



UNIVERSIDAD DE CONCEPCION

Dirección de Postgrado

Facultad de Ciencias Biológicas

Programa de Doctorado

**Evolución de elementos *cis*-regulatorios originados por
elementos transponibles en los genomas de *Xenopus
tropicalis* y *Xenopus laevis***

Tesis para optar al grado de Doctor en Ciencias Biológicas

Área Biología Celular y Molecular

Japhet Josue Caleb Rojas Escobar

Concepción-Chile

2026

Profesor Guía: Dr. Sylvain G. Marcellini Liotaud

Dpto. de Biología Celular, Facultad de Ciencias Biológicas

Universidad de Concepción

Esta tesis ha sido realizada en el Departamento de Biología Celular de la Facultad de Ciencias Biológicas, Universidad de Concepción.

Profesores integrantes Comisión Evaluadora:

Dr. Sylvain Marcellini L.
Profesor Guía de Tesis
Facultad de Ciencias Biológicas
Dpto. Biología celular

Dra. Roxana Pincheira
Comisión interna
Facultad de Ciencias Biológicas
Dpto. Bioquímica y Biología Molecular

Dr. José Leonardo Gutiérrez
Comisión interna
Facultad de Ciencias Biológicas
Dpto. Bioquímica y Biología Molecular

Dr. Fernando Faunes
Comisión externa Universidad Andrés Bello
Facultad de Ciencias de la Vida
Dpto. de Ciencias Biológicas

Directora Dra. Marcela Torrejón
Programa de Doctorado en Ciencias Biológicas,
Área Biología Celular y Molecular. Facultad
Ciencias Biológicas

AGRADECIMIENTOS

Me gustaría agradecer a mi familia: Partiendo por **Efraín Valencia**, mi esposo y pareja desde que comencé mi proyecto de tesis. Estuvo presente en mi defensa de proyecto, me acompañó durante la escritura de esta tesis y estará también a mi lado durante su defensa. Quiero agradecer, además, a toda la familia Arriagada y Valencia, quienes me han recibido como un hijo más. Estoy profundamente contento de formar parte de una familia tan amorosa, humilde, graciosa, cariñosa y linda. Agradezco especialmente a mi madre, **Violeta Escobar**, quien me ha brindado todo su apoyo tanto en los momentos felices como en aquellos más difíciles que atravesé durante mi trayectoria como estudiante de doctorado. A **Carla Schürch**, quien ha sido realmente como una hermana para mí desde que llegué a Concepción, entregándome siempre su apoyo incondicional y confianza. A mis gatos, **Juno**, **Atún** y **Mandarina**, por acompañarme siempre. También a mis sobrinos, **Antonia R. Leiva**, **Tomás R. Orellana**, **León R. Orellana** y **Emma R. Orellana**, quienes han hecho mi vida mucho más feliz. Gracias de corazón por todo su cariño. Quiero agradecer también a **Pamela Orellana**, madre de mis tres últimos sobrinos mencionados, quien ha sido un apoyo fundamental en mi vida. Finalmente, agradezco a mi hermano, **Ángel Rojas Escobar**, por su apoyo y por estar presente en mi vida.

Agradecer a mis amigos: A todos ellos, con quienes he compartido momentos inolvidables: **Leonardo Jara**, **Fabiana Pávez**, **Jaime González**, **Tamara Piña**, **Nidia Torres**, **Felipe Maurelia**, **Monserrat Carrillo**, **Ninoschka Troncoso**, **Sandrines Bruzual**, **Javier Fumeron**, **Valeria Nova**, **Lefney Cumilaf** y **Fabiana Vivar**.

Agradecer a mi laboratorio (LADE): Mi experiencia durante el doctorado, sin duda, no habría sido la misma si no hubiera tenido la oportunidad de conocer y compartir con todos los miembros del LADE: **Héctor Castillo**, **Marco Mundaca**, **Lefney Cumilaf**, **Roberto Amigo**, **Fabiana Vivar**, **Fernanda Flores** y **Guillermo Ditzel**. Gracias por el compañerismo, las experiencias compartidas y por haber sido parte fundamental de esta etapa.

Agradecer a mi tutor, Dr. Sylvain Marcellini: Estoy profundamente agradecido por toda su tutela, apoyo y confianza, que me permitieron desarrollarme y trabajar en el LADE, así como por todo lo que aprendí bajo su supervisión y que hoy se encuentra plasmado en esta tesis. Sin duda, ha sido un tutor detallista, inteligente y comprensivo, además de una gran persona y un excelente científico. De verdad, muchas gracias por todo.

Agradecer al comité del programa de doctorado: Agradezco a la **Dra. Marcela Torrejón**, la **Dra. Teresa Caprile**, la **Dra. María de los Ángeles**, el **Dr. Sylvain Marcellini**, el **Dr. Ariel Castro** y el **Dr. José L. Guzman** por permitirme formar parte de las discusiones y decisiones desarrolladas durante mi periodo como alumno delegado, así como por darme la oportunidad de contribuir al exitoso proceso de reacreditación por ocho años de nuestro programa de doctorado. Muchas gracias por la confianza y por la experiencia.

Agradecer a investigadores colaboradores: Agradezco al **Dr. Braulio Valdebenito** por su ayuda en la generación de datos mediante RepeatMasker y por su disposición para enseñarme análisis bioinformáticos relacionados con la identificación de transposones. Al **Dr. Daniel Buchholz**, por su ayuda en la estandarización de la transgénesis mediada por meganucleasa en *Xenopus*. Asimismo, agradezco a todo el equipo que me recibió durante mi pasantía internacional de investigación en NeuroPSI (Université Paris-Saclay), en especial a **Albert Chesneau**, por enseñarme a realizar transgénesis REMI, cuyos resultados forman parte de esta tesis.

Agradecer a las fuentes de financiamiento que hicieron posible la realización de esta tesis:
Beca de Doctorado Nacional ANID y Gastos Complementarios, N.º **21221189**.
Fondecyt Regular, proyecto N.º **1241095**.
Dirección de Postgrado, Universidad de Concepción.

ÍNDICE

ÍNDICE	IV
ÍNDICE DE FIGURAS:	VI
ÍNDICE DE ANEXOS.	VII
ABREVIATURAS	VIII
RESUMEN.....	1
ABSTRACT	2
1. INTRODUCCIÓN.....	3
1.1. ORGANIZACIÓN FUNCIONAL DE LOS GENOMAS EUCARIOTAS.....	3
1.1.1. Regulación transcripcional en eucariotas: Secuencias <i>cis</i> -regulatorias.	3
1.1.1.1. Los promotores de genes.....	3
1.1.1.2. Potenciadores de genes.....	4
1.1.2. Secuencias repetitivas.....	5
1.1.2.1 Elementos transponibles.	5
1.1.2.1. Retrotransposones.	6
1.1.2.2. Transposones de ADN.	7
1.1.2.3. Elementos transponibles de repetición invertida en miniatura.....	8
1.1.3. Técnicas de secuenciación para el estudio del control transcripcional y elementos transponibles.	9
1.2. DOMESTICACIÓN DE ELEMENTOS TRANSPONIBLES.	10
1.2.1. Domesticación de TEs en promotores.....	10
1.2.2. Domesticación de TEs en enhancers.....	11
1.3. <i>XENOPUS</i> COMO MODELO PARA ESTUDIAR DOMESTICACIÓN DE TES DE ADN.	12
1.3.1. <i>Xenopus tropicalis</i> y <i>Xenopus laevis</i> : dos especies, tres genomas.....	12
1.3.2. Elementos transponibles identificados en los genomas de <i>Xenopus</i>	13
1.3.3 Estudios de elementos regulatorios en <i>Xenopus</i> mediante experimentos de transgénesis.	14
1.4. PREGUNTA DE INVESTIGACIÓN, HIPÓTESIS Y OBJETIVOS.	15
2. MATERIALES Y MÉTODOS	16
2.1. PRESENCIA DE TES EN EL PAISAJE REGULATORIO DE <i>XT</i> Y <i>XL</i>	16
2.1.1. Datos de ATAC-Seq y RNA-Seq.....	16
2.1.2. Cuantificación de la cobertura de elementos transponibles en regiones <i>cis</i> -regulatorias.	17
2.1.3. Asignación de genes a promotores asociados con KN9.	17
2.1.4. Alineamiento múltiple e inferencia filogenética de KN9.	17
2.1.5. Análisis exploratorio de la localización subcelular de proteínas asociadas con KN9.....	18
2.2. PATRÓN DE EXPRESIÓN ESPACIOTEMPORAL DE GENES ASOCIADOS CON PROMOTORES QUE CONTIENEN KN9.....	18
2.2.1. Diseño de sondas para hibridación <i>in situ</i> fluorescente HCR.....	18
2.2.2. Hibridación <i>in situ</i> fluorescente HCR.	19
2.2.3. Microscopía y procesamiento de imágenes obtenidas de <i>in situ</i> HCR.	20
2.3. VALIDACIÓN FUNCIONAL <i>IN VIVO</i> DE LA ACTIVIDAD PROMOTORA DE KN9.....	20
2.3.1. Diseño del vector de transgénesis.	20

2.3.2. Preparación de espermios de <i>Xl</i> .	20
2.3.3. Microinyección de núcleos de espermatozoides y ADN.	21
2.3.4. Microscopía y procesamiento de imágenes de larvas transgénicas vivas.	21
3. RESULTADOS	23
3.1. PRESENCIA DE TRANSPOSONES EN EL PAISAJE REGULATORIO EN <i>XENOPUS</i> .	23
3.2. ELEMENTO TRANSPONIBLE EN MINIATURA (MITE) KN9.	26
3.3. LOCALIZACIÓN SUBCELULAR DE LOS ORTÓLOGOS HUMANOS CODIFICADOS POR GENES ASOCIADOS A PROMOTORES DERIVADOS DE KN9 EN <i>XENOPUS</i> .	31
3.4. EXPRESIÓN ESPACIO-TEMPORAL DE GENES REGULADOS POR PROMOTORES DERIVADOS DE KN9.	31
3.5. ENSAYOS REPORTEROS PARA CORROBORAR LA ACTIVIDAD PROMOTORA DE KN9.	43
4. DISCUSIÓN	46
4.1. TRANSPOSONES ASOCIADOS A REGIONES DE CROMATINA ABIERTA.	46
4.2. ENRIQUECIMIENTO DE TRANSPOSONES DE ADN EN REGIONES REGULATORIAS DE <i>Xt</i> Y <i>XL</i> .	46
4.3. ENRIQUECIMIENTO DE RETROTRANSPOSONES EN REGIONES REGULATORIAS DE <i>XL</i> .	47
4.4. CONSERVACIÓN ANCESTRAL DEL MITE KN9 EN REGIONES PROMOTORAS EN <i>XENOPUS</i> .	48
4.5. ESCENARIOS EVOLUTIVOS DE INSERCIÓN Y DOMESTICACIÓN DE KN9 COMO PROMOTORES.	48
4.6. ELEMENTOS REGULADORES PUTATIVOS IDENTIFICADOS EN LAS SECUENCIAS DE LAS COPIAS DOMESTICADAS DE KN9.	49
4.7. PATRÓN REGULATORIO NEURAL DE KN9 SOBRE GENES CANDIDATOS ANALIZADOS DURANTE EL DESARROLLO EN <i>XENOPUS</i> .	51
4.8. VALIDACIÓN FUNCIONAL <i>IN VIVO</i> DE LA ACTIVIDAD PROMOTORA DE KN9.	53
4.9. MODELO INTEGRATIVO SOBRE EL MECANISMO DE ACCIÓN DE KN9.	53
5. CONCLUSIONES	56
7. ANEXOS	57

ÍNDICE DE FIGURAS:

Figura 1: Filogenia de los anfibios del género <i>Xenopus</i> .-----	13
Figura 2: Esquema de la obtención de bases de datos de elementos cis-regulatorios en <i>Xenopus</i> .-----	23
Figura 3: Enriquecimiento local específico de cinco subfamilias de MITEs en <i>Xt</i> .-----	24
Figura 4: Enriquecimiento local específico de ocho subfamilias de MITEs en <i>Xl</i> . -----	25
Figura 5: Enriquecimiento local específico de siete subfamilias de retrotransposones en <i>Xl</i> . -----	26
Figura 6: Enriquecimiento local del MITE KN9 en promotores ubicuos. -----	27
Figura 7: Resumen de los escenarios evolutivos identificados para los promotores originados a partir del MITE KN9. -----	28
Figura 8: Alineamiento múltiple de las secuencias de copias domesticadas de KN9 de <i>Xt</i> y de los <i>Xl</i> y <i>Xis</i> de <i>Xl</i> respecto de la secuencia consenso. -----	29
Figura 9: Árbol filogenético de los MITEs KN9 de interés que dieron origen a promotores del clúster ubicuo, denominados según su gen diana. -----	30
Figura 10: Localización subcelular reportada para los ortólogos humanos de los genes asociados a promotores derivados de KN9. -----	31
Figura 11: Mapas de calor de los genes candidatos regulados por KN9. -----	32
Figura 12: Hibridación in situ fluorescente mediante HCR para la caracterización espaciotemporal del gen <i>rtn3</i> , regulado por KN9, en <i>Xt</i> . -----	36
Figura 13: Hibridación in situ fluorescente mediante HCR para la caracterización espaciotemporal del gen <i>tmbim6</i> , regulado por KN9, en <i>Xt</i> . -----	37
Figura 14: Hibridación in situ fluorescente mediante HCR para la caracterización espaciotemporal del gen <i>tmbim6.L</i> , regulado por KN9, en <i>Xl</i> . -----	38
Figura 15: Hibridación in situ fluorescente mediante HCR para la caracterización espaciotemporal del gen <i>abitrn</i> , regulado por KN9, en <i>Xt</i> . -----	39
Figura 16: Hibridación in situ fluorescente mediante HCR para la caracterización espaciotemporal de los genes <i>faah2</i> de <i>Xt</i> y <i>faah2.L</i> de <i>Xl</i> , regulados por KN9. -----	40
Figura 17: Hibridación in situ fluorescente mediante HCR para la caracterización espaciotemporal del gen <i>tsopap1</i> , regulado por KN9, en <i>Xt</i> .-----	41
Figura 18: Hibridación in situ fluorescente mediante HCR para la caracterización espaciotemporal del gen <i>nsmce1</i> , regulado por KN9, en <i>Xt</i> .-----	42
Figura 19: Vector utilizado en transgénesis mediada por enzima de restricción en <i>Xl</i> . -----	44
Figura 20: Evaluación de la actividad promotora de KN9 mediante transgénesis REMI en larvas de <i>Xenopus</i> .-----	45
Figura 21: Modelo propuesto para el mecanismo de acción de KN9. -----	55

ÍNDICE DE TABLAS.

Tabla 1: Tamaño de los genomas y porcentaje de elementos transponibles (TEs) presentes en diferentes genomas eucariotas. -----	14
Tabla 2: Secuencias de los sitios de unión para factores de transcripción de las familias Sp/KLF, PBC/MEINOX y ETS presentes en la secuencia consenso de KN9. -----	35

ÍNDICE DE ANEXOS.

Anexo 1: Listado de genes bajo regulación de MITEs Kolobok-N9_XT domesticados en promotores ubicuos. -----	57
Anexo 2: Subfamilias de Transposones de ADN identificadas en <i>X. tropicalis</i> y su presencia en <i>X. laevis</i> . -----	60
Anexo 3: Subfamilias de Transposones de ADN identificadas en <i>X. laevis</i> y su presencia en <i>X. tropicalis</i> . -----	60
Anexo 4: Enlace a todos los mapas de calor elaborados para las copias domesticadas de KN9 para sus respectivos genes blancos. -----	61
Anexo 5: Alineamiento de una copia de la subfamilia K de T2-MITEs identificada por Hikosaka y Kawahara con la secuencia consenso de Kolobok-N9_XT. -----	61
Anexo 6: Enlace hacia la data genómica publicada sobre <i>Nanorana parkeri</i> , incluido archivo de Repeatmasker con la anotación de elementos transponibles. -----	61
Anexo 7: Presencia de KN9 como promotor en los genes homólogos <i>borcs6</i> y <i>borcs7</i> . -----	62
Anexo 8: Presencia de KN9 como promotor de genes vecinos en dirección antisentido <i>atg101</i> y <i>tmbim6</i> . -----	62
Anexo 9: Presencia de KN9 como promotor bidireccional para los genes <i>as3mt.1</i> y <i>borcs7</i> solo en <i>X. tropicalis</i> . -----	62
Anexo 10: Ocupación de las copias NO domesticadas de KN9 a lo largo de la secuencia consenso -----	63
Anexo 11: Perfil de conservación de la región 3' para las copias domesticadas y no domesticadas de KN9 con respecto a la secuencia consenso. -----	63
Anexo 12: Data de chipseq de humano obtenida de https://chip-atlas.org/ para miembros de las familias de TFBs conservadas en KN9. (A) <i>rtn3</i> , (B) <i>tmbim6</i> , (C) <i>abitram</i> , (D) <i>faah2</i> , (E) <i>tspoap1</i> , y (F) <i>nsmce1</i> . -----	64
Anexo 13: Alineamiento de la secuencia consenso Kolobok-N9_XT con el KN9 presente en el promotor de los genes <i>atg101</i> y <i>tmbim6</i> . -----	65

ABREVIATURAS

FT: factor de transcripción

FTs: factores de transcripción

Xt: *Xenopus tropicalis*

Xl: *Xenopus laevis*

XIL: subgenoma L de *Xenopus laevis*

XIS: subgenoma S de *Xenopus laevis*

KN9: Kolobok-N9_XT

RESUMEN

Los elementos transponibles (TEs) constituyen una fuente potencial de innovación regulatoria; sin embargo, la contribución de los transposones de ADN, particularmente de los elementos transponibles de repetición invertida en miniatura (MITEs), ha sido escasamente estudiada en vertebrados. Esta tesis investigó la dinámica de domesticación de TEs en regiones *cis*-regulatorias de *Xenopus tropicalis* (*Xt*) y *Xenopus laevis* (*Xl*), incluidos los subgenomas L (*XIL*) y S (*XIS*). Para ello, se integraron datos de ATAC-seq y ARN-seq de hígado, mesénquima y osteoblastos con anotaciones de secuencias repetitivas. Las regiones de cromatina accesible se clasificaron como promotores o elementos *cis*-regulatorios distales y se analizaron para identificar enriquecimientos locales de subfamilias de TEs. Posteriormente, se caracterizó evolutiva y funcionalmente el MITE Kolobok-N9_XT (KN9), asociado con promotores del clúster ubicuo. KN9 se encontró enriquecido en promotores ubicuos de ambas especies. El análisis de 141 promotores correspondientes a 60 genes permitió clasificar 101 casos como domesticaciones completas, 16 como estados intermedios en los que coexistían el promotor ancestral y el derivado de KN9, y 24 como casos en los que se conservaba únicamente el promotor original. La conservación de inserciones ortólogas y su agrupamiento filogenético respaldan que numerosos eventos de cooptación precedieron a la divergencia entre *Xt* y *Xl*. Las copias domesticadas conservaron principalmente el extremo 3' de KN9, que contiene motivos putativos de unión para factores de transcripción de las familias ETS, SP/KLF y PBC/MEINOX, además de un motivo AGGT compatible con un sitio donador de empalme. La expresión de seis genes candidatos se evaluó mediante HCR. Los cinco genes con señal detectable presentaron una marcada convergencia hacia territorios neurales, aunque varios también mostraron una expresión más amplia. Finalmente, un fragmento de KN9 asociado con *tmbim6* fue suficiente para dirigir la expresión de GFP nuclear en el encéfalo y el epitelio olfatorio de larvas transgénicas de *Xl*. En conjunto, estos resultados muestran que los TEs han contribuido al paisaje regulatorio de *Xenopus* mediante eventos ancestrales y específicos de linaje, y establecen a KN9 como un MITE cooptado con actividad promotora *in vivo*.

ABSTRACT

Transposable elements (TEs) constitute a potential source of regulatory innovation; however, the contribution of DNA transposons, particularly miniature inverted-repeat transposable elements (MITEs), remains poorly characterized in vertebrates. This thesis investigated the dynamics of TE co-option into *cis*-regulatory regions in *Xenopus tropicalis* (*Xt*) and *Xenopus laevis* (*Xl*), including the L (*Xl*) and S (*Xl*) subgenomes. ATAC-seq and RNA-seq datasets from liver, mesenchyme, and osteoblasts were integrated with repeat annotations. Accessible chromatin regions were classified as promoters or distal *cis*-regulatory elements and examined to identify local enrichment of TE subfamilies. The MITE Kolobok-N9_XT (KN9), associated with promoters from the ubiquitous cluster, was subsequently subjected to evolutionary and functional characterization. KN9 was enriched in ubiquitous promoters in both species. Comparative analysis of 141 promoter associated with 60 genes classified 101 cases as complete domestications, 16 as intermediate states in which the ancestral and KN9-derived promoters coexisted, and 24 as cases retaining only the ancestral promoter. The conservation of orthologous insertions and their phylogenetic clustering support the conclusion that multiple co-option events predated the divergence of *Xt* and *Xl*. Domesticated copies predominantly retained the 3' region of KN9, which contains putative binding motifs for ETS, SP/KLF, and PBC/MEINOX transcription factors, together with an AGGT motif compatible with a splice donor site. The spatiotemporal expression of six candidate genes was examined using HCR RNA-FISH. The five genes with detectable signals displayed a marked convergence toward neural territories, although several also showed broader expression. Finally, a KN9 fragment associated with the *tmbim6* promoter was sufficient to drive nuclear GFP expression in the brain and olfactory epithelium of transgenic *Xl* larvae. Collectively, these findings show that TEs have contributed to the *Xenopus* regulatory landscape through both ancestral and lineage-specific events and establish KN9 as an evolutionarily co-opted MITE with promoter activity *in vivo*.

1. INTRODUCCIÓN.

1.1. Organización funcional de los genomas eucariotas.

Los genomas eucariotas están compuestos por secuencias codificantes y una amplia fracción no codificante. Aunque las regiones codificantes han sido históricamente las más estudiadas debido a su relación directa con la producción de proteínas, representan una proporción minoritaria del genoma humano, cercana al 1,2 % (1). En contraste, gran parte del genoma corresponde a secuencias no codificantes, las cuales fueron consideradas durante décadas bajo el concepto de “ADN basura” (2–4). Sin embargo, actualmente se reconoce que una fracción de este ADN participa en procesos clave como el corte y empalme del pre-ARNm (5), la organización tridimensional y reordenamiento del genoma (6–8), producción de ARNs pequeños (9,10), y la presencia de secuencias *cis*-reguladoras (11,12) y secuencias repetitivas (13,14).

1.1.1. Regulación transcripcional en eucariotas: Secuencias *cis*-regulatorias.

La expresión génica en eucariotas depende de la interacción entre secuencias *cis*-reguladoras, factores de transcripción (FTs) y el contexto local de la cromatina (15,16). Las secuencias *cis*-reguladoras corresponden a regiones de ADN que modulan la actividad transcripcional de genes (17). Entre ellas se encuentran promotores, potenciadores (Enhancers, en inglés), silenciadores e insuladores, los cuales contribuyen a definir cuándo, dónde y en qué nivel se expresa un gen. Sin negar la diversidad y complejidad de mecanismos regulatorios, el presente trabajo tiene un enfoque particular hacia promotores y potenciadores, los cuales serán descritos en más detalles a continuación.

1.1.1.1. Los promotores de genes.

Los promotores corresponden a regiones de ADN que definen el sitio de inicio de la transcripción y permiten el reclutamiento de la maquinaria transcripcional. Estas regiones contienen distintos elementos de secuencia, como el iniciador (Inr), la caja TATA y motivos de unión para factores de transcripción, cuya combinación contribuye a determinar la actividad promotora (18–21). En eucariotas, la ARN polimerasa II y los factores generales de transcripción se ensamblan en torno al sitio de inicio de la transcripción (TSS), dentro de una región conocida como promotor basal (Core promoter”, en inglés), generalmente ubicada entre aproximadamente 50 pb río arriba y río abajo del TSS, permitiendo la formación del complejo de pre-iniciación transcripcional (22,23).

Los promotores pueden clasificarse en distintas categorías según su composición de secuencia y su patrón de actividad. En términos generales, algunos promotores se asocian a genes ampliamente expresados, también conocidos como genes ubicuos o de mantenimiento, cuya actividad se detecta en una gran variedad de tipos celulares y que suelen cumplir funciones esenciales para la supervivencia y el mantenimiento celular (24–26). Otros promotores, en cambio, presentan patrones de actividad célula- o tejido-específicos, asociados a genes cuya expresión se

restringe espacial o temporalmente durante el desarrollo. Estos genes contribuyen a definir la identidad y función característica de tejidos o tipos celulares particulares, por lo que su regulación precisa es fundamental para el desarrollo embrionario (11).

A nivel de secuencia, la caja TATA es uno de los elementos promotores más caracterizados y se asocia frecuentemente con genes de expresión regulada o tejido-específica (18–21). En torno a esta región también pueden encontrarse elementos de reconocimiento de FTs conocidos como BRE (“B recognition elements”, en inglés), los cuales participan en el reclutamiento de TFIIB, un componente esencial del complejo de iniciación asociado a la ARN polimerasa II (27). Otro elemento presente en algunos promotores corresponde al DCE (“downstream core element”, en inglés), ubicado río abajo del TSS, el cual puede interactuar con componentes del complejo TFIID, como TAF1 (28). Además, el iniciador (Inr) puede dirigir el inicio de la transcripción incluso en ausencia de una caja TATA (18).

Por otra parte, muchos promotores presentan islas CpG, regiones enriquecidas en dinucleótidos citosina-guanina, una característica conservada entre mamíferos y *Xenopus tropicalis* (29). Los promotores asociados a genes ampliamente expresados suelen presentar propiedades particulares, como un sesgo en la distribución de guanina y citosina a ambos lados del TSS (“GC-skewing” en inglés), con mayor proporción de citosina río arriba y de guanina río abajo del TSS (30). Si bien un promotor basal puede ser suficiente para iniciar la transcripción, en muchos casos la regulación precisa de la expresión génica depende además de la acción combinada de factores de transcripción y potenciadores.

1.1.1.2. Potenciadores de genes.

Los potenciadores de genes corresponden a elementos *cis*-reguladores que aumentan la expresión de sus genes blanco y pueden actuar desde posiciones variables dentro del genoma, incluso a grandes distancias del promotor que regulan (31). Estos elementos suelen presentar longitudes aproximadas entre 200 y 500 pb, y se localizan principalmente en regiones intergénicas o intrónicas (32). Además, contienen múltiples sitios de unión para FTs, cuya combinación contribuye a definir su actividad regulatoria y especificidad celular o tisular (33). Su mecanismo de acción se basa en el reclutamiento de FTs y cofactores transcripcionales, los cuales pueden favorecer la activación de promotores diana ubicados desde una cercanía inmediata, menor a un kilobase, hasta distancias de megabases respecto del potenciador (34–38). Esta interacción depende, en gran medida, de la organización tridimensional de la cromatina, ya que la conformación del locus facilita el contacto físico entre potenciadores y promotores específicos (39).

En vertebrados, se ha propuesto el concepto de paisajes regulatorios (“Regulatory landscapes” en inglés) para describir largos segmentos de ADN que contienen múltiples elementos *cis*-reguladores distales involucrados en la expresión precisa de sus genes blanco (39). Dentro de estos paisajes pueden encontrarse dominios

de asociación topológica (“TADs” por sus siglas en inglés de “topologically associating domains”), los cuales corresponden a regiones genómicas dentro de las cuales las interacciones en la cromatina ocurren con mayor frecuencia (40). Así mismo, los TADs favorecen un entorno ideal para la regulación transcripcional relacionada a elementos *cis*-regulatorios distales. De esta manera, los TADs pueden favorecer un entorno estructural que facilita la comunicación entre promotores y elementos *cis*-reguladores distales. La correcta regulación mediada por estos elementos es fundamental durante el desarrollo embrionario (39), ya que alteraciones en su actividad o en la organización del locus pueden contribuir al origen de patologías y malformaciones (41,42).

1.1.2. Secuencias repetitivas.

Una fracción importante de los genomas eucariotas está compuesta por secuencias repetitivas, las cuales corresponden a regiones de ADN presentes en múltiples copias y distribuidas de distintas formas dentro del genoma (43,44). De manera general, estas secuencias pueden clasificarse en repeticiones en tándem y repeticiones dispersas. Las repeticiones en tándem están formadas por unidades dispuestas consecutivamente e incluyen ADN satélite, minisatélites y microsatélites (43). Grandes acumulaciones de ADN satélite se localizan frecuentemente en regiones centroméricas, mientras que otras repeticiones en tándem pueden encontrarse en regiones teloméricas, génicas o intergénicas. En particular, la variación en el número de copias de minisatélites y microsatélites constituye una fuente importante de polimorfismo y diversidad genética (45,46). Por otro lado, las repeticiones dispersas se localizan en múltiples posiciones del genoma y están constituidas en gran medida por elementos transponibles y sus remanentes (13). Debido a su abundancia, diversidad y capacidad de modificar la arquitectura y regulación del genoma, los elementos transponibles representan uno de los componentes repetitivos más relevantes para comprender la evolución de los genomas eucariotas (47).

1.1.2.1 Elementos transponibles.

Dentro de las secuencias repetitivas presentes en los genomas eucariotas, los elementos transponibles (“TEs”, por sus siglas en inglés de “transposable elements”), también conocidos como transposones, constituyen uno de los componentes repetitivos más abundantes (48). Estos elementos fueron descubiertos por Barbara McClintock, galardonada con el Premio Nobel de Fisiología o Medicina en 1983, quien estudió cómo ciertos elementos genéticos móviles podían modificar la pigmentación de los granos de maíz mediante su movilización entre distintas regiones del genoma (49–51).

