. En lugar de seguir una serie de instrucciones detalladas, la computadora aprende a partir de datos y experiencias anteriores, mejorando su rendimiento al realizar una tarea específica. En el ámbito de la bioinformática y la predicción de *enhancers*,

esto se traduce en modelos entrenados con datos de *enhancers* de genomas bien anotados y respaldados experimentalmente. Una vez entrenado, el modelo es utilizado para predecir *enhancers* en genomas de especies relacionadas a aquella utilizada para el entrenamiento. Los métodos basados en aprendizaje automático detectan patrones en la secuencia de ADN que pueden indicar actividad de *enhancers* (Chen et al, 2018; Minnoye et al, 2020; Hong et al, 2021; MacPhillamy et al, 2022). El entrenamiento de modelos de aprendizaje automático puede ser muy demandante computacionalmente.

En este trabajo de tesis utilizamos dos métodos para predecir *enhancers* transcripcionales entre especies: uno basado en alineamientos y otro "*alignment-free*". La idea de utilizar estos métodos complementarios intenta aportar robustez a nuestros análisis.

1.5 *Enhancers* pleiotrópicos y contexto-específicos

La mayor parte de los genes involucrados en el desarrollo de los metazoos se expresan en distintos contextos espacio-temporales. Este fenómeno, en el que un gen cumple múltiples roles en diferentes tejidos y/o momentos del desarrollo, se denomina "pleiotropía génica" (Paaby y Rockman, 2013). Los genes pleiotrópicos poseen con frecuencia regiones regulatorias complejas, con múltiples *enhancers* transcripcionales.

Históricamente, los enhancers han sido descriptos como elementos que dirigen la expresión génica en un único contexto espacio-temporal (Carroll, 2008; Davidson, 2010). Bajo este paradigma, el patrón de expresión de un gen pleiotrópico sería la sumatoria de los patrones de expresión generados por varios enhancers contexto-específicos (Figura 9A). Un ejemplo de *enhancers* contexto-específicos son aquellos que regulan al gen Pax6 en mamíferos. Pax6 es un gen pleiotrópico que se expresa en páncreas, tubo neural, y ojos, cumpliendo un rol fundamental en el desarrollo de dichos órganos. La expresión de Pax6 en los distintos órganos es regulada por diferentes *enhancers* contexto-específicos; cada *enhancer* genera expresión en un único órgano (Dye, 2012).

La idea de la actividad contexto-específica de los enhancers es atractiva desde el punto de vista evolutivo, ya que provee de una enorme flexibilidad evolutiva a estos elementos. Dentro de este paradigma, una mutación en un enhancer afectaría a sólo uno de los contextos en los que se expresa un gen pleiotrópico, ya que la expresión del gen en otros contextos

estaría codificada en otros *enhancers* (Carroll, 2008; Wittkopp y Kalay, 2011). Por lo tanto, los *enhancers* contexto-específicos estarían libres de las presiones evolutivas impuestas por la pleiotropía (Gompel et al, 2005; McGregor et al, 2007; Chan et al, 2010; Kvon et al, 2016). Un claro ejemplo de este fenómeno es el del gen Pitx1, que es regulado por varios *enhancers* contexto-específicos (Chan et al, 2010). En poblaciones naturales de peces espinosos (*Gasterosteus aculeatus*), la deleción de uno de estos *enhancers* hace que Pitx1 no se exprese en la región pélvica, provocando la pérdida de una espina pélvica. En línea con la noción de *enhancers* contexto-específicos, la expresión de Pitx1 en otros contextos espaciales (dirigida por los otros *enhancers*) permanece inalterada (Chan et al, 2010; Levine et al, 2014).

Estudios de nuestro grupo y otros laboratorios han demostrado que existen *enhancers* que poseen información regulatoria para generar más de un patrón de expresión y, por ende, son *enhancers* pleiotrópicos (McKay y Lieb, 2013; Vizcaya-Molina et al, 2018; Lonfat et al, 2014; Schep et al, 2016; Preger Ben-Noon et al, 2018; Sabaris et al, 2019; Laiker y Frankel, 2022). Por lo tanto, la expresión de un gen pleiotrópico en varios tejidos puede generarse por la actividad de un único *enhancer* pleiotrópico (Figura 9B). Nuestro trabajo sugiere que una proporción grande de los *enhancers* de genomas animales son pleiotrópicos (Sabarís et al, 2019; Laiker y Frankel, 2022).

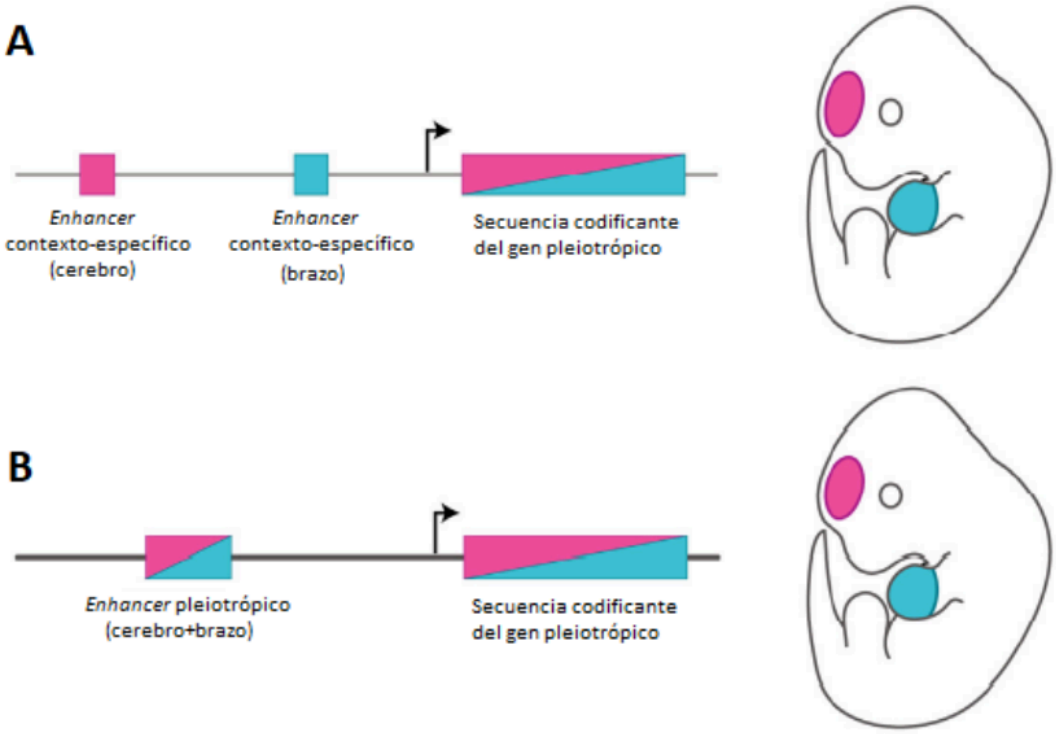


Figura 9. ***Enhancers* pleiotrópicos y contexto-específicos.** Un gen activo en dos órganos distintos (i.e., pleiotrópico) puede estar regulado por dos *enhancers* contexto-específicos (A) o por un único *enhancer* pleiotrópico (B) que genera la expresión del gen en ambos tejidos. Modificado de Sabaris et al, 2019.

En comparación con *enhancers* contexto-específicos, los *enhancers* pleiotrópicos estarían sometidos a mayores restricciones evolutivas, ya que un cambio en su actividad puede producir efectos fenotípicos en varios contextos espacio-temporales. Se ha demostrado que *enhancers* humanos con actividad pleiotrópica presentan una mayor conservación evolutiva que aquellos que son contexto-específicos (Fish et al, 2017; Singh y Yi, 2021). Esto lleva a pensar que un cambio genético en un *enhancer* pleiotrópico podría estar sujeto a selección purificadora más intensa que aquellos cambios genéticos que ocurren en *enhancers* activos en un solo contexto espacio-temporal.

1.6 La función pleiotrópica de los *enhancers* del gen *svb* está conservada entre *D. melanogaster* y *D. virilis*

El gen *shavenbaby* (*svb*) de *Drosophila* es un modelo establecido para el estudio de la estructura, función y evolución de las regiones regulatorias de la transcripción (Stern y Frankel, 2013, Kittelmann, et al., 2021). *svb* es un gen pleiotrópico localizado en el cromosoma X de *Drosophila*, que se expresa en la epidermis embrionaria y en distintos tejidos de la larva y de la pupa (Delon et al. 2003, Pueyo y Couso 2011; Chanut-Delalande et al. 2014; Preger Ben-Noon et al. 2018). *svb* es un gen maestro que codifica para un factor de transcripción que regula la diferenciación de la epidermis y otros procesos del desarrollo en insectos (Ray et al., 2019). La expresión de *svb* en la epidermis embrionaria es necesaria para la diferenciación de tricomas (estructuras cuticulares con forma de pelo) en la epidermis de la larva (Stern y Frankel 2013). En pupa, *svb* es necesario para formar tricomas en alas, patas, tórax y abdomen (Delon et al. 2003, Chanut-Delalande et al. 2014; Preger Ben-Noon et al. 2018). Además, la proteína SVB es necesaria para la correcta formación de los omatidios de los ojos (Delon et al. 2003) y las uniones de los tarsos de la pata (Pueyo y Couso 2011).

La expresión de *svb* en el embrión de *D. melanogaster* es generada por 7 *enhancers* transcripcionales, incluidos en una región de ~90 kb localizada río arriba del TSS del gen (Figura 10A) (Stern and Frankel, 2013). Los patrones de expresión generados por los 7 *enhancers* se superponen parcialmente, confiriendo robustez a la expresión de *svb* frente a perturbaciones ambientales y genéticas (Frankel et al. 2010). Utilizando construcciones reporteras, nuestro laboratorio demostró que los 7 *enhancers* embrionarios están también activos en distintos tejidos de la larva y la pupa y, por ende, son pleiotrópicos (Figura 10B). En la larva, cinco de los *enhancers* dirigen la expresión en la epidermis, seis en el sistema digestivo anterior y cuatro en el sistema nervioso central. En la pupa, los siete *enhancers* de *svb* presentan actividad en la cabeza, el tórax, el abdomen y las alas (Figura 10B).

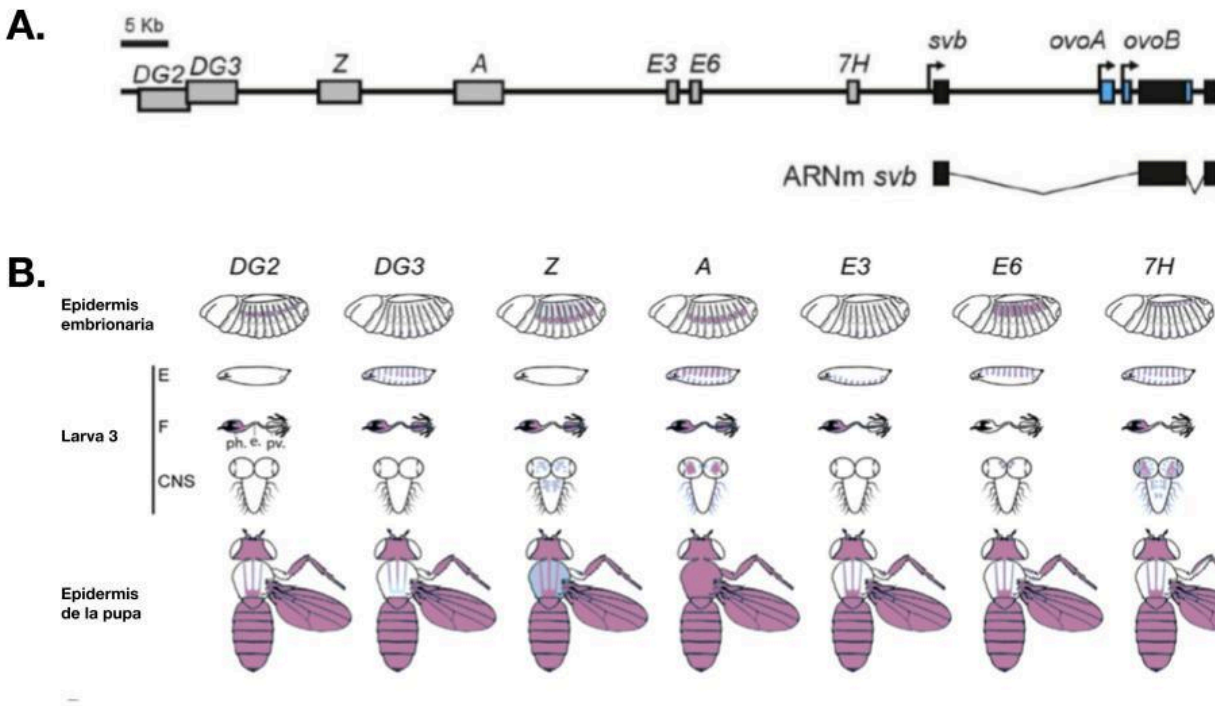


Figura 10. Los 7 *enhancers* de *svb* están activos en distintos estadios del ciclo de vida de *Drosophila melanogaster. (*A) Esquema del locus *svb.* Los 7 *enhancers* se representan como rectangulos grises y los exones del gen como rectángulos negros. **(B)** Patrón de expresión generado por cada uno de los 7 *enhancers* de *sbv* a lo largo de la ontogenia de *D. melanogaster.* La letra E se refiere a la epidermis de la larva, le letra F al sistema digestivo anterior, y la sigla CNS al sistema nervioso central.

Nuestro grupo realizó un estudio comparativo de la arquitectura regulatoria en el locus *svb* entre *D. melanogaster* y *D. virilis*. Para ello, se analizó la región regulatoria completa de *svb* en *D. virilis*, la cual tiene un tamaño de ~132 kb y es ortóloga a la región de ~90 kb que contiene los siete *enhancers* embrionarios de *D. melanogaster.* Utilizando construcciones reporteras, se identificaron seis *enhancers* en la región regulatoria de *svb* en *D. virilis* que están activos en la epidermis embrionaria.

Se demostró que los seis *enhancers* de *D. virilis* son ortólogos a los *enhancers* del locus *svb* en *D. melanogaster*, ya que se ubican en las mismas posiciones relativas dentro del locus (Figura 11), tienen identidad de secuencia baja pero significativa, y además generan patrones de expresión similares en la epidermis embrionaria. Se ha propuesto que la conservación en la ubicación relativa de los *enhancers* podría implicar que el espaciamiento y orden de los *enhancers* son factores importantes para generar correctamente los contactos físicos entre

elementos regulatorios *in vivo* (*enhancer*-promotor y *enhancer-enhancer*) que permiten la correcta regulación de la expresión (Frankel et al 2012).

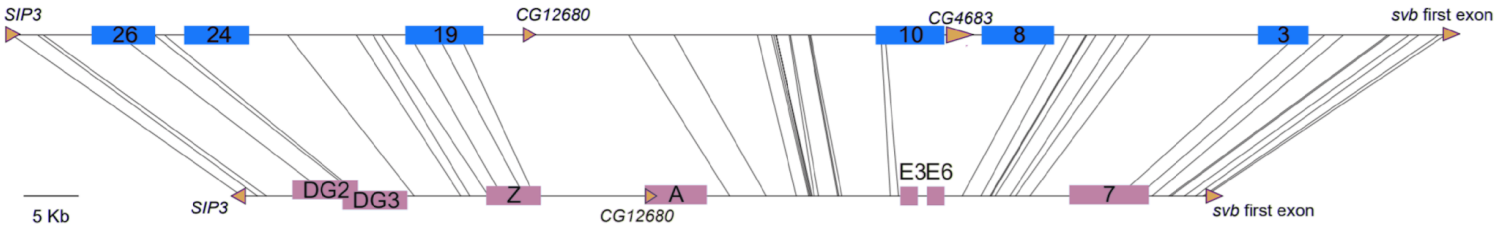


Figura 11. *Enhancers* ortólogos dentro del locus *svb* de *D. melanogaster* y *D. virilis*. Esquema comparativo entre las regiones regulatorias de *svb* de ambas especies. Arriba se esquematiza la región regulatoria de *D. virilis* y abajo la de *D. melanogaster*. Las líneas verticales marcan la posición de secuencias cortas (~30 pb) completamente conservadas entre las dos especies, denominadas "*anchors*". El hecho de que ninguna línea vertical se cruce muestra la conservación estructural del locus *svb.* Los rectángulos corresponden a los *enhancers* de cada especie, *D. virilis* en celeste y *D.* melanogaster en rosa. Los *enhancers* ortólogos están ubicados en las mismas posiciones relativas dentro de la región. Los triángulos amarillos marcan las regiones codificantes. Modificado de Frankel et al. 2012.

Al estudiar la actividad de los seis *enhancers* de *svb* en otros estadios de la ontogenia de *D. virilis,* se demostró que los mismos están activos en tejidos de la larva y de la pupa. Por ende, la actividad pleiotrópica de los *enhancers* transcripcionales de *svb* está conservada entre *D. melanogaster* y *D. virilis* (Figura 12) (Ignacio Mayansky, Tesis de licenciatura, FCEN-UBA, 2020).

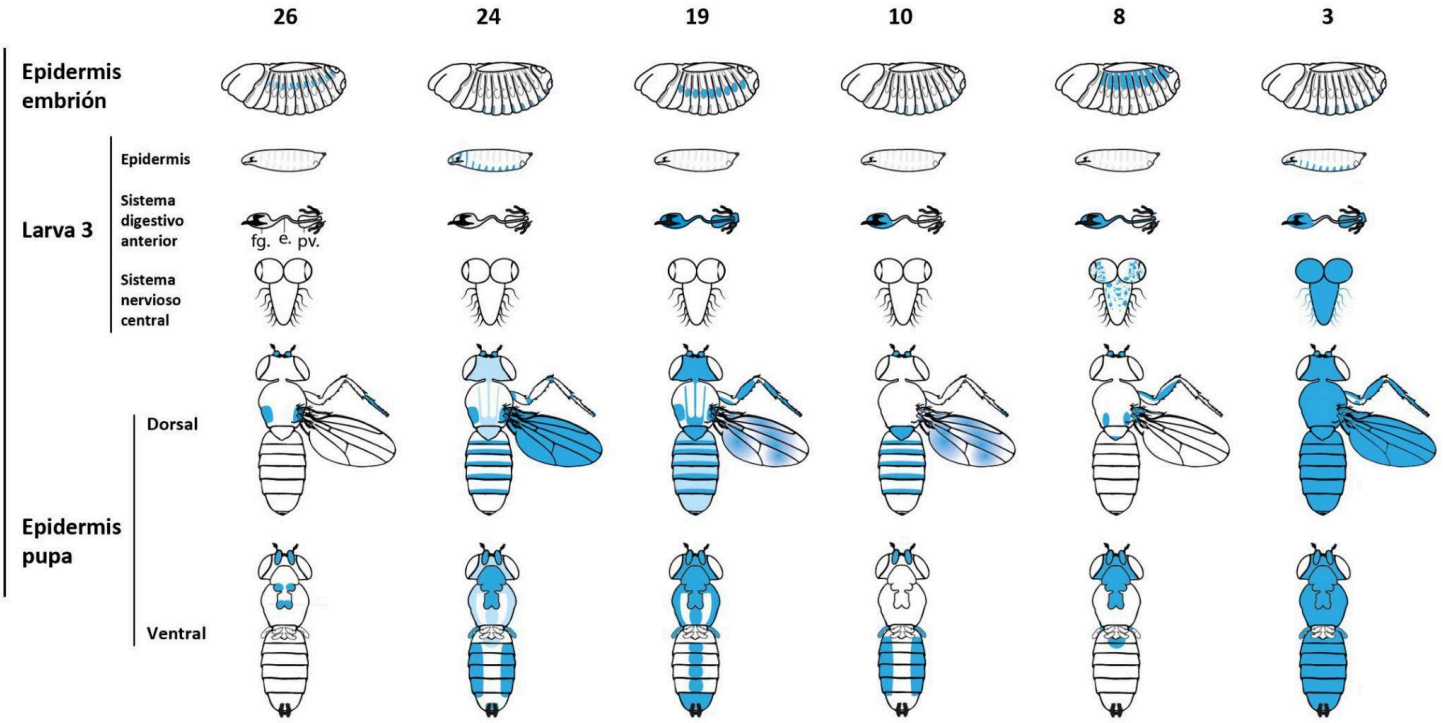


Figura 12. Los 6 *enhancers* de *svb* están activos en distintos estadios del ciclo de vida de *Drosophila virilis*. Esquema del patrón de expresión generado por cada uno de los 6 *enhancers* de *svb* a lo largo de la ontogenia de *D.virilis*.

La conservación de la actividad pleiotrópica de los *enhancer*s de *svb* nos llevó a pensar que la actividad pleiotrópica de los *enhancers* podría estar conservada a nivel genómico entre *D. melanogaster* y *D. virilis.* En otras palabras, creemos que una proporción grande de los *enhancers* que eran pleiotrópicos en el ancestro común de estas dos especies continúan siendo pleiotrópicos en *D. melanogaster* y *D. virilis.* La conservación evolutiva de la actividad pleiotrópica de los *enhancers* sugeriría que la misma es mantenida por selección.

2. Objetivos e Hipótesis

Nuestro laboratorio demostró que la actividad pleiotrópica de *enhancers* transcripcionales está conservada en el locus *svb* entre *Drosophila melanogaster* y *Drosophila virilis.* En este contexto, nos preguntamos si la conservación de la pleiotropía es generalizable para otros genes de *Drosophila*. Nos proponemos como objetivo general estudiar la conservación evolutiva de la pleiotropía de *enhancers* entre *D. melanogaster* y *D. virilis* a nivel genómico. Nos preguntamos si los *enhancers* que son pleiotrópicos en *D. melanogaster* mantienen su actividad pleiotrópica en *D. virilis* más frecuentemente que aquellos con actividad contexto-específica.

Nuestra hipótesis de trabajo es la siguiente:

- La actividad pleiotrópica de numerosos *enhancers* transcripcionales está conservada en el género *Drosophila*.

Los objetivos específicos asociados a esta hipótesis son:

- Identificar en el genoma de *D. virilis* a los ortólogos de dos *sets* de *enhancers* de *D. melanogaster*, utilizando distintos métodos de búsqueda.
- Analizar la actividad de *enhancers* ortólogos de *D. melanogaster* y *D. virilis* en distintos contextos del desarrollo para determinar si la actividad pleiotrópica de los *enhancers* está evolutivamente conservada.