Actualmente, los TEs se definen como segmentos de ADN con la capacidad, actual o ancestral, de movilizarse dentro del genoma mediante un proceso denominado **transposición**. De acuerdo con su mecanismo de movilización, se clasifican en dos grandes grupos: transposones de ARN o clase I, también llamados

retrotransposones, y transposones de ADN o clase II (52,14). A su vez, estos elementos pueden ser autónomos, cuando codifican las proteínas necesarias para su movilización, como transcriptasas reversas o transposasas, o no autónomos, cuando dependen de enzimas codificadas por otros TEs para transponerse (53–55). Además de su capacidad de movilización, algunas copias de TEs pueden adquirir funciones para el hospedero durante la evolución, proceso conocido como **domesticación** o cooptación molecular, el cual será abordado en detalle en la **sección 1.2** (54,55). Los casos de domesticación de TEs han sido ampliamente estudiados en mamíferos, donde diversos retrotransposones se han vinculado con la generación de innovaciones regulatorias y evolutivas (56). Así, aunque históricamente los TEs fueron considerados “parásitos moleculares” o parte del “ADN basura”, actualmente se reconoce que pueden contribuir de manera importante a la arquitectura, regulación y evolución de los genomas eucariotas (2,57).

1.1.2.1. Retrotransposones.

Los retrotransposones, o elementos transponibles de clase I, emplean un proceso de transcripción inversa para movilizarse dentro del genoma. En este mecanismo, una copia de ARN del retrotransposón es retrotranscrita a ADN complementario (ADNc), el cual posteriormente se integra en una nueva región genómica. Por esta razón, los retrotransposones se movilizan mediante un mecanismo de tipo “copiar y pegar” (58). Dentro de los retrotransposones se encuentran los elementos con repeticiones terminales largas, conocidos como LTR (Por “long terminal repeats” en inglés), los cuales presentan secuencias repetidas en ambos extremos y una organización similar a la de los retrovirus (59). Muchos retrotransposones LTR codifican proteínas tipo Gag y Pol; esta última puede incluir actividades de proteasa, retrotranscriptasa, RNasa H e integrasa. Durante su ciclo de movilización, el ARN del retrotransposón es retrotranscrito, generando un ADNc que posteriormente es integrado en el genoma mediante la acción de una integrasa (59).

Por otra parte, existen retrotransposones que carecen de LTR, entre los cuales se encuentran los LINEs (Por “long interspersed nuclear elements” en inglés) y SINEs (Por “short interspersed nuclear elements” en inglés) (60–63). Los LINEs son elementos autónomos que pueden codificar las proteínas necesarias para su movilización. En particular, los elementos LINE-1 presentan generalmente dos marcos de lectura abierta: ORF1, que codifica una proteína de unión a ARN, y ORF2, que codifica una proteína con actividad de endonucleasa y retrotranscriptasa. Estas proteínas se asocian al ARN del propio LINE, formando un complejo ribonucleoproteico que retorna al núcleo. Allí, la endonucleasa genera un corte en el ADN diana y la retrotranscripción ocurre directamente sobre el sitio de inserción, mediante un mecanismo conocido como transcripción inversa cebada por el sitio blanco (“target-primed reverse transcription” en inglés) (62,63). Cabe destacar que es la acción errónea de elementos LINEs sobre ARN mensajeros

celulares que es responsable de la emergencia de retrogenes (64,65). Los SINEs, en cambio, son retrotransposones no autónomos de menor tamaño, derivados frecuentemente de ARN celulares no codificantes, como ARNt o ARN 7SL. Aunque contienen secuencias internas que permiten su transcripción, dependen de la maquinaria enzimática codificada por LINEs para movilizarse en el genoma. Esta dependencia funcional explica que la expansión de SINEs esté estrechamente asociada a la actividad de LINEs en un genoma determinado (60–63).

La transcripción es esencial para la movilización de los retrotransposones, ya que su ARN actúa como intermediario para generar nuevas copias. Sin embargo, muchos corresponden actualmente a fósiles moleculares incapaces de transponerse, aunque pueden conservar actividad transcripcional o secuencias con potencial regulatorio. Así, los retrotransposones representan tanto una fuente de mutaciones como una reserva evolutiva de elementos reguladores principalmente en los genomas de mamíferos (66).

1.1.2.2. Transposones de ADN.

Los transposones de ADN, o elementos transponibles de clase II, se movilizan mediante intermediarios de ADN y, a diferencia de los retrotransposones, no requieren una etapa de retrotranscripción. El mecanismo clásico de movilización corresponde a la transposición de tipo “cortar y pegar”, mediada por una enzima transposasa. En este proceso, la transposasa reconoce secuencias terminales del elemento, escinde el transposón desde su sitio original y cataliza su integración en una nueva región del genoma (67–69). Muchos transposones de ADN de tipo “cortar y pegar” presentan repeticiones terminales invertidas, conocidas como repeticiones terminales invertidas (“TIRs” por sus siglas en inglés de “terminal inverted repeats”), las cuales son reconocidas por la transposasa (67). Estas enzimas suelen pertenecer a la familia de nucleasas DDE, caracterizadas por residuos conservados de ácido aspártico y glutámico que participan en la catálisis de la escisión e integración del elemento. Aunque este mecanismo no genera una copia adicional de manera directa, los transposones de ADN pueden aumentar su número de copias si la transposición ocurre durante la fase S del ciclo celular, por ejemplo, desde una región ya replicada hacia una región aún no replicada (13,48,67–69). Los transposones de ADN presentan una gran diversidad de superfamilias, entre las que se incluyen Tc1/Mariner, hAT, Mutator/MULE, PIF/Harbinger, PiggyBac y Kolobok, entre otras. Estas superfamilias se distinguen por la secuencia y estructura de sus transposasas, sus TIRs, sus preferencias de inserción y las duplicaciones del sitio blanco generadas durante la integración. A nivel evolutivo, las copias autónomas pueden perder regiones codificantes por acumulación de mutaciones o deleciones internas, dando origen a elementos no autónomos que conservan las secuencias necesarias para ser reconocidos por transposasas codificadas por otros elementos relacionados (13,48,70).

Además de los transposones clásicos con TIRs, existen otros tipos de transposones de ADN, como los helitrones. A diferencia de los elementos de “corte y pegado”, los

helitrones no presentan TIRs típicas y se movilizan mediante un mecanismo replicativo relacionado con la replicación por círculo rodante. Estos elementos codifican proteínas con dominios Rep/Hel, asociadas a actividad endonucleasa y helicasa, y su movilización involucra intermediarios de ADN de hebra simple o circular. Además, suelen presentar secuencias terminales conservadas y estructuras palindrómicas cercanas al extremo 3' (69).

La actividad de los transposones de ADN puede tener múltiples consecuencias sobre el genoma hospedero. Sus inserciones pueden ser deletéreas si interrumpen genes o regiones regulatorias esenciales, neutrales si no afectan la función celular, o eventualmente beneficiosas si son cooptadas por el hospedero (47,48,70). En *Xenopus tropicalis*, estudios recientes han sugerido una asociación entre helitrones y la organización tridimensional de la cromatina durante la espermatogénesis, lo que refuerza la idea de que los transposones de ADN pueden influir tanto en la estructura como en la evolución funcional del genoma (71).

1.1.2.3. Elementos transponibles de repetición invertida en miniatura.

Los elementos transponibles de repetición invertida en miniatura (“MITEs”, por sus siglas en inglés de “miniature inverted-repeat transposable elements”) corresponden a elementos transponibles de ADN no autónomos y de pequeño tamaño (70). Los primeros estudios sobre estos elementos se concentraron principalmente en genomas de plantas (72,73), aunque posteriormente también fueron descritos en diversos animales, como *Caenorhabditis elegans* (74), peces teleósteos, incluido pez cebra (75) y humanos (76). Los MITEs generalmente presentan longitudes entre 100 y 500 pb, carecen de marcos de lectura abiertos y poseen TIRs, además de duplicaciones del sitio blanco (“TSDs” por sus siglas en inglés de “Terminal site duplication”) generadas durante su inserción (70). Dado a que estos MITEs no codifican una transposasa, dependen de la maquinaria producida por elementos autónomos relacionados para movilizarse por el genoma. La conservación de TIRs y TSDs entre determinados MITEs y transposones autónomos respalda la existencia de una relación evolutiva entre ambos tipos de elementos, por ejemplo, los transposones autónomos PIF/Pong han sido relacionados con el origen y la amplificación de MITEs tipo Tourist (77). Aunque los mecanismos responsables de su elevada capacidad de proliferación no han sido completamente dilucidados, algunas familias de MITEs han experimentado explosiones de amplificación que han generado cientos o miles de copias dentro de un genoma (70,78). Una vez amplificadas, estas copias pueden permanecer durante largos periodos evolutivos, incluso cuando los elementos autónomos responsables de su movilización han perdido actividad. Además, su distribución no es necesariamente aleatoria, ya que determinadas familias presentan una inserción preferencial en regiones génicas o cercanas a genes (73,78). La identificación y clasificación de los MITEs se basa principalmente en su tamaño, la presencia y similitud de sus TIRs y TSDs, y la semejanza de secuencia entre las copias que integran una familia. Herramientas bioinformáticas como detectMITE y MiteFinderII utilizan estas características

estructurales para identificar posibles MITEs en genomas eucariotas (79,80). En particular, algunos de estos programas emplean criterios operacionales, como TIRs de al menos 10 pb y TSDs de entre 2 y 10 pb, aunque estos valores corresponden a parámetros de detección y no necesariamente a características universales de todos los MITEs.

Una característica relevante de estos elementos es su frecuente asociación con regiones génicas. En plantas, se han identificado MITEs dentro o próximos a genes, donde pueden modificar regiones codificantes, aportar exones alternativos o señales de poliadenilación e influir en la expresión génica, además, algunos MITEs pueden ser transcritos y dar origen a ARN pequeños de interferencia, principalmente siRNAs, capaces de participar en mecanismos de regulación transcripcional y epigenética (73,81,82). Por lo tanto, su pequeño tamaño, elevado número de copias y asociación con genes convierten a los MITEs en una fuente potencial de innovación regulatoria durante la evolución. A pesar de que los MITEs han sido identificados en diversos animales, su posible domesticación como fuente de promotores y otros elementos *cis*-reguladores ha sido estudiada con mucha menor profundidad que en plantas. **En particular, aún se conoce poco sobre su contribución a la evolución de secuencias regulatorias en genomas de vertebrados.**

1.1.3. Técnicas de secuenciación para el estudio del control transcripcional y elementos transponibles.

El estudio del control regulatorio requiere integrar aproximaciones genómicas, transcriptómicas y funcionales (83). Una de las herramientas más utilizadas para identificar regiones regulatorias activas corresponde a la secuenciación de regiones de cromatina accesibles a transposasa, conocida como ATAC-seq (“ATAC-Seq” por sus siglas en inglés de “Assay for transposase-accessible chromatin using sequencing”) (84). Esta técnica permite detectar regiones de cromatina abierta o libres de nucleosomas, las cuales suelen corresponder a promotores, potenciadores y otros elementos *cis*-reguladores activos bajo una condición fisiológica o experimental determinada. En el contexto del estudio de elementos transponibles, la integración de datos de ATAC-seq con anotaciones de secuencias repetitivas, como RepeatMasker, permite identificar copias de TEs localizadas en regiones accesibles del genoma y, por lo tanto, priorizar aquellas con posible actividad regulatoria. En *Xenopus tropicalis*, aproximaciones epigenómicas han permitido estudiar la relación entre transposones y estados de cromatina durante el desarrollo embrionario (85). Además, la combinación de ATAC-seq con secuenciación de ARN, o RNA-seq, permite relacionar la accesibilidad a la cromatina de regiones regulatorias con los patrones globales de expresión génica (86). Sin embargo, estas aproximaciones son principalmente correlativas, por lo que requieren validación experimental para determinar si una secuencia posee actividad regulatoria funcional (87).

1.2. Domesticación de elementos transponibles.

La transposición de un elemento transponible puede tener distintas consecuencias evolutivas para el genoma hospedero. En algunos casos, la inserción de un TE puede ser deletérea, por ejemplo, al interrumpir genes, alterar regiones regulatorias o promover inestabilidad genómica. En otros casos, la inserción puede ser neutra o casi neutra, persistiendo en la población principalmente por deriva genética. Sin embargo, en ciertos contextos, una secuencia derivada de TE puede adquirir una función beneficiosa para el hospedero y ser mantenida por selección natural. Este proceso se conoce como domesticación o cooptación molecular de elementos transponibles (54,55). Durante la domesticación, una secuencia originalmente asociada a la movilidad del TE puede ser reutilizada por el genoma hospedero para cumplir una nueva función molecular. Esta función puede involucrar regiones codificantes, como ocurre con genes derivados de transposasas, o regiones no codificantes con actividad regulatoria, como promotores, potenciadores, sitios de unión para factores de transcripción, señales de poliadenilación o ARN no codificantes (55–57). De esta manera, los TEs representan una fuente importante de innovación molecular, ya que pueden introducir módulos regulatorios completos o secuencias parcialmente funcionales en nuevas posiciones genómicas.

En esta tesis se centra específicamente en la contribución de TEs al origen de elementos *cis*-reguladores, particularmente promotores asociados a sitios de inicio de la transcripción (“TSSs”, por sus siglas en inglés de “transcription start sites”) y elementos *cis*-regulatorios distales como potenciadores. A continuación, se describen ejemplos de domesticación de TEs en promotores y potenciadores.

1.2.1. Domesticación de TEs en promotores.

Una de las formas en que los TEs pueden contribuir a la evolución regulatoria es mediante la generación de promotores o sitios alternativos de inicio de la transcripción. Esto ocurre porque algunas copias de TEs contienen secuencias capaces de reclutar la maquinaria transcripcional o FTs del hospedero, permitiendo iniciar la expresión de genes cercanos o generar nuevos transcritos. Un ejemplo claro se ha descrito en la bacteria *Neisseria meningitidis*, donde elementos tipo Correia, similares a MITEs, contienen promotores fuertes capaces de impulsar la transcripción de genes adyacentes y de ARN pequeños no codificantes (88).

En animales, los TEs pueden aportar motivos de unión para factores de transcripción, señales de inicio transcripcional, sitios de poliadenilación u otros elementos regulatorios. Por ello, una vez insertados en una nueva región genómica, pueden contribuir al origen de promotores, TSSs o nuevos transcritos (89). En este contexto, se han reportado múltiples casos de domesticación de TEs asociados a innovación regulatoria en genomas de mamíferos, particularmente derivados de retrotransposones (56). Un ejemplo relevante corresponde a los retrovirus endógenos HERV-H, cuyos LTRs pueden actuar como elementos promotores en células madre embrionarias humanas y células pluripotentes. En particular, se ha

descrito que HERV-H participa en la transcripción de genes y ARN largos no codificantes asociados al estado pluripotente (90,91). Otro caso funcionalmente validado corresponde a un retrotransposón MT2B2 específico de ratón que actúa como promotor alternativo del gen *Cdk2ap1*, generando una isoforma truncada la cual promueve la proliferación celular durante el desarrollo embrionario temprano, y la eliminación del promotor derivado del retrotransposón reduce la proliferación celular, afecta la implantación y causa defectos embrionarios severos en ratón (92). También, se ha reportado que los promotores NWC presentes en vertebrados con mandíbulas presentan una alta integración del transposón de ADN RAG regulando así la diversidad de inmunoglobulinas y receptores en células T (93).

Estos antecedentes muestran que los promotores derivados de TEs pueden incorporarse funcionalmente a redes transcripcionales del hospedero y contribuir a procesos biológicos esenciales. Sin embargo, la mayoría de los casos funcionalmente caracterizados corresponde a retrotransposones en mamíferos, mientras que la contribución de transposones de ADN, particularmente MITEs, al origen de promotores en animales ha sido mucho menos explorada.

1.2.2. Domesticación de TEs en enhancers

Los TEs, además de originar promotores, también pueden contribuir a la evolución de potenciadores o *enhancers*. Análisis comparativos de potenciadores activos en mamíferos han mostrado que estas regiones presentan una alta dinámica evolutiva, con conservación de algunos elementos antiguos y aparición de potenciadores más recientes asociados a secuencias repetitivas (94). Además, diversos estudios han demostrado que TEs domesticados como potenciadores han contribuido a la evolución de redes regulatorias asociadas a procesos reproductivos y del desarrollo. Por ejemplo, se ha propuesto que TEs antiguos transformaron el paisaje regulatorio uterino durante la evolución del embarazo en mamíferos, aportando elementos *cis*-reguladores a genes relacionados con implantación, comunicación materno-fetal e inmunotolerancia (95). En ratón, también se ha demostrado que ERVs pueden funcionar como potenciadores específicos de especie en placenta, reclutando factores de transcripción esenciales para el desarrollo del trofoblasto (96).

En humanos, se han identificado TEs específicos de primates que actúan como elementos regulatorios durante el desarrollo embrionario, participando en redes transcripcionales específicas de linaje (97). De manera similar, en gibones (primate) se ha reportado la cooptación del retrotransposón compuesto LAVA, específico de este linaje, como elemento regulatorio asociado a potenciadores y a genes relacionados con la reparación del ADN y la mantención de la integridad genómica (98). Estos casos sugieren que la incorporación de TEs en regiones regulatorias puede generar nuevas conexiones entre factores de transcripción y genes blanco, facilitando la evolución de redes génicas específicas de linaje. La domesticación de TEs como potenciadores también ha sido documentada en el sistema inmune. En linfocitos T, subfamilias específicas de TEs contribuyen a distintos dominios de potenciadores activos, sugiriendo que estos elementos pueden modular respuestas

inmunológicas en contextos celulares específicos (99) Asimismo, elementos B2 SINE de ratón pueden funcionar como potenciadores inducibles por interferón, participando en la regulación de genes asociados a la inmunidad innata (100).

En plantas, los MITEs han sido implicados en la amplificación y redistribución de motivos de unión para factores de transcripción, modificando redes regulatorias específicas de linaje relacionadas con respuestas al estrés, floración y maduración de frutos (101). Sin embargo, en animales son escasos los estudios que analizan de manera sistemática la domesticación de MITEs como fuente de promotores o potenciadores. Esta brecha es especialmente relevante para comprender la evolución de elementos *cis*-reguladores derivados de transposones de ADN en vertebrados no mamíferos.

1.3. *Xenopus* como modelo para estudiar domesticación de TEs de ADN.

Los anfibios del género *Xenopus* constituyen un sistema relevante para estudiar la evolución de los genomas de vertebrados, el desarrollo embrionario y la regulación génica. En particular, *Xenopus tropicalis* (*Xt*) y *Xenopus laevis* (*Xl*) han sido ampliamente utilizados como modelos experimentales debido a su fecundación externa, gran tamaño embrionario, facilidad de manipulación y disponibilidad de recursos genómicos y transcriptómicos (102–104). Además, la historia evolutiva particular de estas especies permite comparar genomas relacionados, pero con diferencias importantes en ploidía, arquitectura cromosómica y composición de elementos transponibles.

1.3.1. *Xenopus tropicalis* y *Xenopus laevis*: dos especies, tres genomas.

Xt y *Xl* son dos especies cercanas filogenéticamente que divergieron hace aproximadamente 48 millones de años (**Figura 1**) (102). A pesar de pertenecer al mismo género, ambas especies presentan diferencias relevantes en tamaño genómico, número cromosómico y ploidía. Mientras *Xt* es una especie diploide, *Xl* posee un genoma alotetraploide originado por la hibridación de dos especies diploides ancestrales, seguida de duplicación completa del genoma, hace aproximadamente 17–18 millones de años (102,105). Los eventos de poliploidización han desempeñado un papel importante en la evolución de diversos linajes eucariotas, incluyendo vertebrados, plantas y otros organismos (106,107). En vertebrados, por ejemplo, se han propuesto dos rondas antiguas de duplicación completa del genoma asociadas al origen y diversificación temprana del grupo (108). Estos eventos pueden proporcionar material genético adicional sobre el cual actúan la selección natural y la deriva genética, permitiendo la conservación, pérdida, subfuncionalización o neofuncionalización de genes duplicados (109–112). De esta manera, las duplicaciones genómicas pueden contribuir a la aparición de nuevas funciones, redes génicas y trayectorias evolutivas.

En el caso de *Xl*, la alotetraploidización dio origen a dos subgenomas denominados L (*XlL*) y S (*XlS*), los cuales derivan de los cromosomas de dos especies

progenitoras diploides (**Figura 1**) (102). Estos subgenomas han evolucionado de manera asimétrica: en comparación con el *XIL*, el *XIS* presenta mayor divergencia, más reordenamientos cromosómicos, mayor pérdida de genes y cambios en patrones de expresión y regulación epigenética (102). Por otro lado, *Xt* no experimentó la alotetraploidización reciente observada en *XI*, por lo que conserva un genoma diploide de menor tamaño (113,114). Estas características han permitido utilizarlo como especie de referencia para estudios genómicos, funcionales y evolutivos dentro del género *Xenopus*. En conjunto, la comparación entre el genoma diploide de *Xt* y los *XIL* y *XIS* permite abordar preguntas sobre conservación, divergencia y evolución regulatoria en tres contextos genómicos estrechamente relacionados.

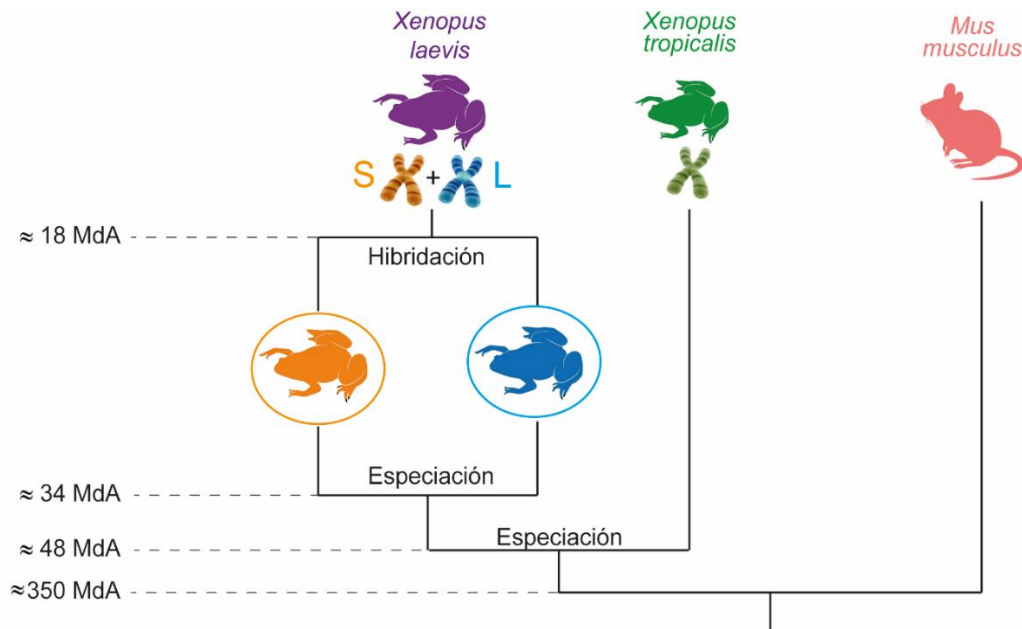


Figura 1: Filogenia de los anfibios del género *Xenopus*. Se indican la divergencia entre *X. tropicalis* y *X. laevis*, ocurrida hace aproximadamente 48 millones de años (Ma); la divergencia entre las especies progenitoras de *X. laevis*, hace aproximadamente 34 Ma; y el evento de poliploidización y duplicación génica por hibridación en el linaje de *X. laevis*, hace aproximadamente 18 Ma. También se muestra la divergencia entre los anfibios del género *Xenopus* y los mamíferos, representados por *Mus musculus* (ratón), ocurrida hace aproximadamente 350 Ma.

1.3.2. Elementos transponibles identificados en los genomas de *Xenopus*.

La composición de elementos transponibles varía ampliamente entre linajes de vertebrados. En mamíferos, una fracción importante de los TEs corresponde a retrotransposones, incluyendo LINEs, SINEs y retrovirus endógenos (66,115). En contraste, en *Xt* y *XI* se ha reportado una mayor representación de transposones de ADN, situación que también se observa en otros vertebrados no mamíferos, como el pez cebra (116). Esta diferencia convierte a *Xenopus* en un modelo especialmente interesante para estudiar la contribución evolutiva de transposones de ADN, incluyendo MITEs, a la regulación génica.

En *Xt*, se ha reportado que una proporción importante de los TEs son de ADN, incluyendo superfamilias como hAT, Harbinger, Mariner, PiggyBac y Kolobok (113,117,118). Además, se han identificado familias de helitrones, como Helitron-N1_XT y Helitron-N2_XT, aunque estas se encuentran escasamente representadas en el genoma (113). En la **Tabla 1** TEs se resumen estas diferencias en la composición de TEs entre mamíferos, *Xenopus* y otros vertebrados, destacando el enriquecimiento relativo de transposones de ADN en *Xt*.

Tabla 1: Tamaño de los genomas y porcentaje de elementos transponibles (TEs) presentes en diferentes genomas eucariotas. %: porcentaje aproximado del genoma ocupado por TEs. Los datos se obtuvieron de las siguientes referencias: *H. sapiens* (6,119,120), *M. musculus* (121), *Xt* (113), *D. rerio* (116). Datos de *Xl* fueron obtenidos durante el desarrollo de esta tesis.

	<i>H. sapiens</i>	<i>M. musculus</i>	<i>X. tropicalis</i>	<i>X. laevis</i>	<i>D. rerio</i>
Tamaño del genoma	3 Gpb	2,5 Gpb	1,7 Gpb	3,1 Gpb	1,5 Gpb
Transposones de ADN	3,6 %	0,9 %	25 %	29 %	46,2 %
Retrotransposones	51 %	37,3 %	9 %	11 %	13,2 %

En *Xl*, la historia alotetraploide de su genoma ha permitido observar diferencias en la distribución y evolución de TEs entre los *XIL* y *XIS*. Se ha reportado la presencia de TEs de la familia PIF/Harbinger en ambos subgenomas, mientras que elementos de la familia Tc1/Mariner se encontraron de manera específica en el subgenoma S (102). Esta observación sugiere que, después de la alotetraploidización, distintas familias de TEs pudieron seguir trayectorias evolutivas diferenciales en cada subgenoma. Por lo tanto, *Xl* ofrece una oportunidad para estudiar eventos de amplificación de TEs comunes o específicos de L o S.

La abundancia y diversidad de TEs de ADN en *Xenopus* es particularmente relevante para esta tesis. A diferencia de los mamíferos, donde gran parte de los ejemplos conocidos de domesticación regulatoria involucran retrotransposones, los genomas de *Xenopus* permiten explorar el potencial regulatorio de transposones de ADN en vertebrados.

1.3.3 Estudios de elementos regulatorios en *Xenopus* mediante experimentos de transgénesis.

Para evaluar funcionalmente la actividad de elementos *cis*-reguladores se utilizan vectores reporteros, en los cuales la secuencia candidata se clona río arriba de un gen reportero, como luciferasa, GFP u otros genes fluorescentes (122). Estos ensayos permiten determinar si una región candidata puede activar o potenciar la expresión génica en un contexto celular o embrionario determinado. Una de las estrategias utilizadas en *Xenopus* corresponde a la transgénesis mediada por la meganucleasa I-SceI, en la cual el constructo reportero contiene sitios de reconocimiento para esta enzima, favoreciendo su integración en el genoma embrionario (123). Este método ha sido implementado en *Xt* y *Xl*, y permite evaluar la actividad de secuencias regulatorias durante el desarrollo embrionario.

De manera alternativa, la transgénesis por integración mediada por enzimas de restricción, conocida como REMI (siglas en inglés de “Restriction enzyme mediated integration”), permite generar embriones transgénicos mediante la incubación de un vector linealizado con núcleos espermáticos, y una enzima de restricción, seguida de microinyección nuclear a ovocitos no fertilizados (124,125). Esta aproximación permite generar embriones F0 transgénicos no mosaicos con alta eficiencia, lo que resulta especialmente útil para evaluar de manera rápida la actividad de promotores y potenciadores en embriones vivos.

1.4. Pregunta de investigación, hipótesis y objetivos.

A modo de resumen, hemos visto que: (1) Principalmente se ha estudiado la domesticación de transposones en elementos *cis*-regulatorios en mamíferos. (2) Anfibios del género *Xenopus* presentan principalmente transposones de ADN. (3) Existe un vacío en el conocimiento en como estos TEs de ADN pueden formar elementos *cis*-regulatorios en animales.

Por lo cual nos planteamos la siguiente Pregunta de Investigación: ¿Cuál es la dinámica de domesticación de MITEs en regiones de control transcripcional en los linajes de *Xt* y *Xl*?

Hipótesis: Los genomas de *Xt* y *Xl* L y S, han sufrido, de forma ancestral e independiente a cada linaje, eventos de domesticación de TEs a promotores transcripcionales ubicuos y elementos *cis*-regulatorios distales, contribuyendo así a la expresión de genes ubicuos y tejido-específicos.

Objetivo General: Determinar la dinámica de domesticación de MITEs hacia promotores ubicuos y elementos *cis*-regulatorios distales ubicuos y tejido-específicos en los genomas de *Xt* y *Xl*.

Objetivos específicos:

1. Identificar la presencia de transposones en el paisaje regulatorio de *Xt* y *Xl*.
2. Caracterizar la expresión de genes regulados por promotores ubicuos derivados del MITE Kolobok-N9_XT (KN9).
3. Caracterizar la función de promotores de genes ubicuos que provienen del MITE KN9.

2. MATERIALES Y MÉTODOS

2.1. Presencia de TEs en el paisaje regulatorio de *Xt* y *Xl*.

2.1.1. Datos de ATAC-Seq y RNA-Seq.

Los conjuntos de regiones reguladoras de *X. tropicalis* correspondientes a hígado, mesénquima, osteoblastos, pulmón y corazón se obtuvieron de Castillo et al. (2024). Para *X. laevis*, los datos crudos de ATAC-seq de hígado, mesénquima y osteoblastos fueron evaluados mediante FastQC (bioinformatics.babraham.ac.uk/projects/fastqc/) y procesados con Cutadapt (cutadapt.readthedocs.io) para eliminar adaptadores y secuencias artefactuales. Las lecturas fueron alineadas al genoma de referencia *X. laevis* v10.1 mediante Bowtie2 (bowtie-bio.sourceforge.net/bowtie2/) utilizando el modo very-sensitive. Los alineamientos fueron posteriormente convertidos, ordenados, indexados y filtrados mediante SAMtools (htslib.org), conservándose únicamente lecturas correctamente pareadas con una calidad de mapeo ≥ 40 y eliminándose las lecturas mitocondriales. Los duplicados fueron eliminados mediante Picard MarkDuplicates (broadinstitute.github.io/picard/). Para el análisis de regiones libres de nucleosomas (NFR) se seleccionaron fragmentos de aproximadamente 30–120 pb y se realizó el peak calling mediante Genrich (github.com/jsh58/Genrich). Los picos obtenidos entre réplicas y tejidos se consolidaron mediante BEDTools (bedtools.readthedocs.io); las lecturas correspondientes a cada región se cuantificaron mediante multicov y los conteos se normalizaron utilizando el método TMM implementado en edgeR (bioconductor.org/packages/edgeR/). Este procedimiento siguió la estrategia general empleada por Castillo et al. (2024) para el procesamiento de datos ATAC-seq de *X. tropicalis*. Las regiones accesibles se clasificaron como promotoras cuando se solapaban con un TSS anotado y como distales cuando no presentaban dicho solapamiento, y posteriormente se agruparon por tejido. Para ambos genomas, las regiones compartidas por hígado, mesénquima y osteoblastos se definieron como ubicuas; en *X. tropicalis*, corazón y pulmón se incorporaron además como conjuntos regulatorios adicionales. La anotación de las regiones y el análisis de sitios de unión a factores de transcripción (TFBS) se realizaron mediante HOMER (homer.ucsd.edu/homer/), utilizando annotatePeaks.pl y findMotifsGenome.pl, respectivamente.

Los datos de RNA-seq de *X. tropicalis* fueron procesados siguiendo el flujo de trabajo descrito por Castillo et al. (2024), y el mismo procedimiento se aplicó a *X. laevis* utilizando el ensamblaje de referencia correspondiente (*X. tropicalis* v9.1 y *X. laevis* v10.1). Brevemente, la calidad de las lecturas crudas se evaluó mediante FastQC (bioinformatics.babraham.ac.uk/projects/fastqc/), los adaptadores y secuencias de baja calidad se eliminaron con Cutadapt (cutadapt.readthedocs.io), y las lecturas resultantes se alinearon contra el transcriptoma de referencia mediante HISAT2 (daehwankimlab.github.io/hisat2/). Posteriormente, utilizando SAMtools (htslib.org), se eliminaron las lecturas con calidad de mapeo inferior a 40

y aquellas correspondientes al genoma mitocondrial. Las lecturas asignadas a cada transcrito se cuantificaron mediante multicov de BEDTools v2.29.2 (bedtools.readthedocs.io), y los conteos resultantes se normalizaron mediante el método TMM implementado en edgeR (bioconductor.org/packages/edgeR/). Finalmente, la concordancia entre réplicas biológicas se evaluó mediante correlación de Pearson.

2.1.2. Cuantificación de la cobertura de elementos transponibles en regiones *cis*-regulatorias.

Para cada especie se utilizaron las coordenadas genómicas de las regiones reguladoras y de los elementos repetitivos anotados mediante RepeatMasker (<https://www.repeatmasker.org/>). Las regiones se agruparon en promotoras (asociadas con el TSS) y distales para las categorías ubicua, hígado, mesénquima y osteoblastos. En *Xt* también se analizaron corazón y pulmón. Antes del análisis, los archivos se restringieron a los cromosomas o contigs presentes en los conjuntos regulatorios y se ordenaron según sus coordenadas genómicas. Se determinó el punto central de cada región reguladora y se construyó una ventana en pares de bases alrededor de este. Como control se generaron ventanas equivalentes alrededor de posiciones genómicas aleatorias. Las subfamilias de elementos transponibles se identificaron a partir de la anotación de RepeatMasker y se analizaron independientemente. La cobertura por posición se calculó mediante bedtools coverage -d, incluido en BEDTools (bedtools.readthedocs.io/en/stable/). Para cada posición relativa (p), la profundidad relativa se definió como: $\text{RelativeDepth}(p) = n_p/N$, donde (N) corresponde al número total de ventanas analizadas y (n_p) al número de ventanas cubiertas por la subfamilia en esa posición. El mismo procedimiento se aplicó a las ventanas aleatorias para obtener una distribución de referencia. Finalmente, se compararon los perfiles centrados en la posición 0 para identificar aumentos locales de cobertura de cada subfamilia respecto del fondo genómico.

2.1.3. Asignación de genes a promotores asociados con KN9.

Las coordenadas de las regiones promotoras solapadas por KN9 se recopilaron en formato BED y se anotaron mediante annotatePeaks.pl, incluido en HOMER (<http://homer.ucsd.edu/homer/>). Para cada especie se utilizó el genoma de referencia correspondiente y su archivo de anotación GTF, incorporado mediante la opción -gtf. A partir de este análisis se recuperaron el gen y el TSS más próximos a cada región.

2.1.4. Alineamiento múltiple e inferencia filogenética de KN9.

Las secuencias de KN9 identificadas en *Xt* y en los *XIL* y *XIS* de *XI* se compilaron en formato FASTA y se alinearon con MAFFT v7 (<https://mafft.cbrc.jp/alignment/server/index.html>) mediante la opción --auto. Cuando se detectaron secuencias en orientación inversa, se incorporó la opción --adjustdirection. El alineamiento se inspeccionó y las secuencias altamente divergentes o con alineamiento deficiente que comprometían su calidad se

excluyeron antes de la reconstrucción filogenética. La inferencia por máxima verosimilitud se realizó con IQ-TREE 2 (<https://iqtree.github.io/>). Los modelos de sustitución se evaluaron con ModelFinder mediante -m TESTONLY y se ordenaron según el criterio de información bayesiano (BIC). Se generaron árboles independientes para los seis modelos con menor BIC, con el propósito de evaluar la estabilidad de las relaciones filogenéticas frente a la selección del modelo. El soporte de las ramas se estimó mediante 1.000 réplicas de ultrafast bootstrap y 1.000 réplicas de SH-like approximate likelihood ratio test (SH-aLRT). También se utilizó -bnni para optimizar los árboles de bootstrap mediante intercambios nearest-neighbor interchange, junto con asignación automática de hilos (-nt AUTO) y una semilla aleatoria fija (-seed 42). Los árboles se visualizaron y anotaron en R (<https://www.r-studio.com/es/>) mediante el paquete ggtree. Las secuencias se identificaron según su origen (*Xt*, *Xl L* o *Xl S*) y el locus asociado. Posteriormente, se examinó el agrupamiento de copias correspondientes a loci ortólogos. Las longitudes de las ramas representan el número estimado de sustituciones por sitio.

2.1.5. Análisis exploratorio de la localización subcelular de proteínas asociadas con KN9.

Los ortólogos humanos de los genes asociados con promotores que contienen KN9 se consultaron en la plataforma Organelle Landscape (https://organelles.sf.czbiohub.org/Subcellular_UMAP). Las proteínas se visualizaron sobre el UMAP a partir de datos de proteómica espacial de células humanas HEK293T (Hein et al., 2025). Este análisis se utilizó como una descripción exploratoria de su distribución subcelular y no como una prueba formal de enriquecimiento.

2.2. Patrón de expresión espaciotemporal de genes asociados con promotores que contienen KN9.

2.2.1. Diseño de sondas para hibridación *in situ* fluorescente HCR.

Los conjuntos de sondas dirigidos a *rtn3*, *tmbim6* y *abitram* fueron diseñados y sintetizados por Molecular Instruments (<https://www.molecularinstruments.com/>). Las sondas dirigidas a *tmbim6.L*, *faah2*, *faah2.L*, *tspoap1* y *nsmce1* se diseñaron para HCR RNA-FISH v3.0 mediante insitu_probe_generator v0.3.2 (https://github.com/rwnull/insitu_probe_generator/blob/v.0.3.2/README.md). El programa se ejecutó mediante UserInterface_v0.3.2.ipynb en Jupyter Notebook (<https://jupyter.org/>), dentro de un entorno Conda (<https://anaconda.org/anaconda/conda>) con Python 3.8 y las dependencias Biopython, pandas y NumPy. Como entrada se utilizó la secuencia sentido del ADNc correspondiente al transcrito de interés. Para cada transcrito se seleccionó un sistema iniciador/amplificador HCR compatible (B1, B2, B3, etc) y se definieron la extensión excluida desde el extremo 5' y las longitudes máximas permitidas para homopolímeros de A/T y C/G. El programa seleccionó regiones adecuadas del transcrito y generó pares de oligonucleótidos dirigidos a regiones adyacentes del ARN diana. Cada oligonucleótido incluyó una región iniciadora, un espaciador y la

secuencia específica de reconocimiento. La especificidad se evaluó localmente con blastn, utilizando la tarea blastn-short de NCBI BLAST+ (<https://blast.ncbi.nlm.nih.gov/doc/blast-help/downloadblastdata.html>) contra el transcriptoma de la especie correspondiente. Se descartaron los pares con posibles alineamientos fuera del ARNm diana. Las sondas seleccionadas y sus coordenadas de hibridación se exportaron en un formato compatible con IDT oPools (<https://eu.idtdna.com/pages/products/custom-dna-rna/dna-oligos/custom-dna-oligos/opools-oligo-pools>).

2.2.2. Hibridación *in situ* fluorescente HCR.

La expresión espacial de los genes de interés se evaluó mediante HCR RNA-FISH v3.0 en embriones completos de *Xenopus*, siguiendo una adaptación de los protocolos de Molecular Instruments. Se utilizó PBST, compuesto por PBS 1× y Tween-20 al 0,1 %, y 5× SSCT, compuesto por SSC 5× y Tween-20 al 0,1 %. Los embriones se transfirieron a viales de vidrio de 1,8 mL y se fijaron durante 1 h en MEMFA, preparada con agua tratada con DEPC, sales MEM 10× y formaldehído. Posteriormente, se realizaron tres lavados de 5 min con PBST y tres de 5 min con metanol (MeOH). Las muestras se mantuvieron en MeOH durante 2 h a temperatura ambiente o durante la noche a -20 °C. Los embriones se rehidrataron mediante lavados sucesivos de 5 min con 75, 50 y 25 % de MeOH en PBST, seguidos de tres lavados con PBST. La pigmentación se eliminó mediante incubación en 1 mL de solución de blanqueamiento (preparada con agua tratada con DEPC, SSC, H₂O₂ y formamida) bajo una fuente de luz intensa. El progreso se examinó cada 10–15 min o cada 3 min en embriones tempranos. Después del blanqueamiento, las muestras se lavaron dos veces durante 5 min con PBST, se refijaron durante 20 min en MEMFA y se lavaron tres veces durante 5 min con PBST. Para la hibridación, los embriones se equilibraron durante 5 min en hielo con una mezcla 1:1 de PBST y 5× SSCT, seguida de incubaciones de 5 min en 5× SSCT y 5 min en 500 µL de tampón de hibridación. Luego se prehibridaron durante 30 min a 37 °C en 500 µL del mismo tampón. Este se reemplazó por tampón de hibridación precalentado a 37 °C que contenía las sondas específicas, y las muestras se incubaron durante 12–16 h a 37 °C. El exceso de sondas se eliminó mediante cuatro lavados de 15 min con 1 mL de tampón de lavado de sondas a 37 °C, seguidos de dos lavados de 5 min con 5× SSCT a temperatura ambiente. Para amplificar la señal, se acondicionaron por separado 6 pmol de cada horquilla h1 y h2 (2 µL de una solución 3 µM) mediante calentamiento a 95 °C durante 90 s y enfriamiento a temperatura ambiente durante 30 min en oscuridad. En paralelo, los embriones se preincubaron durante 30 min a temperatura ambiente en 1 mL de tampón de amplificación. Este se reemplazó por tampón de amplificación que contenía ambas horquillas y las muestras se incubaron durante 12–16 h a temperatura ambiente y en oscuridad. El exceso de horquillas se eliminó con 5× SSCT mediante dos lavados de 5 min, dos de 30 min y uno final de 5 min. Cuando fue necesario teñir los núcleos, se añadió DAPI en una dilución 1:1000 durante el segundo lavado de 30 min. Finalmente, las muestras se lavaron

tres veces con PBS y se procesaron inmediatamente para la adquisición de imágenes o se conservaron a 4 °C hasta su análisis.

2.2.3. Microscopia y procesamiento de imágenes obtenidas de *in situ* HCR.

Todas las imágenes fueron adquiridas a partir de muestras fijadas y montadas utilizando un sistema de microscopía Carl Zeiss Axio Zoom.V16 equipado con ApoTome.2 (FONDEQUIP EQM140146). Para la detección de fluorescencia se utilizaron los canales DAPI, Cy3/red y Cy5/far-red, correspondientes al marcaje nuclear con DAPI y a la detección de Alexa Fluor 546 y Alexa Fluor 647, respectivamente. Las imágenes fueron adquiridas y procesadas mediante el software ZEISS ZEN 3.12 (Carl Zeiss Microscopy). Para todas las imágenes se aplicaron los mismos parámetros de ajuste de brillo y contraste, establecidos a partir de los controles negativos.

2.3. Validación funcional *in vivo* de la actividad promotora de KN9

2.3.1. Diseño del vector de transgénesis.

Para evaluar *in vivo* la actividad promotora de KN9 se utilizó un vector base de 5.876 pb con dos unidades de expresión independientes. La primera, empleada para identificar embriones que expresaban el constructo, contenía el promotor constitutivo del citomegalovirus (CMV), una secuencia Kozak, el gen reportero mCherry y una señal de poliadenilación SV40 late poly(A). La segunda contenía la región 5' UTR de β -globina, una secuencia Kozak, el reportero GFP nuclear (nEGFP) y una segunda señal SV40 late poly(A). Río arriba de nEGFP se dispuso un sitio de clonación múltiple con los sitios PmeI, KpnI, SacII, SphI, Ascl y PacI. El vector también contenía sitios únicos EcoRI, FseI, Swal y NotI para su linearización, además de sitios de reconocimiento para la meganucleasa I-SceI. La secuencia de KN9 asociada al promotor de *tmbim6* se clonó río arriba de nEGFP. La fluorescencia de mCherry se utilizó para identificar embriones que expresaban el constructo, mientras que la fluorescencia de GFP se empleó como indicador de la actividad promotora del fragmento evaluado. Como control se utilizó un vector derivado del mismo constructo mediante digestión con PacI, del cual se esperaba únicamente la fluorescencia dirigida por CMV–mCherry.

2.3.2. Preparación de espermios de *XI*.

Los espermatozoides se obtuvieron de los testículos de un macho adulto de *XI*. Los testículos se hicieron rodar suavemente sobre papel absorbente para retirar los vasos sanguíneos superficiales y se lavaron repetidamente con MMR 1× frío hasta eliminar la mayor cantidad visible de sangre. Cada testículo se transfirió a un tubo de 1,5 mL con 500 μ L de tampón de preparación nuclear frío (NPB 1×: 500 mM de sacarosa, 30 mM de HEPES, 1 mM de trihidrocloruro de espermidina, 0,4 mM de tetrahidrocloruro de espermina, 2 mM de ditioneitol y 2 mM de EDTA) y se disgregó mecánicamente hasta obtener una suspensión homogénea. La suspensión se filtró primero a través de una malla de nailon de 0,5–1 mm. El tubo se enjuagó con 1 mL adicional de NPB 1× frío y el lavado se recuperó a través de la misma malla.

Posteriormente, la suspensión se filtró mediante una malla de 100–200 μm y se centrifugó durante 5 min a 2.000 rpm y 4 °C. Tras descartar el sobrenadante, el pellet se mantuvo aproximadamente 30 min sobre hielo en tampón de almacenamiento de espermatozoides (SSB: NPB 1× con 30 % de glicerol y 0,3 % de BSA) antes de su resuspensión. Cuando persistió contaminación sanguínea, se repitieron la centrifugación y la resuspensión. La concentración se determinó mediante una cámara de Neubauer y la suspensión se distribuyó en alícuotas de aproximadamente 5–8 μL , que se almacenaron a –80 °C.

2.3.3. Microinyección de núcleos de espermatozoides y ADN.

Las alícuotas se descongelaron sobre hielo y cada preparación se mezcló con 0,5–1 μg del plásmido de transgénesis. Después de 15 min, la suspensión se diluyó con 500 μL de agua libre de nucleasas y el volumen final se ajustó según la concentración de espermatozoides. La preparación se mantuvo sobre hielo hasta su utilización. Las agujas de vidrio se prepararon con un puller de micropipetas y sus puntas se ajustaron manualmente hasta alcanzar un diámetro aproximado de 50 μm . Una vez cargadas con la mezcla de espermatozoides y ADN, se examinaron al microscopio para descartar obstrucciones. Cuando se detectaron agregados, la punta se recortó por encima de la zona obstruida. Los huevos obtenidos de hembras de *XI* se desgelatinizaron con cisteína y se transfirieron a placas con Ficoll al 8 % y gentamicina 1× (50 $\mu\text{g}/\text{mL}$) en MMR 0,1×. La mezcla de espermatozoides y ADN se introdujo mediante un sistema de microinyección de flujo continuo, empleando un tiempo inferior a 1 s por huevo. Después de la microinyección, los huevos se mantuvieron a 16 °C en la misma solución. Los ejemplares deteriorados se retiraron después de algunas horas. Al día siguiente, los embriones se lavaron con MMR 0,1× y aquellos con desarrollo normal se transfirieron a placas limpias. Posteriormente, se distribuyeron en placas de seis pocillos, con aproximadamente 5–10 embriones por pocillo, y se mantuvieron en MMR 0,1× con gentamicina 1×. Durante el segundo día post microinyección los embriones en desarrollo se mantuvieron a 18 °C y desde el tercer día post microinyección se mantuvieron a temperatura ambiente. La fluorescencia de mCherry se utilizó para identificar embriones que expresaban el constructo, mientras que la presencia y distribución de GFP se emplearon para evaluar la actividad promotora de KN9. Debido a la elevada autofluorescencia del vitelo en estadios tempranos, la señal de GFP se examinó preferentemente en estadios más avanzados. Los embriones con fluorescencia detectable se separaron de aquellos sin señal y se mantuvieron hasta el estadio larval requerido.

2.3.4. Microscopia y procesamiento de imágenes de larvas transgénicas vivas.

Las larvas transgénicas vivas fueron anestesiadas con una solución diluida de triclaína y visualizadas directamente utilizando un sistema de microscopía Carl Zeiss Axio Zoom.V16 equipado con ApoTome.2 (FONDEQUIP EQM140146). Las imágenes fueron adquiridas en campo claro y utilizando los canales correspondientes a Alexa Fluor 488/GFP y Alexa Fluor 546, para la detección de

fluorescencia verde y roja, respectivamente. La adquisición y el procesamiento de las imágenes se realizaron mediante el software ZEISS ZEN 3.12 (Carl Zeiss Microscopy). Para todas las imágenes se aplicaron los mismos parámetros de ajuste de brillo y contraste, establecidos a partir de los controles negativos.

3. RESULTADOS

3.1. Presencia de transposones en el paisaje regulatorio en *Xenopus*.

Con el propósito de abordar la hipótesis de investigación planteada en esta tesis, se utilizaron los conjuntos de datos sintetizados en la **Figura 2**, los cuales comprenden perfiles de ARN-Seq y ATAC-Seq para tejido hepático, mesénquima y osteoblastos, así como un clúster de accesibilidad a la cromatina ubicuo presente en los tres tejidos. A partir de estos mapas de expresión y cromatina abierta, se identificaron promotores de genes y elementos *cis*-regulatorios distales tanto para *Xt* (*Xt*) como para *Xl* (*Xl*).

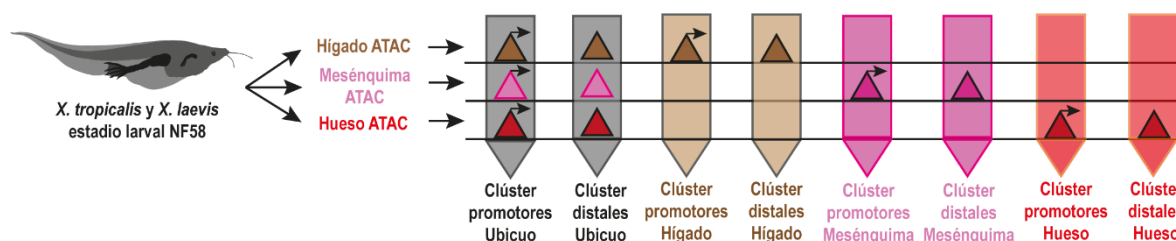


Figura 2: Esquema de la obtención de bases de datos de elementos *cis*-regulatorios en *Xenopus*. Se generaron bibliotecas de secuenciación a partir de tejidos de hígado, mesénquima y hueso mediante ATAC-seq. A partir de los datos obtenidos, se definieron clústeres de promotores y elementos *cis*-regulatorios distales con actividad ubicua (presentes en los tres tejidos) o tejido específica.

Con la finalidad de identificar la contribución de elementos transponibles a estas regiones reguladoras, se realizaron análisis bioinformáticos en los clústeres definidos. Para *Xt*, se identificó un enriquecimiento local específico de cinco subfamilias de transposones de ADN (**Figura 3**): *Kolobok-N9_XT* (KN9) en promotores ubicuos; *Harbinger-N12_XT* en elementos distales ubicuos; *Harbinger-N12A_XT*, *hAT-N16_XT* y *Kolobok-N2_XT* en elementos distales de hígado (126). Por otro lado, en *Xl* se identificaron ocho subfamilias de transposones de ADN (**Figura 4**): *KN9* en promotores ubicuos; *DNA10_XT*, *Harbinger-N12A_XT*, *hAT-N26_XL*, *hAT-N26B_XL*, *hAT-N15_XL* y *hAT-N52_XL* en elementos distales de hígado; y *Kolobok-N3_XT* en elementos distales de mesénquima y osteoblastos. Adicionalmente, en *Xl* se detectaron siete retrotransposones pertenecientes a las subfamilias DIRS (Por sus siglas en inglés de “*Dictyostelium* intermediate repeat sequence”) y Gypsy en elementos distales de hígado **Figura 5**: *DIRS-42_XT*, *DIRS-50_XT*, *DIRS-29_XT*, *DIRS-17_XL*, *DIRS-14_XT*, *DIRS-8A_XT* y *Gypsy-11-LRT_XT*. Finalmente, estos retrotransposones exhibieron una distribución en la señal de cobertura denominada de carácter bimodal, ancho o focalizado, según corresponda.

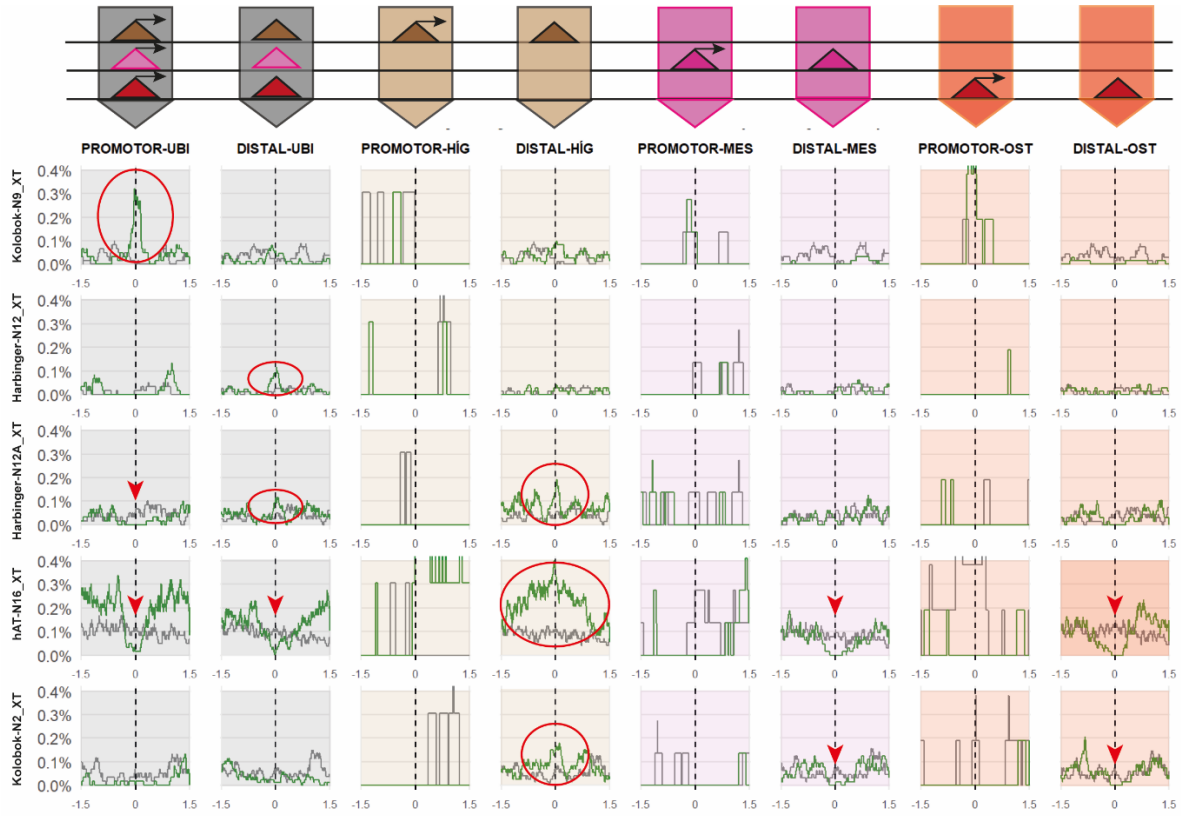


Figura 3: Enriquecimiento local específico de cinco subfamilias de MITEs en Xt. En el eje Y se representa, en verde, el porcentaje de ocupación de MITEs en una ventana de 3 kb centrada en los promotores y elementos distales identificados mediante ATAC-seq para cada tejido. El centro de la ventana se indica con el valor 0. En gris se representa la ocupación de secuencias aleatorias utilizadas como control. Distal: elementos cis-regulatorios distales; UBI: elementos reguladores ubicuos, presentes en los tres tejidos; HIG: elementos reguladores de hígado; MES: elementos reguladores de mesénquima; OST: elementos reguladores de osteoblastos. Los picos señalados con círculos rojos corresponden a subfamilias de MITEs enriquecidas en los respectivos elementos reguladores, mientras que las flechas rojas indican empobrecimientos específicos de las subfamilias correspondientes. El eje X indica la distancia en kilobases.

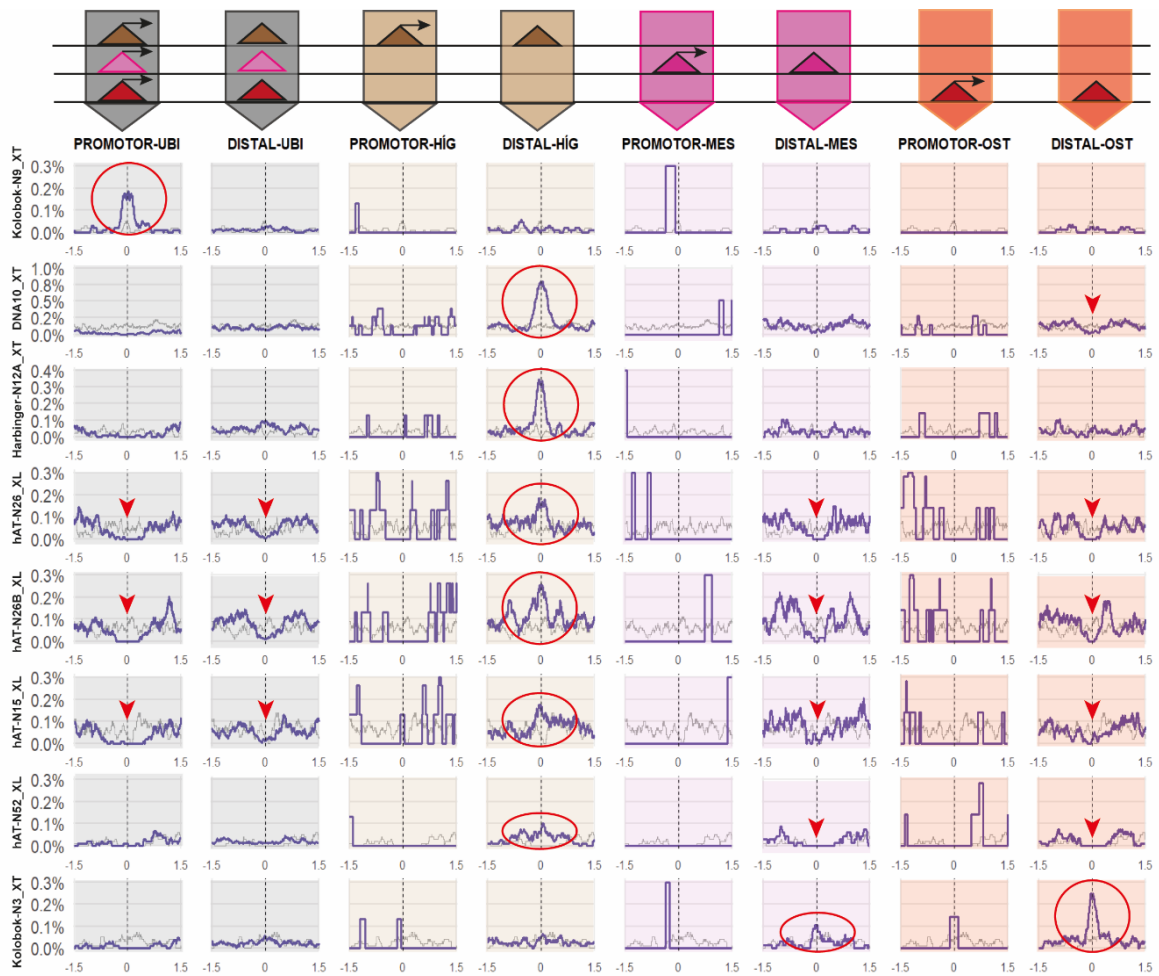


Figura 4: Enriquecimiento local específico de ocho subfamilias de MITEs en XI. En el eje Y se representa, en morado, el porcentaje de ocupación de MITEs en una ventana de 3 kb centrada en los promotores y elementos distales identificados mediante ATAC-seq para cada tejido. En gris se representa la ocupación de secuencias aleatorias utilizadas como control. Distal: elementos cis-regulatorios distales; UBI: elementos reguladores ubicuos, presentes en los tres tejidos; HÍG: elementos reguladores de hígado; MES: elementos reguladores de mesénquima; OST: elementos reguladores de osteoblastos. Los picos señalados con círculos rojos corresponden a subfamilias de MITEs enriquecidas en los respectivos elementos reguladores, mientras que las flechas rojas indican empobrecimientos específicos de las subfamilias correspondientes. El eje X indica la distancia en kilobases.