3. Resultados

Para cumplir con nuestros objetivos, categorizamos *enhancers* de *D. melanogaster* como pleiotrópicos o contexto-específicos utilizando información de la estructura de la cromatina en diferentes contextos del desarrollo de esta especie. Posteriormente, realizamos predicciones de *enhancers* ortólogos (a los definidos en *D. melanogaster*) en el genoma de *D. virilis.* Finalmente, analizamos la estructura de la cromatina de dichos *enhancers* ortólogos a partir del análisis de datos de ATAC-seq y DNAse-seq en distintos contextos del desarrollo de *D. virilis*.

Idealmente, para poner a prueba la hipótesis de la conservación de la pleiotropía, deberíamos contar con datos para predecir la actividad de *enhancers* en al menos una decena de contextos del desarrollo de *D. melanogaster* y *D. virilis*. Al ser *D. melanogaster* una especie ampliamente estudiada, existe información genómica cuantiosa, disponible en bases de datos públicas. La situación es distinta para *D. virilis*, que ha sido menos estudiada hasta el momento. Esto constituye una limitación para realizar el estudio, ya que al contar con datos de de pocos contextos en *D. virilis*, la proporción de *enhancers* con actividad pleiotrópica será subestimada.

3.1. Predicción de *enhancers* en el genoma de *D. melanogaster*

Para evaluar la conservación de la actividad pleiotrópica de *enhancers* entre *D. melanogaster* y *D. virilis*, en primer lugar realizamos predicciones de *enhancers* activos en el genoma de *D. melanogaster.* Para predecir *enhancers* activos usamos datos genómicos de apertura de la cromatina (ATAC-seq y DNAse-seq) y de acetilación de histona 3 en la lisina 27 (ChIP-seq contra H3K27ac) correspondientes a 7 contextos espacio-temporales distintos: blastodermo, embrión 14-16 hs, cerebro de larva, disco imaginal de ojo-antena, disco imaginal de ala, disco imaginal de halterio y cerebro de adulto. Definimos a los *enhancers* putativos como regiones de cromatina abierta que contienen la marca epigenética H3K27ac en al menos

uno de los siete contextos estudiados. Intersectando la información de los siete contextos, consideramos como pleiotrópicos a aquellos *enhancers* que presentan apertura de la cromatina y H3K27ac en más de un contexto espacio temporal (para más detalles de la *pipeline* utilizada ver Materiales y Métodos). De esta manera, identificamos 8062 *enhancers* putativos.

A la vez, en nuestro laboratorio habíamos identificado 85944 *enhancers* a partir de datos de apertura de la cromatina en 13 contextos espacio-temporales: cerebro de adulto, blastodermo, embrión 6-8 hs, embrión 16-18 hs, disco imaginal de ojo-antena, disco imaginal de ala, disco imaginal de halterio, sistema nervioso central de larva, halterio, pata de adulto, ala, pata de pupa y disco imaginal de pata (Ian Laiker, Tesis de Licenciatura, FCEN-UBA, 2020). Para identificar este set de *enhancers* sólo se utilizó apertura de la cromatina como predictor (no se usaron datos de H3K27ac)

Luego, en ambos sets clasificamos los *enhancers* en contexto-específicos o pleiotrópicos. Un *enhancer* fue categorizado como pleiotrópico si y sólo si está activo en más de un contexto espacio-temporal. Para el set definido a partir de datos de apertura de la cromatina y H3K27ac encontramos que ~38%de los *enhancers* son pleiotrópicos y ~62% son contexto-específicos (Figura 13). En el set de *enhancers* definidos sólo a partir de la apertura de la cromatina, estimamos que la mitad de los *enhancers* son pleiotrópicos y la otra mitad contexto-específicos (Figura 13).

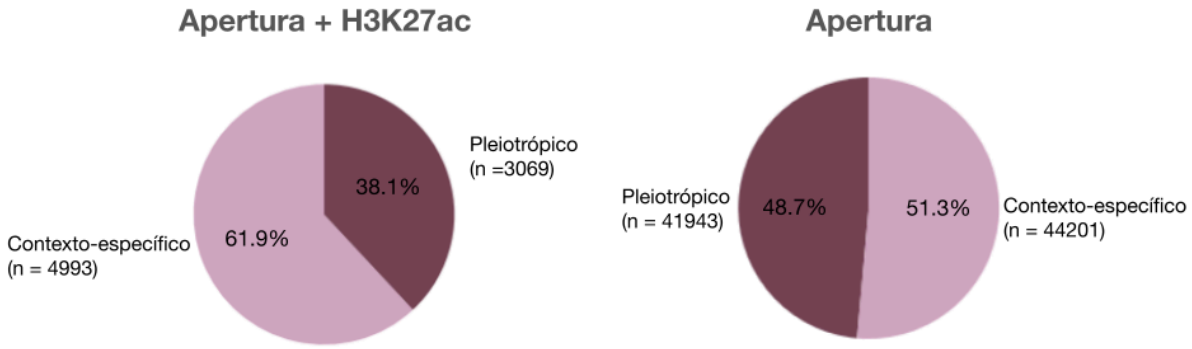


Figura 13. Proporción de *enhancers* pleiotrópicos y contexto-específicos en *D. melanogaster.* A la izquierda se muestran los resultados obtenidos para el set de *enhancers* definidos a partir de datos de apertura de la cromatina y H3K27ac, mientras que el gráfico de la derecha corresponde a los *enhancers* definidos sólo a partir de datos de apertura de la cromatina.

A su vez, estudiamos el grado de pleiotropía (la cantidad de contextos en los que un *enhancer* presenta actividad) de estos *enhancers.* Observamos que la mayoría de los *enhancers* están activos en uno, dos o tres contextos espacio-temporales, mientras que pocos *enhancers* presentan grados de pleiotropía altos (Figura 14).

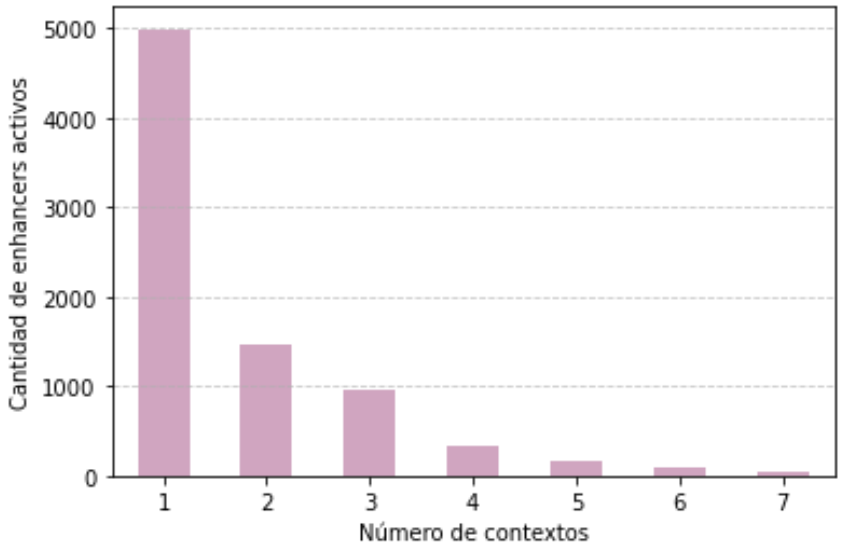


Figura 14. Frecuencia absoluta de enhancers activos por grado de pleiotropía. El gráfico corresponde al set de *enhancers* definidos a partir de datos de apertura de la cromatina y H3K27ac, el grado de pleiotropía (número de contextos) se representa en el eje X.

Asimismo, determinamos cuáles son los contextos en los que descubrimos una mayor cantidad de *enhancers.* Observamos que los discos imaginales de ala y de ojo-antena son los contextos que cuentan con mayor cantidad de *enhancers* activos (Figura 15A), mientras que los tiempos embrionarios (blastodermo y embrión de 14 a 16 hs) fueron los contextos con menor cantidad de *enhancers* predichos (Figura 15A).

Al observar que una proporción importante de *enhancers* en el genoma de *D. melanogaster* son contexto-específicos, nos preguntamos cómo se distribuyen este tipo de *enhancers* entre los distintos contextos estudiados. Observamos que el contexto espacio-temporal con mayor proporción de *enhancers* contexto-específicos es blastodermo, seguido por cerebro de larva, y cerebro del adulto (Figura 15B). Por otro lado, el disco imaginal de halterio, contexto con una frecuencia absoluta de *enhancers* intermedia (Figura 15B), presenta la menor proporción de *enhancers* contexto-específicos (Figura 15B). En cuanto a los

enhancers pleiotrópicos, una gran proporción está activa en los tres discos imaginales estudiados (Figura 15). Además, se observa que cerebro de larva y cerebro de adulto comparten una cantidad alta de *enhancers* activos, lo cual es consistente con lo encontrado utilizando *enhancers* definidos sólo a partir de datos de apertura de la cromatina (Figura Suplementaria 1).

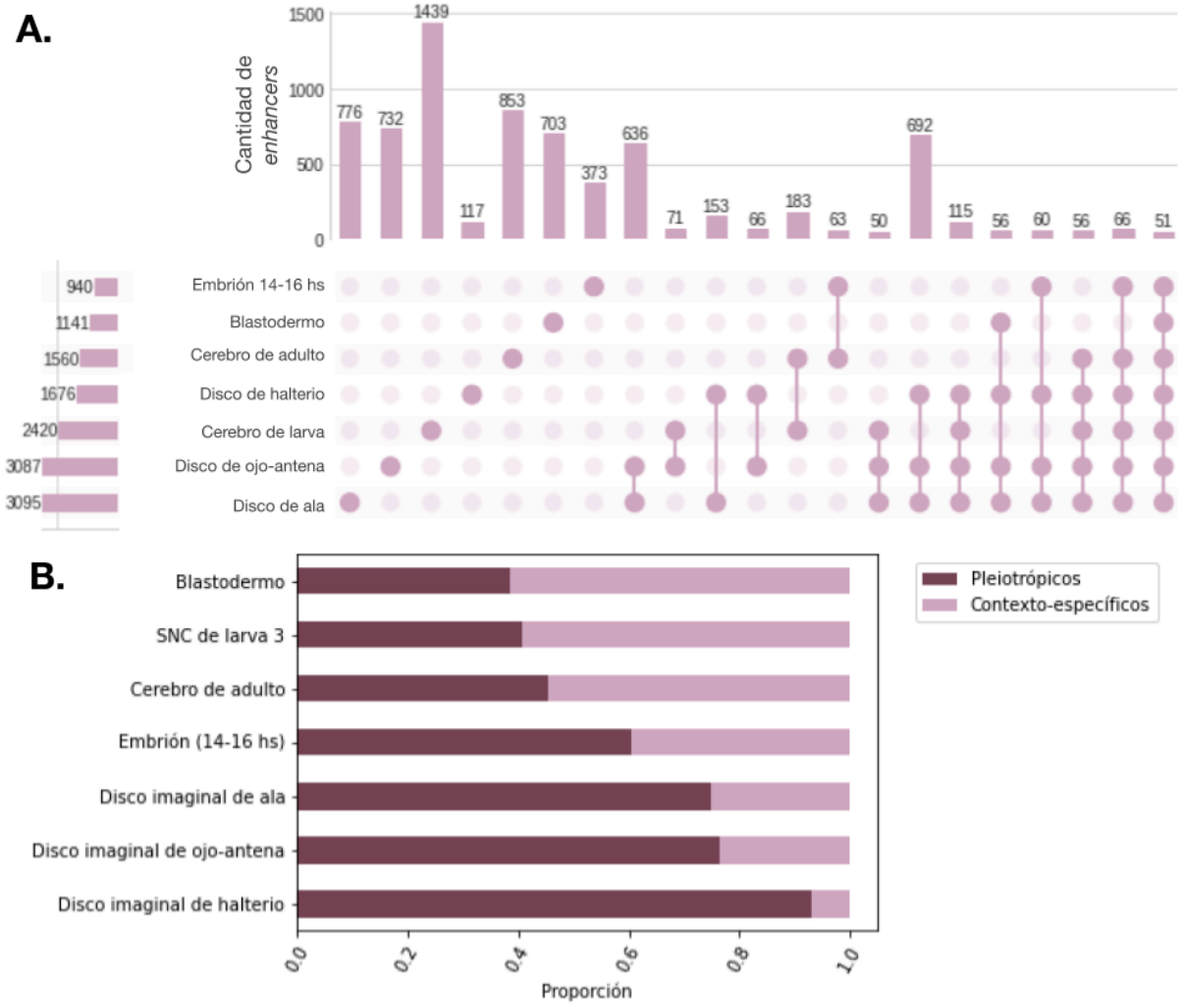


Figura 15. Distribución de *enhancers* pleiotrópicos y contexto-específicos por contexto. (A) Upset plot de *enhancers* de *D. melanogaster* definidos a partir de apertura de la cromatina y H3K27ac. Las barras verticales representan la cantidad de *enhancers* compartidos por los contextos indicados en la matriz de puntos (cantidad de *enhancers* por subset), en el gráfico solo se muestran cantidades mayores a 50. La cantidad de *enhancers* totales activos en cada contexto se muestra con las barras horizontales a la izquierda. **(B)** Proporción de *enhancers* pleiotrópicos y contexto-específicos por contexto espacio-temporal analizado.

3.2 Búsqueda de *enhancers* ortólogos en *D. virilis*

Una vez definidos los *enhancers* de *D. melanogaster* buscamos los *enhancers* ortólogos en el genoma de *D. virilis.* Para ello empleamos dos métodos distintos: *reciprocal-liftOver* y un método "*Alignment-free*". En las siguientes secciones, se detallan los resultados obtenidos utilizando cada uno de los métodos.

3.2.1 Identificación de *enhancers* ortólogos entre *D. melanogaster* y *D. virilis* utilizando *reciprocal-liftOver*

En primer lugar, intentamos identificar *enhancers* ortólogos a los de *D. melanogaster* en el genoma de *D. virilis* con un método basado en alineamientos. El método, al que llamamos *reciprocal-liftOver,* aprovecha la herramienta *liftOver* de manera de encontrar relaciones uno-a-uno entre secuencias de distintas especies. En otras palabras, al utilizar *reciprocal-liftOver*, cada *enhancer* de *D. melanogaster* es asignado a un único segmento en el genoma de *D. virilis* y viceversa. Este método permitió encontrar en el genoma de *D. virilis* al 74,66% de los *enhancers* de *D. melanogaster* definidos a partir de información de apertura de la cromatina y H3K27ac (Figura 16). Para el set de *enhancers* de *D. melanogaster* definidos usando datos de apertura de la cromatina, el porcentaje de *enhancers* que cuentan con un ortólogo en *D. virilis* fue muy similar, alcanzando el 70% (Figura 16).

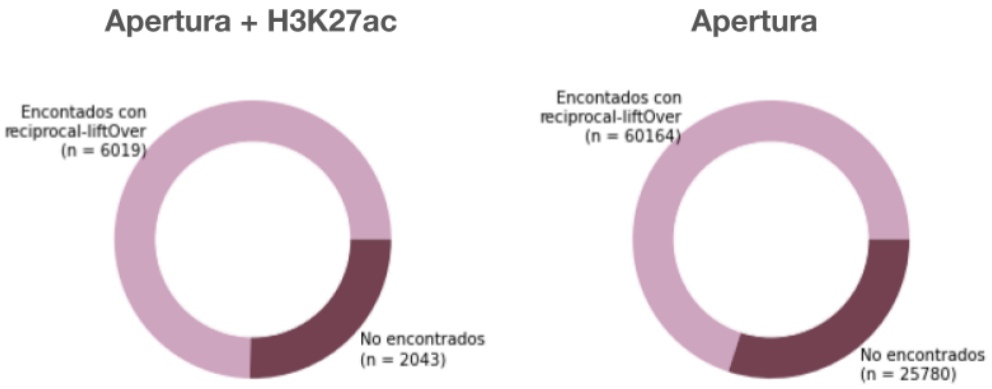


Figura 16. Cantidad de *enhancers* de *D. melanogaster* encontrados en el genoma de *D. virilis* utilizando *reciprocal-liftOver.* A la izquierda se muestra la cantidad de *enhancers* encontrados y no encontrados para el set de *enhancers* definidos a partir de datos de apertura de la cromatina y H3K27ac. El gráfico de la derecha corresponde a los *enhancers* definidos a partir de datos de apertura.

Al observar que cerca de un cuarto de los *enhancers* de *D. melanogaster* no podían ser encontrados en el genoma de *D. virilis*, nos preguntamos cuál o cuáles son las causas de que *reciprocal-liftOver* no logre hallarlos. Utilizando información generada por el mismo programa, observamos que la mayoría de los *enhancers* de *D. melanogaster* que *reciprocal-liftOver* no encuentra están presentes sólo parcialmente (la cantidad de bases encontradas no es suficiente para considerarlo una predicción exitosa) o no están presentes en el genoma de *D. virilis* (Figura 17). A su vez, para un número de casos similares no se obtuvo la ubicación original en *D. melanogaster* al realizar el mapeo recíproco, de modo que no es posible garantizar una relación de uno-a-uno para estos *enhancers* (Figura 17).

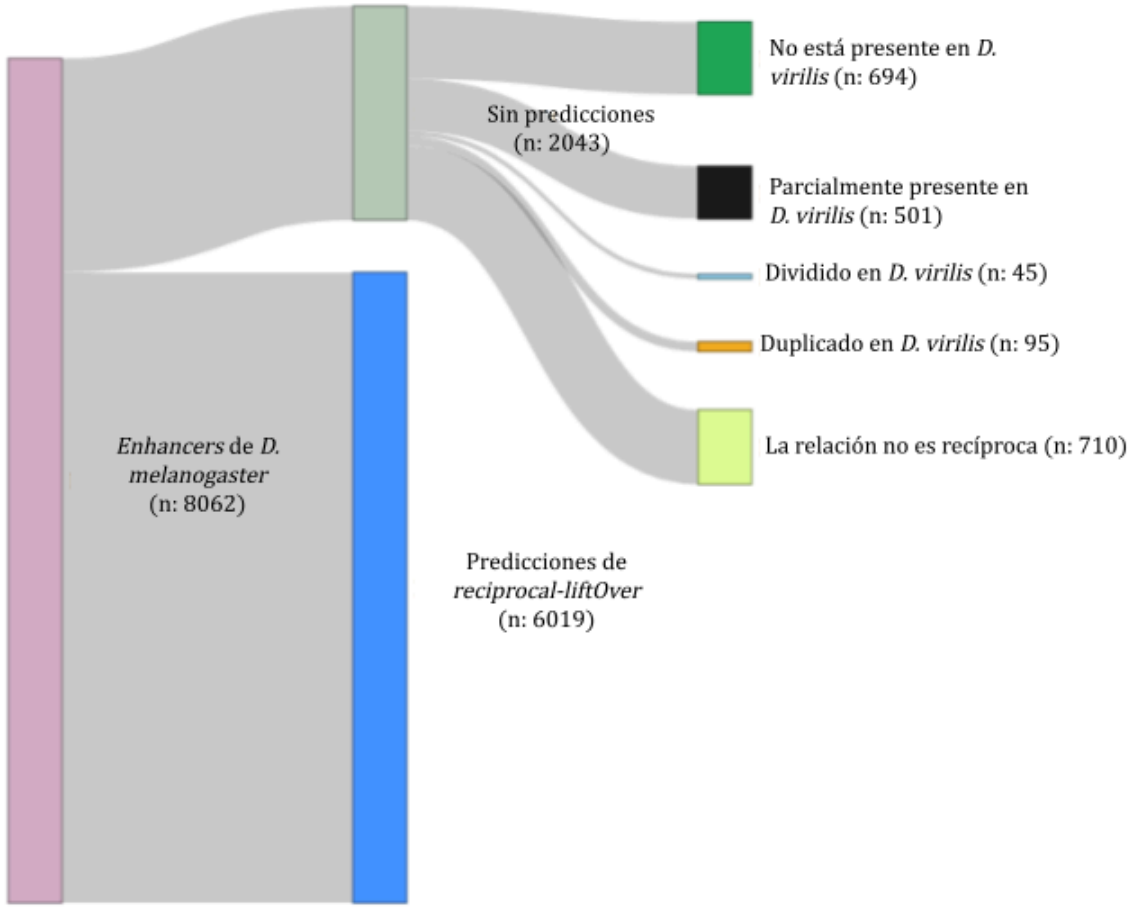


Figura 17. *Sankey plot* para *enhancers* de *D. melanogaster* con y sin ortólogos predichos por *reciprocal-liftOver* en el genoma de *D. virilis*. Cada barra vertical representa un conjunto. El sombreado gris indica la proporción del conjunto de la izquierda que forma parte del conjunto representado a la derecha. Los datos representados corresponden al set de *enhancers* definidos a partir de datos de apertura de la cromatina y H3K27ac.

3.2.1.1 Ubicación en el genoma de *D. melanogaster* de los *enhancers* que no poseen un ortólogo en *D. virilis*

Dado que para cerca de un cuarto de los *enhancers* de *D. melanogaster* no encontramos un posible ortólogo en *D. virilis*, nos preguntamos si la distribución de estos elementos es homogénea a lo largo del genoma de *D. melanogaster.* En otras palabras, nos preguntamos si los *enhancers* de *D. melanogaster* a los que no fue posible asignarles un ortólogo en *D. virilis* se concentran en una región específica del genoma. En términos generales, observamos que la distribución de los *enhancers* es relativamente homogénea a lo largo de los cromosomas de *D. melanogaster,* a excepción del cromosoma X, el cual presenta una mayor densidad de *enhancers* sin ortólogo que el resto de los autosomas (Figura 18). También se observa un único *enhancer* no encontrado en el cromosoma Y, y una alta densidad de *enhancers* que no están presentes en *D. virilis* en el cromosoma 4 (Figura 18).