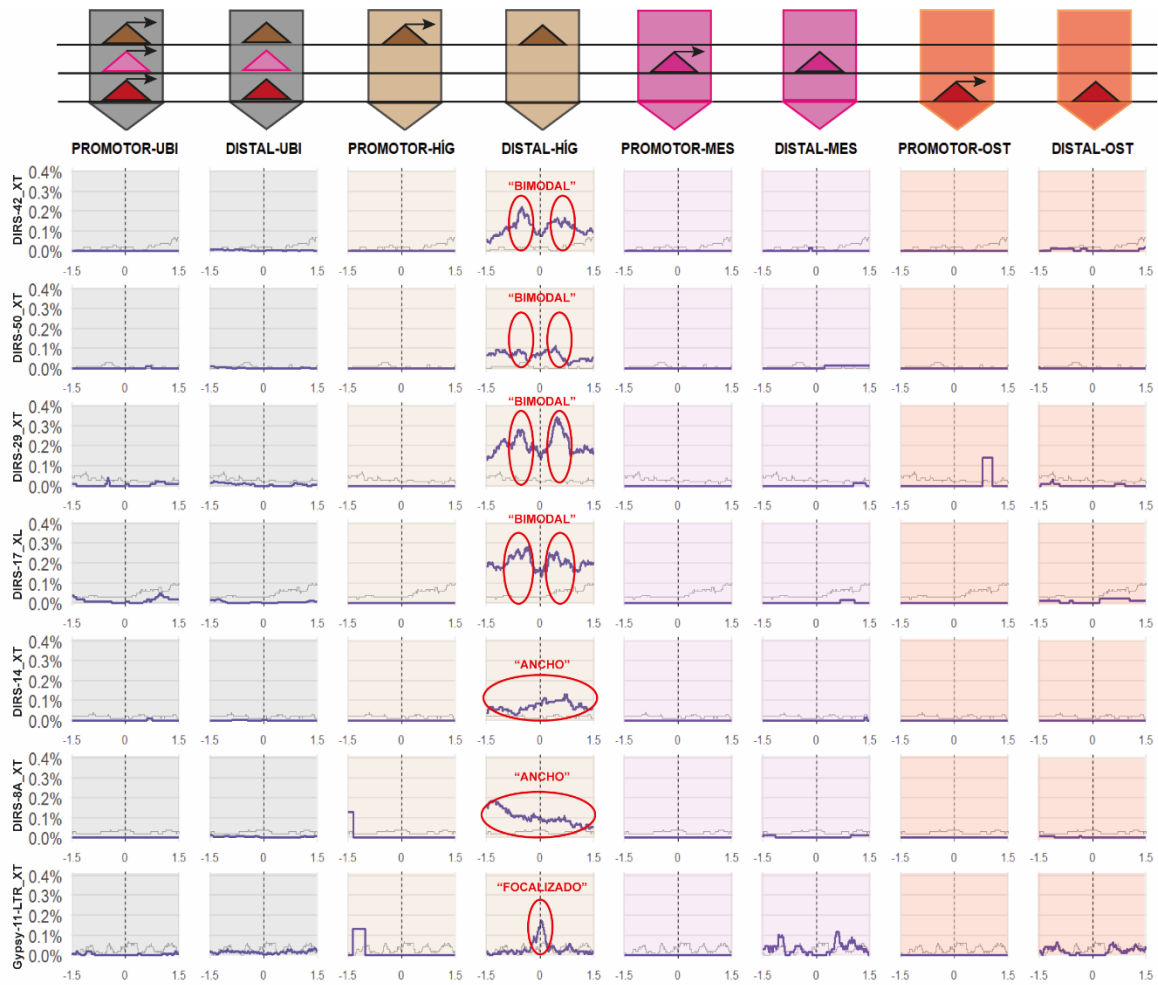


Figura 5: Enriquecimiento local específico de siete subfamilias de retrotransposones en XI. En el eje Y se representa, en morado, el porcentaje de ocupación de cada retrotransposón en una ventana de 3 kb centrada en los promotores y elementos distales identificados mediante ATAC-seq para cada tejido. En gris se representa la ocupación de secuencias aleatorias utilizadas como control. Distal: elementos cis-regulatorios distales; UBI: elementos reguladores ubicuos, presentes en los tres tejidos; HÍG: elementos reguladores de hígado; MES: elementos reguladores de mesénquima; OST: elementos reguladores de osteoblastos. Los picos señalados con círculos rojos corresponden a subfamilias de retrotransposones enriquecidas en los respectivos elementos reguladores, cuyos perfiles se clasifican como bimodales, anchos o focalizados. El eje X indica la distancia en kilobases.

3.2. Elemento transponible en miniatura (MITE) KN9.

Para investigar la función reguladora de los transposones en las regiones genómicas proximales de *Xenopus*, se seleccionó la subfamilia *Kolobok-N9_XT* (KN9) identificada en el clúster de promotores ubicuos. La elección de este elemento se fundamentó en los siguientes criterios: (1) El escaso conocimiento sobre la domesticación de transposones de ADN en regiones promotoras. (2) Su condición de evento de domesticación compartido entre *Xt* y los *XtL* y *XtS* de *Xl* (**Figura 6**), lo que le otorga un alto valor evolutivo y comparativo. (3) La inequívoca asignación del gen blanco dada la contigüidad posicional de KN9 al sitio de inicio de la

transcripción. (4) La factibilidad técnica para realizar ensayos reporteros en promotores derivados de KN9.

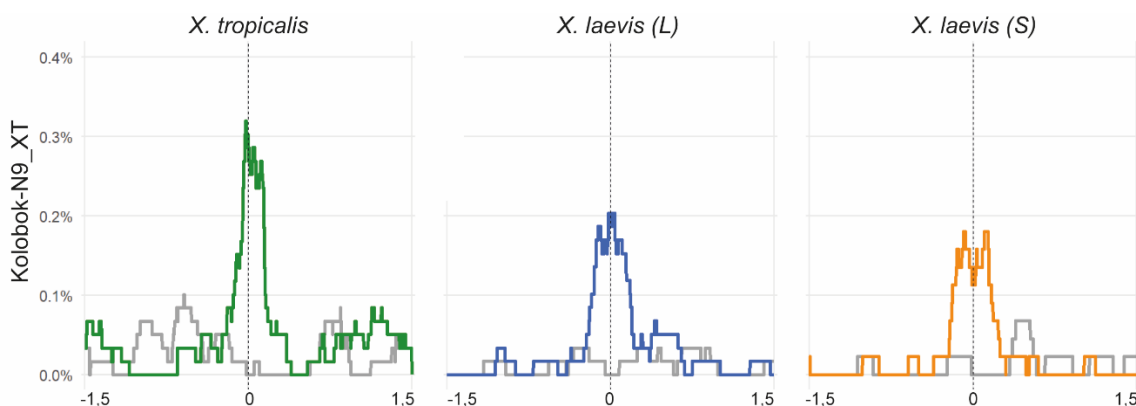


Figura 6: Enriquecimiento local del MITE KN9 en promotores ubicuos. Se muestran en verde los datos de Xt, en azul los del subgenoma L y en naranja los del subgenoma S de Xi. En el eje Y se representa el porcentaje de ocupación de KN9 en una ventana de 3 kb centrada en los promotores del clúster ubicuo identificados mediante ATAC-seq. En gris se representa la ocupación de secuencias aleatorias utilizadas como control. El eje X indica la distancia en kilobases.

KN9 fue identificado en los promotores de 60 genes; considerando los tres genomas y eventuales pérdidas génicas por delección, se contabilizó un total de 141 promotores derivados de este MITE (**Figura 7A**). El análisis detallado de estas regiones reveló que 101 promotores presentan una domesticación completa (es decir, KN9 originó el promotor funcional y el promotor ancestral fue reemplazado). Asimismo, otros 16 promotores derivados de KN9 se encuentran en un estado evolutivo intermedio, coexistiendo el promotor original y el derivado de KN9 (**Figura 7**). Finalmente, se observaron casos donde predomina el promotor original, con pérdida secundaria de KN9. Estos escenarios evolutivos se ilustran en la **Figura 7B**. En los intermediarios evolutivos (**Figura 7C**) se aprecian dos regiones de accesibilidad cromatínica en el perfil de ATAC-Seq, correspondientes al promotor ancestral y al promotor originado por KN9. En contraste, los eventos de domesticación completa muestran únicamente señal de cromatina abierta en la región del promotor derivado de KN9. Cabe destacar que, en ambos casos, en el extremo 3' de KN9 se detecta señal de transcripción en los perfiles de ARN-Seq, sugiriendo que el MITE aporta adicionalmente el primer exón, como es esperable para una función promotora la cual, por definición, contiene el inicio de transcripción. Estos datos indican una tendencia evolutiva en la cual KN9 se insertó en las cercanías de un promotor y, a lo largo del tiempo macroevolutivo, el promotor original fue progresivamente reemplazado. El listado completo de genes regulados por KN9 se detalla en el **Anexo 1**.

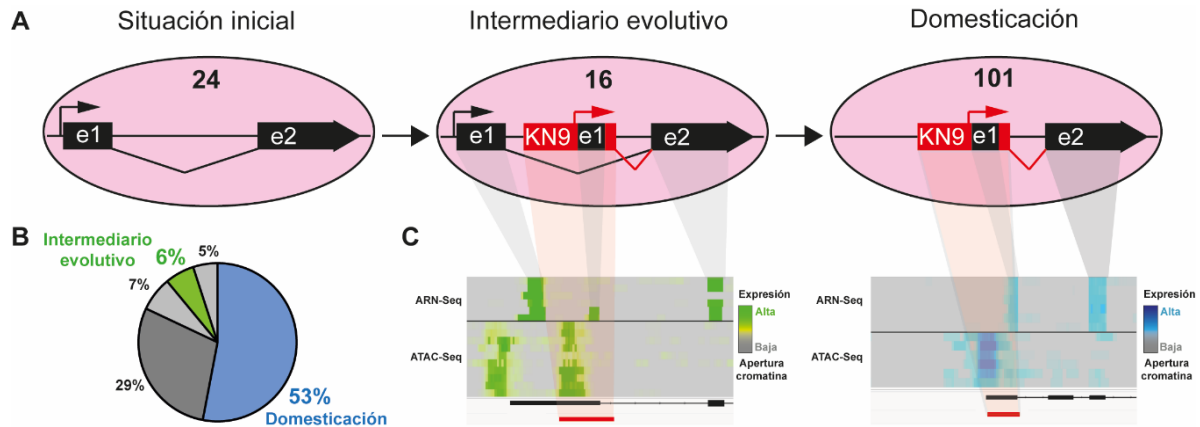


Figura 7: Resumen de los escenarios evolutivos identificados para los promotores originados a partir del MITE KN9. **(A)** De los 141 promotores pertenecientes al clúster ubicuo, 101 corresponden a casos de domesticación; 16 representan intermediarios evolutivos, en los que el promotor original coexiste con el promotor domesticado; y 24 corresponden a casos sin domesticación del transposón, en los que solo está presente el promotor original. Los rectángulos negros representan los exones e1 y e2, mientras que el rectángulo rojo representa el MITE KN9. La flecha negra situada en e1 indica el sitio de inicio de la transcripción. **(B)** Gráfico circular que muestra la frecuencia de cada escenario evolutivo. Del total de casos, el 5 % corresponde a la situación inicial presente únicamente en *Xl*, el 7 % a eventos de domesticación exclusivos de *Xt* y el 29 % a casos que aún no han sido resueltos. **(C)** Mapas de calor que ejemplifican los escenarios de intermediario evolutivo y domesticación. Se muestran las señales de RNA-seq y ATAC-seq. Cada fila representa un tejido y su respectivo duplicado. La escala indica el nivel de expresión génica o de accesibilidad de la cromatina, según corresponda: el gris representa ausencia de señal, mientras que el verde o el azul representan la intensidad máxima. Los rectángulos negros representan la anotación del gen y el rectángulo rojo, el MITE KN9. El ejemplo mostrado en verde corresponde al gen *ubxn1* de *Xt*, mientras que el ejemplo mostrado en azul corresponde al gen *borcs6* de *Xl*.

A continuación, se caracterizó exhaustivamente la secuencia de las copias domesticadas de KN9 en ambas especies de *Xenopus*. La **Figura 8A** muestra la secuencia consenso del transposón de ADN Kolobok-N9_XT, señalando las repeticiones invertidas terminales (TIRs) y la región en el extremo 3' en rojo que constituye las copias domesticadas, observable en la ocupancia de las copias domesticadas sobre el consenso (**Figura 8B**). El alineamiento de estas copias permitió identificar elementos conservados (**Figura 8C**), entre los cuales destacan sitios de unión para FTs de las familias Sp/KLF, PBC/MEINOX y ETS, además del motivo putativo de empalme AGGT. Además, una reconstrucción filogenética a partir de estas secuencias (**Figura 9**) agrupó en clados monofiléticos las copias de KN9 que regulan al mismo gen homólogo en los tres genomas (*Xt*, *XIL* y *XIS*). Esto confirma que los eventos de domesticación ocurrieron en el ancestro común de ambas especies mediante un único evento de transposición por gen. Asimismo, la uniformidad en las distancias evolutivas entre la mayoría de los clados sugiere un episodio principal de amplificación y domesticación de KN9 previo a la divergencia entre *Xt* y *Xl*.



Figura 8: Alineamiento múltiple de las secuencias de copias domesticadas de KN9 de Xt y de los XIL y XIS de XI respecto de la secuencia consenso. **(A)** Secuencia consenso de KN9. En negro se indican las repeticiones terminales invertidas (TIRs); en gris, la región de la secuencia consenso ausente en las copias domesticadas de KN9; y en rojo, la fracción del consenso correspondiente a las copias domesticadas de Xt y de los XIL y XIS de XI. **(B)** Ocupación de las copias domesticadas de KN9 a lo largo de la secuencia consenso. La barra representa la longitud completa de la secuencia consenso y conserva el código de colores descrito en A. Debajo se muestra el gráfico de ocupación a lo largo de toda la secuencia consenso, representado en la misma escala (1:1). **(C)** Alineamiento múltiple de las secuencias de las copias domesticadas de KN9. La barra representa la región domesticada de la secuencia consenso y, debajo, se muestra el alineamiento en la misma escala (1:1). Bajo el alineamiento se indican motivos de secuencia conservados, incluidos los sitios de unión para FTs de las familias Sp/KLF, PBC/MEINOX y ETS, además del motivo putativo de empalme AGGT.

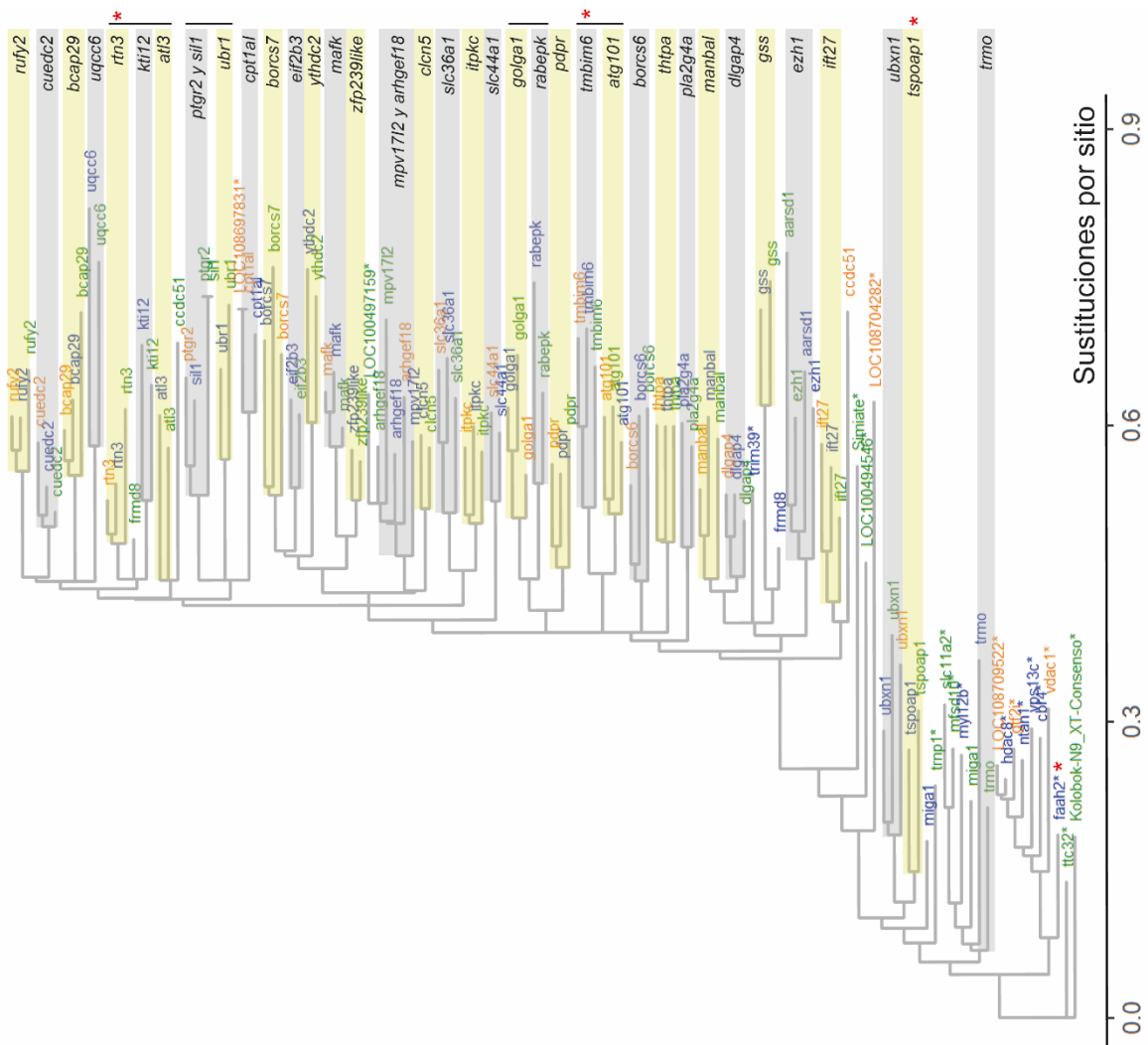


Figura 9: Árbol filogenético de los MITEs KN9 de interés que dieron origen a promotores del clúster ubicuo, denominados según su gen diana. Los KN9 domesticados de Xt se muestran en verde, mientras que aquellos pertenecientes a los XIL y XIS de XI se muestran en azul y naranja, respectivamente. Los sombreados amarillo y gris señalan clados monofiléticos. La escala representa la distancia evolutiva, expresada como sustituciones por sitio. Las líneas negras indican genes vecinos. Los asteriscos rojos señalan los genes seleccionados como candidatos para estudiar su patrón de expresión espaciotemporal. Los asteriscos verdes, azules y naranjos indican MITEs específicos de Xt y de los XIL y XIS de XI, respectivamente.

3.3. Localización subcelular de los ortólogos humanos codificados por genes asociados a promotores derivados de KN9 en *Xenopus*.

Para examinar la distribución subcelular de las proteínas codificadas por genes asociados a promotores derivados de KN9, se consultó la base de datos de proteómica subcelular de Hein et al. (2025), generada en células humanas HEK293T. Por lo tanto, este análisis representa la localización reportada para los ortólogos humanos de los genes estudiados. Los resultados indican que estas proteínas se distribuyen en diversos compartimentos subcelulares, con una mayor representación en el retículo endoplásmico, el aparato de Golgi, los lisosomas, el citosol y las mitocondrias (**Figura 10**). Así mismo, destaca una muy baja representación de proteínas nucleares, de FTs, y de componentes de vías de señalización. Esta observación es coherente con lo que se espera para genes que presentarían una expresión y una función ubicua.

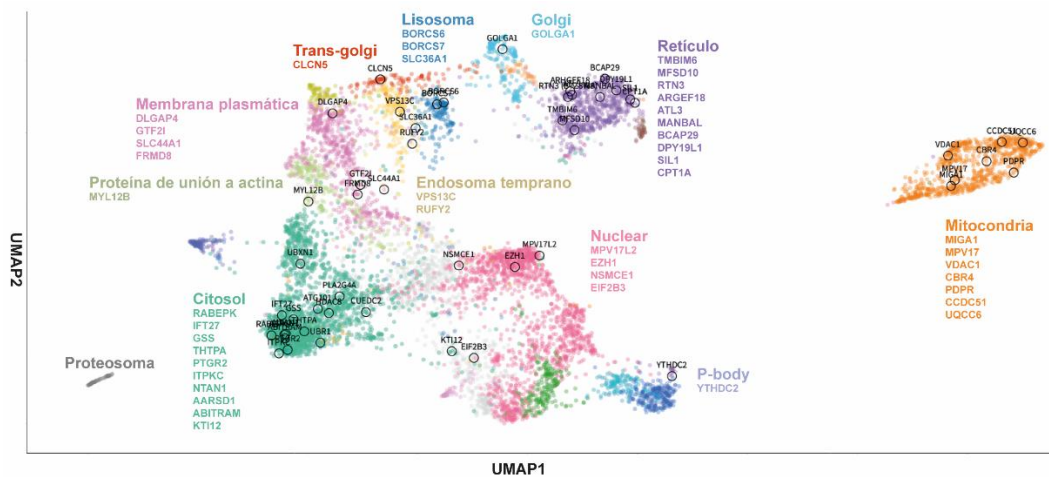


Figura 10: Localización subcelular reportada para los ortólogos humanos de los genes asociados a promotores derivados de KN9. Cada color y grupo representa una región, estructura o compartimento celular, según se indica en la figura. La información fue obtenida de una base de datos de acceso abierto generada por Hein et al. (Cell, vol. 188, n.º 4, 2025).

3.4. Expresión espacio-temporal de genes regulados por promotores derivados de KN9.

Si bien los análisis de secuenciación utilizados aportan un marco global sobre el paisaje transcripcional y regulatorio ubicuo, carecen de resolución espacial y temporal a nivel tisular. Por ello, se caracterizó la dinámica de expresión espacio-temporal de seis genes candidatos durante el desarrollo embrionario mediante hibridación *in situ* fluorescente por reacción en cadena de hibridación (HCR).

Los genes candidatos (**Figura 11**) se seleccionaron de acuerdo con los siguientes criterios: (1) representación de las localizaciones subcelulares identificadas en la **Figura 10**, incluyendo *rtn3* (**Figura 11A**) y *tmbim6* (**Figura 11B**), asociados al retículo endoplásmico, y *abitrham* (**Figura 11C**), asociado al citosol; (2) evaluación de una asociación entre KN9 y el promotor específica del linaje de *Xl*, seleccionando *faah2*, para el cual KN9 actúa como promotor únicamente en el subgenoma L (**Figura 11D**); (3) evaluación de un promotor asociado a KN9 que carece de señal

de ATAC-seq en los tejidos examinados, seleccionando *tspoap1* (Figura 11E); y (4) evaluación del posible efecto de una mutación en el sitio ETS conservado, seleccionando *nsmce1* (Figura 11F).

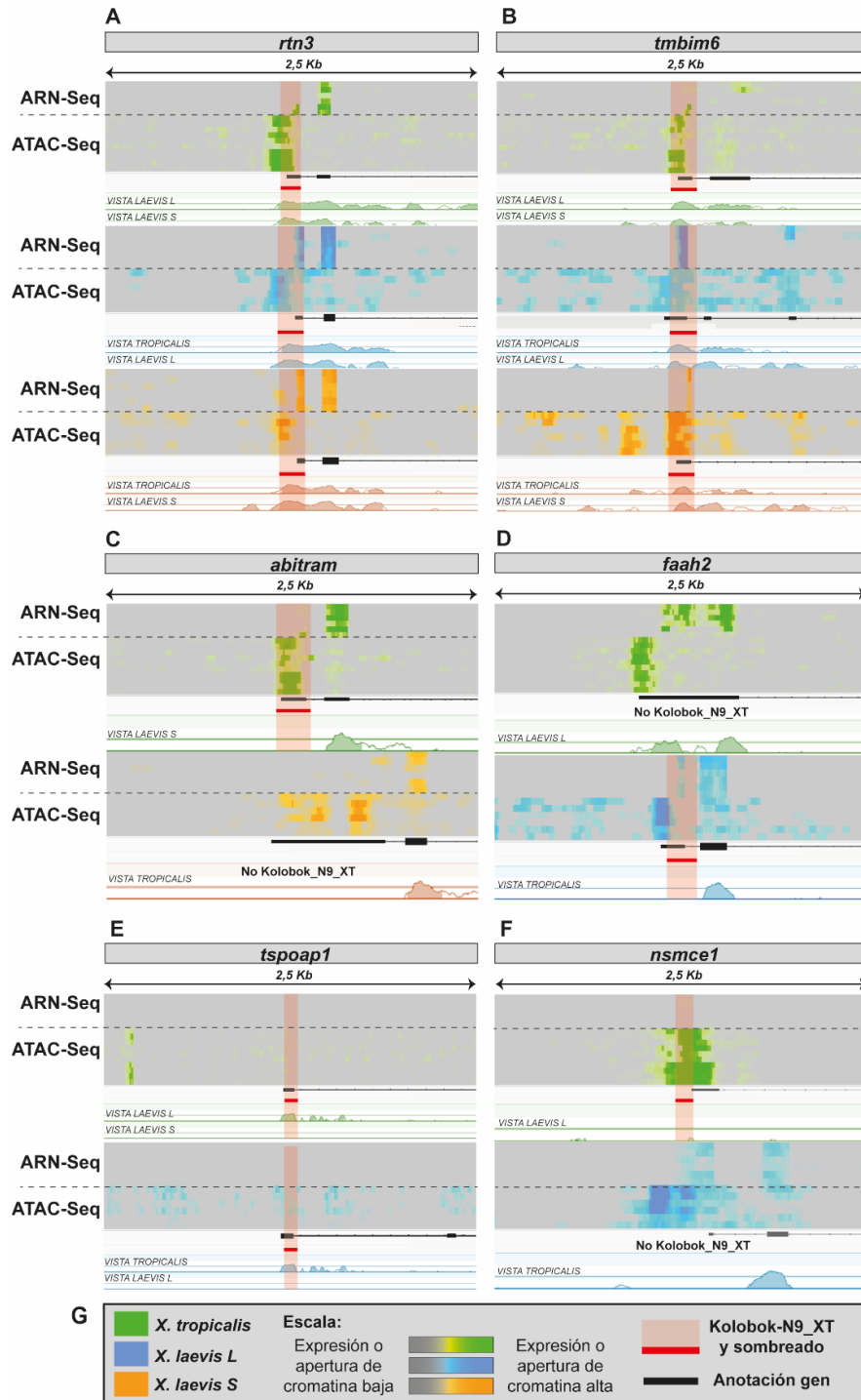


Figura 11: Mapas de calor de los genes candidatos regulados por KN9. (A) Mapa de calor de una ventana de 2,5 kb que muestra la presencia de KN9 como promotor del gen *rtn3* en el genoma de *Xt* y en los XIL y XIS de *Xl*. (B) Mapa de calor de una ventana de 2,5 kb que muestra la presencia de KN9 como promotor del gen *tmbim6* en el genoma de *Xt* y en los XIL y XIS de *Xl*. (C) Mapa de calor

de una ventana de 2,5 kb que muestra la presencia de KN9 como promotor del gen *abitrans* en el genoma de *Xt*. El gen *abitrans* fue eliminado del subgenoma L de *Xl*, mientras que, en el subgenoma S, KN9 no actúa como promotor de este gen. **(D)** Mapa de calor de una ventana de 2,5 kb que muestra la presencia de KN9 como promotor del gen *faah2* en el subgenoma L de *Xl*. El gen *faah2* fue eliminado del subgenoma S de *Xl*, mientras que, en *Xt*, KN9 no actúa como promotor de este gen. **(E)** Mapa de calor de una ventana de 2,5 kb que muestra la presencia de KN9 como promotor del gen *tspoap1* en el genoma de *Xt* y en el subgenoma L de *Xl*. **(F)** Mapa de calor de una ventana de 2,5 kb que muestra la presencia de KN9 como promotor del gen *nsmce1* en el genoma de *Xt*. El gen *nsmce1* fue eliminado del subgenoma S de *Xl*, mientras que, en el subgenoma L, KN9 no actúa como promotor de este gen. **(G)** Esquema de la simbología utilizada en los mapas de calor. El genoma de *Xt* y los *XIL* y *XIS* de *Xl* se representan en verde, azul y naranja, respectivamente.

En primer lugar, para caracterizar el patrón regulatorio de KN9 se realizó la HCR del gen *rtn3* **(Figura 12)**. La señal se detectó durante la neurulación en la placa neural (Np, NF18) y en células de la cresta neural (Nc, NF18–NF20). En NF34, la expresión se observó en el tubo neural (Nt), la vesícula ótica (Ot), el ojo (Ey), los arcos hioideos y branquiales (Ha y Ba), el pronefros (Pn) y la yema caudal (Tb). La señal en la vesícula ótica y el ojo persistió hasta NF40. En este último estadio, la expresión se detectó ampliamente en el sistema nervioso central, incluyendo el encéfalo (Br, Fb, Mb, Hb) y la médula espinal (Sc), además de células de la cresta neural posmigratoria (Pmnc). La expresión de *tubb2* se utilizó como control positivo para la identificación de las estructuras neurales. En conjunto, estos resultados muestran que el patrón regulatorio de KN9 es amplio y presenta un evidente componente neural.

En segundo lugar, para evaluar si el patrón de expresión asociado a KN9 se conserva entre *Xt* y *Xl*, se caracterizó mediante HCR la expresión del gen *tmbim6*. En *Xt* **(Figura 13)**, la señal se detectó desde la neurulación en la placa neural (Np, NF18) y en células de la cresta neural (Nc, NF18–NF20). Durante los estadios posteriores, la expresión se observó en el tubo neural (Nt, NF34), la vesícula ótica y el ojo (Ot y Ey, NF34–NF40), los arcos hioideos y branquiales (Ha y Ba, NF34), el pronefros y riñón (Pn y Kd, NF34–NF40) y la yema caudal (Tb, NF34). En NF40, la señal se detectó ampliamente en el sistema nervioso central, incluyendo el encéfalo (Br, Fb, Mb, Hb) y la médula espinal (Sc), además de células de la cresta neural posmigratoria (Pmnc). En *Xl* **(Figura 14)** no se dispuso de datos para NF18–NF20; sin embargo, desde NF34 el patrón coincidió con el observado en *Xt* en todas las estructuras evaluadas. Asimismo, ambas presentaron un patrón amplio de expresión y principalmente neural, corroborado con la expresión de *tubb2*, utilizada como control positivo del sistema nervioso en *Xt*. En conjunto, la correspondencia espacial observada entre NF34 y NF40 sugiere la conservación del patrón de expresión de *tmbim6* entre ambas especies y es consistente con la conservación de la actividad regulatoria asociada a KN9, además de la similitud de expresión de *rtn3*, *tmbim6* Y *tmbim6.L*, ambas del clúster de proteínas de retículo **(Figura 10)**.

En tercer lugar, se caracterizó mediante la expresión del gen *abitrans*, seleccionado como representante de los genes cuya proteína codificada presenta una localización predominantemente citosólica **(Figura 10; Figura 15)**. En NF28, la

señal se detectó en células de la cresta neural (Nc), el tubo neural (Nt) y el ojo (Ey). La expresión en el tubo neural y el ojo persistió en NF34, estadio en el que también se observó señal en la vesícula ótica (Ot), los arcos hioideos y branquiales (Ha y Ba), el pronefros (Pn) y la yema caudal (Tb). En NF40, la expresión se mantuvo en la vesícula ótica y el ojo, y se detectó ampliamente en el sistema nervioso central, incluyendo el encéfalo (Br, Fb, Mb y Hb) y la médula espinal (Sc), además de células de la cresta neural posmigratoria (Pmnc). En conjunto, *abitram* presentó un patrón de expresión amplio entre NF28 y NF40, concordante con lo observado en *rtn3* y *tmbim6*. Aunque para este caso no se incluyó *tubb2* como marcador neural, estos resultados son consistentes con la actividad de un promotor ubicuo derivado de KN9.

En cuarto lugar, para evaluar un evento de domesticación de KN9 específico del linaje de *Xl*, se caracterizó la expresión de *faah2* en *Xt* y *Xl* (**Figura 16**). En *Xt*, la señal se detectó desde la neurulación en la placa neural (Np, NF18) y en células de la cresta neural (Nc, NF18–NF20), mientras que para estos estadios no se dispuso de datos en *Xl*. Desde NF34, ambas especies presentaron un patrón coincidente en el tubo neural (Nt), la vesícula ótica (Ot), el ojo (Ey), los arcos hioideos y branquiales (Ha y Ba), el pronefros y riñón (Pn y Kd), y la yema caudal (Tb). En NF40, la señal se mantuvo en la vesícula ótica, el ojo y el riñón, y se detectó en el encéfalo (Br) y en células de la cresta neural posmigratoria (Pmnc) de ambas especies. La principal diferencia se observó en la notocorda (No), donde no se detectó expresión en *Xt*, mientras que en *Xl* se observó señal durante NF34–NF40. En conjunto, *faah2* presentó un patrón ampliamente conservado entre ambas especies. Esto apoya la idea que KN9 tiene tropismo para genes que ya previamente presentan una expresión ubicua.

En quinto lugar, se caracterizó mediante HCR la expresión de *tspoap1*, seleccionado como un gen muy atípico por ser asociado a un promotor derivado de KN9 que carece de señal de ATAC-seq en los tejidos evaluados (**Figura 11E; Figura 17**). Durante NF18–NF20 se detectó expresión en la placa neural (Np) ni en células de la cresta neural (Nc). En NF34, la señal se restringió al ojo (Ey), sin observarse expresión en el tubo neural (Nt), la vesícula ótica (Ot), los arcos hioideos y branquiales (Ha y Ba), el pronefros (Pn) o la yema caudal (Tb). En NF40, la expresión se mantuvo en el ojo y se detectó en el sistema nervioso central, incluyendo el encéfalo (Br, Fb, Mb, Hb) y la médula espinal (Sc). No se observó señal en células de la cresta neural posmigratoria (Pmnc), el riñón (Kd) ni la notocorda (No). La expresión de *tubb2* se utilizó como control positivo para la identificación de las estructuras neurales. En conjunto, *tspoap1* presentó un patrón de expresión restringido a territorios oculares y neurales, acorde en lo observado en la expresión neural para los genes *rtn3*, *tmbim6*, *abitram* y *faah2*, pero sin la presencia de una expresión amplia. Este patrón es concordante con la ausencia de señal de ATAC en los tejidos no neurales evaluados y es consistente con la

actividad de un promotor derivado de un elemento KN9 divergente con regulación neural.

Finalmente, se caracterizó mediante la expresión de *nsmce1*, cuyo promotor derivado de KN9 presenta una mutación en el sitio ETS conservado (**Figura 11F; Figura 18**). A diferencia de los demás genes analizados, no se detectó señal en ninguno de los estadios evaluados. Esta ausencia de expresión contrasta con la señal ubicua de ATAC en la región de KN9 asociada a *nsmce1*, lo que indica que la accesibilidad de la cromatina no se traduce necesariamente en actividad transcripcional detectable. Aunque no se dispuso de datos de *tubb2* como control positivo, este resultado es consistente con la hipótesis de que la mutación del sitio ETS afecta la actividad del promotor derivado de KN9. El efecto causal de esta mutación deberá evaluarse mediante ensayos reporteros que comparen las versiones conservada y mutada del sitio.

Tabla 2: Secuencias de los sitios de unión para FTs de las familias Sp/KLF, PBC/MEINOX y ETS presentes en la secuencia consenso de KN9 y en los promotores originados a partir de este elemento correspondientes a los genes candidatos seleccionados. Las bases indicadas en rojo representan cambios respecto de la secuencia consenso.

	SP/KLF	PBC/MEINOX	ETS
Consenso	AGGGAGGGG	TGTCAGTCA	CACTTCCTGT
<i>rtn3</i>	GGGG C GGGG	TGTCA C TCA	A ACTTCCTGT
<i>tmbim6</i>	GGGG C GGGG	TGTCA TTT A	CACTTCCTGT
<i>tmbim6.L</i>	GGGG C GT G A	TG A CA C TCA	CACTTCCTGT
<i>abitrām</i>	AGG A AGG A C	TGT T ATT G C	CACTTCCTGT
<i>faah2.L</i>	TGGGAGGGG	TGTC T GTCA	CACTTCCTGT
<i>tspoap1</i>	TGGGAGG A G	TGTCA A TCA	CACTTCCTGT
<i>nsmce1</i>	CCCCGCCCA	TGAG T GA C T	GTAAGCAGC

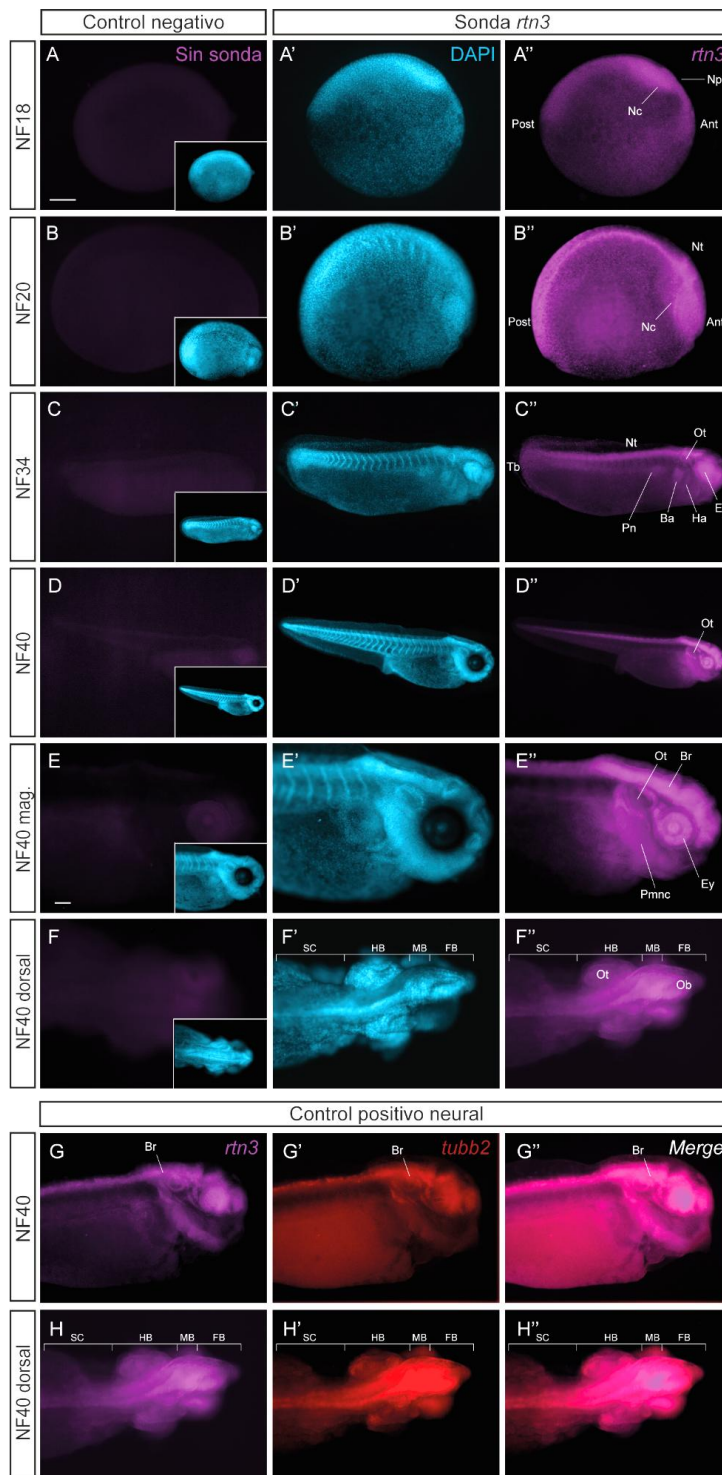


Figura 12: Hibridación *in situ* fluorescente mediante HCR para la caracterización espaciotemporal del gen *rtn3*, regulado por KN9, en *Xt*. Los paneles A, A' y A'' muestran el control negativo, la tinción con DAPI y la señal de la sonda dirigida contra *rtn3*, respectivamente, en el estadio NF18. Siguiendo la misma disposición, los paneles B–B'', C–C'' y D–D'' corresponden a los estadios NF20, NF34 y NF40, respectivamente. Los paneles E–E'' muestran una magnificación del estadio NF40 y los paneles F–F'', una vista dorsal del mismo estadio. Los paneles G, G' y G'' muestran la señal de la sonda dirigida contra *rtn3*, la señal del control neural *tubb2* y la combinación de ambos canales, respectivamente, en el estadio NF40. Los paneles H–H'' muestran las mismas señales en una vista dorsal del estadio NF40. Post: región posterior del embrión; Ant: región anterior del embrión; Np: placa neural; Nc: células de la cresta neural; Nt: tubo neural; Tb: yema caudal; Ot: vesícula ótica; Ey: ojo; Ha: arcos hioideos; Ba: arcos branquiales; Pn: pronefros; Br: encéfalo; Pmnc: células de la cresta neural postmigratorias; SC: médula espinal; HB: rombencéfalo; MB: mesencéfalo; FB: prosencéfalo; Ob: bulbo olfatorio. Escala en series A, B, C, y D corresponde a 200 μ m. Escala en series E, F, G y H corresponde a 100 μ m.

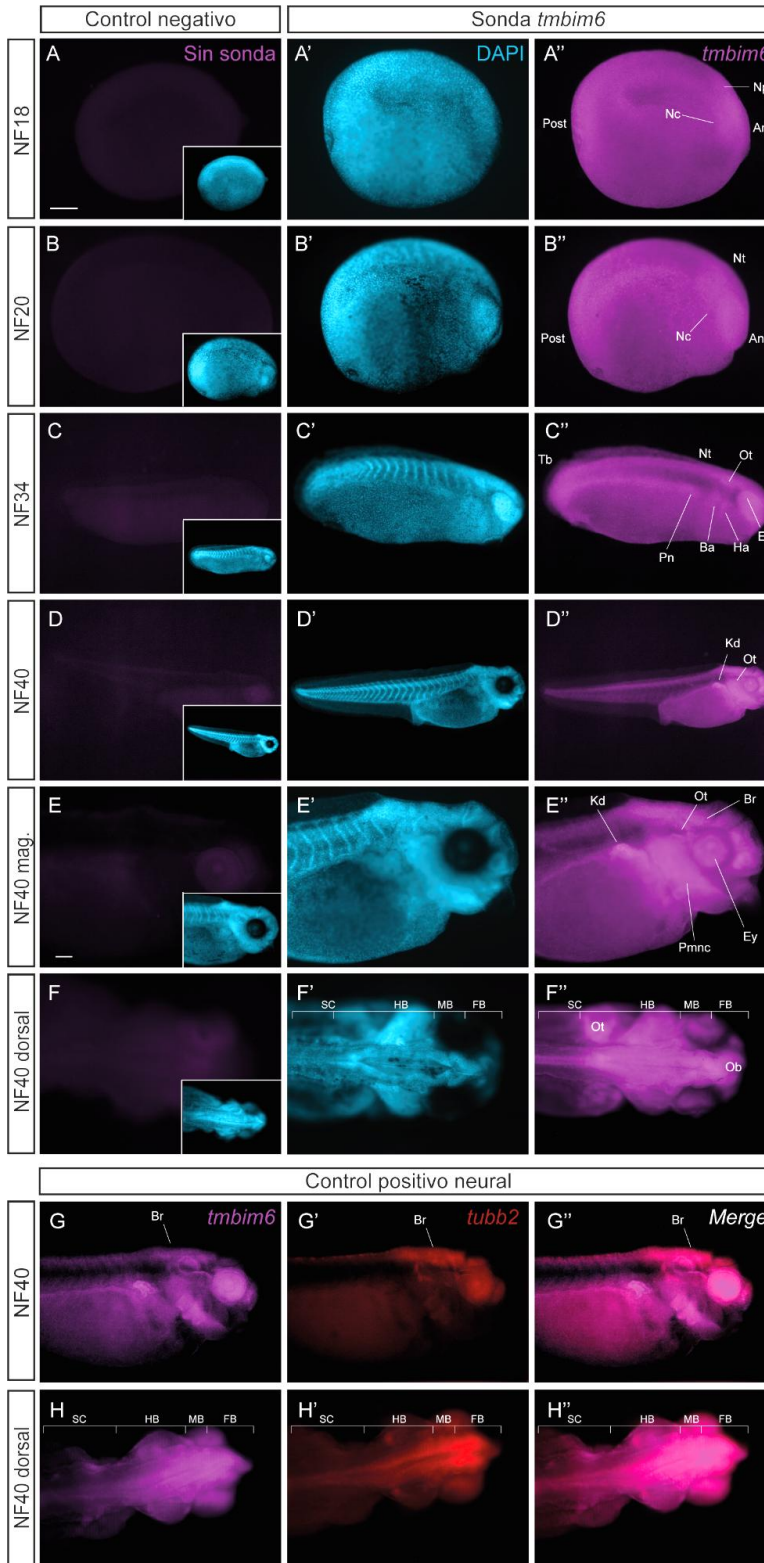


Figura 13: Hibridación *in situ* fluorescente mediante HCR para la caracterización espaciotemporal del gen *tmbim6*, regulado por KN9, en *Xt*. Los paneles A, A' y A'' muestran el control negativo, la tinción con DAPI y la señal de la sonda dirigida contra *tmbim6*, respectivamente, en el estadio NF18. Siguiendo la misma disposición, los paneles B–B'', C–C'' y D–D'' corresponden a los estadios NF20, NF34 y NF40, respectivamente. Los paneles E–E'' muestran una magnificación del estadio NF40 y los paneles F–F'', una vista dorsal del mismo estadio. Los paneles G, G' y G'' muestran la señal de la sonda dirigida contra *tmbim6*, la señal del control neural *tubb2* y la combinación de ambos canales, respectivamente, en el estadio NF40. Los paneles H–H'' muestran las mismas señales en una vista dorsal del estadio NF40. Post: región posterior del embrión; Ant: región anterior del embrión; Np: placa neural; Nc: células de la cresta neural; Nt: tubo neural; Tb: yema caudal; Ot: vesícula ótica; Ey: ojo; Ha: arcos hioideos; Ba: arcos branquiales; Pn: pronefros; Br: encéfalo; Pmnc: células de la cresta neural postmigratorias; SC: médula espinal; HB: rombencéfalo; MB: mesencéfalo; FB: prosencéfalo; Ob: bulbo olfatorio. Escala en series A, B, C, y D corresponde a 200 μ m. Escala en series E, F, G y H corresponde a 100 μ m.

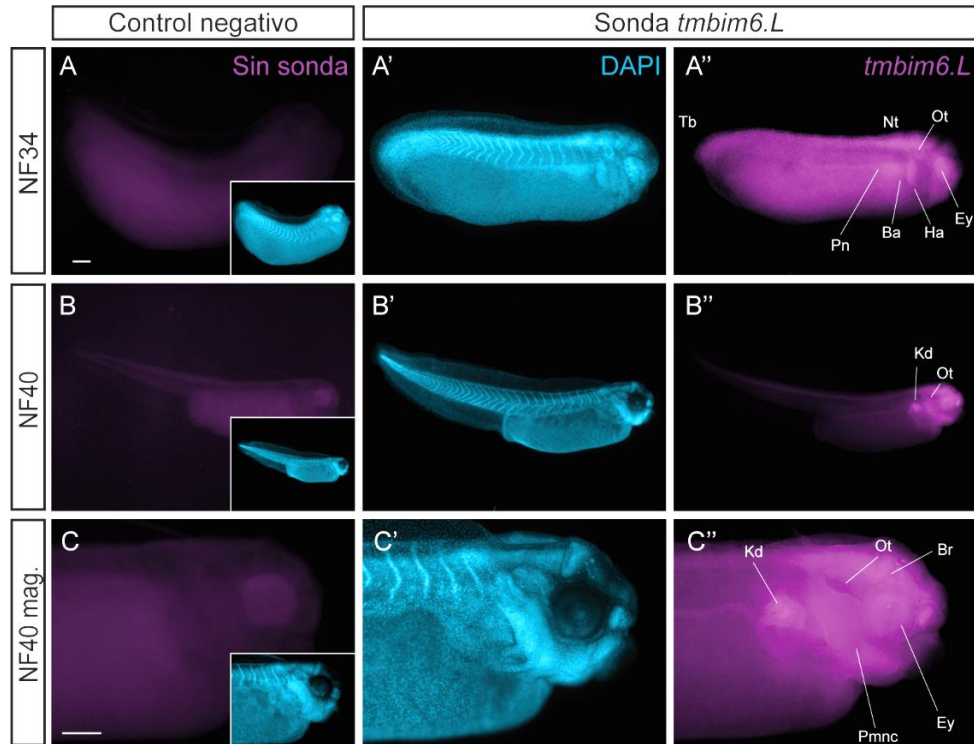


Figura 14: Hibridación *in situ* fluorescente mediante HCR para la caracterización espaciotemporal del gen *tmbim6.L*, regulado por KN9, en *Xl*. Los paneles A, A' y A'' muestran el control negativo, la tinción con DAPI y la señal de la sonda dirigida contra *tmbim6.L*, respectivamente, en el estadio NF34. Siguiendo la misma disposición, los paneles B–B'' corresponden al estadio NF40 y los paneles C–C'' muestran una magnificación del mismo estadio. Nt: tubo neural; Tb: yema caudal; Kd: riñón; Ot: vesícula ótica; Ey: ojo; Ha: arcos hioideos; Ba: arcos branquiales; Pn: pronefros; Br: encéfalo; Pmnc: células de la cresta neural postmigratorias. Escala en series A y B corresponde a 200 μ m. Escala en serie C corresponde a 200 μ m.

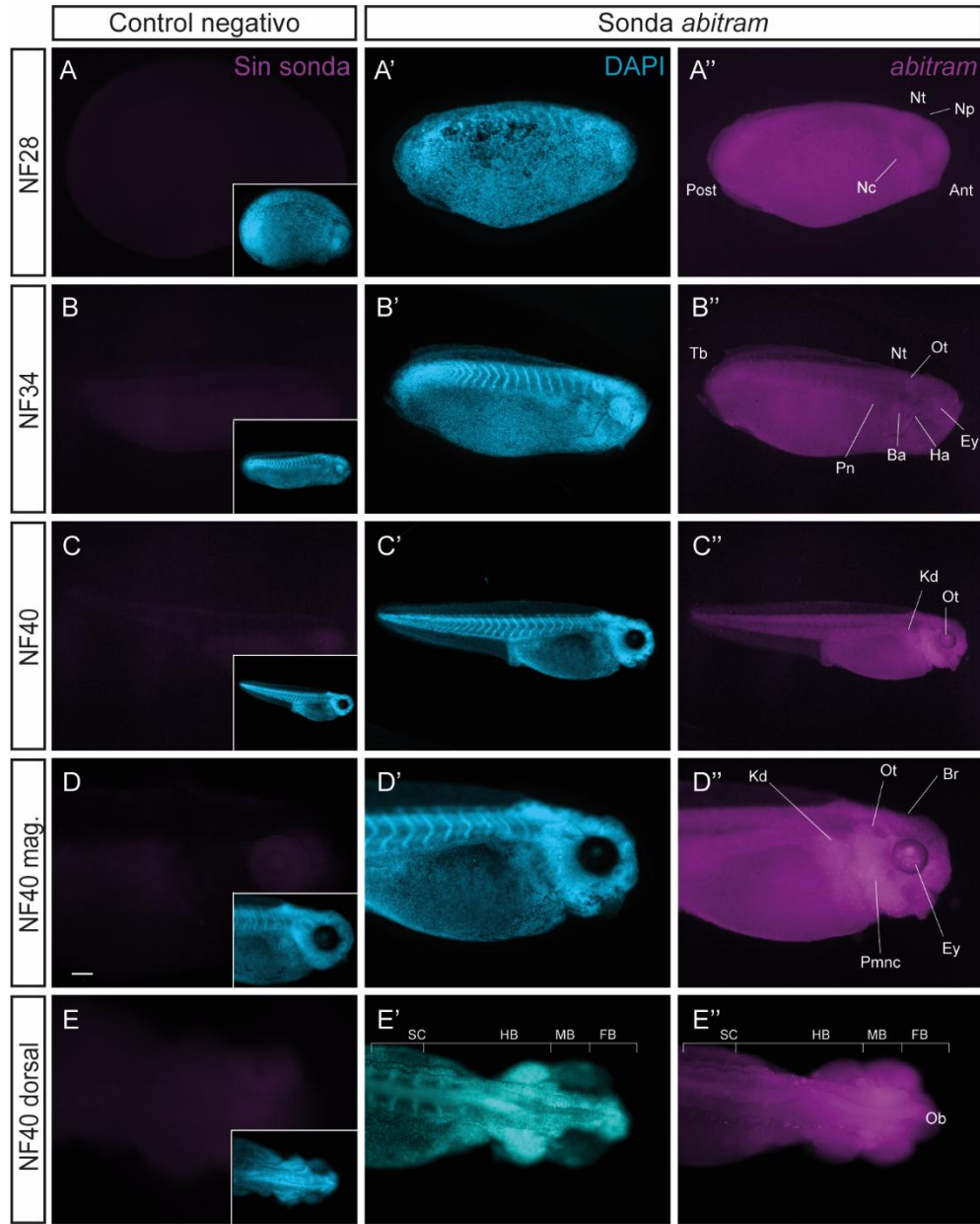


Figura 15: Hibridación *in situ* fluorescente mediante HCR para la caracterización espaciotemporal del gen *abitr*am, regulado por KN9, en *Xt*. Los paneles A, A' y A'' muestran el control negativo, la tinción con DAPI y la señal de la sonda dirigida contra *abitr*am, respectivamente, en el estadio NF28. Siguiendo la misma disposición, los paneles B–B'' y C–C'' corresponden a los estadios NF34 y NF40, respectivamente. Los paneles D–D'' muestran una magnificación del estadio NF40, mientras que los paneles E–E'' corresponden a la vista dorsal del mismo estadio. Post: región posterior del embrión; Ant: región anterior del embrión; Np: placa neural; Nc: células de la cresta neural; Nt: tubo neural; Tb: yema caudal; Kd: riñón; Ot: vesícula ótica; Ey: ojo; Ha: arcos hioideos; Ba: arcos branquiales; Pn: pronefros; Br: encéfalo; Pmnc: células de la cresta neural postmigratorias. Escala en series A, B, y C corresponde a 200 μ m. Escala en series D y E corresponde a 100 μ m.

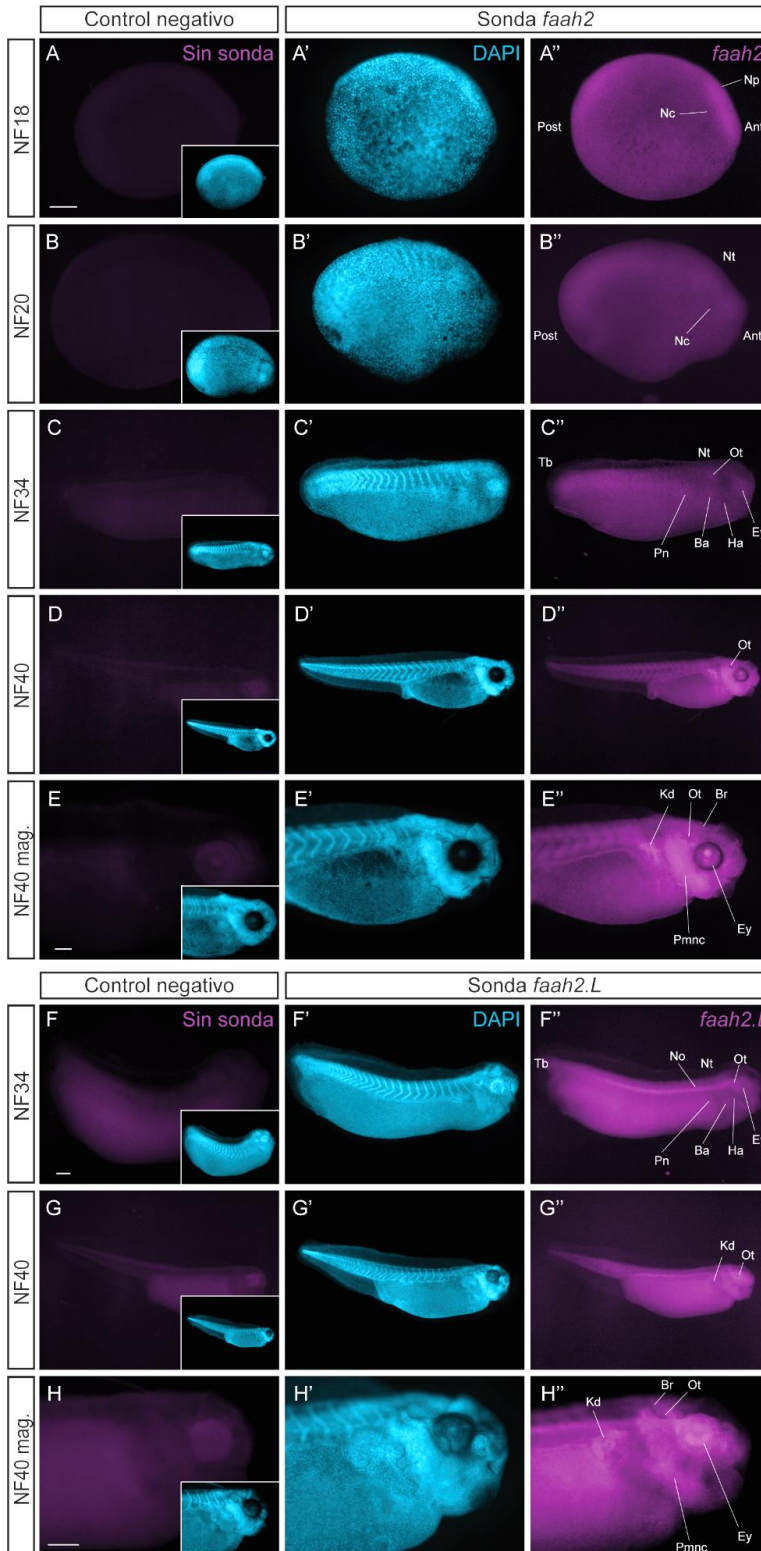


Figura 16: Hibridación *in situ* fluorescente mediante HCR para la caracterización espaciotemporal de los genes *faah2* de Xt y *faah2.L* de XI, regulados por KN9. Los paneles A, A' y A'' muestran el control negativo, la tinción con DAPI y la señal de la sonda dirigida contra *faah2*, respectivamente, en el estadio NF18 de Xt. Siguiendo la misma disposición, los paneles B–B'', C–C'' y D–D'' corresponden a los estadios NF20, NF34 y NF40, respectivamente, mientras que los paneles E–E'' muestran una magnificación del estadio NF40. En XI, los paneles F, F' y F'' muestran el control negativo, la tinción con DAPI y la señal de la sonda dirigida contra *faah2.L*, respectivamente, en el estadio NF34. Los paneles G–G'' corresponden al estadio NF40 y los paneles H–H'' muestran una magnificación del mismo estadio. Post: región posterior del embrión; Ant: región anterior del embrión; Np: placa neural; Nc: células de la cresta neural; Nt: tubo neural; Tb: yema caudal; Ot: vesícula ótica; Ey: ojo; Kd: riñón; Ha: arcos hioideos; Ba: arcos branquiales; Pn: pronefros; Br: encéfalo; Pmnc: células de la cresta neural postmigratorias; No: notocorda. Escala en series A, B, C, y D corresponde a 200 μ m. Escala en serie E corresponde a 100 μ m. Escala en series F y G corresponde a 200 μ m. Escala en serie H corresponde a 200 μ m.