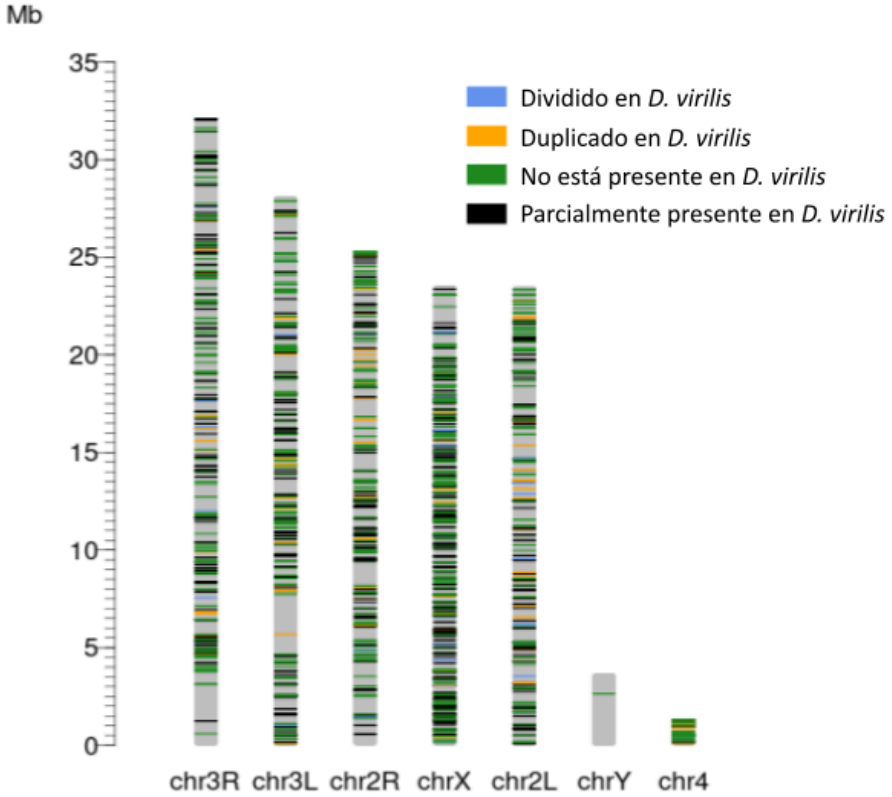


Figura 18. Ubicación de los *enhancers* de *D. melanogaster* que no tienen ortólogo en *D. virilis.* Las barras verticales grises representan a 7 cromosomas/brazos cromosómicos de *D. melanogaster,* donde su altura representa su tamaño. Las líneas horizontales muestran la ubicación de los elementos no encontrados en *D. virilis,* y su color representa el motivo por el cual no fueron encontrados.

Posteriormente cuantificamos la proporción de *enhancers* de *D. melanogaster* sin ortólogo en *D. virilis* por cromosoma (número de *enhancers* sin ortólogo dividido el número total de enhancers en el cromosoma). Los resultados fueron consistentes con lo observado anteriormente (Figura 18); aproximadamente el 30% de los *enhancers* ubicados en el cromosoma X no tienen un ortólogo asignado por *reciprocal-liftOver* en el genoma de *D. virilis.* Este porcentaje es mayor que el de los brazos cromosómicos 2R, 2L, 3R y 3L, que presentaron porcentajes de *enhancers* sin ortólogos predichos de aproximadamente 15% (Figura 19). Esto podría implicar que las regiones regulatorias del cromosoma X evolucionan a mayor velocidad que aquellas ubicadas en los autosomas 2 y 3.

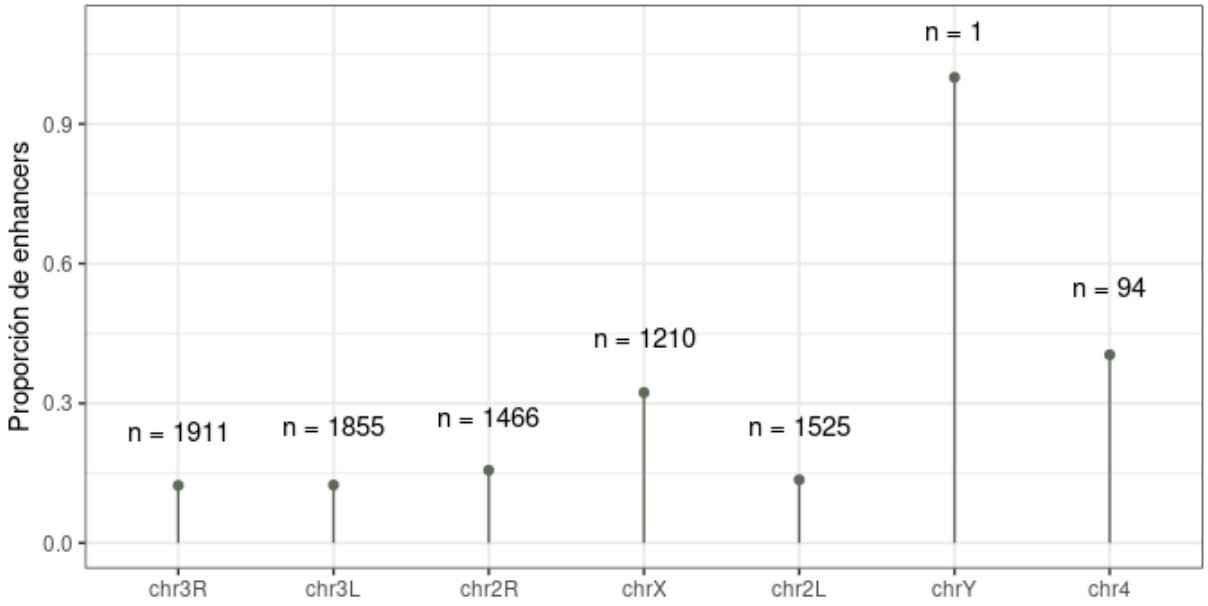


Figura 19. Proporción de *enhancers* sin ortólogos predichos por cromosoma/brazo cromosómico de *D. melanogaster.* La proporción fue calculada como número de *enhancers* sin ortólogo dividido el número total de *enhancers* por cromosoma. Arriba de cada barra se incluye la frecuencia absoluta de *enhancers* por cromosoma/brazo cromosómico.

A su vez, se observa que un único *enhancer* se ubica en el cromosoma Y de *D. melanogaster*, y el mismo no fue encontrado en *D. virilis.* También es importante destacar que aproximadamente el 40% de los *enhancers* del cromosoma 4 no parecen poseer un ortólogo en *D. virilis* (Figura 19).

3.2.1.2 Un gran porcentaje de los *enhancers* ortólogos está flanqueado por los mismos genes en las dos especies

Se sabe que los *enhancers* de *Drosophila melanogaster* suelen regular al gen más cercano a ellos (Massouras et al, 2012; Kvon et al, 2014). Teniendo esto en mente, nos preguntamos si los *enhancers* ortólogos están flanqueados por los mismos genes en los dos genomas analizados. Para responder esta pregunta, determinamos no sólo la ubicación de *enhancers* ortólogos en las dos especies, sino también la ubicación de genes ortólogos en los dos genomas utilizando *BLASTP* recíproco. En el análisis se tuvieron en cuenta solamente genes de *D. melanogaster* que posean ortólogos en *D. virilis* y viceversa. Para cada *enhancer,* se determinó el par de genes flanqueantes (el gen más cercano río arriba del *enhancer* y el gen más cercano río abajo del mismo) en las dos especies. Posteriormente, analizamos si estos genes eran los mismos en las dos especies. Observamos que más del 90% de los *enhancers* con ortólogo predicho por *reciprocal-liftOver* conserva al menos uno de sus dos genes flanqueantes, obteniéndose el mismo resultado para los dos sets de *enhancers* de *D. melanogaster* estudiados (Figura 20). A su vez, aproximadamente la mitad de los *enhancers* de los dos *sets* están flanqueados por los mismos genes en las dos especies (Figura 20). Nuestros resultados sugieren que la ubicación de los *enhancers* en relación a sus genes más cercanos posee restricciones evolutivas, dado que se encuentra conservada entre especies muy distantes.

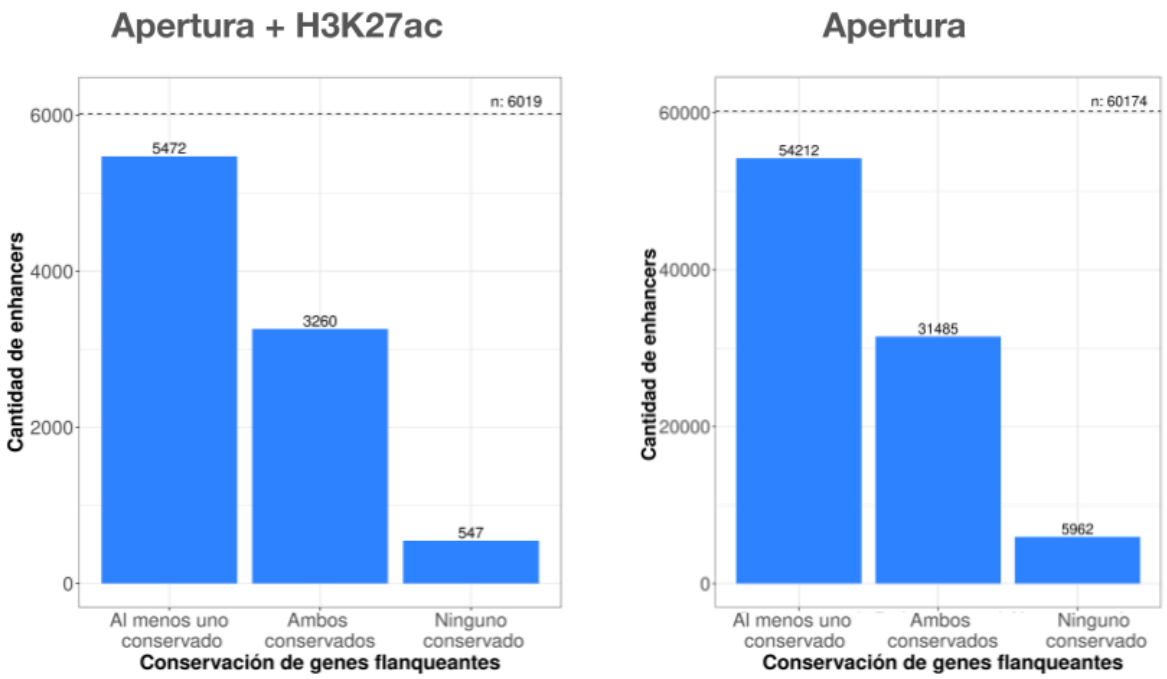


Figura 20. La mayoría de los *enhancers* de *D. melanogaster* que poseen un ortólogo en *D. virilis* están flanqueados por los mismos genes en las dos especies. El gráfico de la izquierda corresponde al set de *enhancers* definidos a partir de datos de apertura de la cromatina y H3K27ac, mientras que el de la derecha corresponde a los *enhancers* definidos sólo a partir de datos de apertura de la cromatina. La altura a la que se ubica la línea punteada representa la cantidad de *enhancers* de *D. melanogaster* con un ortólogo en *D. virilis* predicho por *reciprocal-liftOver.*

3.2.2 Identificación de *enhancers* ortólogos entre *D. melanogaster* y *D. virilis* utilizando un método basado en la presencia de secuencias cortas conservadas (método "*Alignment-free*")

Para buscar *enhancers* ortólogos a los de *D. melanogaster* en el genoma de *D. virilis,* implementamos un método de predicción que no utiliza alineamientos, al que nos referiremos como método "*Alignment-free*" (Arunachalam et al, 2010). Este método analiza la similitud en la composición de palabras (*k-mers*) entre las secuencias, independientemente del orden y orientación de los *k-mers*. La predicción de secuencias ortólogas se realiza en un espacio de búsqueda delimitado por la ubicación de los genes flanqueantes a cada *enhancer* de *D. melanogaster* en el genoma de *D. virilis.* Por ende, para realizar la predicción, se define para cada *enhancer* su región intergénica correspondiente en el genoma de *D. virilis.* Para

aproximadamente el 30% de los *enhancers* de *D. melanogaster* de ambos sets estudiados, no fue posible identificar una región intergénica ortóloga en el genoma de *D. virilis* (Figura 21). La incapacidad de definir una región intergénica ortóloga en *D. virilis* radica en la imposibilidad de encontrar a los genes flanqueantes al *enhancer* de *D. melanogaster* en el genoma de *D. virilis*.

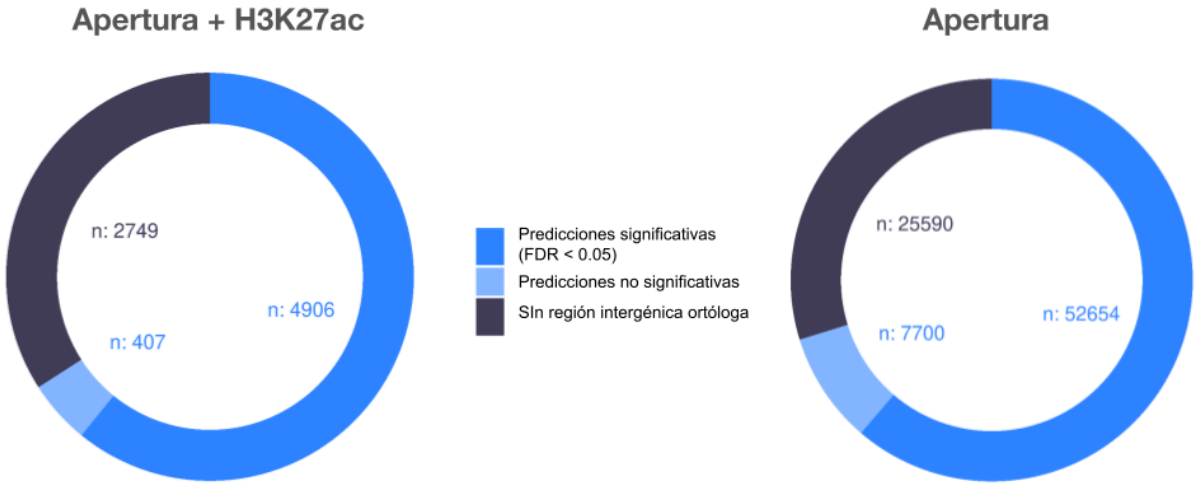


Figura 21. Predicción de *enhancers* ortólogos a partir del método "*Alignment-free*". A la izquierda se muestran los resultados obtenidos para el set de *enhancers* definidos a partir de datos de apertura de la cromatina y H3K27ac (Apertura + H3K27ac), mientras que a la derecha se muestran los resultados del set definido a partir de datos de apertura de la cromatina.

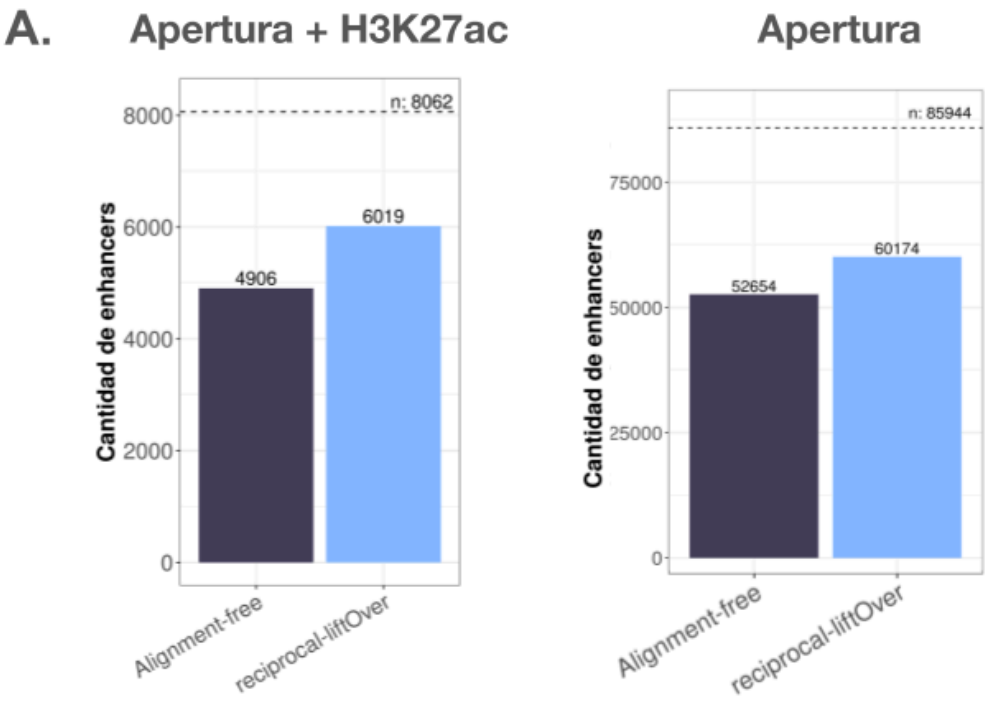
Para la mayoría de los *enhancers* de ambos sets, el método "*Alignment-free*" permitió identificar una secuencia ortóloga en el genoma de *D. virilis* (61,27% para apertura y 60,85% para apertura+H3K27ac) (Figura 21).

3.2.3 Comparación de los resultados obtenidos con *reciprocal-liftOver* y el método "*Alignment-free*"

Habiendo predicho *enhancers* ortólogos en el genoma de *D. virilis* utilizando dos métodos distintos, nos propusimos comparar los resultados de las predicciones. Resumiendo información mostrada en las secciones anteriores, *reciprocal-liftOver* logra predecir ortólogos en el genoma de *D. virilis* para una mayor cantidad de *enhancers* de *D. melanogaster* que el método "*Alignment-free*" (Figura 22A). Esto ocurre tanto para el set de *enhancers* definidos a

partir de datos de apertura de la cromatina, como para aquellos definidos a partir de información de apertura y acetilación. También cuantificamos la cantidad de casos en los que los dos métodos coinciden en las predicciones de ortólogos. Observamos que las predicciones coinciden para el 41,46% de los *enhancers* del set de apertura y para el 43,96% de los *enhancers* del set de apertura+acetilación (Figura 22B). Por ende, las predicciones no coinciden para un porcentaje importante de los *enhancers* de ambos sets (Figura 22B). Esto implica que los dos métodos podrían complementarse entre sí (*reciprocal-liftOver* podría predecir *enhancers* ortólogos que el método "*Alignment-free*" no predice y viceversa).

Es importante mencionar que para 420 *enhancers* de *D. melanogaster* pertenecientes al set obtenido a partir de datos de apertura de la cromatina y H3K27ac se obtuvieron predicciones no coincidentes entre los dos métodos. Este número fue de 4931 *enhancers* para aquellos definidos a partir de la apertura de la cromatina en *D. melanogaster.* Es decir, los casos en los que los métodos predicen ortólogos distintos en el genoma de *D. virilis* para un mismo *enhancer* de *D. melanogaster* son minoritarios.



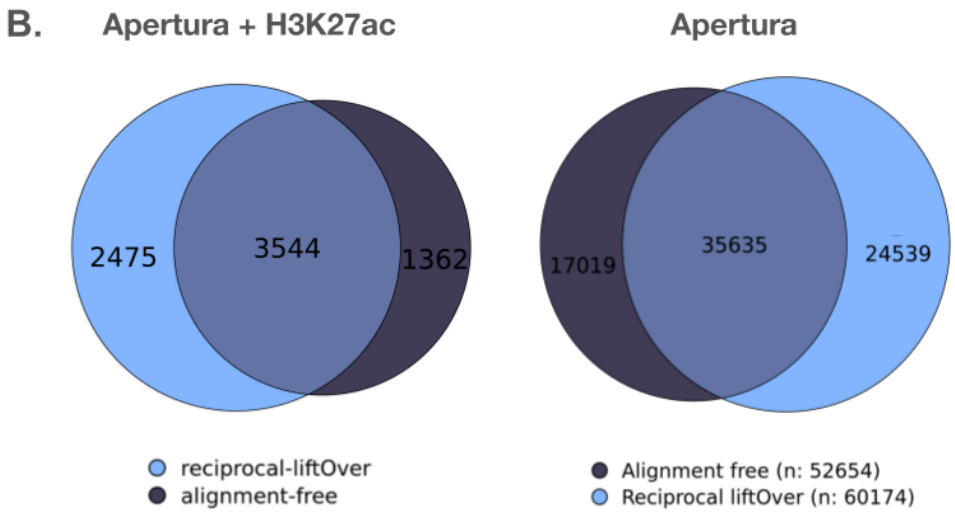


Figura 22. Comparación de las predicciones obtenidas utilizando *reciprocal-liftOver* y el método "*Alignment-free*". Tanto en (A) como en (B) se muestran a la izquierda los resultados obtenidos utilizando los *enhancers* de *D. melanogaster* definidos a partir de datos de apertura y acetilación de la cromatina, y a la derecha los obtenidos para los *enhancers* definidos a partir de datos de apertura de la cromatina únicamente. **(A)** Cantidad de *enhancers* con un ortólogo predicho por cada uno de los métodos. La altura a la que se ubica la línea punteada representa la cantidad total de *enhancers* totales del set correspondiente. **(B)** Diagrama de Venn que ilustra el solapamiento entre los dos conjuntos de predicciones.

3.2.4 El grado de pleiotropía de un *enhancer* de *D. melanogaster* no afecta la probabilidad de encontrar su ortólogo en *D. virilis*

Al estar activos en varios contextos espacio-temporales, es lógico pensar que los *enhancers* pleiotrópicos estén sometidos a mayor selección purificadora que aquellos *enhancers* contexto-específicos. En la misma línea, es posible pensar que las restricciones evolutivas serían más fuertes cuanto mayor sea el grado de pleiotropía de un *enhancer.* Nos preguntamos si estas restricciones evolutivas se traducirían en una conservación de secuencia en los *enhancers* pleiotrópicos que facilite la detección de sus *enhancers* ortólogos, sesgando así el análisis. Para responder esta pregunta estudiamos el grado de pleiotropía de los *enhancers* de *D. melanogaster* para los cuales fue posible encontrar un ortólogo en *D. virilis.* Si el grado de pleiotropía afectase la probabilidad de encontrar *enhancers* ortólogos, esperaríamos que los *enhancers* con un ortólogo predicho en *D. virilis* no respeten las proporciones de entre pleiotrópicos y contexto-específicos de *D. melanogaster.* Sin embargo,

observamos que la proporción de *enhancers* pleiotrópicos en *D. melanogaster* para los que se encontró un ortólogo en *D. virilis* es muy similar a la proporción original de *enhancers* pleiotrópicos definidos en *D. melanogaster,* independientemente del método utilizado para realizar la búsqueda de secuencias ortólogas (Figura 23). En otras palabras, el conjunto de *enhancers* que poseen un ortólogo en *D. virilis* no está enriquecido en *enhancers* con actividad pleiotrópica en *D. melanogaster.* Se obtuvieron los mismos resultados para el set de *enhancers* de *D. melanogaster* definido a partir de información de apertura y H3K27ac (Figura 23) y para el set definido sólo a partir de datos de apertura de la cromatina (datos no mostrados).