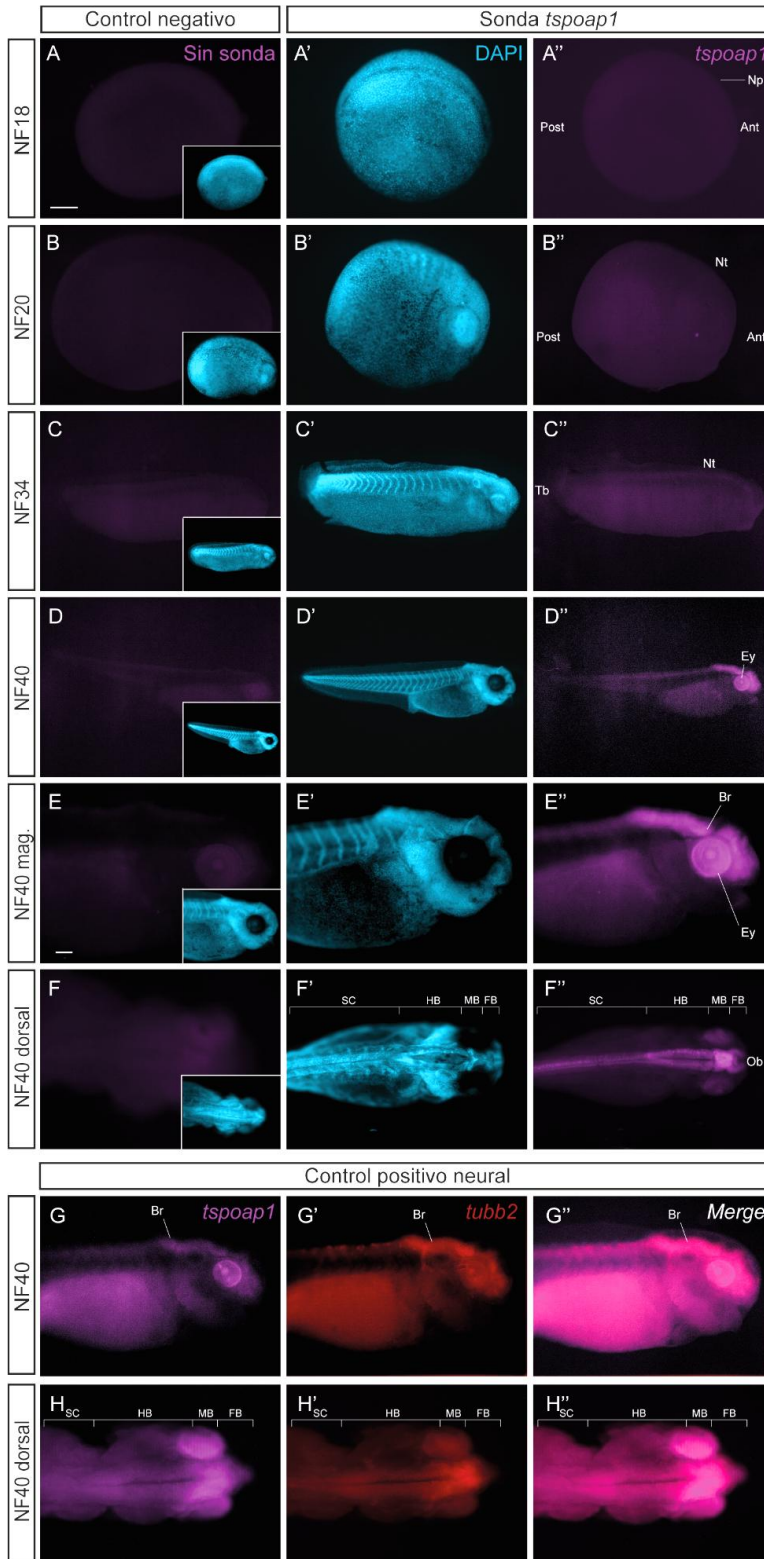


Figura 17: Hibridación *in situ* fluorescente mediante HCR para la caracterización espaciotemporal del gen *tspoap1*, regulado por KN9, en *Xt*. Los paneles A, A' y A'' muestran el control negativo, la tinción con DAPI y la señal de la sonda dirigida contra *tspoap1*, respectivamente, en el estadio NF18. Siguiendo la misma disposición, los paneles B–B'', C–C'' y D–D'' corresponden a los estadios NF20, NF34 y NF40, respectivamente. Los paneles E–E'' muestran una magnificación del estadio NF40 y los paneles F–F'', una vista dorsal del mismo estadio. Los paneles G, G' y G'' muestran la señal de la sonda dirigida contra *tspoap1*, la señal del control neural *tubb2* y la combinación de ambos canales, respectivamente, en el estadio NF40. Los paneles H–H'' muestran las mismas señales en una vista dorsal del estadio NF40. Post: región posterior del embrión; Ant: región anterior del embrión; Np: placa neural; Nc: células de la cresta neural; Nt: tubo neural; Tb: yema caudal; Ot: vesícula ótica; Ey: ojo; Ha: arcos hioideos; Ba: arcos branquiales; Pn: pronefros; Br: encéfalo; Pmnc: células de la cresta neural postmigratorias; SC: médula espinal; HB: rombencéfalo; MB: mesencéfalo; FB: prosencéfalo; Ob: bulbo olfatorio. Escala en series A, B, C, y D corresponde a 200 μ m. Escala en series E, F, G y H corresponde a 100 μ m.

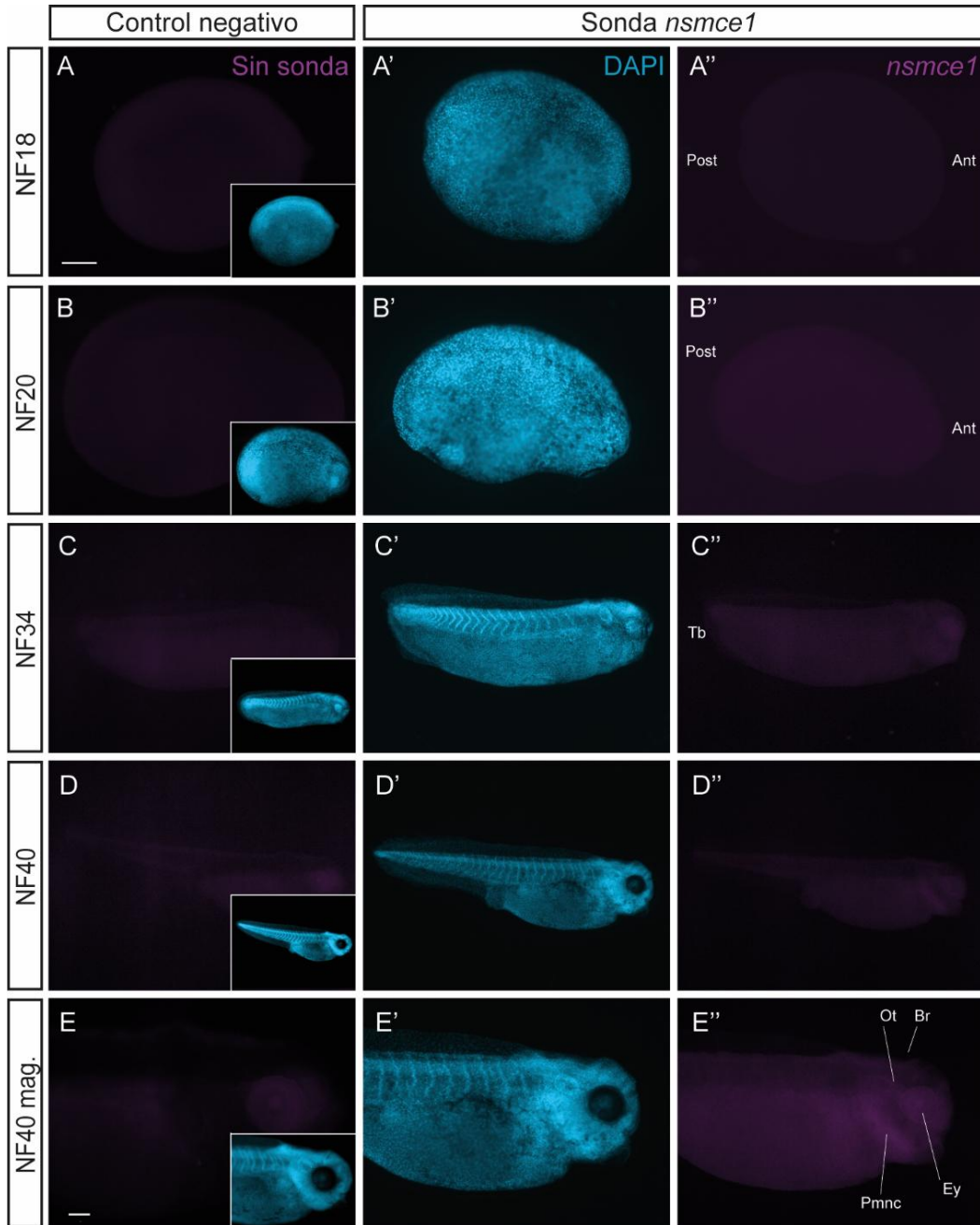


Figura 18: Hibridación in situ fluorescente mediante HCR para la caracterización espaciotemporal del gen *nsmce1*, regulado por KN9, en *Xt*. Los paneles A, A' y A'' muestran el control negativo, la tinción con DAPI y la señal de la sonda dirigida contra *nsmce1*, respectivamente, en el estadio NF18. Siguiendo la misma disposición, los paneles B–B'', C–C'' y D–D'' corresponden a los estadios NF20, NF34 y NF40, respectivamente. Los paneles E–E'' muestran una magnificación del estadio NF40. Post: región posterior del embrión; Ant: región anterior del embrión; Ot: vesícula ótica; Ey: ojo; Br: encéfalo; Pmnc: células de la cresta neural postmigratorias. Escala en series A, B, C, y D corresponde a 200 μ m. Escala en serie E corresponde a 100 μ m.

3.5. Ensayos reporteros para corroborar la actividad promotora de KN9.

Para evaluar la actividad promotora de KN9, se diseñó un vector para ensayos de transgénesis mediante integración mediada por enzimas de restricción (REMI; del inglés *restriction enzyme-mediated integration*). Para ello, se utilizó el vector presentado en la **Figura 19A**, que contiene un control interno de mCherry bajo el control del promotor ubicuo del citomegalovirus (CMV). En el constructo experimental, la secuencia de KN9 que actúa como promotor del gen *tmbim6* en *Xt* se ubicó río arriba del gen reportero que codifica GFP nuclear (**Figura 19B**). Como control, se utilizó un vector equivalente que carece de la secuencia de KN9 río arriba de GFP nuclear (**Figura 19C**).

En la **Figura 20** se presentan imágenes representativas de los resultados obtenidos mediante transgénesis REMI. En el primer conjunto de imágenes, se muestran dos larvas transgénicas en estadio NF40 portadoras del vector KN9 (**A-A'' y B-B''**), una larva transgénica portadora del vector control (**C-C''**) y una larva silvestre no transgénica (**D-D''**). En el segundo conjunto, se muestran vistas dorsales de larvas en estadio NF44 portadoras del vector KN9 (**E-E''**), del vector control (**G-G''**) y del control silvestre (**I-I''**), junto con sus respectivas vistas ventrales (**F-F''**, **H-H'' y J-J''**). Finalmente, se presentan magnificaciones del encéfalo y del epitelio olfatorio de las larvas portadoras del vector KN9 (**K-K'' y L-L''**), del vector control (**M-M'' y N-N''**) y del control silvestre (**O-O'' y P-P''**), respectivamente. Las larvas portadoras del constructo experimental presentaron señal de GFP nuclear en los estadios NF40 y NF44, particularmente en el encéfalo y el epitelio olfatorio. En contraste, no se detectó señal equivalente en las larvas portadoras del vector control ni en las larvas silvestres no transgénicas. La expresión de mCherry permitió identificar las larvas transgénicas y controlar la incorporación y expresión del constructo. En conjunto, estos resultados indican que el fragmento de KN9 asociado al promotor de *tmbim6* de *Xt* es suficiente para dirigir la expresión de un gen reportero *in vivo* en larvas de *Xl*.

En conjunto, la expresión de GFP nuclear observada en las larvas portadoras del vector KN9 respalda que la secuencia derivada de este elemento posee actividad promotora *in vivo* durante el desarrollo embrionario de *Xenopus*.

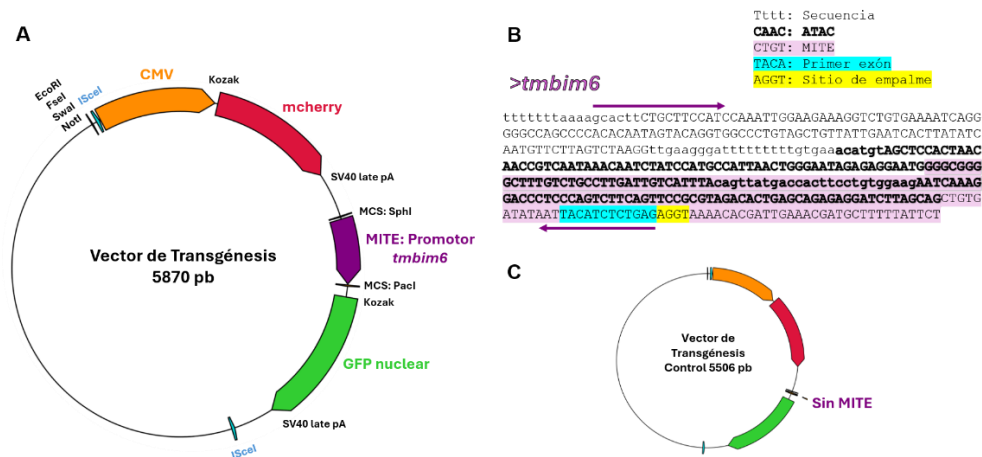


Figura 19: Vector utilizado en transgénesis mediada por enzima de restricción en *Xl*. (A) Vector MITE de transgénesis con presencia de mCherry como control interno de transgénesis y la secuencia de KN9 domesticado en el promotor de *tmbim6* río arriba de un gen reportero GFP nuclear. Se indican adicionalmente sitio de corte de enzimas de restricción y meganucleasa. (B) Secuencia de KN9 clonada en el vector de transgénesis. Las flechas indican la región insertada, la secuencia en **negrita** indica la señal del ATAC, el subrayado lila a KN9, el celeste el primer exón presente en KN9, y en amarillo el sitio putativo conservado de empalme. (C) Vector control, sin presencia de KN9 río arriba de GFP nuclear, solo el promotor CMV río arriba de mCherry.

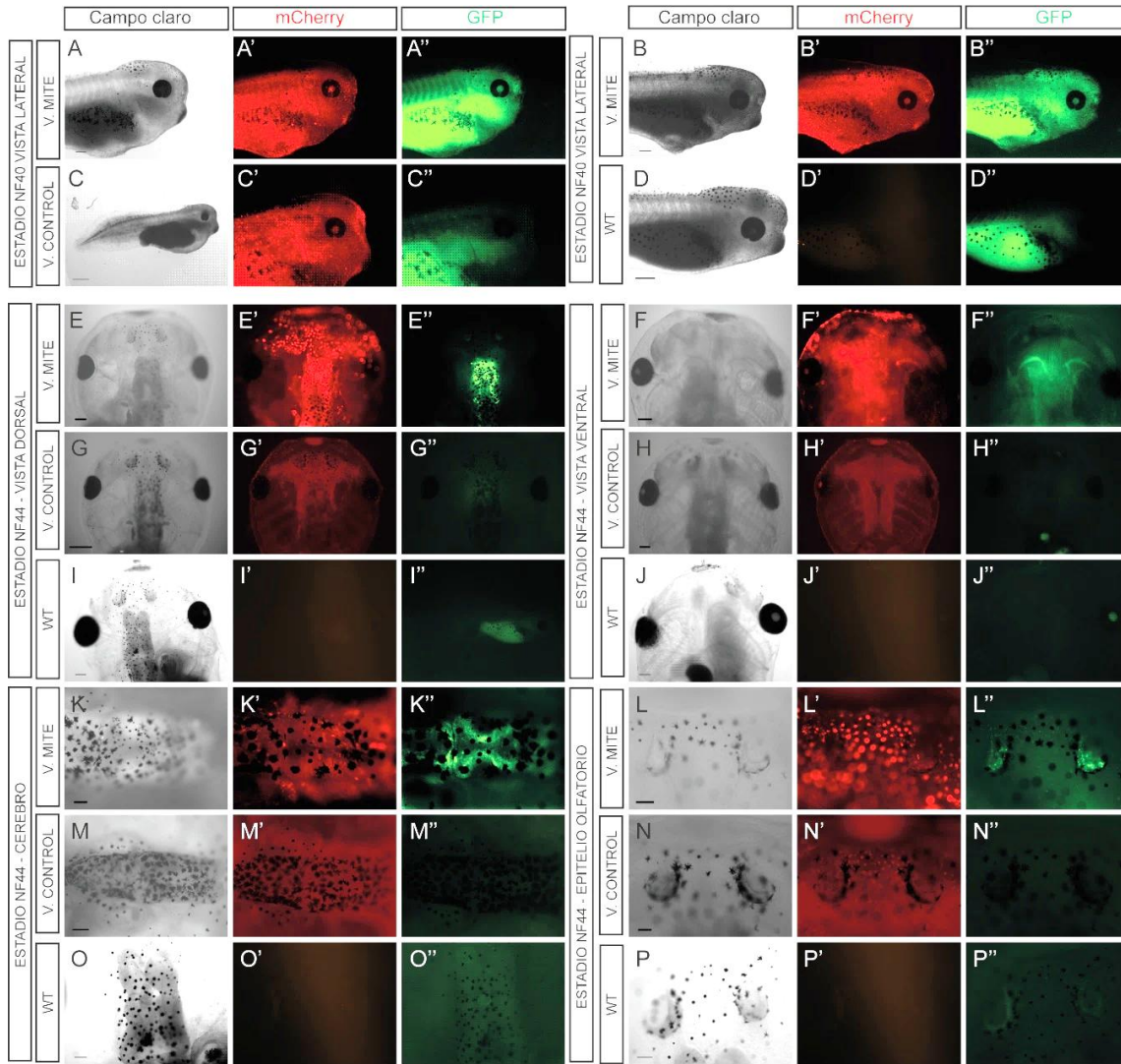


Figura 20: Evaluación de la actividad promotora de KN9 mediante transgénesis REMI en larvas de *Xenopus*. (A–A" y B–B") Larvas transgénicas en estadio NF40 portadoras del vector KN9, en el cual la secuencia promotora derivada de este elemento controla la expresión de GFP nuclear. (C–C") Larva en estadio NF40 portadora del vector control, carente de la secuencia de KN9 río arriba de GFP nuclear. (D–D") Larva silvestre no transgénica en estadio NF40. (E–E", G–G" e I–I") Vistas dorsales de larvas en estadio NF44 portadoras del vector KN9, del vector control y del control silvestre, respectivamente. (F–F", H–H" y J–J") Vistas ventrales de las mismas larvas. (K–K" y L–L") Magnificaciones del encéfalo y del epitelio olfatorio, respectivamente, de la larva portadora del vector KN9. (M–M" y N–N") Regiones equivalentes en la larva portadora del vector control. (O–O" y P–P") Regiones equivalentes en la larva silvestre no transgénica. En cada serie se muestran, de izquierda a derecha, campo claro, fluorescencia de mCherry y fluorescencia de GFP nuclear. La expresión de mCherry, regulada por el promotor CMV, se utilizó como control interno de transgénesis, mientras que la señal de GFP nuclear indica la actividad promotora de KN9. NF: estadio de Nieuwkoop y Faber. Escala en todas las series corresponde a 200 μ m.

4. DISCUSIÓN

Los resultados de esta tesis muestran que los elementos transponibles han contribuido al paisaje regulatorio de *Xenopus* en distintas escalas evolutivas. Esta contribución se examinó desde dos perspectivas complementarias: la asociación de diferentes subfamilias de TEs con regiones de cromatina abierta específicas de tejido y la domesticación evolutivamente conservada de KN9 en promotores de genes. En las secciones siguientes se discuten estos resultados.

4.1. Transposones asociados a regiones de cromatina abierta.

La integración de datos de ATAC-seq y RNA-seq permitió construir, para *Xt* y *Xl*, un catálogo de promotores candidatos y posibles elementos *cis*-regulatorios distales en hígado, mesénquima y osteoblastos. Además, se definió operacionalmente como “ubicuo” el conjunto de regiones que presentaron accesibilidad compartida entre los tres tejidos analizados. Por consiguiente, en esta tesis el término “ubicuo” se refiere exclusivamente a la accesibilidad detectada en estos tejidos y no implica actividad en todos los tejidos o estadios del organismo.

El análisis partió de datos generados en un trabajo previo del laboratorio, en el cual se estableció la metodología para identificar regiones de cromatina abierta en *Xt* (29). Posteriormente, este catálogo regulatorio sirvió de base para el trabajo derivado de la presente tesis, en el que se caracterizó la asociación de MITEs con elementos *cis*-regulatorios distales de hígado y se evaluó su posible cooptación regulatoria (126). Estrategias comparables, basadas en la integración de perfiles de accesibilidad, transcripción, marcas epigenéticas y anotaciones de TEs, han sido aplicadas durante el desarrollo de ratón (127), en pez cebra (128) y en hígado y cerebro de salmón Atlántico (129). En conjunto, estos estudios respaldan el uso de esta estrategia para identificar y priorizar copias de TEs candidatas a cooptación *cis*-regulatoria.

4.2. Enriquecimiento de transposones de ADN en regiones regulatorias de *Xt* y *Xl*.

La asociación de transposones de ADN con regiones regulatorias presentó diferencias entre ambas especies. *Xl* mostró un repertorio más amplio, particularmente en elementos distales hepáticos, mientras que KN9 se encontró enriquecido en promotores del clúster ubicuo de ambas especies (**Figuras 3 y 4**). Estos resultados sugieren la coexistencia de contribuciones regulatorias potencialmente ancestrales, como la representada por KN9, y asociaciones más recientes o específicas de linaje en elementos distales.

En *Xt*, las tres subfamilias de MITEs enriquecidas en regiones distales hepáticas presentaron motivos putativos de unión a FTs relacionados con la regulación del hígado, la hematopoyesis y la respuesta inmunitaria. Además, dos de ellas se localizaron con mayor frecuencia cerca de genes con expresión específica de hígado (126). Estas asociaciones respaldan su identificación como candidatos a

elementos *cis*-regulatorios, aunque no demuestran directamente su actividad como potenciadores ni permiten asignar un gen diana a cada copia.

La presencia de motivos relacionados con la hematopoyesis y la inmunidad resulta biológicamente plausible, dado que el hígado larval de *Xenopus* contiene poblaciones eritroides, linfoides y mieloides y mantiene actividad hematopoyética durante el desarrollo (130–133). Sin embargo, este antecedente también introduce una limitación interpretativa: los perfiles obtenidos a partir de hígado completo integran señales procedentes de hepatocitos y de células hematopoyéticas e inmunitarias. Por ello, el enriquecimiento de motivos inmunitarios podría reflejar que una fracción de las regiones accesibles corresponde a estas últimas poblaciones y no necesariamente a hepatocitos. Esta interpretación es coherente con la cooptación previamente documentada de secuencias derivadas de TEs en la regulación de genes de respuesta inmunitaria (99).

Harbinger-N12A_*Xt* constituye un caso particularmente informativo, ya que se encontró enriquecido en regiones hepáticas de ambas especies, pero la mayoría de las inserciones analizadas en *Xt* no se conserva en posiciones ortólogas de *Xl* (126). Por lo tanto, el enriquecimiento compartido a nivel de subfamilia no implica la conservación de las mismas copias y es compatible con el reclutamiento regulatorio paralelo en ambos linajes.

Finalmente, se evaluó la presencia de los TEs específicos de cada especie en la contraria para corroborar que corresponden efectivamente a inserciones linaje específicas en *Xt* (**Anexo 2**) y *Xl* (**Anexo 3**). Este análisis comparativo revela que el enriquecimiento del MITE Kolobok-N3_XT en potenciadores específicos de hueso solo se encuentra presente en *Xl* y no en *Xt*. Dicho hallazgo no solamente demuestra que eventos de domesticación son sumamente dinámicos, sino que además, es de suma relevancia para nuestro laboratorio que ha consolidado una línea de investigación enfocada en el control transcripcional de la osificación en *Xenopus* (29).

4.3. Enriquecimiento de retrotransposones en regiones regulatorias de *Xl*.

En *Xl* también se identificaron retrotransposones enriquecidos en elementos *cis*-regulatorios distales de hígado, correspondientes a una subfamilia Gypsy y seis subfamilias del orden DIRS (**Figura 5**). Aunque ambos grupos pertenecen a la clase I de elementos transponibles, presentan arquitecturas y mecanismos de integración diferentes. Gypsy pertenece al orden de los retrotransposones LTR, cuya integración depende de una integrasa de tipo DDE. En contraste, los elementos DIRS carecen de esta integrasa y codifican una tirosina recombinasa asociada con su incorporación al genoma (134,135). El enriquecimiento de subfamilias DIRS únicamente en las regiones regulatorias analizadas de *Xl* nos podría indicar una dinámica evolutiva distinta de estos elementos en ambas especies. La literatura apunta a que en el linaje de *Xl*, ocurrieron dos ondas de amplificación de este orden

de retrotransposones hace 60-80 y 110-135 millones de años, la cuales no ocurrieron en el linaje de *Xt* (136).

4.4. Conservación ancestral del MITE KN9 en regiones promotoras en *Xenopus*.

La superfamilia Kolobok se encuentra ampliamente distribuida entre animales, hongos, plantas y protistas. Sus elementos autónomos codifican una transposasa con un núcleo catalítico DDD/E, cuya actividad genera duplicaciones de cuatro pares de bases (TTAA) en el sitio de inserción (137). En los genomas de *Xt* y *Xl* se han identificado candidatos a transposasas Kolobok que conservan motivos catalíticos y de unión al ADN (138). Debido a que los MITEs no codifican su propia transposasa, su movilización requiere que una transposasa compatible sea proporcionada en trans por un elemento autónomo. Sin embargo, todavía no se ha demostrado experimentalmente qué transposasa moviliza cada subfamilia de Kolobok en *Xenopus*.

La subfamilia KN9 fue descrita anteriormente como la subfamilia K de los T2-MITEs en *Xt* por Hikosaka y Kawahara (117) (**Anexo 5**). Este estudio indica que no ha existido una amplificación reciente de esta subfamilia, lo cual es compatible con la evidencia obtenida en esta tesis identificando inserciones de KN9 en posiciones ortólogas entre *Xt* y *Xl* (**Figura 9; Figura 11; Anexo 1; Anexo 4**). Este patrón apoya que dichas inserciones ya estaban presentes en el ancestro común de ambos linajes, antes de su divergencia, estimada en aproximadamente 48 millones de años (102).

Xt y *Xl* pertenecen, respectivamente, a los subgéneros hermanos *Silurana* y *Xenopus*, ambos perteneciente al grupo de ranas Pipidae (139,140). Por ello, la conservación de las inserciones de KN9 en ambas especies permite predecir que algunas de ellas también podrían encontrarse en posiciones ortólogas de otras especies del género (141). La elevada abundancia de KN9 tampoco parece ser una característica general de los genomas de anfibios. En contraste con *Xenopus*, el genoma de la rana tibetana *Nanorana parkeri* presenta una mayor contribución de retrotransposones LTR, y su anotación publicada contiene solamente nueve copias asignadas a KN9, frente a las 1.726 identificadas en *Xt* (142) (**Anexo 6**). Esta diferencia es compatible con una amplificación diferencial de KN9 en el linaje de *Xenopus*, sin embargo, aún no existe la suficiente data de secuenciación en distintos géneros de anfibios que nos pueda responder esta interrogante.

4.5. Escenarios evolutivos de inserción y domesticación de KN9 como promotores.

Los datos obtenidos en esta tesis (**Figura 7; Anexo 4**) sugieren una transición evolutiva en la que: (1) una copia de KN9 con potencial regulatorio se insertó en el microambiente molecular de un promotor; (2) en algún momento de la evolución, KN9 y el promotor original regularon conjuntamente la expresión del gen diana; y (3) el promotor originado a partir de KN9 prevaleció sobre el promotor original y se

conservó a lo largo de la evolución. De cumplirse esta hipótesis, KN9 podría representar un caso particularmente interesante en el que una secuencia móvil fue incorporada como un núcleo promotor estable.

Se identificaron varios casos interesantes de domesticación: (1) la domesticación de KN9 como promotor de los genes homólogos *borc6* y *borcs7* (**Anexo 7**), cuya función involucra el transporte de lisosomas a través de los microtúbulos (143). Estos genes se encuentran ubicados en cromosomas distintos, lo que, junto con los resultados de la **Figura 10**, sugiere una posible preferencia de KN9 por la regulación de genes con funciones celulares comunes; (2) la domesticación de KN9 en los promotores de genes vecinos, ejemplificada en el **Anexo 8** por *atg101* y *tmbim6*, ambos asociados con funciones de autofagia celular (144,145). En la **Figura 9** se presentan casos similares. En ellos, la proximidad de las copias de KN9 asociadas con ambos genes sugiere un origen común del MITE y aporta indicios de una posible duplicación de KN9, fenómeno descrito para otras subfamilias de transposones de ADN (146,147); y (3) la domesticación de KN9 como componente de promotores bidireccionales (**Anexo 9**), que resulta particularmente interesante, ya que una misma copia podría incorporarse a una arquitectura regulatoria compartida por genes transcritos en orientaciones opuestas. Existen antecedentes en humanos, para un retrotransposon LTR derivado de ERV1, el cual funciona como promotor bidireccional (148).

4.6. Elementos reguladores putativos identificados en las secuencias de las copias domesticadas de KN9.

El análisis exhaustivo de las secuencias de las copias domesticadas de KN9 (**Figura 8A-B**) reveló una preferencia por la domesticación de copias que solo presentaban la región 3' de la secuencia consenso. Al analizar las copias no domesticadas de KN9 en *Xt*, se observó que estas se distribuyen de manera homogénea a lo largo del consenso. En cambio, en *Xl*, las copias no domesticadas también se agrupan en la región 3' del consenso en ambos subgenomas (**Anexo 10**). Esta distribución impide concluir que la localización en el extremo 3' sea una característica exclusiva de las copias domesticadas. Por esta razón, para comprender qué ocurre a nivel de secuencia en esta región de KN9, se realizó el análisis presentado en el **Anexo 11**. Este análisis, junto con el alineamiento múltiple de secuencias (**Figura 8C**), mostró que las copias domesticadas presentan una mayor conservación de los motivos de unión a FTs de las familias ETS, PBC/MEINOX y SP/KLF respecto a la secuencia consenso, en comparación con las copias no domesticadas de KN9. Estos resultados sugieren que la domesticación de los elementos KN9 podría depender de características asociadas con estos motivos. Sin embargo, su relevancia debe comprobarse experimentalmente, por ejemplo, mediante ensayos reporteros y la eliminación o mutación de los sitios de unión, con el propósito de determinar cómo estas alteraciones modifican la actividad reguladora de KN9.