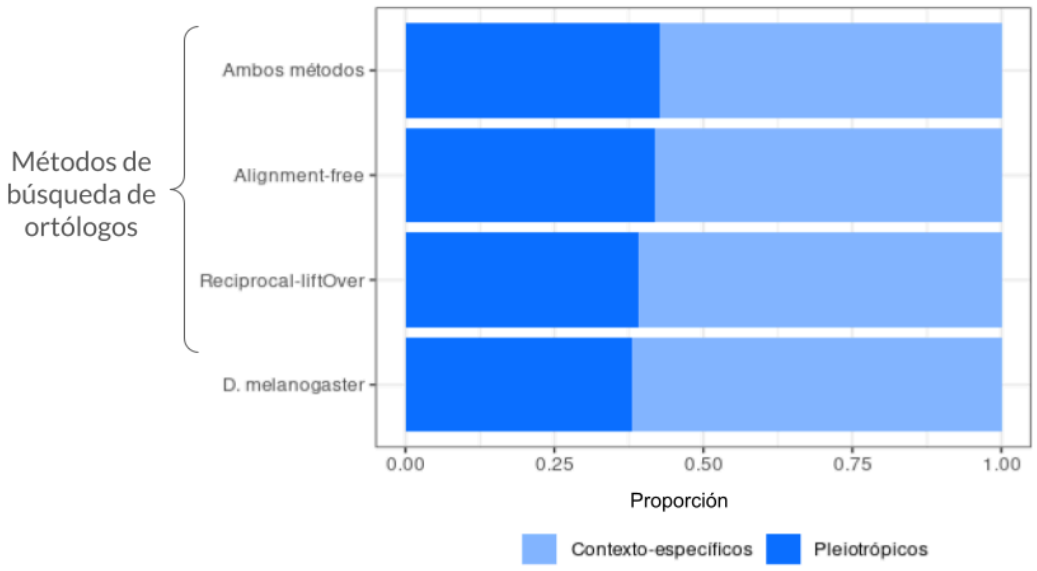


Figura 23. La actividad pleiotrópica en un *enhancer* de *D. melanogaster* no afecta la probabilidad de encontrar su ortólogo en *D. virilis.* La barra inferior muestra la proporción de *enhancers* pleiotrópicos pertenecientes al set de *enhancers* consenso de *D. melanogaster* definido a partir de datos de apertura y H3K27ac. El resto de las barras representan la proporción de *enhancers* pleiotrópicos que poseen un ortólogo en *D. virilis*, encontrado utilizando el método que se indica en el eje Y.

Al estudiar el grado de pleiotropía en *D. melanogaster* de aquellos *enhancers* que poseen un ortólogo en *D. virilis*, nuevamente encontramos que la proporción de *enhancers* para cada grado de pleiotropía se mantiene. Las proporciones correspondientes a los distintos métodos empleados para predecir ortólogos no presentan grandes diferencias entre sí (Figura 24). Una posible explicación para estos resultados es que la conservación a nivel secuencia de

cualquier *enhancer,* solo por el hecho de ser un elemento funcional, alcanza un umbral mínimo que permite la detección de ortólogos, independientemente de su grado de pleiotropía.

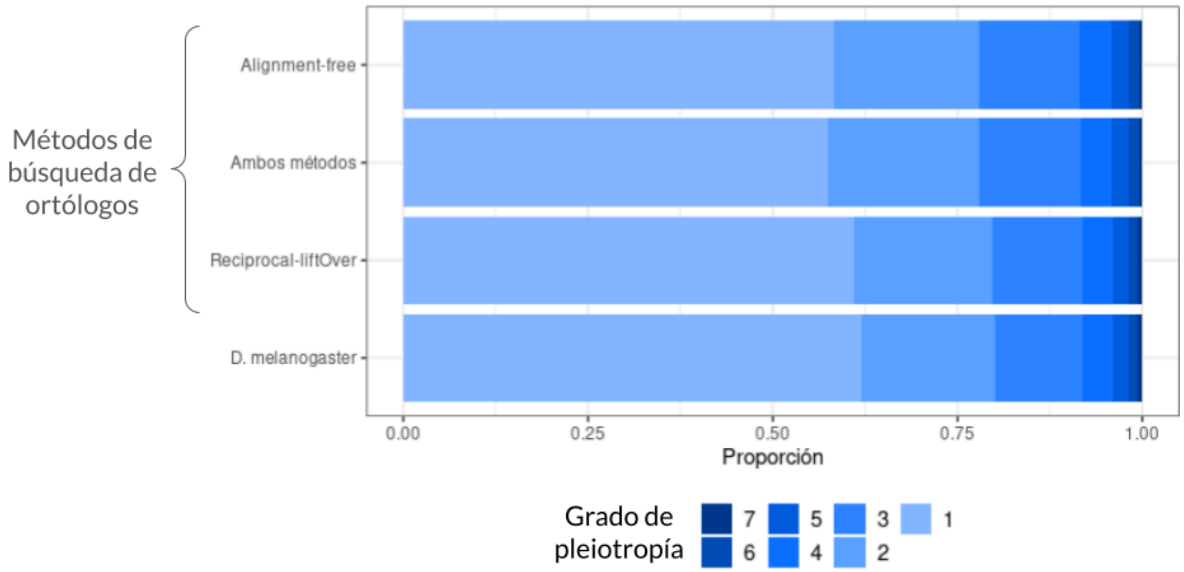


Figura 24. El grado de pleiotropía en un *enhancer* de *D. melanogaster* no afecta la probabilidad de encontrar su ortólogo en *D. virilis.* La barra inferior muestra la proporción de enhancers por grado de pleiotropía en el set de *enhancers* consenso de *D. melanogaster* definido a partir de datos de apertura y H3K27ac. El resto de las barras representan la proporción de *enhancers* que poseen un ortólogo para cada grado de pleiotropía, habiendo sido el ortólogo encontrado utilizando el método que se indica en el eje Y. Tonos más oscuros de azul representan grados de pleiotropía más altos.

3.3 Definición de *enhancers* consenso en *D. virilis*

Empleando una metodología similar a la utilizada para definir *enhancers* consenso en *D. melanogaster,* definimos elementos consenso activos en el genoma de *D. virilis* utilizando datos de apertura de la cromatina. Para estudiar la actividad pleiotrópica de *enhancers* es necesario identificar elementos activos en más de un contexto espacio-temporal. Al definir *enhancers* consenso, integramos la información de la estructura de la cromatina en distintos contextos del desarrollo, permitiéndonos categorizar *enhancers* como pleiotrópicos o contexto-específicos. Los datos de apertura de la cromatina utilizados corresponden a experimentos de DNAse-seq en embrión de 2 a 5 hs (blastodermo) y embrión de 19 a 22 hs

(estadio equivalente a embrión de 14-16 hs en *D. melanogaster*), y experimentos de ATAC-seq en discos imaginales de ojo-antena. Para la especie *D. virilis* no contamos con datos de H3K27ac en ninguno de estos contextos, por lo que consideramos como posibles *enhancers* activos a regiones de cromatina abierta. Logramos predecir la presencia de 28943 *enhancers* activos en el genoma de *D. virilis,* los cuales fueron categorizados como pleiotrópicos (activos en más de un contexto) o contexto-específicos (activos en un único contexto). Encontramos que un porcentaje importante de los *enhancers* son pleiotrópicos, representando casi el 40% del total (Figura 25A). Aproximadamente el 85% de los *enhancers* pleiotrópicos está activo en dos contextos espacio-temporales distintos, mientras que los restantes presentan actividad en los tres contextos estudiados (Figura 25B).

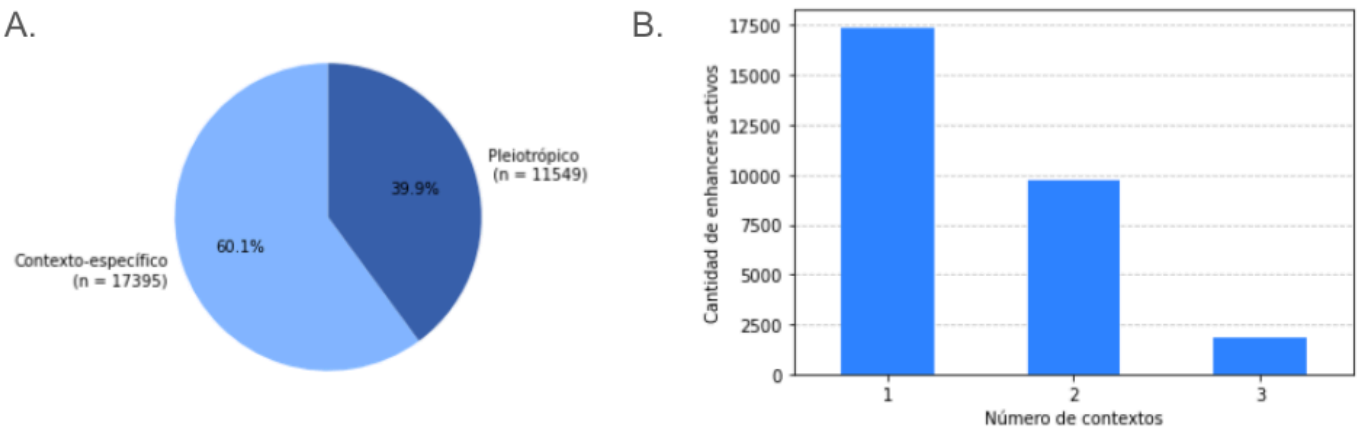


Figura 25. Actividad pleiotrópica de los *enhancers* consenso de *D. virilis.* (A) Gráfico de torta que ilustra la proporción de *enhancers* pleiotrópicos y contexto-específicos. **(B)** Cantidad de *enhancers* (eje Y) por grado de pleiotropía (número de contextos con actividad, eje X)

Al analizar cuántos *enhancers* están activos en las distintas combinaciones de los tres contextos analizados, encontramos que la mayoría de los *enhancers* contexto-específicos están activos en embrión de 2-5hs (blastodermo) ó en disco imaginal de ojo-antena (Figura 26). A su vez, observamos que una gran proporción de los *enhancers* pleiotrópicos corresponden a *enhancers* activos en los dos estadios de la embriogénesis estudiados (Figura 26).

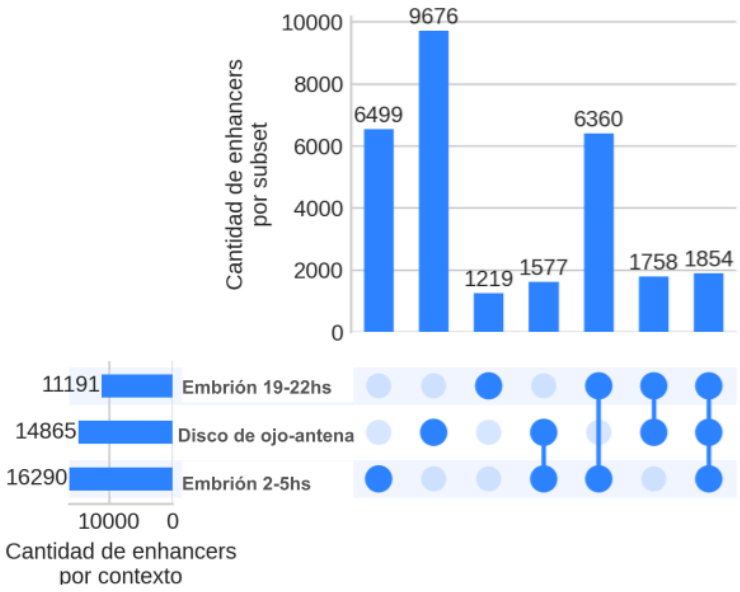


Figura 26. Upset plot de *enhancers* de *D. virilis* definidos a partir de información de apertura de la cromatina. Las barras verticales representan la cantidad de *enhancers* compartidos por los contextos indicados en la matriz de puntos. La cantidad de *enhancers* totales activos por contexto se muestra con las barras horizontales a la izquierda.

3.4 Actividad de los *enhancers* predichos en *D. virilis* y *D. melanogaster*

Teniendo las predicciones de *enhancers* en el genoma de *D. virilis* (sección 3.2) y habiendo definido elementos consenso activos de *D. virilis* (sección 3.3.), estudiamos la apertura de la cromatina en cada uno de los *enhancers* predichos. ¿Cuál es la proporción de elementos predichos que presenta cromatina abierta en *D. virilis?* ¿Alguno de los métodos utilizados funciona mejor para predecir *enhancers* ortólogos activos?

Para responder estas preguntas analizamos la superposición entre los *enhancers* predichos en *D. virilis* por los distintos métodos y los elementos consenso activos de *D. virilis.* Encontramos que ~30% de los *enhancers* predichos por ambos métodos están activos en al menos un contexto de los analizados en *D. virilis* (Figura 27). *Reciprocal-liftover* es el método que individualmente logra detectar una mayor proporción de *enhancers* activos, siendo la performance del método *"Alignment-free"* ligeramente menor (Figura 27).

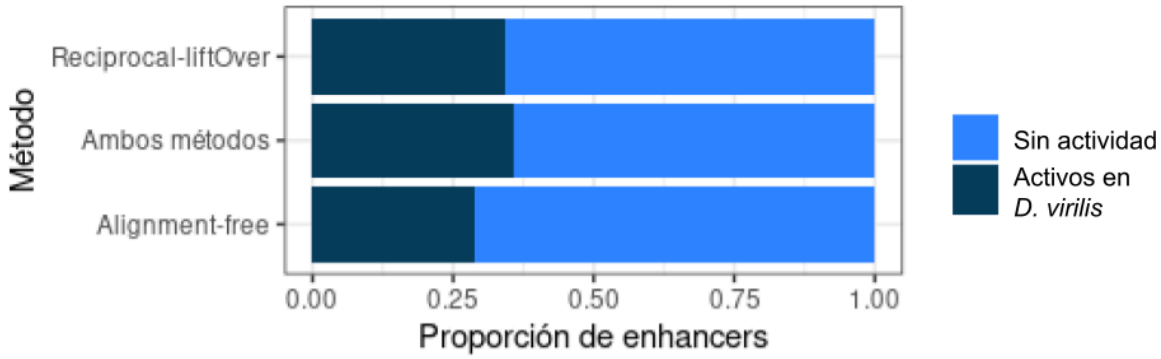


Figura 27. Actividad de los *enhancers* ortólogos predichos en *D. virilis*. Cada una de las barras representa un método distinto utilizado para buscar ortólogos. La porción de la barra color azul oscuro representa la proporción de predicciones con actividad en al menos uno de los tres contextos estudiados en *D. virilis.*

Posteriormente, nos preguntamos qué porcentaje de *enhancers* activos en un determinado contexto en *D. melanogaster* mantiene su actividad en el mismo contexto en *D. virilis.* Este análisis nos permitiría inferir cuál de los contextos estudiados tiene una mayor conservación de su repertorio de *enhancers*. Para realizar este análisis utilizamos los *enhancers* de *D. melanogaster* predichos con apertura y acetilación. Al intersectar los *enhancers* activos en el mismo contexto en las dos especies, observamos que el disco imaginal de ojo-antena es el contexto con mayor conservación de *enhancers* activos (Figura 28). Este resultado es independiente del método de predicción de ortólogos utilizado. Nuevamente, los porcentajes de *enhancers* activos en el mismo contexto en ambas especies fueron mayores cuando el método utilizado para predecir ortólogos en *D. virilis* fue *reciprocal-liftOver.* Según estas predicciones, casi la mitad de los *enhancers* de disco imaginal de ojo-antena están activos tanto en *D. virilis* como en *D. melanogaster.*

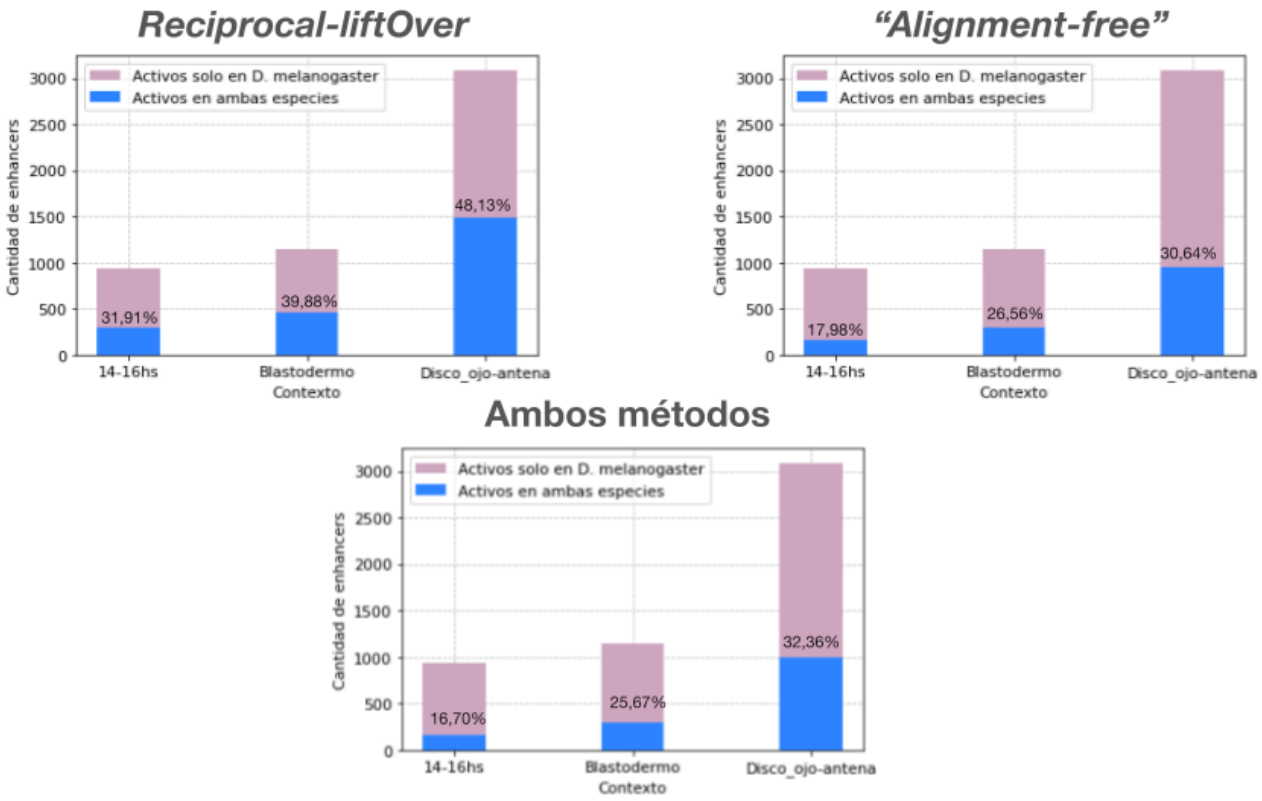


Figura 28. Cantidad de *enhancers* activos en un mismo contexto en las dos especies. Los porcentajes incluidos en los gráficos corresponden a los *enhancers* activos tanto en *D. melanogaster* como en *D. virilis* en el contexto indicado. La altura de la barra representa la cantidad total de *enhancers* de *D. melanogaster.*

3.5 Conservación de la actividad pleiotrópica de *enhancers* entre *D. melanogaster* y *D. virilis*

El objetivo principal de esta tesis es determinar si la actividad pleiotrópica de los *enhancers* está evolutivamente conservada entre *D. melanogaster* y *D. virilis*. Nos preguntamos si los *enhancers* que son pleiotrópicos en *D. melanogaster* mantienen su actividad pleiotrópica en *D. virilis* más frecuentemente que aquellos con actividad contexto-específica.

Para dar respuesta a este interrogante, analizamos a los *enhancers* de *D. melanogaster* que pueden ser categorizados como pleiotrópicos estudiando su actividad en blastodermo, embrión tardío y disco imaginal de ojo-antena. De esta manera, serían considerados pleiotrópicos aquellos enhancers activos en al menos dos de estos contextos. Elegimos categorizar a los *enhancers* a partir de su actividad en estos tres contextos puesto que

contamos con información de estructura de la cromatina en las dos especies. De esta manera, los enhancers ortólogos de *D. virilis* también fueron categorizados en pleiotrópicos o contexto-específicos utilizando información de contextos equivalentes.

Para el set de *enhancers* de *D. melanogaster* definidos a partir de datos de apertura de la cromatina y H3K27ac utilizamos información de tres contextos equivalentes entre *D. melanogaster* y *D. virilis*. Sin embargo, para el set de *enhancers* de *D. melanogaster* definido a partir de información de apertura de la cromatina (que fue obtenido antes de la realización de esta tesis) no contábamos con información equivalente a embrión de 18-22 hs en *D. virilis* (14-16 hs en *D. melanogaster)*. Esto representa un problema a la hora de evaluar la conservación de la pleiotropía utilizando estadios equivalentes entre las dos especies. Para solucionar este problema, analizamos la actividad de *enhancers* en embrión de *D. melanogaster* de 14-16 hs a partir de la superposición entre los elementos consenso previamente definidos y los picos de apertura correspondientes a este estadio. Así, un *enhancer* incluído en el set que se superponga en al menos un 50% de su extensión con un pico de apertura correspondiente a embrión de 14-16 hs es considerado un *enhancer* activo en dicho estadio. Entonces, definimos como *enhancers* contexto-específicos de embrión 14-16 hs a aquellos picos de apertura de la cromatina que no se superponen con ninguno de los elementos consenso previamente definidos.

Por otro lado, consideramos que un *enhancer* es contexto-específico cuando está activo solamente en blastodermo o embrión tardío o disco de ojo-antena, incorporando también la información de todos los contextos estudiados en *D. melanogaster*. Es decir, para el set de *enhancers* de *D. melanogaster* definidos sólo a partir de apertura de la cromatina, consideramos como contexto-específico a los *enhancers* activos en sólo uno de los 3 contextos y que no estuviesen activos en los 10 contextos estudiados, y para los definidos con datos de apertura y acetilación consideramos como contexto-específico a los *enhancers* activos en sólo uno de los 3 contextos y que no estuviesen activos en los otros 4 contextos estudiados. De esta manera, logramos clasificar a 14126 *enhancers* como contexto-específicos para el set de apertura (Figura 29A) y 1808 para el set de apertura y H3K27ac (Figura 29C).

A partir de este análisis, determinamos que la proporción de *enhancers* con actividad pleiotrópica conservada es mayor que la proporción de *enhancers* contexto-específicos

(definidos como contexto-específicos en *D. melanogaster*) con actividad conservada en las dos especies (Figura 29B y D). Este resultado es obtenido independientemente del set de *enhancers* empleado y el método de búsqueda de ortólogos utilizado. Que la proporción de *enhancers* con actividad pleiotrópica conservada sea mayor a la proporción de *enhancers* contexto-específicos que están activos en las dos especies demuestra que la pleiotropía de *enhancer* es una función evolutivamente conservada entre especies distantes.

Para el set de apertura, observamos que el porcentaje de *enhancers* con actividad pleiotrópica conservada alcanza un 19,89% cuando se emplea *reciprocal-liftOver* para predecir *enhancers* ortólogos en *D. virilis* (Figura 29B). Este porcentaje es considerablemente mayor que el 12,38% que representan los *enhancers* contexto-específicos de *D. melanogaster* que también están activos en *D. virilis* (Figura 29B).

Para el set de *enhancers* definidos a partir de apertura de la cromatina y H3K27ac las mayores proporciones de *enhancers* conservados también se obtienen cuando *reciprocal-liftOver* es el método de predicción de ortólogos utilizado. En este caso, el 41,92% de los *enhancers* pleiotrópicos de *D. melanogaster* conservan su actividad pleiotrópica en *D. virilis*, mientras que el 32,46% de *enhancers* contexto-específico en *D. melanogaster* también están activos en *D. virilis* (Figura 29D).

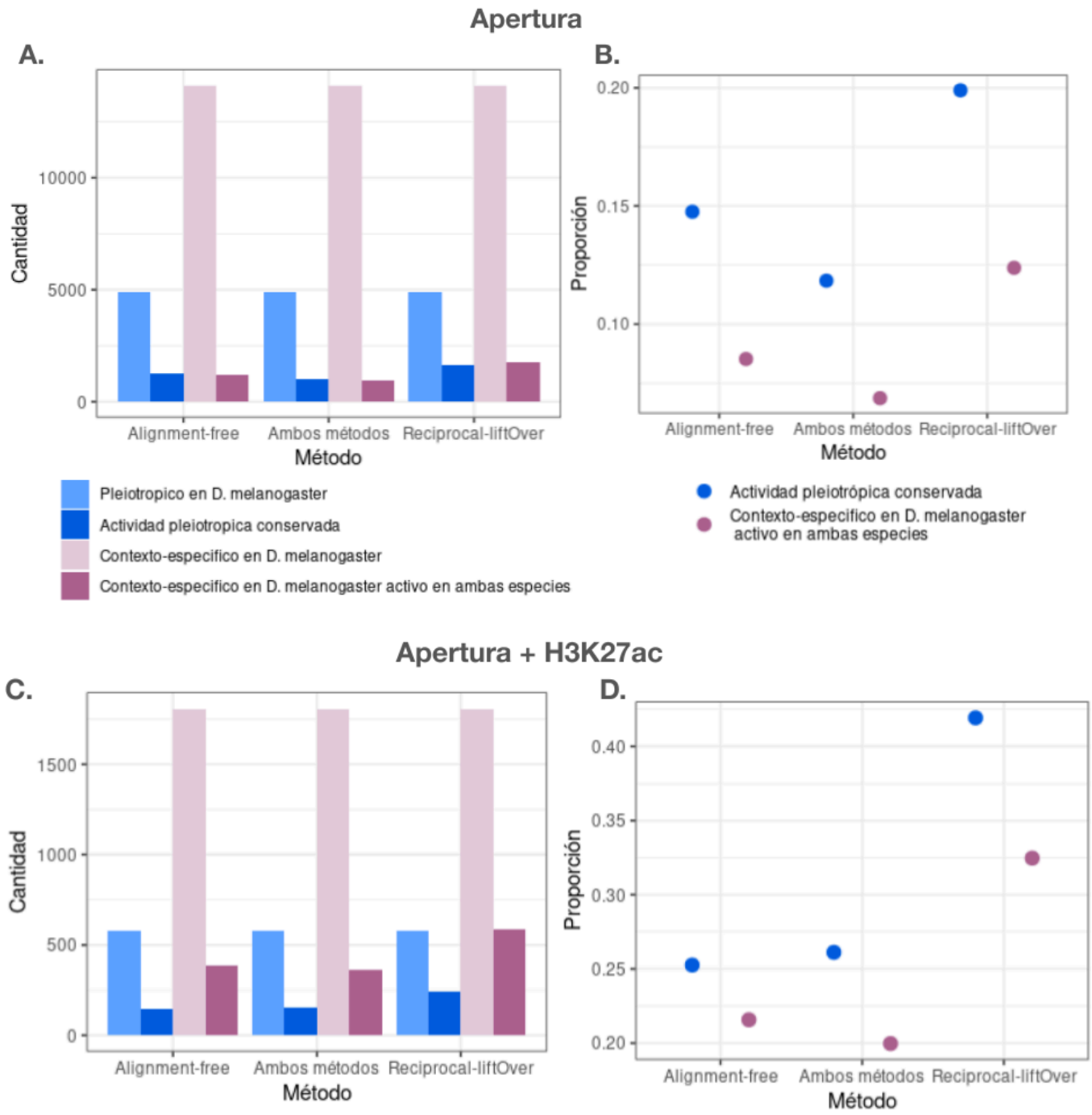


Figura 29. Conservación de la actividad pleiotrópica de *enhancers* entre *D. melanogaster* y *D. virilis.* (A) y **(B)** Resultados para *enhancers* de *D. melanogaster* definidos por apertura de cromatina. **(C) y (D)** Resultados para enhancers definidos por apertura de cromatina y presencia de H3K27ac. En los cuatro paneles, el eje X representa el método de predicción de *enhancers* ortólogos en *D. virilis.* **(A) y (C)** Se muestra en tonos de azul la cantidad de *enhancers* pleiotrópicos en *D. melanogaster* junto con los que son pleiotrópicos en ambas especies,y en tonos de rosa la cantidad de *enhancers* contexto-específicos y los que están activos en ambas especies. **(B) y (D)** En azul se indica la proporción de *enhancers* pleiotrópicos en *D. melanogaster* con actividad conservada en ambas especies, y en magenta la proporción de *enhancers* contexto-específicos de *D. melanogaster* con actividad también en *D. virilis*.

4. Discusión y conclusiones

En esta tesis corroboramos que un porcentaje alto de los *enhancers* de *D. melanogaster* son pleiotrópicos. Observamos que aproximadamente el 40% de los *enhancers* identificados a partir de información de apertura de la cromatina y H3K27ac tienen actividad pleiotrópica en *D. melanogaster* (Figura 13). Anteriormente en nuestro laboratorio habíamos calculado que aproximadamente la mitad de los *enhancers* son pleiotrópicos, usando datos de apertura de la cromatina (Ian Laiker, Tesis de licenciatura, FCEN-UBA, 2020). Esta diferencia puede ser explicada por la cantidad de contextos espacio-temporales considerados en cada análisis. El set de *enhancers* definido en esta tesis corresponde a información de 7 tejidos y/o momentos del desarrollo distintos, mientras que el otro set de *enhancers* estudiado incorpora información de 13 contextos diferentes. Teniendo en cuenta que se ha demostrado que la proporción de *enhancers* pleiotrópicos aumenta con el número de contextos espacio-temporales incorporados al análisis (Laiker y Frankel, 2022), es esperable que el set definido en esta tesis tenga una menor proporción de *enhancers* con actividad pleiotrópica.

Observamos que tener en cuenta datos de acetilación para identificar *enhancers* cambia sustancialmente la cantidad de *enhancers* putativos obtenidos. El set de *enhancers* definido en esta tesis incorpora información de H3K27ac, lo cual constituye un requisito adicional para considerar que una región de cromatina abierta es un *enhancer* putativo. Al ser más exigentes en la predicción de *enhancers* y estar considerando información correspondiente a un número menor de contextos, no es sorprendente que el set definido en esta tesis esté compuesto por una cantidad menor de elementos consenso.

En el set de *enhancers* definido con información de apertura de la cromatina y de H3K27ac, observamos que un gran porcentaje de los *enhancers* pleiotrópicos están activos en los discos imaginales de ala, de halterio y de ojo-antena (Figura 15). Estos resultados son consistentes con datos reportados previamente, que indican que los discos imaginales cuentan con un repertorio compartido de *enhancers* (McKay y Lieb, 2013). Es decir, a pesar de

dar lugar a estructuras morfológicas muy distintas, los discos imaginales tienen redes de regulación genética que usan mayormente los mismos *enhancers* (McKay y Lieb, 2013).

Al buscar *enhancers* ortólogos a los de *D. melanogaster* en el genoma de *D. virilis* usando *reciprocal-liftOver*, notamos que el cromosoma X de *D. melanogaster* posee una mayor proporción de *enhancers* sin ortólogo en *D. virilis.* Comparando el cromosoma X con los brazos cromosómicos 2R, 2L, 3R y 3L, que tienen tamaños y cantidades de *enhancers* similares, encontramos que la proporción de *enhancers* sin ortólogo es el doble en el cromosoma X (Figura 19). Nuestros hallazgos son consistentes con la idea de que los cambios en la expresión génica en *Drosophila* son más rápidos para los genes en el cromosoma X que para los genes en los autosomas (el efecto "*faster-X*") (Meisel et al, 2012). La evolución de la expresión génica más rápida en el X también fue observada en mamíferos terios (Brawand et al, 2011). La mayor divergencia en la expresión génica de genes del cromosoma X de *Drosophila* podría estar explicada por la ganancia/pérdida o la divergencia a nivel secuencia de elementos regulatorios en el cromosoma X a lo largo de la evolución. Por otro lado, debido a las características particulares del cromosoma 4 (elemento de Muller F), como su alto contenido de heterocromatina y su alto contenido repetitivo, no es sorprendente que el mismo presente una proporción alta de *enhancers* sin ortólogo en *D. virilis* (Figura 19).

La búsqueda de ortólogos con *reciprocal-liftOver* también nos permitió determinar que la mayoría de los *enhancers* están flanqueados por los mismos genes en las dos especies, a pesar de los ~40 millones de años de divergencia entre ellas. Sabiendo que los *enhancers* de *D. melanogaster* suelen regular al gen más cercano (Massouras et al, 2012; Kvon et al, 2014), la conservación posicional de los *enhancers* con respecto a sus genes flanqueantes podría indicar la presencia de restricciones evolutivas. Es decir, los *enhancers* mantendrían sus posiciones a lo largo de la evolución "para" conservar interacciones con el promotor basal de su(s) gen(es) blanco.

Al buscar *enhancers ortólogos* mediante el método "*Alignment-free*", en muchos casos no fue posible establecer un espacio de búsqueda (una región ortóloga) en *D. virilis.* Este es el principal motivo por el cual encontramos menos *enhancers* ortólogos con este método, comparado con *reciprocal-liftOver.* A futuro, podrían introducirse mejoras en el método "*Alignment-free*" que permitan definir una región de búsqueda para una mayor cantidad de *enhancers*. Por ejemplo, se podría tener en cuenta una mayor cantidad de genes flanqueantes

(no solamente los dos genes más cercanos al *enhancer),* para definir esta región. A su vez, contar con mejores anotaciones de genes en las especies estudiadas también podría aumentar el número de predicciones exitosas.

En términos generales, para un determinado *enhancer* de *D. melanogaster* los dos métodos suelen predecir la misma ubicación en *D. virilis*, a pesar de estar basados en propiedades diferentes de las secuencias genómicas. Esto confiere robustez a nuestras predicciones. A su vez, los métodos poseen una performance similar en cuanto a la detección de *enhancers* ortólogos activos (Figura 27). En ambos casos, más del 25% de los ortólogos predichos tienen actividad en *D. virilis*. A priori, este porcentaje de *enhancers* activos en *D. virilis* puede parecer bajo, pero es importante recordar que las predicciones provienen de estudiar la actividad de *enhancers* de *D. melanogaster* en 7 contextos distintos (set apertura + H3K27ac), mientras que en *D. virilis* solo contamos con información de 3 contextos. Claramente, dicho porcentaje sería mayor si se estudiase la actividad de estos *enhancers* predichos en una mayor cantidad de contextos en *D. virilis.* Es decir*,* para obtener un número preciso de *enhancers* ortólogos activos en las dos especies, debería estudiarse el mismo número de contextos espacio-temporales en ambas especies.

Resultados previos de nuestro laboratorio y de otros grupos sugieren que cuanto más pleiotrópico es un *enhancer,* mayor es su conservación evolutiva (Fish et al, 2017; Singh y Yi, 2021, Ian Laiker, Tesis de Licenciatura, FCEN-UBA, 2020) . Utilizando scores de *PhastCons*, se observó que el grado de pleiotropía de un *enhancer* está correlacionado positivamente con la conservación evolutiva de su secuencia (Ian Laiker, Tesis de Licenciatura, FCEN-UBA, 2020). Teniendo esto en mente, esperábamos que los *enhancers* pleiotrópicos de *D. melanogaster* fuesen más fáciles de encontrar en el genoma de *D. virilis.* Sin embargo, independientemente del método utilizado para buscar *enhancers* ortólogos, se observó que la probabilidad de encontrar un ortólogo es independiente de si el *enhancer* es pleiotrópico o contexto-específico. Estos resultados, que en un principio pueden parecer contrapuestos, fueron obtenidos mediante metodologías muy distintas: el *score* de *PhastCons* incorpora información de 27 especies distintas (*phastCons27way*), mientras que en esta tesis analizamos solo dos especies. Nuestros resultados podrían ser explicados por el hecho de que la conservación a nivel secuencia de cualquier *enhancer*, solo por el hecho de ser un elemento funcional, alcanza

un umbral mínimo que permite la detección de ortólogos, independientemente de su grado de pleiotropía.

Nuestros resultados demuestran que la proporción de *enhancers* con actividad pleiotrópica conservada es mayor que la proporción de *enhancers* contexto-específicos que están activos en las dos especies (Figura 29). Este resultado sugiere que la actividad pleiotrópica de los *enhancers* es importante en los genomas animales, ya que está conservada evolutivamente en especies distantes. Creemos que la proporción de *enhancers* con actividad pleiotrópica conservada calculada en esta tesis es una subestimación, puesto que solamente disponemos de información correspondiente a tres contextos en *D. virilis.* Resultados previos de nuestro laboratorio demuestran que la proporción de *enhancers* pleiotrópicos de *D. melanogaster* aumenta al considerar más contextos espacio-temporales en el análisis (Ian Laiker, Tesis de Licenciatura, FCEN-UBA, 2020). Es razonable suponer que esto también ocurre en *D. virilis.* Si al estudiar más contextos la proporción de *enhancers* con actividad pleiotrópica aumenta en *D. melanogaster* y *D. virilis*, es plausible que también aumente la proporción de *enhancers* pleiotrópicos conservados entre ambas especies. Creemos que la proporción de *enhancers* contexto-específicos con actividad conservada en las dos especies podría mantenerse, o inclusive disminuir, al incorporar más contextos al análisis. Esto se explicaría por dos efectos contrapuestos: por un lado al incorporar un nuevo contexto estaríamos también incorporando *enhancers* contexto-específicos activos en ese contexto y, por otro lado, *enhancers* previamente categorizados como contexto-específicos que estén activos en este nuevo contexto pasarían a ser considerados pleiotrópicos.

Al identificar *enhancers* con actividad pleiotrópica conservada podemos suponer que estos *enhancers* eran pleiotrópicos en el ancestro común de *D. melanogaster* y *D. virilis*. Sin embargo, un *enhancer* podría ser pleiotrópico en *D. melanogaster* y *D. virilis* por convergencia. Dado que estudiamos la pleiotropía de *enhancers* ortólogos en los mismos contextos en las dos especies, la conservación de la actividad pleiotrópica ancestral es el escenario más probable, pero, dado que estudiamos solamente dos especies, no puede descartarse un escenario de convergencia. Para un análisis riguroso de la conservación deberían incorporarse más especies al análisis, de forma de poder inferir el estado ancestral en la actividad de cada uno de los *enhancers* estudiados. Incorporar más especies al análisis ayudaría a entender si los *enhancers* contexto-específicos surgen evolutivamente como tales, o si derivan de

enhancers pleiotrópicos que van perdiendo actividad a lo largo de su evolución. En este último caso, esperaríamos encontrar que la proporción de *enhancers* ancestralmente pleiotrópicos que pasan a ser contexto-específicos en las especies derivadas sea mayor que la proporción de *enhancers* contexto-específicos que se convierten en pleiotrópicos.

5. Materiales y métodos

5.1 Datos públicos utilizados

Todos los datos utilizados en esta tesis fueron extraídos de bases de datos públicas. Para la predicción de *enhancers* transcripcionales activos de *Drosophila melanogaster* se utilizaron datos de secuenciación masiva (NGS) de apertura de cromatina y de acetilación de la lisina 27 de la histona 3 (H3K27ac) de siete tejidos o estadios del desarrollo. La búsqueda de estos datos de estructura de la cromatina fue realizada utilizando ChIP-Atlas (Zou et al 2021), y todos los datos están depositados en la base de datos *Sequence Read Archive* (https://www.ncbi.nlm.nih.gov/sra).

Por otra parte, para predecir *enhancers* activos en *Drosophila virilis*, se utilizaron datos de apertura de la cromatina de tres contextos espacio-temporales obtenidos de la misma manera que los datos correspondientes a *D. melanogaster*.

A continuación se muestra información correspondiente a todos los datos analizados:

Especie	**Estadio del desarrollo**	**Tejido/Órgano**	**Experimento**	**ID SRA**
D. melanogaster	Adulto	Cerebro	H3K27ac (ChIP-seq)	SRA543291
D. melanogaster	Embrión	Embrión 14-16hs	H3K27ac (ChIP-seq)	SRA082456
D. melanogaster	Embrión	Blastodermo	H3K27ac (ChIP-seq)	SRA1547818
D. melanogaster	Embrión	Blastodermo	H3K27ac (ChIP-seq)	SRA174152
D. melanogaster	Embrión	Blastodermo	H3K27ac (ChIP-seq)	SRA998002
D. melanogaster	Larva	Cerebro	H3K27ac (ChIP-seq)	SRA538131
D. melanogaster	Larva	Disco imaginal de ojo-antena	H3K27ac (ChIP-seq)	SRA092341

D. melanogaster	Larva	Disco imaginal de ojo-antena	H3K27ac (ChIP-seq)	SRA685120
D. melanogaster	Larva	Disco imaginal de halterio	H3K27ac (ChIP-seq)	SRA610778
D. melanogaster	Larva	Disco imaginal de ala	H3K27ac (ChIP-seq)	SRA092341
D. melanogaster	Larva	Disco imaginal de ala	H3K27ac (ChIP-seq)	SRA610778
D. melanogaster	Adulto	Cerebro	ATAC-Seq	SRA1294292
D. melanogaster	Adulto	Cerebro	ATAC-Seq	SRA1176878
D. melanogaster	Adulto	Cerebro	ATAC-Seq	SRA634591
D. melanogaster	Embrión	Embrión 14-16hs	DNase-Seq	ERA2870388
D. melanogaster	Embrión	Blastodermo	ATAC-Seq	SRA1087749
D. melanogaster	Embrión	Blastodermo	ATAC-Seq	ERA2399713
D. melanogaster	Embrión	Blastodermo	DNase-Seq	ERA2399718
D. melanogaster	Embrión	Blastodermo	ATAC-Seq	SRA1159760
D. melanogaster	Larva	Cerebro	ATAC-Seq	SRA638171
D. melanogaster	Larva	Disco imaginal de ojo-antena	ATAC-Seq	SRA1135888
D. melanogaster	Larva	Disco imaginal de ojo-antena	ATAC-Seq	SRA1294292
D. melanogaster	Larva	Disco imaginal de ojo-antena	ATAC-Seq	SRA638171
D. melanogaster	Larva	Disco imaginal de ojo-antena	ATAC-Seq	SRA590858
D. melanogaster	Larva	Disco imaginal de ojo-antena	ATAC-Seq	SRA1007316
D. melanogaster	Larva	Disco imaginal de ojo-antena	ATAC-Seq	SRA597907
D. melanogaster	Larva	Disco imaginal de halterio	ATAC-Seq	SRA1195874
D. melanogaster	Larva	Disco imaginal de ala	ATAC-Seq	SRA1195874
D. melanogaster	Larva	Disco imaginal de ala	ATAC-Seq	SRA1422043
D. melanogaster	Larva	Disco imaginal de ala	ATAC-Seq	SRA1294292
D. melanogaster	Larva	Disco imaginal de ala	ATAC-Seq	SRA999651
D. melanogaster	Larva	Disco imaginal de ala	ATAC-Seq	SRA1210731

D. melanogaster	Larva	Disco imaginal de ala	ATAC-Seq	SRA1474857
D. melanogaster	Larva	Disco imaginal de ala	ATAC-Seq	SRA600592
D. melanogaster	Larva	Disco imaginal de ala	ATAC-Seq	SRA1470826
D. virilis	Embrión	Blastodermo	DNAse-Seq	PRJEB10089
D. virilis	Embrión	19-22 hs	DNAse-Seq	PRJEB40271
D. virilis	Larva	Disco imaginal de ojo-antena	ATAC-Seq	PRJNA397746

5.2 Procesamiento de secuencias y alineamientos a los genomas de referencia

Todos los archivos en formato SRA fueron descargados y transformados a formato FASTQ a través de un script de BASH, utilizando el programa *parallel-fastq-dump* (https://github.com/rvalieris/parallel-fastq-dump). El control de calidad de las lecturas fue realizado utilizando el programa *fastqc* (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/). El control de calidad incluyó la evaluación de la longitud de las lecturas, la calidad base por base, el contenido de GC, los niveles de duplicación y el contenido de adaptadores, entre otros aspectos importantes. Los archivos FASTQ fueron procesados utilizando el programa *bbduk y* un archivo conteniendo secuencias conocidas de adaptadores de Illumina*,* ambos incluidos en el paquete de programas *bbmap* (https://sourceforge.net/projects/bbmap/). Como resultado, las lecturas son recortadas de forma de eliminar la secuencia correspondiente a adaptadores, o filtradas en caso de que, al eliminar los adaptadores, la secuencia remanente sea muy corta. Los parámetros utilizados para correr *bbduk* fueron **ktrim=r k=23 hdist=1 mink=11 mlf=.5 rcomp=t** para experimentos *single end* ó **ktrim=r k=23 hdist=1 tbo tpe mink=11 mlf=.5 rcomp=t** en el caso de experimentos *paired end.* Para verificar que las secuencias correspondientes a adaptadores hayan sido eliminadas, se realizó un segundo control de calidad utilizando *fastqc*.