La familia ETS comprende un amplio repertorio de FTs, con 28 miembros identificados en humanos, caracterizados por poseer un dominio conservado de

unión al ADN. Además, se ha demostrado que la delección de diferentes miembros de esta familia en ratón provoca alteraciones inmunitarias y neurales, entre otras (149). Los defectos neurales descritos constituyen un antecedente relevante para esta tesis, ya que, como se discutirá en la **sección 4.7**, los genes regulados por KN9 presentan patrones de expresión y de actividad reguladora predominantemente neurales (**Figuras 12-17; 20**). En concordancia con lo anterior, *nsmce1*, gen regulado por KN9 cuyo sitio de unión a factores ETS se encuentra completamente mutado (**Tabla 2; Figura 18**), no presentó expresión neural. Asimismo, se han descrito los patrones de expresión de los factores *ets1* y *ets2* durante el desarrollo de *XI* (150–152). Los transcritos de ambos factores se detectan en la placa neural, las células de la cresta neural y sus derivados, la yema caudal y los arcos branquiales. Adicionalmente, *ets2* se expresa en el pronefros y el cerebro. Esto coincide parcialmente con los patrones espaciotemporales observados para los genes regulados por KN9 (**Figuras 12-17**). Otro miembro de la familia ETS es GABPA, un FT de expresión ubicua, lo que concuerda con la identificación de KN9 en el clúster de promotores ubicuos (**Figuras 2-4**). Recientemente, se publicaron dos preprints en células madre embrionarias de ratón que sugieren que factores de la familia ETS, entre ellos GABPA y ETS1, pueden actuar como reguladores dominantes de determinados promotores asociados con islas CpG, promoviendo la apertura de la cromatina y contribuyendo a definir la posición del TSS (25,26). En conjunto, estos antecedentes sugieren que los motivos de unión a factores ETS podrían contribuir a la actividad reguladora de KN9 y a los patrones de expresión neural asociados con sus copias domesticadas.

Dentro de la familia de FTs SP/KLF existen tanto activadores como represores transcripcionales, cuya actividad depende del contexto celular (153,154). Durante el desarrollo de *Xenopus* se han caracterizado los patrones de expresión de varios miembros de esta familia (155,156). Por ejemplo, *klf2* se expresa en la yema caudal; *klf4*, en los pliegues neurales; *klf5*, en el prosencéfalo y la vesícula ótica; *klf6*, en la placa neural; y *klf7*, en el cerebro, entre otros territorios. Estos patrones coinciden parcialmente con los datos de expresión espaciotemporal obtenidos para los genes regulados por KN9 (**Figuras 12-17**). Sin embargo, resulta difícil determinar cómo las distintas mutaciones presentes en los motivos de unión a factores SP/KLF de los genes candidatos afectan sus patrones de expresión espaciotemporal (**Tabla 2; Figura 11**). Es posible que diferentes miembros de esta familia, tanto activadores como represores, actúen de manera diferencial según la secuencia exacta del motivo en cada copia de KN9 y el contexto celular en el que se encuentra. Una situación similar podría ocurrir con los motivos de unión a factores PBC/MEINOX identificados en KN9 (**Figura 8**). Los factores de estas familias suelen formar complejos entre sí y con proteínas HOX, lo que aumenta la especificidad de reconocimiento del ADN (157). Miembros de estas familias, como *pknox1*, *Xpbx1* y *Xmeis1*, presentan expresión en el encéfalo de *XI* (158) (*pknox1* disponible en *Xenbase*).

Adicionalmente, se analizó data de chipseq de humano obtenida de <https://chip-atlas.org/> para miembros de estos motivos de unión a FTs conservados en KN9 (**Anexo 12**). En este anexo, se observa que para los genes *rtn3*, *tmbim6*, *abitrans*, *faah2*, *tspoap1* y *nsmce1* en humanos presentan sitios de unión para ETS1, ETS2, GABPA, KLF1, KLF4, PBX1, y MEIS1. Esto sugiere que KN9 mimetiza el microambiente regulatorio presente en los genes en los que actúa como promotor, siendo necesaria esta compatibilidad para su inserción. En consecuencia, una posibilidad atractiva es que los tres tipos de motivos de unión a FTs identificados en KN9 (ETS, SP/KLF y PBC/MEINOX) actúen de manera combinatoria para determinar su actividad reguladora.

Finalmente, la presencia de un motivo AGGT conservado en la región 3' de KN9 (**Figura 8**) plantea un posible mecanismo para su incorporación al transcrito. El límite de la cobertura de ARN-seq observado dentro de KN9 coincide con la posición de este motivo (**Anexo 12**), lo que es compatible con su utilización como sitio donador de empalme 5'. Esta disposición corresponde a una unión exón–intrón canónico AGGT, en la que AG representaría el extremo del exón y GT el inicio del intrón reconocido por el espliceosoma (159). Además, se han descrito sitios con esta arquitectura que, dependiendo del contexto de secuencia y transcripcional, pueden presentar doble especificidad y ser reconocidos alternativamente como sitios donadores 5' o aceptores 3' (160,161). Presentamos evidencia directa que, en los loci de *tmbim6* y *atg101* (**Anexo 13**), el lugar de empalme de las secuencias de ARNm disponibles para ambos transcritos corresponde exactamente a los nucleótidos AG del motivo AGGT. En conjunto, esta identidad y el perfil de cobertura de ARN-seq respaldan que una porción de este MITE se incorpora al transcrito maduro como un primer exón no codificante. Según este modelo, KN9 no solo contribuiría a la actividad promotora, sino también al extremo 5' de los transcritos asociados (**Figura 11**). La incorporación de secuencias derivadas de TEs principalmente dentro de regiones UTR, constituye un fenómeno documentado en mamíferos y otros vertebrados (162,163). Una vez incorporadas al ARNm, estas secuencias también pueden adquirir funciones postranscripcionales. En mamíferos, por ejemplo, secuencias derivadas de retrotransposones presentes principalmente en regiones 3' UTR pueden aportar sitios de reconocimiento para microARN (164,165). Estos antecedentes ilustran el potencial funcional de las UTR derivadas de TEs, aunque no implican que KN9 participe necesariamente en este tipo de regulación. Esta hipótesis debería confirmarse mediante 5'-RACE, para localizar el TSS dentro de KN9, y mediante RT-PCR a través de la unión entre KN9 y el exón siguiente, seguida de la secuenciación del producto, para comprobar directamente la incorporación de KN9 al primer exón y la utilización del sitio donador propuesto.

4.7. Patrón regulatorio neural de KN9 sobre genes candidatos analizados durante el desarrollo en *Xenopus*.

El análisis conjunto de los genes candidatos (**Figura 11**), cuya expresión se evaluó mediante hibridación *in situ* fluorescente por HCR (**Figuras 12–18**), reveló una

marcada convergencia hacia patrones predominantemente neurales. No obstante, *tmbim6*, *abitram* y *faah2* (**Figuras 13, 15 y 16**) además, presentaron un componente de expresión ubicua, en concordancia con lo discutido en la sección 4.6.

El componente neural observado en *rtn3*, *tmbim6* y *abitram* (**Figuras 12–15**) resulta biológicamente coherente con las funciones descritas para las proteínas que codifican. RTN3 es una proteína del retículo endoplasmático (**Figura 10**) relacionada con la organización de membranas y la respuesta al estrés. En modelos murinos, también se ha asociado con respuestas tempranas de la glía de Müller y con efectos neuroprotectores (166,167). Por su parte, TMBIM6 regula la homeostasis del Ca^{2+} , el estrés del retículo endoplasmático, la apoptosis y la autofagia (168,169). En modelos murinos, TMBIM6 (denominado como BI-1) favorece la supervivencia y diferenciación neuronal de células madre embrionarias y protege las neuronas corticales frente al estrés del retículo endoplasmático y el daño cerebral (170,171). Además, la correspondencia espacial entre los patrones de expresión de *tmbim6* y *tmbim6.L* (**Figuras 13 y 14**) es compatible con la conservación de un sistema regulador ancestral. Finalmente, se ha descrito que Simiate, la proteína codificada por *abitram*, interactúa con la actina, modifica la dinámica de los filopodios y favorece la ramificación dendrítica en cultivos primarios de neuronas murinas (172).

El caso de *faah2* resulta particularmente interesante (**Figura 16**), ya que tanto en *Xt* (Sin KN9) como en *Xl* (Con KN9) se observó un patrón de expresión con componentes neural y ubicuo. Este resultado plantea la hipótesis de que las copias de KN9 podrían haberse insertado o retenido preferentemente en regiones promotoras de genes con estas características de expresión. Sin embargo, los datos disponibles aún no permiten sustentar esta hipótesis. En cuanto a su rol neural, en pez cebra, la inactivación de *faah2a* altera la respuesta conductual frente al estrés, mientras que, en humanos, se ha descrito una variante con pérdida parcial de función asociada con manifestaciones neurológicas y psiquiátricas (173,174).

Otro caso relevante corresponde a *tspoap1* (**Figura 17**), que presentó un patrón de expresión restringido principalmente al sistema nervioso, en concordancia con la ausencia de señal de ATAC-Seq en los tejidos no neurales evaluados (**Figura 11E**). Este gen codifica RIMBP1, una proteína que participa en la organización de complejos presinápticos y en la neurotransmisión. En humanos, la pérdida de su función se ha relacionado con alteraciones neurológicas (175).

Finalmente, no se detectó señal de HCR, neural ni ubicua, para *nsmce1* (**Figura 18**). La copia de KN9 asociada con el promotor de este gen presenta una mutación en el motivo ETS, lo que plantea la posibilidad de que este sitio contribuya a la actividad reguladora de KN9, incluido su componente neural. El gen *nsmce1* (*nse1*) no es específicamente neural, sino que codifica un componente ampliamente expresado, implicado en la progresión de la replicación, la reparación del ADN y el mantenimiento de la estabilidad cromosómica (176,177). Este perfil funcional es

concordante con la accesibilidad ubicua observada mediante ATAC-Seq en los tejidos evaluados (**Figura 11F**), aunque la accesibilidad de la cromatina no implica necesariamente una expresión ubicua detectable mediante HCR en los estadios analizados.

4.8. Validación funcional *in vivo* de la actividad promotora de KN9.

El uso exclusivo de datos genómicos y del análisis de los patrones de expresión de genes asociados a KN9 no demuestra por sí solo su domesticación como promotor. Por ello, uno de los principales aportes de este trabajo consiste en integrar la identificación y caracterización de KN9 con ensayos funcionales que respaldan experimentalmente su actividad promotora.

Mediante transgénesis REMI, se obtuvo evidencia funcional de que el fragmento de KN9 asociado a *tmbim6* es suficiente para dirigir *in vivo* la expresión de un gen reportero que codifica GFP nuclear en el encéfalo y el epitelio olfatorio de larvas transgénicas de *Xl* (**Figura 20**). La ausencia de una señal equivalente en el constructo control, que carecía de KN9 río arriba de GFP (**Figura 19**), y en las larvas silvestres respalda que, en el contexto del ensayo, la actividad observada depende del fragmento de KN9. Además, el patrón de expresión del reportero reproduce el componente neural observado en los patrones de expresión de *tmbim6* y *tmbim6.L* (**Figuras 13-14**). REMI y las variantes de transgénesis basadas en la transferencia de núcleos espermáticos se han utilizado ampliamente en *Xenopus* para evaluar *in vivo* la suficiencia de promotores y enhancers, así como para identificar motivos *cis*-regulatorios funcionales mediante delecciones y mutaciones dirigidas (178–181).

La elaboración de constructos reporteros con mutaciones específicas en los motivos ETS, SP/KLF y PBC/MEINOX permitiría determinar la contribución individual de cada uno a la actividad promotora de KN9. De manera complementaria, la eliminación de KN9 y la inactivación independiente del marco de lectura de *tmbim6* mediante CRISPR, seguidas de la comparación de los fenotipos resultantes, permitirían evaluar si la pérdida del elemento regulador reproduce parcial o totalmente los efectos de la pérdida del gen. Esta aproximación contribuiría a establecer la relevancia funcional de estos eventos ancestrales de domesticación en *Xenopus*.

4.9. Modelo integrativo sobre el mecanismo de acción de KN9.

En su conjunto, los resultados de esta tesis nos permiten proponer un modelo evolutivo para el mecanismo de acción de KN9, enfocado en la fuerte presión de selección que actúa en distintas etapas del proceso de domesticación (**Figura 21**). En primer lugar, planteamos que KN9 presenta características de secuencias excepcionales que lo diferencian de otros TEs y que explican porque solo esta subfamilia ha sido domesticada en promotores en *Xt* y *Xl* (**Figura 21A**). Estas secuencias incluyen, como elementos fundamentales, un sitio ETS (25,26), una región propicia para el inicio de transcripción (**Anexo 13**) y un sitio dador de empalme que permita “capturar” el segundo exón de genes del huésped (**Figura**

21A). Según nuestro modelo, una vez KN9 es amplificado, existen dos condiciones que se deben cumplir para que exista una posibilidad de domesticación: que la copia de KN9 se inserte en la cercanía de un promotor endógeno (**Figura 21B**) y, además, en la orientación sentido del gen blanco (**Figura 21C**). Sin embargo, una vez un gen es capturado, otros criterios dictan si KN9 será efectivamente domesticado o si será descartado por selección natural. Por ejemplo, argumentamos que la captura de un gen cuyo codón ATG de inicio de traducción se ubica en el primer exón sería perjudicial para el organismo, dado que el transcrito resultante carecería de un marco de lectura integro (**Figura 21D**). Adicionalmente, tanto la expresión como la función de los genes capturados deben ser compatibles con la actividad transcripcional de KN9, la cual, como hemos visto, es principalmente ubicua con un componente neural pronunciado (**Figura 21E**). De esta forma, se espera que genes altamente especializados, como genes estrechamente relacionados con el desarrollo (por ejemplo, factores de transcripción, componentes de vías de señalización) o con fenotipos celulares específicos (por ejemplo, colágeno, enzimas digestivas, hormonas, etc.) causen graves defectos en el organismo y sean eliminados por un proceso de selección natural (**Figura 21E**). Si bien el número de copias domesticadas puede parecer bajo, proponemos que estos fuertes criterios de selección explican de forma satisfactoria porque la mayoría de las copias KN9 no fueron domesticadas, o lograron capturar genes y luego fueron desechadas por su incompatibilidad por la biología del organismo.

Una vez un gen es capturado, nuestros análisis demuestran que el promotor original tiende a ser eliminado, por lo que, al parecer el uso conjunto de dos promotores alternativos representaría un intermediario evolutivo poco favorable (**Figura 21F**). Finalmente, hemos visto que los KN9 domesticados son sumamente estables y tienden a conservarse en ambas especies de *Xenopus* examinadas (**Figura 21G**). Sin embargo, en algunos casos, se observan grandes modificaciones funcionales en ciertas copias KN9 que adquirieron características de promotores bidireccionales o de promotores altamente tejido-específicos (**Figura 21G**). Esta última observación probablemente tiene implicancias profundas para la biología evolutiva de elementos transponibles, según la cual, un bloque molecular conservado (el KN9) representa un sustrato moldeable que permita producir variabilidad fenotípica y diversidad biológica en organismos vertebrados.

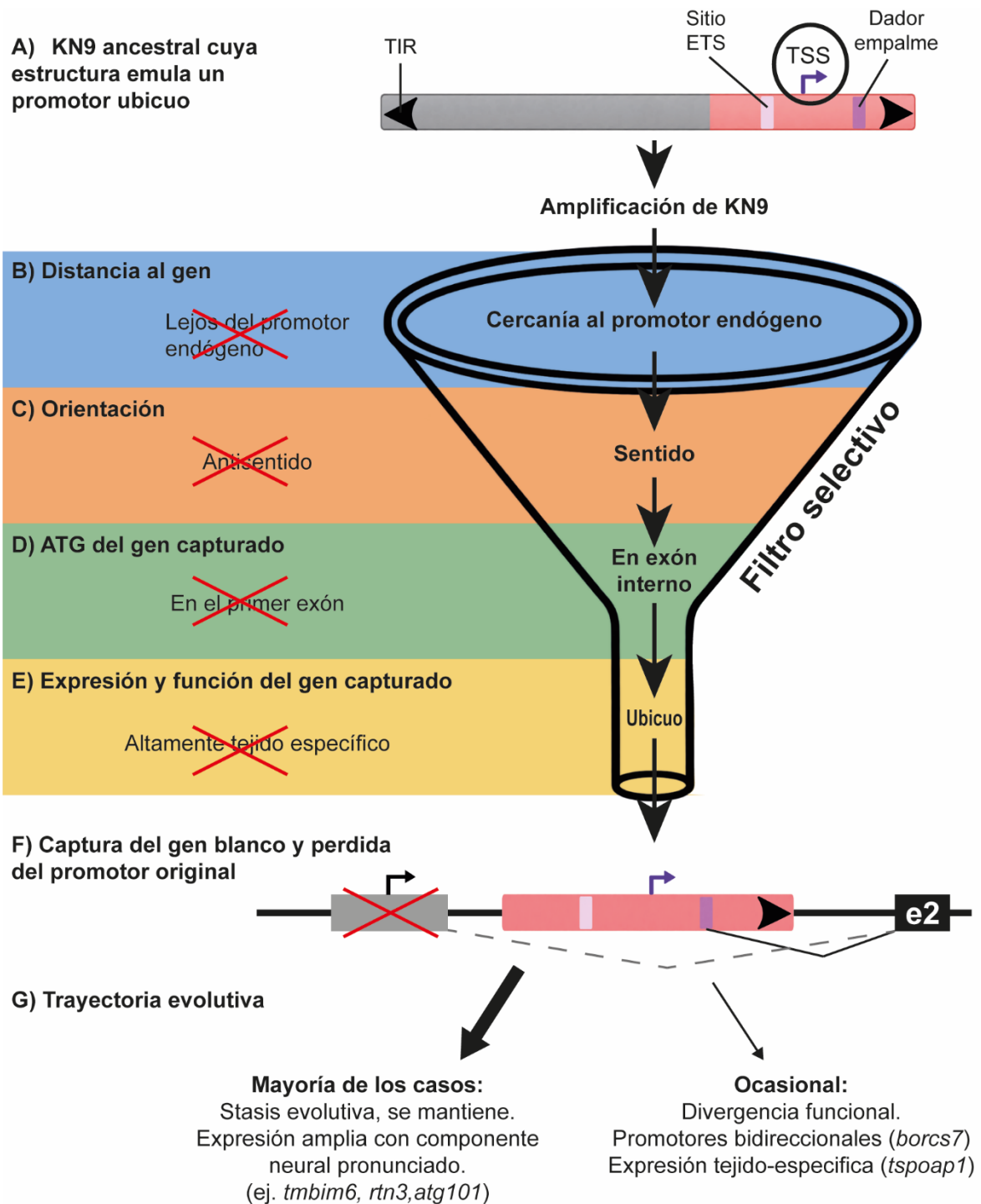


Figura 21: Modelo propuesto para el mecanismo de acción de KN9. Ver el texto principal para las explicaciones detalladas.

5. CONCLUSIONES

1. La integración de datos de ATAC-Seq, RNA-Seq y Repeatmasker permitió identificar asociaciones específicas entre subfamilias de TEs y promotores o elementos *cis*-regulatorios distales en *Xt* y *Xl*.
2. Se identificaron cinco subfamilias de transposones de ADN enriquecidas en regiones regulatorias de *Xt* y ocho en *Xl*. En *Xl* también se detectaron seis subfamilias DIRS y una Gypsy asociadas con elementos distales hepáticos.
3. KN9 representa una contribución regulatoria ancestral compartida por *Xt*, *Xl* y *X/S*.
4. La conservación de motivos putativos de unión para factores ETS, SP/KLF y PBC/MEINOX sugiere una arquitectura regulatoria combinatoria. El motivo AGGT y la correspondencia de algunas copias con regiones 5' UTR apoyan además una posible contribución de KN9 al primer exón no codificante.
5. Los genes candidatos asociados con KN9 mostraron una convergencia predominante hacia patrones de expresión neurales. El ensayo mediante transgénesis REMI demostró que el fragmento asociado con *tmbim6* es suficiente para dirigir actividad promotora neural *in vivo*.

En conjunto, los datos presentados durante esta tesis confirman la hipótesis.

7. ANEXOS

Anexo 1: Listado de genes bajo regulación de MITes KN9 domesticados en promotores ubicuos. En amarillo se identifica dentro de que especie y sub-genoma corresponde cada gen. La tabla indica el nombre del gen (orden alfabético), breve descripción, función general, y el ID de la base de datos Xenbase (<https://www.xenbase.org/xenbase/>) para cada gen.

<i>X. tropicalis</i> – <i>X. laevis</i> L – <i>X. laevis</i> S			
Gene name	Description	Function	Xenbase ID
<i>arhgef18</i>	Rho guanine nucleotide exchange factor 18	Cytoskeleton regulation	XB-GENEPAGE-990866
<i>atg101</i>	Autophagy related 101	Autophagosome biogenesis	XB-GENEPAGE-5873848
<i>bcap29</i>	B-cell receptor-associated protein 29	Role in anterograde transport	XB-GENEPAGE-950368
<i>borcs6</i>	BLOC-1 related complex subunit 6	Lysosome positioning	XB-GENEPAGE-6045472
<i>borcs7</i>	BLOC-1 related complex subunit 7	Lysosome positioning (BORC complex)	XB-GENE-1006807
<i>cuedc2</i>	CUE domain containing 2	Ubiquitin-binding adaptor; protein degradation	XB-GENE-1015746
<i>dlgap4</i>	Discs large-associated protein 4	Postsynaptic scaffold; synapse organization	XB-GENE-1010901
<i>ezh1</i>	Enhancer of zeste homolog 1	Histone H3K27 methyltransferase; transcriptional repression	XB-GENE-100493532
<i>golga1</i>	Golgin A1	Golgi vesicle tethering/trafficking	XB-GENE-995783
<i>ift27</i>	Intraflagellar transport 27	Ciliary transport (IFT-B)	XB-GENE-1001555
<i>itpkc</i>	Inositol-trisphosphate 3-kinase C	IP3 phosphorylation; Ca ²⁺ signaling	XB-GENE-992155
<i>mafk</i>	Small Maf protein K	bZIP transcription factor (CNC partner)	XB-GENE-959430
<i>manbal</i>	Mannosidase beta-like	Carbohydrate/glycoprotein metabolism	XB-GENE-971373
<i>pdpr</i>	Pyruvate dehydrogenase phosphatase regulatory	PDH complex regulation	XB-GENE-5791461
<i>rtn3</i>	Reticulon 3	ER membrane shaping/trafficking	XB-GENE-493567
<i>rufy2</i>	RUN and FYVE domain-containing 2	Endosomal trafficking/signaling	XB-GENE-1015917
<i>slc36a1</i>	Solute carrier family 36A1	Amino acid transport	XB-GENE-976345
<i>thtpa</i>	Thiamine triphosphatase	Vitamin B1 metabolism (ThTP hydrolysis)	XB-GENE-974251
<i>tmbim6</i>	Transmembrane BAX inhibitor motif 6	Apoptosis/ER-stress regulator	XB-GENE-952610
<i>ubxn1</i>	UBX domain protein 1	p97/VCP adaptor; proteostasis	XB-GENE-979857

<i>X. tropicalis</i> – <i>X. laevis</i> L			
Gene name	Description	Function	Xenbase ID
<i>aarsd1</i>	Alanyl-tRNA synthetase domain-containing 1	tRNA metabolism	XB-GENE-5787115
<i>atl3</i>	Atlastin 3	ER membrane fusion	XB-GENE-1009682
<i>clcn5</i>	Chloride channel 5	Ion transport	XB-GENE-5957240
<i>eif2b3</i>	Eukaryotic translation initiation factor 2B subunit gamma	Translation regulation	XB-GENE-989022
<i>frmd8</i>	FERM domain-containing protein 8	Protease regulation	XB-GENE-5786941
<i>gss</i>	Glutathione synthetase	Redox metabolism	XB-GENE-982236
<i>kti12</i>	KTI12 homolog	Elongator complex cofactor	XB-GENE-5787131
<i>miga1</i>	Mitoguardin 1	Mitochondrial fusion	XB-GENE-5787053
<i>mpv17l2</i>	MPV17-like 2	Mitochondrial inner membrane	XB-GENE-5787224
<i>pla2g4a</i>	Cytosolic phospholipase A2 alpha	Lipid metabolism	XB-GENE-1011792
<i>rabepk</i>	Rab9 effector protein with Kelch motifs	Golgi trafficking	XB-GENE-1001182
<i>sil1</i>	SIL1 nucleotide exchange factor	ER chaperone regulator	XB-GENE-5787168
<i>trmo</i>	tRNA (guanine-N7-)-methyltransferase	tRNA modification	XB-GENE-5787175
<i>tspoap1</i>	TSPO associated protein 1	Synaptic scaffolding	XB-GENE-5787184
<i>ubr1</i>	Ubiquitin protein ligase N-recognin 1	Protein degradation	XB-GENE-6052363
<i>uqcc6</i>	Ubiquinol-cytochrome c reductase complex chaperone 6	Respiratory chain assembly	XB-GENE-5787198
<i>ythdc2</i>	YTH domain containing 2	RNA metabolism	XB-GENE-5787207
<i>LOC100488006</i> <i>LOC108699309</i>	Zinc finger protein 239-like	Predicted TF	XB-GENE-5787213
<i>X. tropicalis</i> – <i>X. laevis</i> S			
Gene name	Description	Function	Xenbase ID
<i>ccdc51</i>	Coiled-coil domain containing 51	Mitochondrial membrane protein	XB-GENE-5786943
<i>ptgr2</i>	Prostaglandin reductase 2	Lipid metabolism	XB-GENE-5787087
<i>X. laevis</i> L – <i>X. laevis</i> S			
Gene name	Description	Function	Xenbase ID

<i>cpt1a1</i>	Carnitine palmitoyltransferase 1A (liver) like	Carnitine metabolic process	XB-GENE-1000855
<i>slc44a1</i>	Solute carrier family 44A1	Choline transporter	XB-GENE-1000057
<i>X. tropicalis</i>			
Gene name	Description	Function	Xenbase ID
<i>mfsd10</i>	Major facilitator superfamily 10	Transporter, organic anion/drug efflux	XB-GENE-1000222
<i>nsmce1</i>	Non-SMC element 1	Structural maintenance of chromosomes complex	XB-GENE-1000988
<i>simiate</i> (<i>abitram</i>)	Actin bundling protein	Actin cytoskeleton regulation in synapses; dendritic spine dynamics	XB-GENE-5787094
<i>slc11a2</i> LOC105946559	Iron transporter	Divalent metal transporter; iron homeostasis	XB-GENE-5787147
<i>trnp1</i>	Neural progenitor proliferation protein	Neural progenitor proliferation, cortical development	XB-GENE-5787172
<i>ttc32</i>	Tetratricopeptide repeat protein 32	Scaffold protein, protein–protein interactions	XB-GENE-5787202
<i>X. laevis L</i>			
Gene name	Description	Function	Xenbase ID
<i>cbr4</i>	Carbonyl reductase 4	Fatty acid metabolism; mitochondrial NADPH-dependent reductase	XB-GENE-1000822
<i>dpy19l1</i>	Dpy-19 like 1	Protein glycosylation (C-mannosylation of WxxW motifs)	XB-GENE-1000844
<i>faah2</i>	Fatty acid amide hydrolase 2	Endocannabinoid metabolism; hydrolysis of fatty acid amides (anandamide)	XB-GENE-1000898
<i>hdac8</i>	Histone deacetylase 8	Chromatin remodeling; histone H3/H4 deacetylation, transcription regulation	XB-GENE-1000920
<i>myl12b</i>	Myosin light chain 12B	Actin cytoskeleton regulation; regulatory light chain for myosin II	XB-GENE-1000967
<i>ntan1</i>	N-terminal asparagine amidase	N-end rule pathway; converts Asn to Asp for ubiquitin-mediated degradation	XB-GENE-1001014
<i>trim39</i>	Tripartite motif containing 39	RING-type ubiquitin ligase; apoptosis and cell cycle regulation	XB-GENE-1001052
<i>vps13c</i>	Vacuolar protein sorting 13 homolog C	Organelle contact sites; lipid transport between ER, mitochondria, lysosomes	XB-GENE-1001076
<i>X. laevis S</i>			
Gene name	Description	Function	Xenbase ID

gtf2i

General transcription factor Iii

Multifunctional TF; regulates RNA
polymerase II transcription, signaling
pathways

XB-GENE-1000917

vdac1

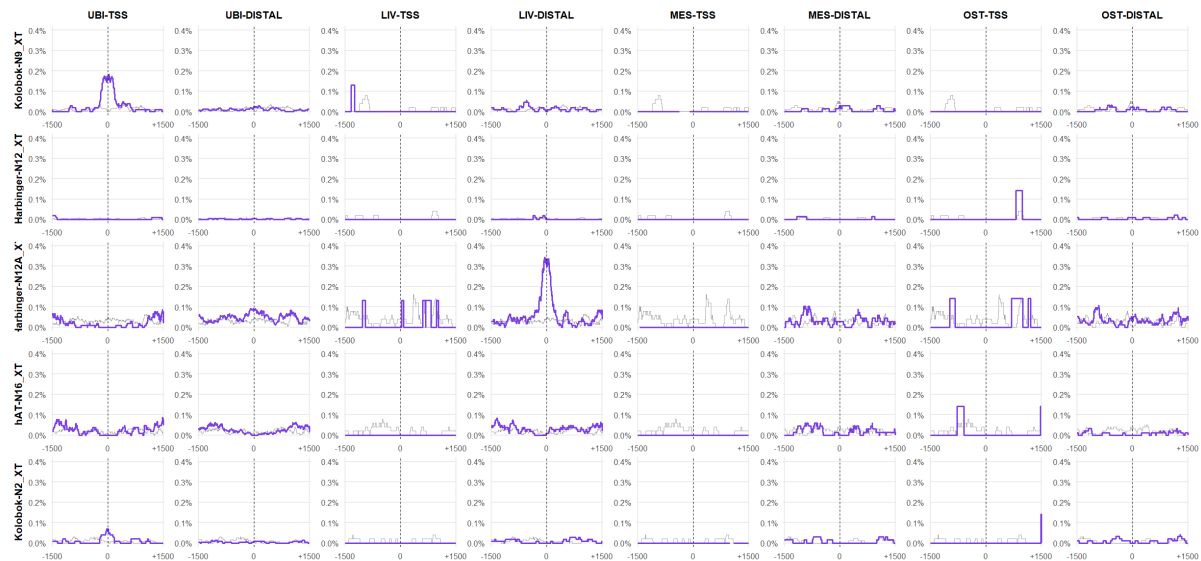
Voltage dependent anion channel

Mitochondrial porin

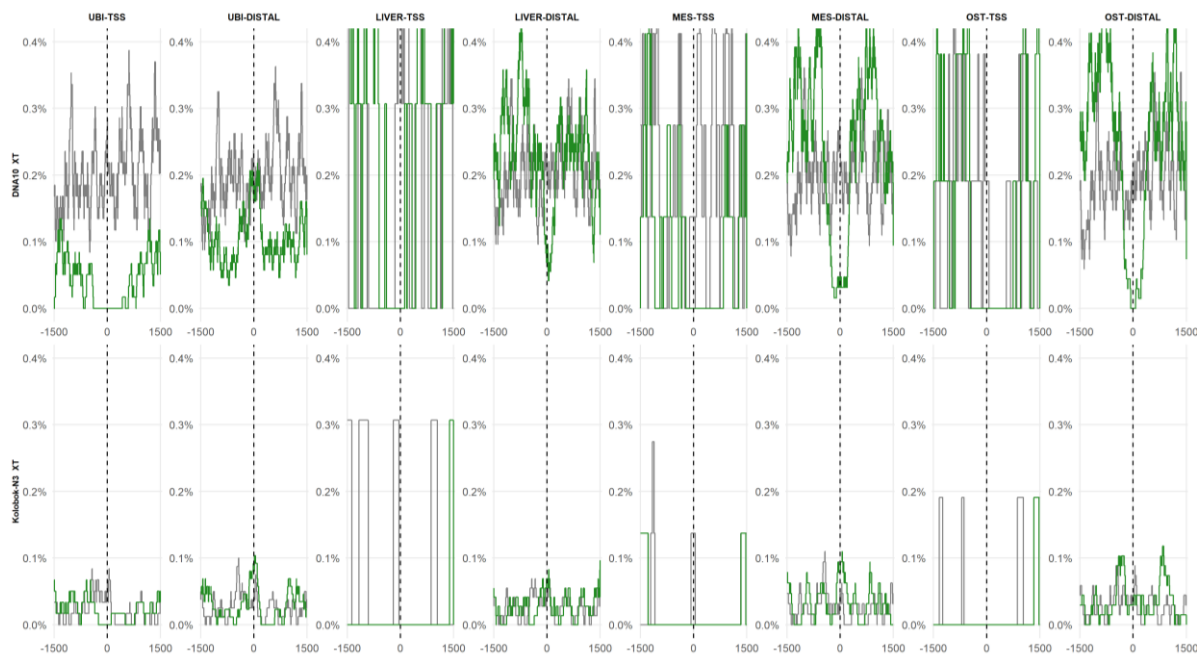
XB-GENE-1001036

1

Anexo 2: Subfamilias de Transposones de ADN identificadas en *X. tropicalis* y su presencia en *X. laevis*. LIV: hígado, MES: mesénquima, OST: Osteoblastos.