Los archivos FASTQ fueron alineados al genoma de referencia de *Drosophila melanogaster* versión dm6 (https://hgdownload.soe.ucsc.edu/goldenPath/dm6/bigZips/) o al genoma de referencia de *Drosophila virilis* version droVir3

(https://hgdownload.soe.ucsc.edu/goldenPath/droVir3/bigZips/) según corresponda. Todos los alineamientos fueron realizados utilizando el programa *Bowtie2* (Langmead et al., 2009) con los parámetros por defecto.

Los archivos de alineamiento en formato SAM fueron ordenados según cromosoma y posición en el genoma y comprimidos a formato binario BAM utilizando el programa *samtools* (http://www.htslib.org/). Las lecturas duplicadas fueron marcadas usando el programa *picard MarkDuplicates* (https://broadinstitute.github.io/picard/). Las lecturas duplicadas, las de baja calidad y las no alineadas fueron eliminadas utilizando *samtools* con los parámetros **-q 30 -F 3332** para experimentos *single-end* y **-q 30 -F 3332 -f 3** para experimentos *paired-end.*

La *pipeline* utilizada fue automatizada utilizando el manager de *pipilines* llamado *snakemake* (https://snakemake.readthedocs.io/en/stable/) y se encuentra disponible online (https://github.com/AilenAlt/single_end_fastq_to_bam para experimentos *single-end* y https://github.com/AilenAlt/paired_end_fastq_to_bam/ para experimentos *paired-end*).

5.3 *Peak-calling*

El proceso de *peak-calling* permite encontrar regiones enriquecidas en lecturas con respecto a un *background* genómico. A estas regiones se las llama picos. En todos los casos, el *peak-calling* fue realizado utilizando el programa *MACS2* (Zhang et al, 2008). Los archivos BAM correspondientes a *D. melanogaster* fueron analizados utilizando los parámetros **-f BAM -g dm --keep-dup all -B --SPMR --nomodel --extsize 100 --shift -50 -q 0.01** para experimentos *single-end* y **-f BAMPE -g dm --keep-dup all -B --SPMR --nomodel -q 0.01** en el caso de experimentos *paired-end*. Para *D. virilis* los parámetros utilizados fueron **-f BAM -g 1.9E8 --keep-dup all 3 -B --SPMR --nomodel --extsize 100 --shift -50** para experimentos *single-end* y **-f BAMPE -g 1.9E8 --keep-dup all -B --SPMR --nomodel** para experimentos *paired-end*. El valor 1.9E8 corresponde al tamaño de genoma efectivo de *D. virilis,* estimado como la cantidad de bases del genoma de referencia que no corresponden a regiones repetitivas (mismo enfoque utilizado por Peng et al, 2019). Al analizar los datos correspondientes a experimentos de ChIP-seq, se realizó el *peak*-calling normalizando con el pegado inespecífico de un anticuerpo IgG (control o *input*).

Una de las salidas de *MACS2* es un archivo en formato *narrowPeak*, que proporciona información sobre la ubicación de los picos en el genoma. El formato *narrowPeak* sigue la estructura BED 6+4, lo que significa que incluye las primeras 6 columnas de un archivo BED estándar (cromosoma, base inicial, base final, nombre, *score*, y hebra) con 4 campos adicionales (valor de la señal, p-valor, q-valor, y posición de máxima señal relativa a la base inicial del pico). Además, dentro de cada pico, existe una base donde la cobertura de secuenciación es máxima, denominada *summit*. La información sobre los *summits* es generada por *MACS2* en un archivo formato BED 3+1, donde el campo adicional corresponde a la intensidad de señal en esa base. Los comandos **-B --SPMR** indican la creación de un archivo de densidad en formato *bedGraph* normalizado por la profundidad de secuenciación de cada experimento, que indica en cada base del genoma la intensidad de señal (densidad de lecturas normalizada). Los mismos fueron comprimidos en formato *bigWig* (binario) con el programa *bedGraphToBigWig* (Kent et al, 2010) para su rápido acceso y visualización utilizando el programa *Genome Browser IGV* (Robinson et al, 2011).

5.4 Definición de *enhancers* consenso en *D. melanogaster*

Para identificar *enhancers* putativos en el genoma de *D. melanogaster* y su actividad a lo largo de los 7 contextos espacio-temporales estudiados se utilizó una *pipeline* desarrollada por nuestro laboratorio (Laiker y Frankel, 2022) (Figura 30). La *pipeline* fusiona elementos de diferentes tejidos mediante la búsqueda de superposiciones entre los *summits* de regiones accesibles, en lugar de utilizar superposiciones entre elementos completos. Para lograr esto, se crea un intervalo de confianza para la ubicación del *summit* para cada contexto, considerando la variación en la posición del *summit* entre réplicas. En primer lugar, para contextos que cuentan con réplicas de DNAse-seq ó ATAC-seq se calcula la relación señal-ruido (SNR) para cada réplica, al calcular la proporción de lecturas dentro de los picos (N° lecturas dentro de picos/N° lecturas totales). Luego, se intersectan los picos correspondientes a la muestra con mayor SNR con los picos de H3K27ac disponibles para el mismo tejido, conservando solos los casos en los que existe una distancia igual o menor a 150 bp. Posteriormente, se calcula la variación en la posición del *summit* del mismo pico en diferentes réplicas (regiones con más de un 50% de superposición), generando un intervalo de

confianza para la ubicación del *summit.* Al superponer los intervalos de confianza para la ubicación de los *summits* de distintos tejidos se generan *clusters* de *summits.* Los *enhancers* consenso son creados a partir de los *clusters* de *summits,* fusionando todos los picos de MACS que contribuyen a cada *cluster.* De esta manera, se obtiene un elemento consenso por cada *cluster* de *summits* analizado. A partir de estos datos se construye una matriz binaria en la que cada fila corresponde a un elemento consenso y cada columna a su actividad en un determinado contexto espacio-temporal.

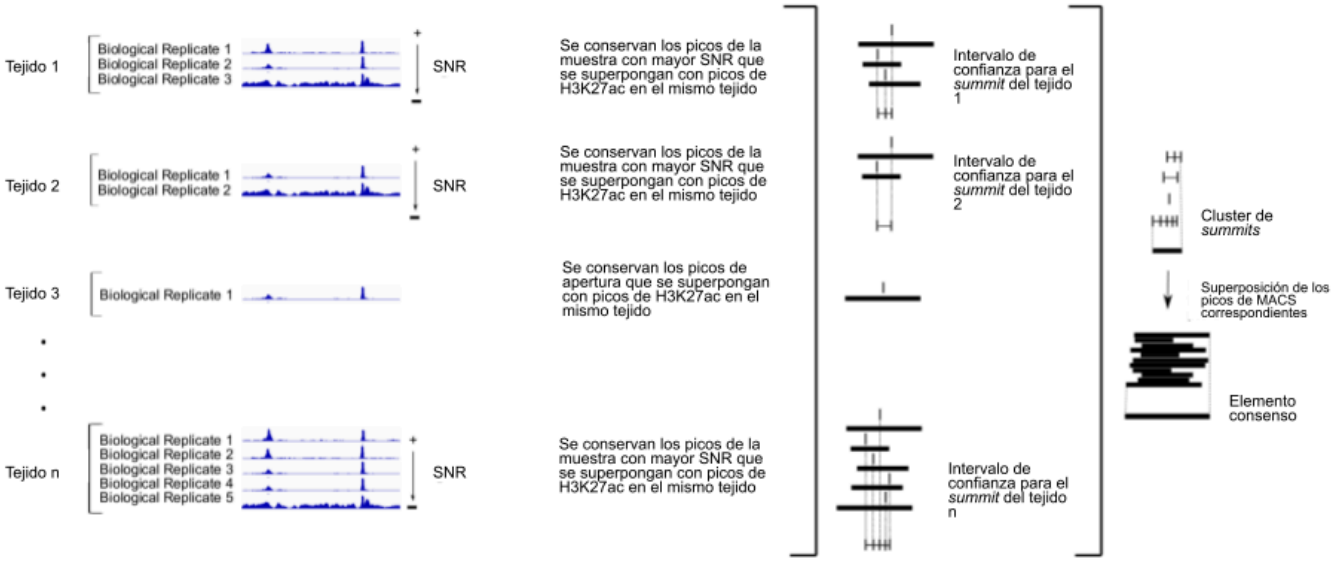


Figura 30. *Pipeline* para definir *enhancers* consenso a partir de datos de apertura de la cromatina y H3K27ac. Se esquematiza el paso por paso para definir la ubicación de un *enhancer* pleiotrópico, considerando la información de la estructura de la cromatina en múltiples contextos en los que está activo. Modificado de Laiker y Frankel 2022.

Los elementos consenso ubicados a 100 bp o menos de un TSS anotado fueron eliminados del conjunto, ya que estos podrían corresponder a regiones promotoras. El tamaño promedio de los elementos consenso obtenidos es de 932 pb, siendo el percentil 75 del tamaño de dichos elementos de 1288 pb (Figura Suplementaria 2).

Se consideran *enhancers* pleiotrópicos en *D. melanogaster* a aquellos elementos en los que se observó actividad (apertura de la cromatina y presencia de la marca H3K27ac) en más de un contexto espacio-temporal.

A su vez, en nuestro laboratorio definimos un set de 85944 *enhancers* sólo con de información de apertura de la cromatina en 13 contextos espacio-temporales: cerebro de adulto, blastodermo, embrión (6-8 hs), embrión tardío (16-18 hs), disco de ojo-antena, disco de ala, disco de halterio, sistema nervioso central de larva, halterio, pata de adulto, ala, pata de pupa y disco imaginal de pata (Ian Laiker, Tesis de Licenciatura, 2020). Todos estos elementos presentan un tamaño estandarizado de 400 bp.

5.5 Búsqueda de *enhancers* ortólogos en *D. virilis*

Una vez identificados *enhancers* putativos en el genoma de *D. melanogaster* buscamos a los ortólogos en *D. virilis*. Esta búsqueda se realizó tanto para los *enhancers* de *D. melanogaster* definidos sólo a partir de datos de apertura de la cromatina como para los *enhancers* consenso definidos a partir de datos de apertura y H3K27ac. Como se mencionó en la introducción, la predicción de la ubicación de *enhancers* entre especies no es una tarea fácil, sobre todo al tratarse de un par de especies con ~40 millones de años de divergencia. Es por ello que se decidió utilizar más de un método para realizar la predicción de *enhancers* en *D. virilis* a partir de los *enhancers* definidos en *D. melanogaster.* Se utilizó un método basado en alineamientos y un método "*alignment-free*" para realizar las predicciones en *D. virilis,* de forma de obtener resultados más robustos que los que se obtendrían utilizando un único método. En las siguientes secciones se describen los dos métodos utilizados.

5.5.1 *Reciprocal-liftOver*

El programa *liftOver* permite convertir coordenadas genómicas entre distintas versiones del genoma de una especie empleando para ello un alineamiento precomputado entre los genomas involucrados en la conversión. Siguiendo esta misma lógica, *liftOver* también es utilizado para la conversión de coordenadas genómicas entre especies. Para ello, *liftOver* utiliza un archivo en formato *chain*, el cual describe un alineamiento de a pares que permite *gaps* en las dos secuencias simultáneamente. El archivo en formato *chain* es construido a partir de bloques de alineamientos realizados con BLASTZ. Los bloques pueden fusionarse si se superponen o están lo suficientemente cerca, y posteriormente, los bloques pueden

agruparse para formar tramos más largos de sintenia. El agrupamiento de bloques es realizado utilizando un sistema de puntuaciones más permisivo que el de los alineamientos tradicionales, permitiendo la presencia de *gaps* en las dos secuencias simultáneamente.

Un archivo *chain* entre dos especies se construye de manera que un bloque de la especie 1 solo puede convertirse en una coordenada de la especie 2, pero varios bloques de la especie 1 podrían corresponder a un mismo bloque en la especie 2 (en la nomenclatura de base de datos, se dice que la relación es uno a muchos). Esto podría traer problemas a la hora de identificar secuencias ortólogas, en donde la relación debería ser uno-a-uno. Es por ello que, mediante un script de bash, implementamos una conversión bidireccional de *liftOver* (especie A → especie B → especie A) a la que llamamos *reciprocal-liftOver,* para evitar definiciones de ortología ambiguas que podrían resultar de eventos de duplicación genómica.

Para buscar los *enhancers* ortólogos a los de *D. melanogaster* en el genoma de *D. virilis* usando *reciprocal-liftover* descargamos los archivos *chain* correspondientes a las conversiones *D. melanogaster* → *D. virilis* (https://hgdownload-test.gi.ucsc.edu/goldenPath/dm6/liftOver/) y *D. virilis* → *D. melanogaster* (https://hgdownload.soe.ucsc.edu/goldenPath/droVir3/liftOver/). El programa fue ejecutado utilizando el parámetro **-minMatch 0.25**, por lo que solo se consideraron conversiones exitosas aquellas en las que mapean el 25% de las bases o más. Por lo tanto, cuanto mayor es el valor de *minMatch* utilizado, más exigente es la búsqueda y menos regiones del genoma de *D. virilis* son consideradas como ortólogas (Figura suplementaria 3). Para determinar el valor óptimo de *minMatch* a utilizar, usamos 11 *enhancers* ortólogos validados experimentalmente en las dos especies: 6 *enhancers* de *svb* (Frankel et al, 2012) y 5 *enhancers* que dirigen la expresión en el neuroectodermo embrionario (Crocker, 2008). El valor del parámetro minMatch seleccionado fue el porcentaje más alto que permitió encontrar al 100% de los *enhancers* validados en las dos especies. Este método nos permitió encontrar un total de 6019 *enhancers* en *D.virilis* para el set definido a partir de datos de apertura de la cromatina y H3K27ac en *D. melanogaster*, y 60174 *enhancers* para el set definido sólo con datos de apertura. En este último caso, filtramos a los *enhancers* encontrados en *D. virilis* por su tamaño, conservando sólo aquellos con tamaños entre 100 y 1000 bp.

Reciprocal-liftOver devuelve un archivo BED con la ubicación de los *enhancers* en *D. virilis* y un BED con la información de los *enhancer* de *D. melanogaster* que no lograron ser encontrados en el genoma de *D. virilis.*

5.5.2 Método *Alignment-free*

El segundo método utilizado para predecir la ubicación de *enhancers* ortólogos a los de *D. melanogaster* en el genoma de *D. virilis* es el llamado "*Alignment-free*", el cual se basa en la co-ocurrencia de palabras (*k-mers*) en un espacio de búsqueda limitado. Este enfoque esta basado en los trabajos de van Helden (2004) y Arunachalam (2010), y su esquema básico es el siguiente: dada la secuencia de un *enhancer* en un genoma, se escanea la región intergénica correspondiente en un segundo genoma con una ventana móvil y se calcula un puntaje de similitud entre el *enhancer* y cada una de las ventanas de la región intergénica. Luego se identifica cuál es la ventana con el puntaje de similitud máximo y se calcula la significancia de dicho puntaje. A continuación se explica en detalle cada uno de los pasos a seguir.

Para encontrar *enhancers* de *D. melanogaster* en el genoma de *D. virilis*, el primer paso es identificar la región intergénica ortóloga. Para definir este espacio de búsqueda, asumimos que los *enhancers* estarán presentes en regiones sinténicas flanqueadas por los mismos genes en las dos especies. Para identificar pares de genes ortólogos utilizamos BLASTP recíproco, empleando para ello las anotaciones de genes en formato GTF disponibles para D*. melanogaster* (https://ftp.flybase.org/genomes/dmel/current/gtf/) y *D. virilis* (https://ftp.flybase.org/genomes/Drosophila_virilis/dvir_r1.07_FB2018_05/gtf/). Se logró identificar una región intergénica ortóloga para 5313 *enhancers* de *D. melanogaster* definidos a partir de datos de apertura de la cromatina y H3K27ac, y para 60354 *enhancers* de *D. melanogaster* definidos con información de apertura de la cromatina.

Para asignarle una puntuación a la ocurrencia de diferentes *k-mers*, necesitamos definir su frecuencia de ocurrencia base en los genomas estudiados. Dada la frecuencia (f_i) para el k-mer *i* en el genoma de la especie *j*, el número esperado de ocurrencias m_i del k-mer *i* en una secuencia de longitud *L* se obtiene mediante:

(Ec. 1) $$m_i^j = f_i . T = f_i (L - w + 1)$$

donde *w* es el largo del *k-mer* y *T* la cantidad de ubicaciones posibles que puede ocupar el *k-mer* en la secuencia de largo *L*. Se utilizaron *k-mers* de largo 6 y las frecuencias genómicas precalculadas para dichos k-mers (f_i) en los genomas de *D. melanogaster* y *D. virilis* se obtuvieron del trabajo de Arunachalam (2010).

Para dos secuencias dadas, se calcula un puntaje de similitud basado en distribuciones de Poisson (*S*) en función de la probabilidad de co-ocurrencia de *k-mers*. La divergencia de secuencia se refleja en un puntaje de disimilitud basado en distribuciones de Poisson (*D)*, que refleja la diferencia entre las ocurrencias de palabras en las dos secuencias. Finalmente, una métrica mixta M es definida como una combinación ponderada del puntaje de similitud (*S*) y el puntaje de disimilitud (*D*) (Figura 31).

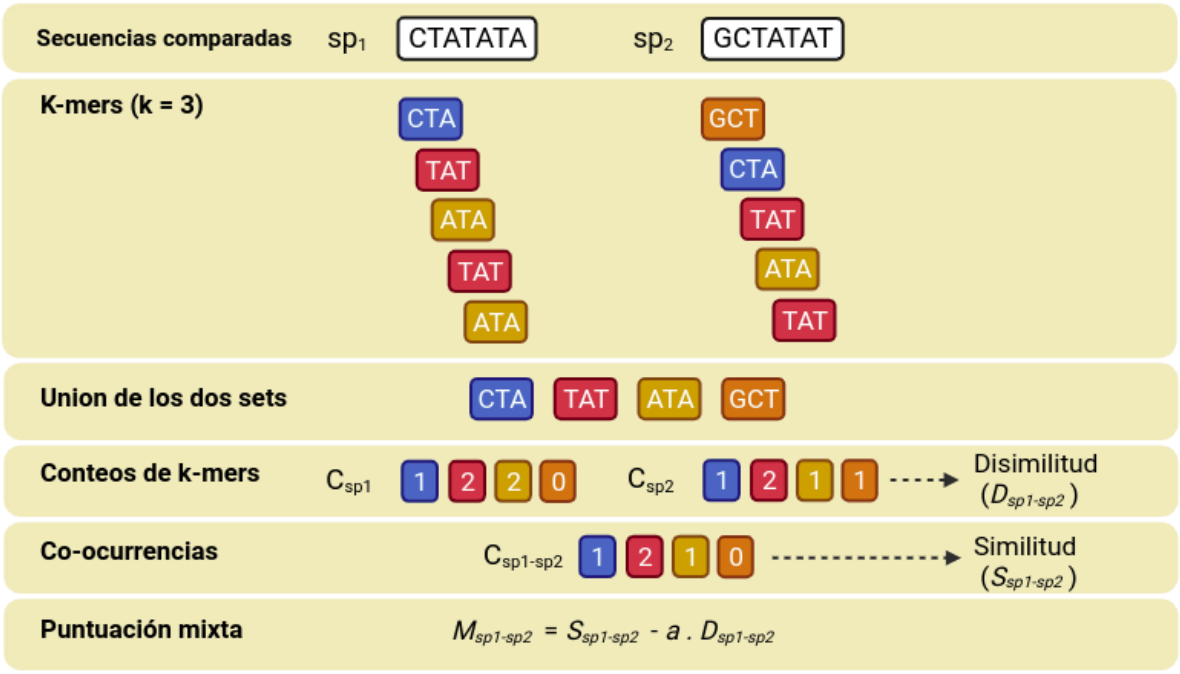


Figura 31. Método "*alignment-free*" para la búsqueda de secuencias ortólogas. El esquema resume los pasos a seguir y las métricas a calcular para comparar a la secuencia de la especie 1 (sp_1) con la secuencia de la especie 2 (sp_2). La figura fue confeccionada utilizando *biorender* (https://app.biorender.com/)

Comparamos la secuencia de un *enhancer e* de *D. melanogaster* (especie 1) contra una región intergénica ortóloga candidata *c* de *D. virilis* (especie 2). El puntaje de similitud $S^{ec}{}_i$ para un solo *k-mer i* se calcula de la siguiente manera:

(Ec. 2) $$S_i^{ec} = \left[1 - P_1(x \geq C_i^{ec})\right] + \left[1 - P_2(x \geq C_i^{ec})\right]$$

donde C_i^{ec} es la cantidad de co-ocurrencias del *k-mer i* en las secuencias *e* y *c* (Ec. 3).

(Ec . 3) $$C_i^{ec} = min(N_i^e, N_i^c)$$

donde $N^e{}_i$ y $N^c{}_i$ es la cantidad de veces que el *k-mer* i aparece en el *enhancer* (*e*) ó en la región intergénica (*c*) respectivamente. A partir de C_i^{ec} es posible calcular la probabilidad de observar al menos C_i^{ec} ocurrencias del *k-mer* en la especie *j* ($P_j(x \geq C_i^{ec})$, *Ec. 4*):

(Ec. 4) $$P_j(x \geq C_i^{ec}) = \begin{cases} [1 - F(C_i^{ec} - 1, m_i^j)] & \text{si } C_i^{ec} > 0, \\ 1 & \text{si } C_i^{ec} \leq 0 \end{cases}$$

donde la función de probabilidad acumulada de Poisson $F(C_i^{ec} - 1, m_i^j)$ representa la probabilidad de observar menos de C_i^{ec} ocurrencias, cuando el valor esperado es m_i^j.

La similitud multivariada entre *e* y *c* (S^{ec}) teniendo en cuenta *p* cantidad de *k-mers* es obtenida según la Ec. 5.

(Ec. 5) $$S^{ec} = (1/p) \sum S_i^{ec}$$

El puntaje de disimilitud para el *k-mer i* entre las secuencias *e* y *c* se calcula según:

(Ec. 6) $$D_i^{ec} = |F(N_i^e - 1, m_i^1) - F(N_i^c - 1, m_i^2)|$$

La disimilitud multivariada teniendo en cuenta todos los posibles *p k-mers* es obtenida según la Ec. 7:

(Ec. 7) $$D^{ec} = (1/p) \sum D_i^{ec}$$

Finalmente el puntaje M^{ec} que combina la similitud (Ec. 5) y disimilitud multivariada (Ec. 7) fue definido como:

(Ec. 8) $$M^{ec} = S^{ec} - aD^{ec} + b$$

donde *a* es un parámetro de ponderación positivo que se puede ajustar para dar más énfasis a las ocurrencias comunes o distintas entre dos secuencias, y b es un desplazamiento para que la métrica resulte positiva. Utilizamos un valor de *a* de 0.3 para dar mayor énfasis a la co-ocurrencia de *k-mers.* Cuando dos secuencias tienen exactamente los mismos conteos para todos los patrones, su disimilitud es 0. Este puntaje aumenta cuanto mayor es la diferencia entre los conteos de *k-mers* de *e* y *c*.

Entonces, para cada región intergénica se determinó cual es la ventana *c* que posee el mayor puntaje M^{ec}. Para evaluar la significancia de dicho puntaje M^{ec}, calculamos una distribución *background* empírica de puntajes mediante el muestreo de secuencias intergénicas aleatorias de *D. melanogaster* y *D. virilis.* Se seleccionó aleatoriamente un par ortólogo y luego se muestrearon ventanas de igual longitud en ubicaciones aleatorias en cada región, y se calculó el M^{ec} entre las ventanas (utilizando la Ec. 8). Para construir la distribución *background*, este proceso se repitió 100000. Una vez obtenida la distribución, se calculó el p-valor como la probabilidad de obtener un valor de M^{ec} igual o más extremo que el observado bajo esta distribución. Como se realizaron múltiples comparaciones para encontrar la ubicación de miles de *enhancers* de *D. melanogaster* en el genoma de *D. virilis*, los p-valores fueron corregidos calculando un FDR. Se consideraron predicciones exitosas aquellas cuyo FDR sea menor a 0.05, determinando como ubicación del *enhancer* en *D. virilis* la ubicación de la ventana con mayor puntuación.

Para el set de *enhancers* de *D. melanogaster* definido a partir de datos de apertura de la cromatina y H3K27ac se utilizó un tamaño de ventana de 800 bp (aproximadamente la mediana del tamaño de los elementos consenso obtenidos, Figura Suplementaria 3). Los elementos consenso de tamaño menor a 800 bp fueron extendidos desde su base central hasta alcanzar dicho tamaño. Por otro lado, para el set de *enhancers* de *D. melanogaster* definidos sólo a partir de información de apertura de la cromatina se utilizaron ventanas de 400 bp (tamaño estandarizado de todos los elementos del set).

El método fue implementado en *Python 3.9.7,* haciendo uso de las bibliotecas *Biopython* (https://biopython.org/docs/1.75/api/Bio.html), *scipy* (https://scipy.org/) y *scikit-learn* (https://scikit-learn.org/stable/).

5.5.3 Comparación de predicciones de ambos métodos

Contando con un set de predicciones de *enhancers* de *D. virilis* realizada con el método "*Alignment-free*" y un set obtenido usando *reciprocal-liftOver*, evaluamos en qué casos los métodos coinciden en la predicción. Es decir, para un determinado *enhancer* de *D. melanogaster* buscamos a su ortólogo en *D. virilis* usando los dos métodos y conservamos aquellos casos en los que ambos métodos predicen la misma ubicación. Consideramos predicciones coincidentes aquellas en las que existe superposición de al menos el 10% de la secuencia del *enhancer*. Este análisis fue realizado utilizando la herramienta *bedtools* (https://bedtools.readthedocs.io/en/latest/) y un script de *BASH* personalizado. Bajo este criterio, las predicciones coinciden en la mayoría de los casos.

5.6 Análisis de genes flanqueantes a los *enhancers* en las dos especies

Para analizar si los *enhancers* se encuentran flanqueados por los mismos genes en las dos especies, estudiamos las predicciones realizadas con *reciprocal-liftOver*, ya que este método no requería encontrar una región intergénica ortóloga. Para ello, se generó una tabla de genes ortólogos entre *D. melanogaster* y *D. virilis* usando los mejores *hits* recíprocos de *BLAST.* Como primer paso, se obtuvieron las secuencias de las proteínas codificadas por los genes anotados de *Drosophila melanogaster* (*https://ftp.flybase.org/genomes/dmel/current/gtf/*) *y Drosophila virilis* (*https://ftp.flybase.org/genomes/Drosophila_virilis/dvir_r1.07_FB2018_05/gtf/*). A partir de estas secuencias, se generaron bases de datos de *BLAST* para las proteínas anotadas de los dos genomas. Luego, se realizó la comparación de proteínas con *BLASTP* utilizando las bases de datos generadas anteriormente. Se compararon tanto las proteínas de *D. melanogaster* con la base de datos de *D. virilis*, como las proteínas de *D. virilis* con la base de datos de *D. melanogaster.* Luego, se seleccionaron únicamente los mejores *hits* para cada una de las comparaciones utilizando la función *sort* de BASH y estos son comparados de forma de conservar sólo los resultados recíprocos. Finalmente se obtiene una tabla conteniendo la información de los pares de genes ortólogos. Todo este análisis es realizado a través de un *script* de BASH.

Una vez obtenida la tabla de relaciones de ortología entre los genes de *D. melanogaster* y de *D. virilis,* identificamos los dos genes flanqueantes más cercanos para cada *enhancer* de *D. melanogaster* utilizando el programa *bedtools closest*. De igual manera, identificamos los genes flanqueantes para cada *enhancer* ortólogo predicho en *D. virilis.* Utilizando la tabla de pares de genes ortólogos, evaluamos si cada *enhancer* está flanqueado por los mismos genes en las dos especies. Dicho análisis fue realizado utilizando *scripts* de BASH.

5.6 Definición de *enhancers* consenso en *D. virilis*

Los *enhancers* consenso de *D. virilis* fueron definidos de manera casi idéntica a los *enhancers* de *D. melanogaster*, utilizando la misma *pipeline* desarrollada por nuestro laboratorio (Figura 28).

En primer lugar, a partir de los archivos BED correspondientes a los picos de *D. virilis* en 3 contextos espacio-temporales y los archivos BAM, se determinó cual es la muestra con mayor SNR (número de lecturas en picos / número de lecturas totales) para cada contexto. Para esta especie no contamos con datos de H3K27ac, por lo que los picos de la réplica con mayor SNR para cada contexto no fueron filtrados, sino que todos ellos fueron utilizados para construir los intervalos de confianza para los *summits.* Nuevamente se realiza una superposición de intervalos de confianza para la ubicación de los *summits* obtenidos para distintos tejidos, obteniendo *clusters* de *summits*. Los *enhancers* consenso nuevamente son creados a partir de los clusters de *summits*, fusionando todos los picos de MACS que contribuyen a cada *cluster*. Finalmente se obtiene una matriz binaria donde cada fila corresponde a un elemento consenso y cada columna a su actividad en un determinado contexto espacio-temporal.

Se filtraron elementos a distancias menores a 100 pb de cualquier TSS extraído de la anotación de genes de *D. virilis* en formato GTF (https://ftp.flybase.org/genomes/Drosophila_virilis/dvir_r1.07_FB2018_05/gtf/), de manera de descartar regiones promotoras.

5.7 Análisis de la apertura de la cromatina en los *enhancers* predichos de *D. virilis*

Una vez predichas las ubicaciones de los *enhancers* ortólogos a los de *D. melanogaster* en el genoma de *D. virilis,* evaluamos la actividad de cada predicción en los tres contextos espacio-temporales estudiados en *D. virilis.* Para ello estudiamos la superposición entre los *enhancers* ortólogos predichos y los elementos consenso definidos a partir de información de apertura de la cromatina en *D. virilis.* Consideramos que un *enhancer* ortólogo predicho se encuentra activo en *D. virilis* si presenta una superposición de al menos el 50% de su secuencia con un elemento consenso. Dicho análisis fue realizado utilizando el programa *bedtools intersect* (https://bedtools.readthedocs.io/en/latest/content/tools/intersect.html) con el parámetro **-wo** (*bedtools* reporta cada superposicion entre dos elementos junto con la cantidad de bases superpuestas entre ellos). El archivo de salida de *bedtools* fue analizado en *Python* utilizando la biblioteca *pandas*. Para los casos en que un *enhancer* ortólogo predicho se superpone con más de un elemento consenso definido en *D. virilis*, se le asignó a ese *enhancer* la información de apertura de cromatina (actividad en los tres contextos) correspondiente al elemento con mayor cantidad de bases solapadas.

5.8 Gráficos

Todos los gráficos se realizaron en *R* con *ggplot2* (https://ggplot2.tidyverse.org/) o en *python* utilizando *matplotlib (https://matplotlib.org/), seaborn (https://seaborn.pydata.org/)* y *upsetplot (https://upsetplot.readthedocs.io/en/stable/).*

Figuras suplementarias

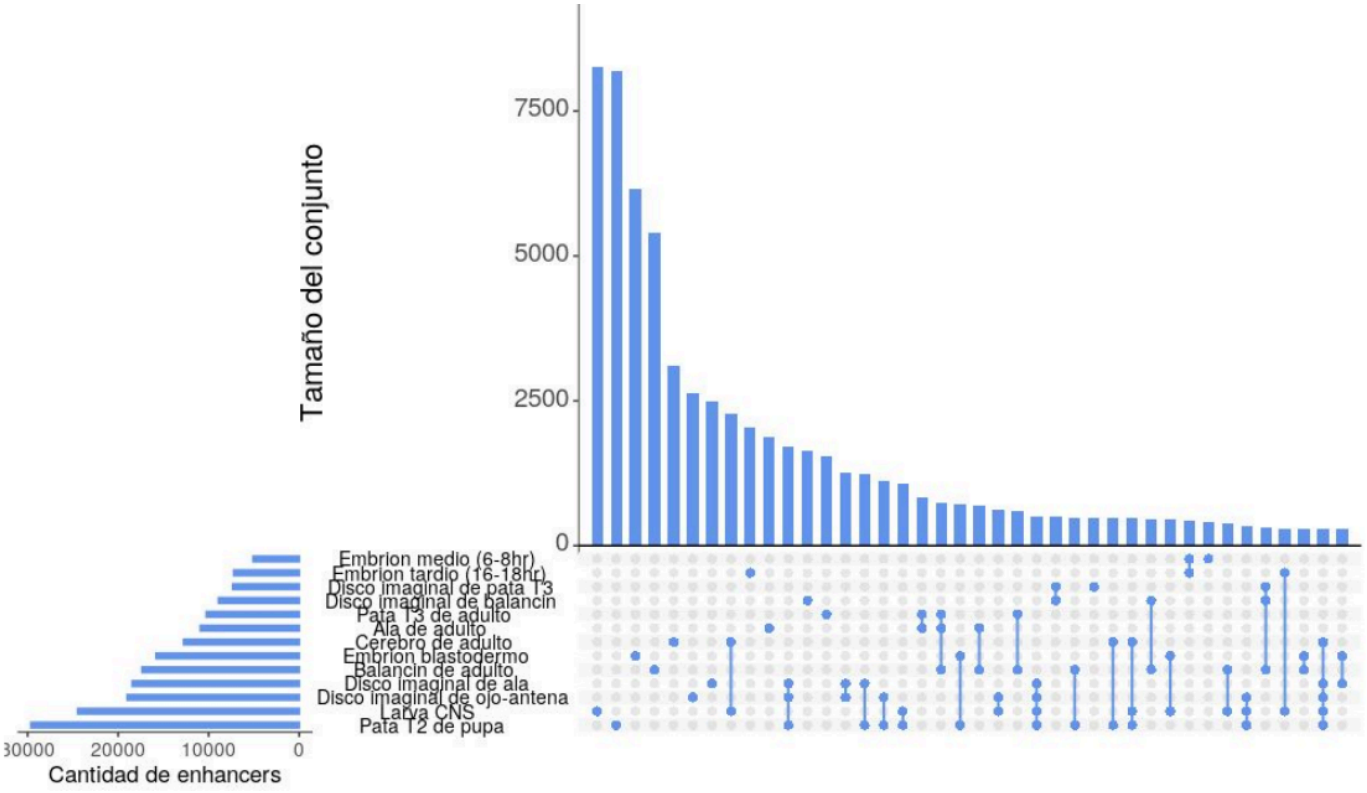


Figura suplementaria 1. Upset plot de *enhancers* de *Drosophila* definidos a partir de información de apertura de la cromatina. Modificado de la Tesis de Licenciatura de Ian Laiker, 2020, FCEN (UBA).

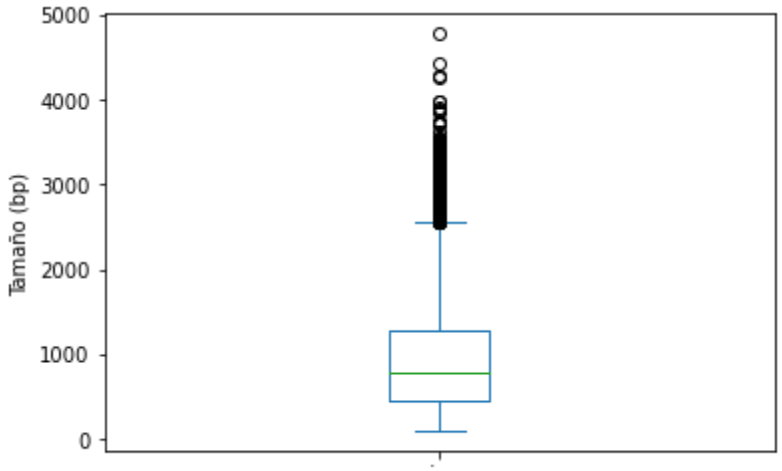


Figura Suplementaria 2. *Boxplot* correspondiente al tamaño de los elementos consenso de *D. melanoagster.*

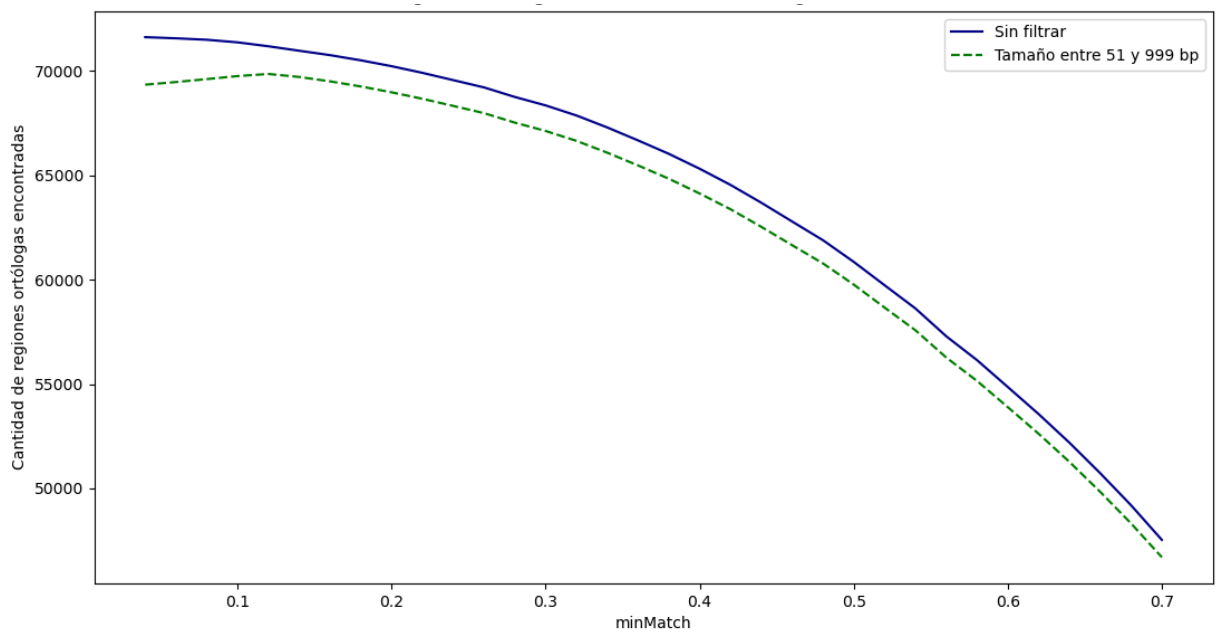


Figura suplementaria 3. Cantidad de regiones ortólogas a las de *D. melanogaster* encontradas por liftOver en el genoma de *D. virilis,* según el valor de *minMatch* utilizado. Los resultados corresponden a la búsqueda de *enhancers* de *D. melanogaster* definidos a partir de información de apertura de la cromatina en el genoma de *D. virilis.* La línea azul continua representa la cantidad de regiones ortólogas contenidas, mientras que la línea punteada representa la cantidad de regiones encontradas cuyo tamaño es de entre 51 y 999 bp.

Bibliografía

- Adams, M. D., Celniker, S. E., Holt, R. A., Evans, C. A., Gocayne, J. D., Amanatides, P. G., Scherer, S. E., Li, P. W., Hoskins, R. A., Galle, R. F., George, R. A., Lewis, S. E., Richards, S., Ashburner, M., Henderson, S. N., Sutton, G. G., Wortman, J. R., Yandell, M. D., Zhang, Q., … Craig Venter, J. (2000). The genome sequence of Drosophila melanogaster. Science (New York, N.Y.), 287(5461), 2185–2195. https://doi.org/10.1126/SCIENCE.287.5461.2185
- Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, *215*(3), 403–410. https://doi.org/10.1016/S0022-2836(05)80360-2
- Andersson, R., & Sandelin, A. (2019). Determinants of enhancer and promoter activities of regulatory elements. *Nature Reviews Genetics 2019 21:2*, *21*(2), 71–87. https://doi.org/10.1038/s41576-019-0173-8
- Arunachalam, M., Jayasurya, K., Tomancak, P., & Ohler, U. (2010). An alignment-free method to identify candidate orthologous enhancers in multiple Drosophila genomes. *Bioinformatics (Oxford, England)*, *26*(17), 2109–2115. https://doi.org/10.1093/BIOINFORMATICS/BTQ358
- Ashburner, M. (1989). *Drosophila* (Vol. 1). Cold Spring Harbor Laboratory.
- Bainbridge, S. P., & Bownes, M. (1981). Staging the metamorphosis of Drosophila melanogaster. *Development*, *66*(1), 57–80. https://doi.org/10.1242/DEV.66.1.57
- Barolo, S., & Posakony, J. W. (2002). Three habits of highly effective signaling pathways: principles of transcriptional control by developmental cell signaling. *Genes & Development*, *16*(10), 1167–1181. https://doi.org/10.1101/GAD.976502
- Benson, K. R. (2001). T. H. Morgan's resistance to the chromosome theory. *Nature Reviews Genetics*, *2*(6), 469–474. https://doi.org/10.1038/35076532
- Bopp, D., Calhoun, G., Horabin, J. I., Samuels, M., & Schedl, P. (1996). Sex-specific control of Sex-lethal is a conserved mechanism for sex determination in the genus Drosophila. *Development (Cambridge, England)*, *122*(3), 971–982. https://doi.org/10.1242/DEV.122.3.971
- Brawand, D., Soumillon, M., Necsulea, A., Julien, P., Csárdi, G., Harrigan, P., Weier, M., Liechti, A., Aximu-Petri, A., Kircher, M., Albert, F. W., Zeller, U., Khaitovich, P., Grützner, F., Bergmann, S., Nielsen, R., Pääbo, S., & Kaessmann, H. (2011). The evolution of gene expression levels in mammalian organs. *Nature*, *478*(7369), 343–348. https://doi.org/10.1038/NATURE10532
- Bulger, M., & Groudine, M. (1999). Looping versus linking: toward a model for long-distance gene activation. *Genes & Development*, *13*(19), 2465–2477. https://doi.org/10.1101/GAD.13.19.2465
- Burtis, K. C., Thummel, C. S., Jones, C. W., Karim, F. D., & Hogness, D. S. (1990). The Drosophila 74EF early puff contains E74, a complex ecdysone-inducible gene that encodes two ets-related proteins. *Cell*, *61*(1), 85–99. https://doi.org/10.1016/0092-8674(90)90217-3
- Campos-Ortega, J. A., & Hartenstein, V. (1985). *The Embryonic Development of Drosophila melanogaster - Jose A. Campos-Ortega, Volker Hartenstein - Google Libros*. Springer Science & Business Media.
- Carelli, F. N., Liechti, A., Halbert, J., Warnefors, M., & Kaessmann, H. (2018). Repurposing of promoters and enhancers during mammalian evolution. *Nature Communications 2018 9:1*, *9*(1), 1–11. https://doi.org/10.1038/s41467-018-06544-z
- Carroll, S. B. (2008). Evo-Devo and an Expanding Evolutionary Synthesis: A Genetic Theory of Morphological Evolution. *Cell*, *134*(1), 25–36. https://doi.org/10.1016/J.CELL.2008.06.030

- Chan, Y. F., Marks, M. E., Jones, F. C., Villarreal, G., Shapiro, M. D., Brady, S. D., Southwick, A. M., Absher, D. M., Grimwood, J., Schmutz, J., Myers, R. M., Petrov, D., Jónsson, B., Schluter, D., Bell, M. A., & Kingsley, D. M. (2010). Adaptive evolution of pelvic reduction in sticklebacks by recurrent deletion of a Pitx1 enhancer. *Science (New York, N.Y.)*, *327*(5963), 302–305. https://doi.org/10.1126/SCIENCE.1182213
- Chanut-Delalande, H., Hashimoto, Y., Pelissier-Monier, A., Spokony, R., Dib, A., Kondo, T., Bohère, J., Niimi, K., Latapie, Y., Inagaki, S., Dubois, L., Valenti, P., Polesello, C., Kobayashi, S., Moussian, B., White, K. P., Plaza, S., Kageyama, Y., & Payre, F. (2014). Pri peptides are mediators of ecdysone for the temporal control of development. *Nature Cell Biology*, *16*(11), 1035–1044. https://doi.org/10.1038/NCB3052
- Chen, H., Li, C., Zhou, Z., & Liang, H. (2018). Fast-Evolving Human-Specific Neural Enhancers Are Associated with Aging-Related Diseases Article Fast-Evolving Human-Specific Neural Enhancers Are Associated with Aging-Related Diseases. *Cell Systems*, *6*, 604–611. https://doi.org/10.1016/j.cels.2018.04.002
- Chen, L., Fish, A. E., & Capra, J. A. (2018). Prediction of gene regulatory enhancers across species reveals evolutionarily conserved sequence properties. *PLOS Computational Biology*, *14*(10), e1006484. https://doi.org/10.1371/JOURNAL.PCBI.1006484
- Claringbould, A., & Zaugg, J. B. (2021). Enhancers in disease: molecular basis and emerging treatment strategies. *Trends in Molecular Medicine*, *27*(11), 1060–1073. https://doi.org/10.1016/J.MOLMED.2021.07.012
- Clark, A. G., Eisen, M. B., Smith, D. R., Bergman, C. M., Oliver, B., Markow, T. A., Kaufman, T. C., Kellis, M., Gelbart, W., Iyer, V. N., Pollard, D. A., Sackton, T. B., Larracuente, A. M., Singh, N. D., Abad, J. P., Abt, D. N., Adryan, B., Aguade, M., Akashi, H., … MacCallum, I. (2007). Evolution of genes and genomes on the Drosophila phylogeny. *Nature*, *450*(7167), 203–218. https://doi.org/10.1038/NATURE06341
- Corradin, O., & Scacheri, P. C. (2014). Enhancer variants: evaluating functions in common disease. *Genome Medicine*, *6*(10). https://doi.org/10.1186/S13073-014-0085-3
- Creyghton, M. P., Cheng, A. W., Welstead, G. G., Kooistra, T., Carey, B. W., Steine, E. J., Hanna, J., Lodato, M. A., Frampton, G. M., Sharp, P. A., Boyer, L. A., Young, R. A., & Jaenisch, R. (2010). Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(50), 21931–21936. https://doi.org/10.1073/PNAS.1016071107/-/DCSUPPLEMENTAL/PNAS.201016071SI.PDF
- Crocker, J., Tsai, A., & Stern, D. L. (2017). A Fully Synthetic Transcriptional Platform for a Multicellular Eukaryote. *Cell Reports*, *18*(1), 287–296. https://doi.org/10.1016/J.CELREP.2016.12.025
- D'Aurizio, R., Catona, O., Pitasi, M., Li, Y. E., Ren, B., & Nicolis, S. K. (2022). Bridging between Mouse and Human Enhancer-Promoter Long-Range Interactions in Neural Stem Cells, to Understand Enhancer Function in Neurodevelopmental Disease. *International Journal of Molecular Sciences*, *23*(14). https://doi.org/10.3390/IJMS23147964/S1
- Dahn, R. D., Davis, M. C., Pappano, W. N., & Shubin, N. H. (2006). Sonic hedgehog function in chondrichthyan fins and the evolution of appendage patterning. *Nature 2006 445:7125*, *445*(7125), 311–314. https://doi.org/10.1038/nature05436
- Davidson, E. H. (2010). The Regulatory Genome: Gene Regulatory Networks In Development And Evolution - Eric H. Davidson - Google Books. *The Regulatory Genome: Gene Regulatory Networks In Development And Evolution*, 304.
- Delon, I., Chanut-Delalande, H., & Payre, F. (2003). The Ovo/Shavenbaby transcription factor specifies actin remodelling during epidermal differentiation in Drosophila. *Mechanisms of Development*, *120*(7), 747–758. https://doi.org/10.1016/S0925-4773(03)00081-9
- Demerec, M., Spradling, A., Kaufmann, B. P., & of Washington. Department of Genetics, C. I. (1996). *Drosophila Guide: Introduction to the Genetics and Cytology of Drosophila Melanogaster*. Carnegie Institution of Washington. https://books.google.com.ar/books?id=b8sonwEACAAJ

- Dye, F. J. (2012). Dictionary of Developmental Biology and Embryology. https://doi.org/10.1002/9781118196649
- Edwards, S. L., Beesley, J., French, J. D., & Dunning, M. (2013). Beyond GWASs: illuminating the dark road from association to function. *American Journal of Human Genetics*, *93*(5), 779–797. https://doi.org/10.1016/J.AJHG.2013.10.012
- Fish, A., Chen, L., & Capra, J. A. (2017). Gene Regulatory Enhancers with Evolutionarily Conserved Activity Are More Pleiotropic than Those with Species-Specific Activity. *Genome Biology and Evolution*, *9*(10), 2615. https://doi.org/10.1093/GBE/EVX194
- Frankel, N., Wang, S., & Stern, D. L. (2012). Conserved regulatory architecture underlies parallel genetic changes and convergent phenotypic evolution. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(51), 20975–20979. https://doi.org/10.1073/PNAS.1207715109
- Frankel, N., Davis, G. K., Vargas, D., Wang, S., Payre, F., & Stern, D. L. (2010). Phenotypic robustness conferred by apparently redundant transcriptional enhancers. *Nature*, *466*(7305), 490–493. https://doi.org/10.1038/NATURE09158
- Göke, J., Schulz, M. H., Lasserre, J., & Vingron, M. (2012). virilisEstimation of pairwise sequence similarity of mammalian enhancers with word neighbourhood counts. *Bioinformatics*, *28*(5), 656. https://doi.org/10.1093/BIOINFORMATICS/BTS028
- Hales, K. G., Korey, C. A., Larracuente, A. M., & Roberts, D. M. (2015). Genetics on the Fly: A Primer on the Drosophila Model System. *Genetics*, *201*(3), 815–842. https://doi.org/10.1534/GENETICS.115.183392
- Hardison, R. C., & Taylor, J. (2012). Genomic approaches towards finding cis-regulatory modules in animals. *Nature Reviews. Genetics*, *13*(7), 469–483. https://doi.org/10.1038/NRG3242
- Hong, J., Gao, R., & Yang, Y. (2021). CrepHAN: cross-species prediction of enhancers by using hierarchical attention networks. *Bioinformatics*, *37*(20), 3436–3443. https://doi.org/10.1093/BIOINFORMATICS/BTAB349
- Inoue, F., Kircher, M., Martin, B., Cooper, G. M., Witten, D. M., McManus, M. T., Ahituv, N., & Shendure, J. (2017). A systematic comparison reveals substantial differences in chromosomal versus episomal encoding of enhancer activity. *Genome Research*, *27*(1), 38–52. https://doi.org/10.1101/GR.212092.116
- Kantorovitz, M. R., Robinson, G. E., & Sinha, S. (2007). A statistical method for alignment-free comparison of regulatory sequences. *Bioinformatics (Oxford, England)*, *23*(13). https://doi.org/10.1093/BIOINFORMATICS/BTM211
- Kent, W. J., Zweig, A. S., Barber, G., Hinrichs, A. S., & Karolchik, D. (2010). BigWig and BigBed: enabling browsing of large distributed datasets. *Bioinformatics (Oxford, England)*, *26*(17), 2204–2207. https://doi.org/10.1093/BIOINFORMATICS/BTQ351
- Kim, B. Y., Wang, J. R., Miller, D. E., Barmina, O., Delaney, E., Thompson, A., Comeault, A. A., Peede, D., D'agostino, E. R. R., Pelaez, J., Aguilar, J. M., Haji, D., Matsunaga, T., Armstrong, E. E., Zych, M., Ogawa, Y., Stamenković-Radak, M., Jelić, M., Veselinović, M. S., … Petrov, D. A. (2021). Highly contiguous assemblies of 101 drosophilid genomes. *ELife*, *10*. https://doi.org/10.7554/ELIFE.66405
- King, M. C., & Wilson, A. C. (1975). Evolution at two levels in humans and chimpanzees. *Science*, *188*(4184), 107–116. https://doi.org/10.1126/SCIENCE.1090005/ASSET/72CE3BB0-9EC2-4B40-9A60-8C0781D1AB65/ASSETS/SCIENCE.1090005.FP.PNG
- Kittelmann, S., Preger-Ben Noon, E., McGregor, A. P., & Frankel, N. (2021). A complex gene regulatory architecture underlies the development and evolution of cuticle morphology in Drosophila. *Current Opinion in Genetics & Development*, *69*, 21–27. https://doi.org/10.1016/J.GDE.2021.01.003
- Kornberg, R. D. (2005). Mediator and the mechanism of transcriptional activation. *Trends in Biochemical Sciences*, *30*(5), 235–239. https://doi.org/10.1016/J.TIBS.2005.03.011

- Kvon, E. Z., Kazmar, T., Stampfel, G., Yáñez-Cuna, J. O., Pagani, M., Schernhuber, K., Dickson, B. J., & Stark, A. (2014). Genome-scale functional characterization of Drosophila developmental enhancers in vivo. *Nature*, *512*(7512), 91–95. https://doi.org/10.1038/NATURE13395
- Laiker, I., & Frankel, N. (2022). Pleiotropic Enhancers are Ubiquitous Regulatory Elements in the Human Genome. *Genome Biology and Evolution*, *14*(6). https://doi.org/10.1093/GBE/EVAC071
- Langley, A. R., Smith, J. C., Stemple, D. L., & Harvey, S. A. (2014). New insights into the maternal to zygotic transition. *Development (Cambridge, England)*, *141*(20), 3834–3841. https://doi.org/10.1242/DEV.102368
- Langmead, B., Trapnell, C., Pop, M., & Salzberg, S. L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology*, *10*(3), 1–10. https://doi.org/10.1186/GB-2009-10-3-R25/TABLES/5
- Lettice, L. A., Heaney, S. J. H., Purdie, L. A., Li, L., de Beer, P., Oostra, B. A., Goode, D., Elgar, G., Hill, R. E., & de Graaff, E. (2003). A long-range Shh enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly. *Human Molecular Genetics*, *12*(14), 1725–1735. https://doi.org/10.1093/HMG/DDG180
- Leung, W., Shaffer, C. D., Cordonnier, T., Wong, J., Itano, M. S., Tempel, E. E. S., Kellmann, E., Desruisseau, D. M., Cain, C., Carrasquillo, R., Chusak, T. M., Falkowska, K., Grim, K. D., Guan, R., Honeybourne, J., Khan, S., Lo, U., McGaha, R., Plunkett, J., … Elgin, S. C. R. (2010). Evolution of a Distinct Genomic Domain in Drosophila: Comparative Analysis of the Dot Chromosome in Drosophila melanogaster and Drosophila virilis. *Genetics*, *185*(4), 1519. https://doi.org/10.1534/GENETICS.110.116129
- Levine, M., Cattoglio, C., & Tjian, R. (2014). Looping Back to Leap Forward: Transcription Enters a New Era. *Cell*, *157*(1), 13–25. https://doi.org/10.1016/J.CELL.2014.02.009
- Li, S., Hannenhalli, S., & Ovcharenko, I. (2023). De novo human brain enhancers created by single-nucleotide mutations. *Science Advances*, *9*(7).
- Lonfat, N., Montavon, T., Darbellay, F., Gitto, S., & Duboule, D. (2014). Convergent evolution of complex regulatory landscapes and pleiotropy at Hox loci. *Science*, *346*(6212), 1004–1006.
- Long, H. K., Prescott, S. L., & Wysocka, J. (2016). Ever-Changing Landscapes: Transcriptional Enhancers in Development and Evolution. *Cell*, *167*(5), 1170–1187. https://doi.org/10.1016/J.CELL.2016.09.018
- MacPhillamy, C., Alinejad-Rokny, H., Pitchford, W. S., & Low, W. Y. (2022). Cross-species enhancer prediction using machine learning. *Genomics*, *114*(5). https://doi.org/10.1016/J.YGENO.2022.110454
- Markow, T. A., & O'Grady, P. M. (2005). Evolutionary genetics of reproductive behavior in Drosophila: connecting the dots. *Annual Review of Genetics*, *39*, 263–291. https://doi.org/10.1146/ANNUREV.GENET.39.073003.112454
- Markow, T. A., & O'Grady, P. M. (2007). Drosophila Biology in the Genomic Age. *Genetics*, *177*(3), 1269. https://doi.org/10.1534/GENETICS.107.074112
- Markstein, M., & Levine, M. (2002). Decoding cis-regulatory DNAs in the Drosophila genome. *Current Opinion in Genetics and Development*, *12*(5), 601–606. https://doi.org/10.1016/S0959-437X(02)00345-3
- Massouras, A., Waszak, S. M., Albarca-Aguilera, M., Hens, K., Holcombe, W., Ayroles, J. F., Dermitzakis, E. T., Stone, E. A., Jensen, J. D., Mackay, T. F. C., & Deplancke, B. (2012). Genomic Variation and Its Impact on Gene Expression in Drosophila melanogaster. *PLOS Genetics*, *8*(11), e1003055. https://doi.org/10.1371/JOURNAL.PGEN.1003055
- McKay, D. J., & Lieb, J. D. (2013). A common set of DNA regulatory elements shapes Drosophila appendages. *Developmental Cell*, *27*(3), 306–318. https://doi.org/10.1016/J.DEVCEL.2013.10.009
- Meisel, R. P., Malone, J. H., & Clark, A. G. (2012). Faster-X Evolution of Gene Expression in Drosophila. *PLOS Genetics*, *8*(10), e1003013. https://doi.org/10.1371/JOURNAL.PGEN.1003013
- Mellerick, D. M., & Nirenberg, M. (1995). Dorsal-ventral patterning genes restrict NK-2 homeobox gene expression to the ventral half of the central nervous system of Drosophila embryos. *Developmental Biology*, *171*(2), 306–316. https://doi.org/10.1006/DBIO.1995.1283

- Mikhaylichenko, O., Bondarenko, V., Harnett, D., Schor, I. E., Males, M., Viales, R. R., & Furlong, E. E. M. (2018). The degree of enhancer or promoter activity is reflected by the levels and directionality of eRNA transcription. *Genes & Development*, *32*(1), 42–57. https://doi.org/10.1101/GAD.308619.117
- Minnoye, L., Taskiran, I. I., Mauduit, D., Fazio, M., van Aerschot, L., Hulselmans, G., Christiaens, V., Makhzami, S., Seltenhammer, M., Karras, P., Primot, A., Cadieu, E., van Rooijen, E., Marine, J. C., Egidy, G., Ghanem, G. E., Zon, L., Wouters, J., & Aerts, S. (2020). Cross-species analysis of enhancer logic using deep learning. *Genome Research*, *30*(12), 1815–1834. https://doi.org/10.1101/GR.260844.120
- Mirol, P. M., Routtu, J., Hoikkala, A., & Butlin, R. K. (2020). Signals of demographic expansion in Drosophila virilis. *BMC Evolutionary Biology*, *8*(1). https://doi.org/10.1186/1471-2148-8-59
- Mirzoyan, Z., Sollazzo, M., Allocca, M., Valenza, A. M., Grifoni, D., & Bellosta, P. (2019). Drosophila melanogaster: A Model Organism to Study Cancer. *Frontiers in Genetics*, *10*. https://doi.org/10.3389/FGENE.2019.00051
- Mohr, S. E. (2018). *First in fly : Drosophila research and biological discovery*. Harvard University Press. https://www.hup.harvard.edu/books/9780674971011
- Morgan, T. H., & Bridges, C. B. (1916). *Sex-linked Inheritance in Drosophila*. Carnegie Institution of Washington. https://books.google.com.ar/books?id=JNQGAAAAYAAJ
- Morgan, T. H. (1910). Sex Limited Inheritance in *Drosophila*. *Science*, *32*(812), 120–122. https://doi.org/10.1126/science.32.812.120
- O'Grady, P. M., & DeSalle, R. (2018). Phylogeny of the Genus Drosophila. *Genetics*, *209*(1), 1–25. https://doi.org/10.1534/GENETICS.117.300583
- Paaby, A. B., & Rockman, M. v. (2013). The many faces of pleiotropy. *Trends in Genetics : TIG*, *29*(2), 66–73. https://doi.org/10.1016/J.TIG.2012.10.010
- Peng, P. C., Khoueiry, P., Girardot, C., Reddington, J. P., Garfield, D. A., Furlong, E. E. M., & Sinha, S. (2019). The Role of Chromatin Accessibility in cis-Regulatory Evolution. *Genome Biology and Evolution*, *11*(7), 1813–1828. https://doi.org/10.1093/GBE/EVZ103
- Preger-Ben Noon, E., Sabarís, G., Ortiz, D. M., Sager, J., Liebowitz, A., Stern, D. L., & Frankel, N. (2018). Comprehensive Analysis of a cis-Regulatory Region Reveals Pleiotropy in Enhancer Function. *Cell Reports*, *22*(11), 3021–3031. https://doi.org/10.1016/J.CELREP.2018.02.073
- Pueyo, J. I., & Couso, J. P. (2011). Tarsal-less peptides control Notch signalling through the Shavenbaby transcription factor. *Developmental Biology*, *355*(2), 183–193. https://doi.org/10.1016/J.YDBIO.2011.03.033
- Ramirez, M., Badayeva, Y., Yeung, J., Wu, J., Abdalla-Wyse, A., Yang, E., Trost, B., Scherer, S. W., & Goldowitz, D. (2022). Temporal analysis of enhancers during mouse cerebellar development reveals dynamic and novel regulatory functions. *ELife*, *11*. https://doi.org/10.7554/ELIFE.74207
- Ray, S., Rosenberg, M. I., Chanut-Delalande, H., Decaras, A., Schwertner, B., Toubiana, W., Auman, T., Schnellhammer, I., Teuscher, M., Valenti, P., Khila, A., Klingler, M., & Payre, F. (2019). The mlpt/Ubr3/Svb module comprises an ancient developmental switch for embryonic patterning. *ELife*, *8*. https://doi.org/10.7554/ELIFE.39748
- Rebeiz, M., & Tsiantis, M. (2017). Enhancer evolution and the origins of morphological novelty. *Current Opinion in Genetics & Development*, *45*, 115–123. https://doi.org/10.1016/J.GDE.2017.04.006
- Robinson, J. T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E. S., Getz, G., & Mesirov, J. P. (2011). Integrative Genomics Viewer. *Nature Biotechnology*, *29*(1), 24. https://doi.org/10.1038/NBT.1754
- Sabarís, G., Laiker, I., Preger-Ben Noon, E., & Frankel, N. (2019). Actors with Multiple Roles: Pleiotropic Enhancers and the Paradigm of Enhancer Modularity. *Trends in Genetics : TIG*, *35*(6), 423–433. https://doi.org/10.1016/J.TIG.2019.03.006

- Salz, H. K., Cline, T. W., & Schedl, P. (1987). Functional changes associated with structural alterations induced by mobilization of a P element inserted in the Sex-lethal gene of Drosophila. *Genetics*, *117*(2), 221–231. https://doi.org/10.1093/GENETICS/117.2.221
- Schep, R., Necsulea, A., Rodríguez-Carballo, E., Guerreiro, I., Andrey, G., Huynh, T. H. N., Marcet, V., Zákány, J., Duboule, D., & Beccari, L. (2016). Control of Hoxd gene transcription in the mammary bud by hijacking a preexisting regulatory landscape. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(48), E7720–E7729.
- Siepel, A., Bejerano, G., Pedersen, J. S., Hinrichs, A. S., Hou, M., Rosenbloom, K., Clawson, H., Spieth, J., Hillier, L. D. W., Richards, S., Weinstock, G. M., Wilson, R. K., Gibbs, R. A., Kent, W. J., Miller, W., & Haussler, D. (2005). Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Research*, *15*(8), 1034–1050. https://doi.org/10.1101/GR.3715005
- Singh, D., & Yi, S. v. (2021). Enhancer Pleiotropy, Gene Expression, and the Architecture of Human Enhancer–Gene Interactions. *Molecular Biology and Evolution*, *38*(9), 3898–3909. https://doi.org/10.1093/MOLBEV/MSAB085
- Spitz, F., & Furlong, E. E. M. (2012). Transcription factors: from enhancer binding to developmental control. *Nature Reviews. Genetics*, *13*(9), 613–626. https://doi.org/10.1038/NRG3207
- Stern, D. L., & Franke, N. (2013). The structure and evolution of cis-regulatory regions: the shavenbaby story. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *368*(1632). https://doi.org/10.1098/RSTB.2013.0028
- Sucena, E., Delon, I., Jones, I., Payre, F., & Stern, D. L. (2003). Regulatory evolution of shavenbaby/ovo underlies multiple cases of morphological parallelism. *Nature 2003 424:6951*, *424*(6951), 935–938. https://doi.org/10.1038/nature01768
- Szabo, Q., Jost, D., Chang, J. M., Cattoni, D. I., Papadopoulos, G. L., Bonev, B., Sexton, T., Gurgo, J., Jacquier, C., Nollmann, M., Bantignies, F., & Cavalli, G. (2018). TADs are 3D structural units of higher-order chromosome organization in Drosophila. *Science Advances*, *4*(2). https://doi.org/10.1126/SCIADV.AAR8082/SUPPL_FILE/AAR8082_SM.PDF
- van Berkum, N. L., & Dekker, J. (2009). Determining spatial chromatin organization of large genomic regions using 5C technology. *Methods in Molecular Biology (Clifton, N.J.)*, *567*, 189–213. https://doi.org/10.1007/978-1-60327-414-2_13
- Vizcaya-Molina, E., Klein, C. C., Serras, F., Mishra, R. K., Guigó, R., & Corominas, M. (2018). Damage-responsive elements in Drosophila regeneration. *Genome Research*, *28*(12), 1841–1851. https://doi.org/10.1101/GR.233098.117
- Weirauch, M. T., & Hughes, T. R. (2010). Conserved expression without conserved regulatory sequence: the more things change, the more they stay the same. *Trends in Genetics : TIG*, *26*(2), 66–74. https://doi.org/10.1016/J.TIG.2009.12.002
- Wittkopp, P. J., & Kalay, G. (2011). Cis-regulatory elements: molecular mechanisms and evolutionary processes underlying divergence. *Nature Reviews. Genetics*, *13*(1), 59–69. https://doi.org/10.1038/NRG3095
- Wong, E. S., Zheng, D., Tan, S. Z., Bower, N. I., Garside, V., Vanwalleghem, G., Gaiti, F., Scott, E., Hogan, B. M., Kikuchi, K., McGlinn, E., Francois, M., & Degnan, B. M. (2020). Deep conservation of the enhancer regulatory code in animals. *Science*, *370*(6517). https://doi.org/10.1126/SCIENCE.AAX8137/SUPPL_FILE/AAX8137_WONG_SM.PDF
- Zou, Z., Ohta, T., Miura, F., & Oki, S. (2022). ChIP-Atlas 2021 update: a data-mining suite for exploring epigenomic landscapes by fully integrating ChIP-seq, ATAC-seq and Bisulfite-seq data. *Nucleic Acids Research*, *50*(W1), W175–W182. https://doi.org/10.1093/NAR/GKAC199