Anexo 3: Subfamilias de Transposones de ADN identificadas en *X. laevis* y su presencia en *X. tropicalis*. LIV: hígado, MES: mesénquima, OST: Osteoblastos.



Anexo 4: Enlace a todos los mapas de calor elaborados para las copias domesticadas de KN9 para sus respectivos genes blancos.

https://1drv.ms/b/c/87844a323760020d/IQCyRee2dA_ARKnKnODdQT_-AbSeq3ofYOAvitZB_rQMcaq?e=Xh1ecH

Anexo 5: Alineamiento de una copia de la subfamilia K de T2-MITEs identificada por Hikosaka y Kawahara con la secuencia consenso de KN9.

K_TS_subfamily_Xt_JGlv4.1_scaffold11_1621369-1621974_RC_606bp
Sequence ID: **Query_7517139** Length: **606** Number of Matches: **1**

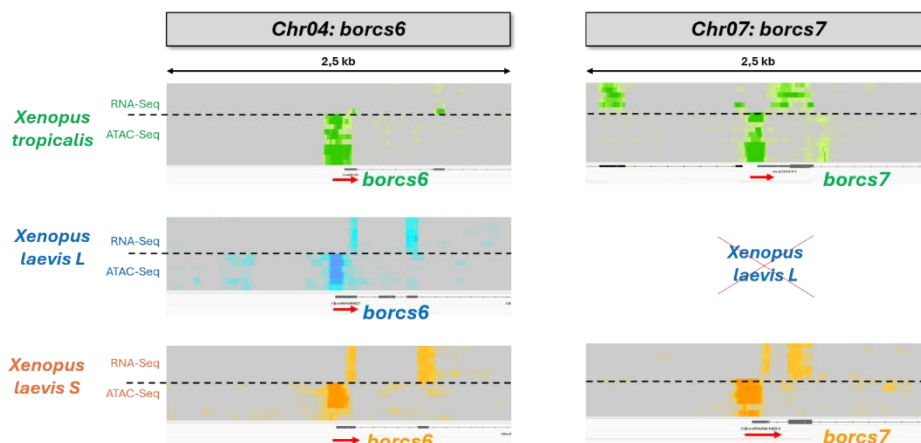
Range 1: 1 to 606 [Graphics](#) [Next Match](#)

Score	Expect	Identities	Gaps	Strand
975 bits(1080)	0.0	587/615(95%)	10/615(1%)	Plus/Plus
Query 1		AGGGGATATATACCATAAATTTTACAATGCACATAAACCATGTAAATAATGTAA/		
Sbjct 1		AGGGGATATATAGCATACATTTTACAATGCACATAAACCATGTAAATAATGTAA/		
Query 61		AAGGAAGTGCTTTTATTTTAATACCTTTGGGTTCAAATATGTCATAACTGCTTG/		
Sbjct 61		AAGGAAGTGCTTTTATTTTAATACCTTTGGGTTCAAATATGTCATAACTGCTTG/		
Query 121		TGTGCTCTACCCAGACCTTATGCCAGCTTCCGGTTTAGACTCTTCAGTATATCGG/		
Sbjct 121		TGTGCTCTACCCAGACCTTATGCCAGCTTCCGGTTTAGACTCTTCAGTATAG-GG/		
Query 181		AGCCGCCGGCCCTCATAGACTTCTATGACACCA-CCTCACCTACTTAAAGTGAAG/		
Sbjct 180		AGCCGCCGACCTCATAGACTTCTATGACACCTCTCACCTACTTAAAGTGAAG/		
Query 240		ATTCTAGAGTGGGCGGGCTTAGCGAGGTCACAGACAGTCCGTTCTGCTCTAC/		
Sbjct 240		ATTCTAGAGTGGGCGGGCTTAGCGAGGTCACAGACAGTCCGTTCTGCTCTAC/		
Query 300		CCTCTCCCCCCCCCCCCCTCCCTGGCAGCAGCACAAGCAGAGACACAGAAG/		
Sbjct 300		CCTCTCCCCCCC-----TCCCTGGCAGCAGCACAAGCAGAGACACAGAAG/		
Query 360		AGGGGTTTTCCCTGCTGCTGCTGAGTAAAGCCTGTCACTCAGTGTGACCACT/		
Sbjct 353		AGGGGTTTTCCCTGCTGCTGCTGAGTAAAGCCTGTCACTCAGTGTGACCACT/		
Query 420		TGGGAGACTGAAGCGACACTCCCATGTCTCATTTGGCTCTGGCTTCAGACTAGAG/		
Sbjct 413		TGGGAGACTGAAG-GACACTCCCATGTCTCATTTACTCTGGCTTCAGACTAGAG/		
Query 480		GTGTATGCAGCTCTCATATAATGGCAGTTTTAGGAGGTAAAATGCTTTCAAAACA/		
Sbjct 472		GTCTCTGCAGCTCTCATATAATGGCAGTTTTAGGAGGTAAAATGCTTTCAAAACA/		
Query 540		CTTCTGCATCATTTACATTTTGGAGGAGGATGAATATTATAACTGAATATGAA/		
Sbjct 532		CTTCTGCATCATTTACATTTTGGAGGAGGATGAATATTATAACTGAATATGAA/		
Query 600		TATAAAATGTCTCCT 614		
Sbjct 592		TATAAAATGTCTCCT 606		

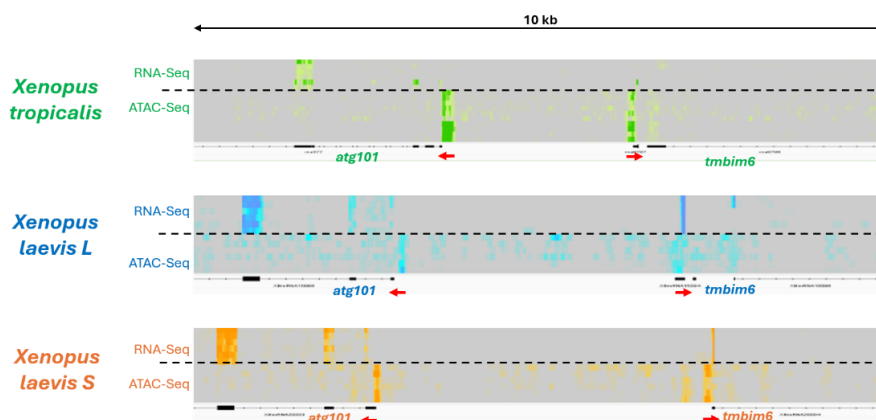
Anexo 6: Enlace hacia la data genómica publicada sobre *Nanorana parkeri*, incluido archivo de Repeatmasker con la anotación de elementos transponibles.

https://genome.ucsc.edu/cgi-bin/hgGateway?hgsid=949829428_RA5BcRnwFLbMZnyfXzBYcQQiTm1h&redirect=manual&source=genome-asia.ucsc.edu&token=1_v_or487jmxBrzNnxVG5AOu6OfbIPbnCQbEhgAVAbps-FFOcUmBv07uHuTIXb1jJGbox-dBzkM7h5n2DD6ZGp_o0q01RkAVLBO7jOkL6vRXUXsqBQO6GWIWzO513CP2XCCLkngTU00Vh23xuE4LAjh41GKbLzDjZhCpNhmsq1BP5e2Jq14wK-XSOaaNtSRddq5n-rRuRIhww96bJXqNICfVMCI84Ef7eBTosfa4B9xwhM3QW7iiQUB5QDS7PO_6_YBq6l6J0DwRzlsG5_XQ0mWrW841ci33zFD50WFnkCryrWHi4IYleAQroRWbFvFbPO2_cHpiq-EFabHYZHFeC9YpOo_ONI0UZWPkImL2IzOiN3gviXNxyOnQQIJuUJicd4aSTh7yzUxAell-60C_8HvVbZiyEhPiUusyVAawVeelRKglcQL6Xs_3Hz_CoNp5I2OMmQR06l8l2D8cr_xjwh8841wumb2hazRivrBTdza_vS2GjAYPoM7sRfRWYU_DrXDETERIA0KIVgAHbrHbdaAkWOGLE21yE2zHuTtTVRkuFqsrWgQ_e_NIN0VlqT3Zab4c1nJ4KrnXGANXmsadPhkPt-cj8iGGHbxQd5ZRkxQ5QL-iMjSNxqbYBVZ7x2OCOBWvADFuzz00JhIhuQbcA.EuMAqD3mUlmzOesxNo77ng.bb2bfe5193e7715d3aec7677116f6e12f7cc403667f13e9c81ec7d0d5260bd00

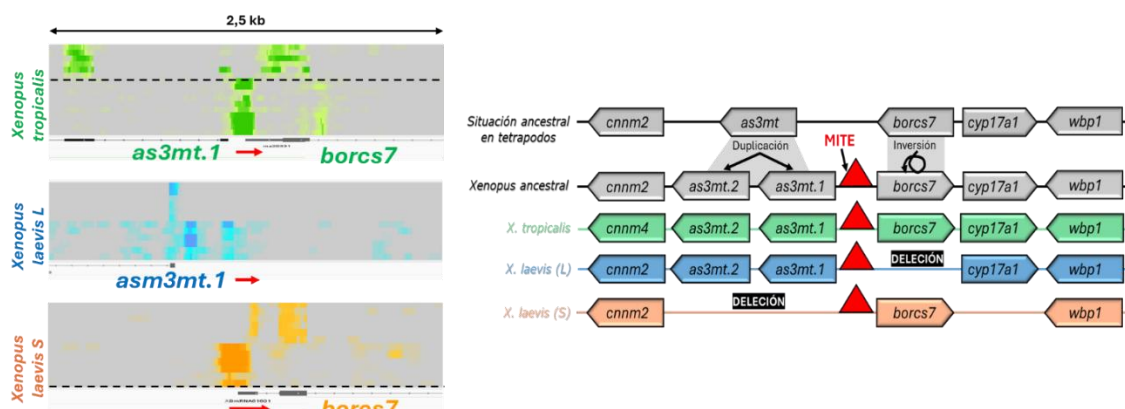
Anexo 7: Presencia de KN9 como promotor en los genes homólogos *borcs6* y *borcs7*. Chr: cromosoma. La flecha roja indica la posición y sentido de KN9. *X. laevis* L no presenta el gen *borcs7*.



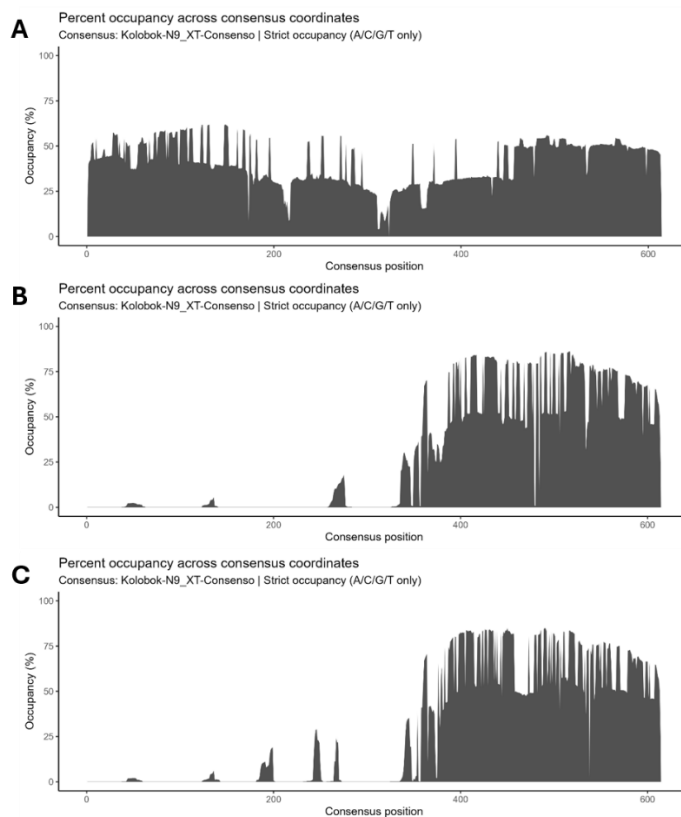
Anexo 8: Presencia de KN9 como promotor de genes vecinos en dirección antisentido *atg101* y *tmbim6*. La flecha roja indica la posición y sentido de KN9.



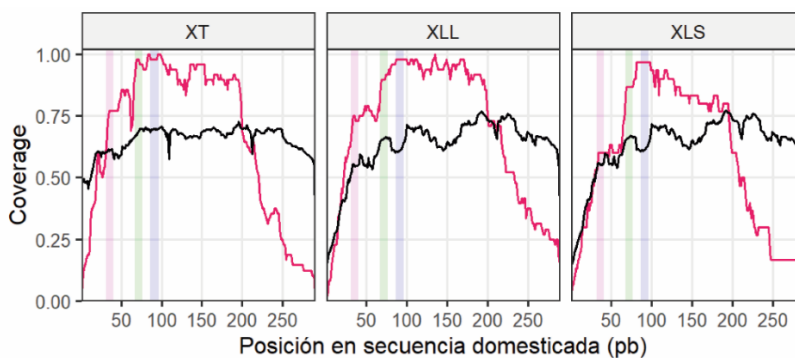
Anexo 9: Presencia de KN9 como promotor bidireccional para los genes *as3mt.1* y *borcs7* solo en *X. tropicalis*. A la derecha, se muestra la historia evolutiva del locus evaluado.



Anexo 10: Ocupación de las copias NO domesticadas de KN9 a lo largo de la secuencia consenso para (A) *X. tropicalis*, y subgenomas (B) L y (C) S de *X. laevis*.



Anexo 11: Perfil de conservación de la región 3' para las copias domesticadas (línea roja) y no domesticadas (línea negra) de KN9 con respecto a la secuencia consenso. Cobertura igual a 1 indica conservación total con la secuencia consenso. Xt: *X. tropicalis*, XLL: Subgenoma L *X. laevis*, XLS: Subgenoma S *X. laevis*. Se colorea de color rosado, verde y azul la posición de los TFBs SP/KLF, PBC/MEINOX, y ETS respectivamente.



Anexo 12: Data de chipseq de humano obtenida de <https://chip-atlas.org/> para miembros de las familias de TFBs conservadas en KN9. (A) *rtn3*, (B) *tmbim6*, (C) *abitram*, (D) *faah2*, (E) *tspoap1*, y (F) *nsmce1*. Bajo cada letra se observa la anotación del gen, ETS1, ETS2, GABPA, KLF1, KLF4, PBX1, MEIS1. En el cuadrado rojo se encierran los FTs que se unen a la región promotora.



Anexo 13: Alineamiento de la secuencia consenso KN9 con el KN9 presente en el promotor de los genes *atg101* y *tmbim6*.

Sitio ETS

CONSENSO	taaagcctgtcagtcagtcagtcgtgtgac caatttcctgt gggagactgaagcgcaactcccat	120
MITE-ATG101	---TGACTGTCACCTCACagtttatgac caatttcctgt aggAGAAAAGAAAGGACACGCCAG	57
MITE-TMBIM6	---TGATTGTCATTTAcagtttatgac caatttcctgt ggaagAATCAAGGACCCCTCCAG	57
	* * * * *	
CONSENSO	gtctctatttggtctggtccttcagactagagaggttgtgtatgcagctctcatataatggc	180
MITE-ATG101	CTCTCATTTCCGTGTAAAC ACTGAGCAGAGGGGATCTTTCAGCAGCTGTG--TTC-TAAT	114
MITE-TMBIM6	TCTTCAGTTCCGCGTAGACACTGAGCAGAGAGGATCTTAGCAGCTGTGAT--ATAAT TAC	115
	* * * * *	
splicing consensus		
CONSENSO	aggttttagg agg taaaaatgctttcaaaacatctttctttctgcattcatttcacattttga	240
MITE-ATG101	ACATTTCCTGAG GTAAAAACGCTATT-----	138
MITE-TMBIM6	ATCCTGAGAG GTAAAAACAGATT-----	139
	* * * ***** *	
aggt splicing consensus		
AMARILLO, PRIMER EXON DE ATG101		
NEGRO, PRIMER EXON DE TMBIM6		

8. REFERENCIAS

1. The ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature*. 2012 Sep;489(7414):57–74. doi:10.1038/nature11247
2. Biémont C, Vieira C. Junk DNA as an evolutionary force. *Nature*. 2006 Oct;443(7111):521–4. doi:10.1038/443521a
3. Palazzo AF, Gregory TR. The Case for Junk DNA. Akey JM, editor. *PLoS Genet*. 2014 May 8;10(5):e1004351. doi:10.1371/journal.pgen.1004351
4. Havstad JC, Palazzo AF. Not functional yet a difference maker: junk DNA as a case study. *Biol Philos*. 2022 Aug;37(4):29. doi:10.1007/s10539-022-09854-1
5. Rogalska ME, Vivori C, Valcárcel J. Regulation of pre-mRNA splicing: roles in physiology and disease, and therapeutic prospects. *Nat Rev Genet*. 2023 Apr;24(4):251–69. doi:10.1038/s41576-022-00556-8
6. Cordaux R, Batzer MA. The impact of retrotransposons on human genome evolution. *Nat Rev Genet*. 2009 Oct;10(10):691–703. doi:10.1038/nrg2640
7. Schmidt D, Schwalie PC, Wilson MD, Ballester B, Gonçalves Â, Kutter C, et al. Waves of Retrotransposon Expansion Remodel Genome Organization and CTCF Binding in Multiple Mammalian Lineages. *Cell*. 2012 Jan;148(1–2):335–48. doi:10.1016/j.cell.2011.11.058
8. Lawson HA, Liang Y, Wang T. Transposable elements in mammalian chromatin organization. *Nat Rev Genet*. 2023 Oct;24(10):712–23. doi:10.1038/s41576-023-00609-6
9. Bartel DP. Metazoan MicroRNAs. *Cell*. 2018 Mar;173(1):20–51. doi:10.1016/j.cell.2018.03.006
10. Chen LL, Kim VN. Small and long non-coding RNAs: Past, present, and future. *Cell*. 2024 Nov;187(23):6451–85. doi:10.1016/j.cell.2024.10.024
11. Wittkopp PJ, Kalay G. Cis-regulatory elements: molecular mechanisms and evolutionary processes underlying divergence. *Nat Rev Genet*. 2012 Jan;13(1):59–69. doi:10.1038/nrg3095
12. Newman SA. Cell differentiation: What have we learned in 50 years? *Journal of Theoretical Biology*. 2020 Jan;485:110031. doi:10.1016/j.jtbi.2019.110031
13. Bourque G, Burns KH, Gehring M, Gorbunova V, Seluanov A, Hammell M, et al. Ten things you should know about transposable elements. *Genome Biol*. 2018 Dec;19(1):199. doi:10.1186/s13059-018-1577-z
14. Wells JN, Feschotte C. A Field Guide to Eukaryotic Transposable Elements. *Annu Rev Genet*. 2020 Nov 23;54(1):539–61. doi:10.1146/annurev-genet-040620-022145
15. Spitz F, Furlong EEM. Transcription factors: from enhancer binding to developmental control. *Nat Rev Genet*. 2012 Sep;13(9):613–26. doi:10.1038/nrg3207
16. Klemm SL, Shipony Z, Greenleaf WJ. Chromatin accessibility and the regulatory epigenome. *Nat Rev Genet*. 2019 Apr;20(4):207–20. doi:10.1038/s41576-018-0089-8
17. Wittkopp PJ, Kalay G. Cis-regulatory elements: molecular mechanisms and evolutionary processes underlying divergence. *Nat Rev Genet*. 2012 Jan;13(1):59–69. doi:10.1038/nrg3095

18. Smale ST, Baltimore D. The “initiator” as a transcription control element. *Cell*. 1989 Apr;57(1):103–13. doi:10.1016/0092-8674(89)90176-1
19. Chalkley GE, Verrijzer CP. DNA binding site selection by RNA polymerase II TAFs: a TAFII250–TAFII150 complex recognizes the Initiator. *EMBO J*. 1999 Sep 1;18(17):4835–45. doi:10.1093/emboj/18.17.4835
20. Carninci P, Sandelin A, Lenhard B, Katayama S, Shimokawa K, Ponjavic J, et al. Genome-wide analysis of mammalian promoter architecture and evolution. *Nat Genet*. 2006 Jun 1;38(6):626–35. doi:10.1038/ng1789
21. Louder RK, He Y, López-Blanco JR, Fang J, Chacón P, Nogales E. Structure of promoter-bound TFIID and model of human pre-initiation complex assembly. *Nature*. 2016 Mar 31;531(7596):604–9. doi:10.1038/nature17394
22. Haberle V, Stark A. Eukaryotic core promoters and the functional basis of transcription initiation. *Nat Rev Mol Cell Biol*. 2018 Oct;19(10):621–37. doi:10.1038/s41580-018-0028-8
23. Brázda V, Bartas M, Bowater RP. Evolution of Diverse Strategies for Promoter Regulation. *Trends in Genetics*. 2021 Aug;37(8):730–44. doi:10.1016/j.tig.2021.04.003
24. Levine M, Davidson EH. Gene regulatory networks for development. *Proc Natl Acad Sci USA*. 2005 Apr 5;102(14):4936–42. doi:10.1073/pnas.0408031102
25. Capriati M, Nestmann S, Akan IS, Tuschmann FB, Druffel-Augustin S, Grand RS. Essential genes are dominantly activated by single transcription factors [Internet]. *Molecular Biology*; 2026 [cited 2026 Aug 14]. Available from: <http://biorxiv.org/lookup/doi/10.64898/2026.04.27.721063> doi:10.64898/2026.04.27.721063
26. Moribe F, Iskar M, Durdu S, Wirbelauer C, Hoerner L, Smallwood SA, et al. Modular control of CpG islands by chromatin-opening transcription factors specifies transcription start sites [Internet]. *Molecular Biology*; 2026 [cited 2026 Aug 14]. Available from: <http://biorxiv.org/lookup/doi/10.64898/2026.05.04.722487> doi:10.64898/2026.05.04.722487
27. Deng W, Roberts SGE. A core promoter element downstream of the TATA box that is recognized by TFIIB. *Genes Dev*. 2005 Oct 15;19(20):2418–23. doi:10.1101/gad.342405
28. Functional Characterization of Core Promoter Elements: the Downstream Core Element Is Recognized by TAF1. *Molecular and Cellular Biology*. 2005 Dec 1;25(24):11192–11192. doi:10.1128/MCB.25.24.11192.2005
29. Castillo H, Hanna P, Sachs LM, Buisine N, Godoy F, Gilbert C, et al. *Xenopus tropicalis* osteoblast-specific open chromatin regions reveal promoters and enhancers involved in human skeletal phenotypes and shed light on early vertebrate evolution. *Cells & Development*. 2024 Sep;179:203924. doi:10.1016/j.cdev.2024.203924
30. Hartono SR, Korf IF, Chédin F. GC skew is a conserved property of unmethylated CpG island promoters across vertebrates. *Nucleic Acids Res*. 2015 Aug 7;gkv811. doi:10.1093/nar/gkv811
31. Andersson R, Sandelin A. Determinants of enhancer and promoter activities of regulatory elements. *Nat Rev Genet*. 2020 Feb;21(2):71–87. doi:10.1038/s41576-019-0173-8

32. Welboren W, Van Driel MA, Janssen-Megens EM, Van Heeringen SJ, Sweep FC, Span PN, et al. ChIP-Seq of ER α and RNA polymerase II defines genes differentially responding to ligands. *EMBO J.* 2009 May 20;28(10):1418–28. doi:10.1038/emboj.2009.88
33. García-González E, Escamilla-Del-Arenal M, Arzate-Mejía R, Recillas-Targa F. Chromatin remodeling effects on enhancer activity. *Cell Mol Life Sci.* 2016 Aug;73(15):2897–910. doi:10.1007/s00018-016-2184-3
34. Johnson SP, Warner JR. Unusual Enhancer Function in Yeast rRNA Transcription. *Molecular and Cellular Biology.* 1989 Nov 1;9(11):4986–93. doi:10.1128/mcb.9.11.4986-4993.1989
35. Müller HP, Sogo JoséM, Schaffner W. An enhancer stimulates transcription in Trans when attached to the promoter via a protein bridge. *Cell.* 1989 Aug;58(4):767–77. doi:10.1016/0092-8674(89)90110-4
36. Pennacchio LA, Bickmore W, Dean A, Nobrega MA, Bejerano G. Enhancers: five essential questions. *Nat Rev Genet.* 2013 Apr;14(4):288–95. doi:10.1038/nrg3458
37. Robson MI, Ringel AR, Mundlos S. Regulatory Landscaping: How Enhancer-Promoter Communication Is Sculpted in 3D. *Molecular Cell.* 2019 Jun;74(6):1110–22. doi:10.1016/j.molcel.2019.05.032
38. Yokoshi M, Fukaya T. Dynamics of transcriptional enhancers and chromosome topology in gene regulation. *Dev Growth Differ.* 2019 Jun;61(5):343–52. doi:10.1111/dgd.12597
39. Bolt CC, Duboule D. The regulatory landscapes of developmental genes. *Development.* 2020 Feb 1;147(3):dev171736. doi:10.1242/dev.171736
40. Dixon JR, Selvaraj S, Yue F, Kim A, Li Y, Shen Y, et al. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature.* 2012 May;485(7398):376–80. doi:10.1038/nature11082
41. Zhang G, Shi J, Zhu S, Lan Y, Xu L, Yuan H, et al. DiseaseEnhancer: a resource of human disease-associated enhancer catalog. *Nucleic Acids Research.* 2018 Jan 4;46(D1):D78–84. doi:10.1093/nar/gkx920
42. Kragestein BK, Spielmann M, Paliou C, Heinrich V, Schöpflin R, Esposito A, et al. Dynamic 3D chromatin architecture contributes to enhancer specificity and limb morphogenesis. *Nat Genet.* 2018 Oct;50(10):1463–73. doi:10.1038/s41588-018-0221-x
43. Richard GF, Kerrest A, Dujon B. Comparative Genomics and Molecular Dynamics of DNA Repeats in Eukaryotes. *Microbiol Mol Biol Rev.* 2008 Dec;72(4):686–727. doi:10.1128/MMBR.00011-08
44. Biscotti MA, Olmo E, Heslop-Harrison JS. Repetitive DNA in eukaryotic genomes. *Chromosome Res.* 2015 Sep;23(3):415–20. doi:10.1007/s10577-015-9499-z
45. Gemayel R, Vences MD, Legendre M, Verstrepen KJ. Variable Tandem Repeats Accelerate Evolution of Coding and Regulatory Sequences. *Annu Rev Genet.* 2010 Dec 1;44(1):445–77. doi:10.1146/annurev-genet-072610-155046
46. Hannan AJ. Tandem repeats mediating genetic plasticity in health and disease. *Nat Rev Genet.* 2018 May;19(5):286–98. doi:10.1038/nrg.2017.115

47. Chambeyron S, Pélisson A, Koga Y, Grimaud C, Siomi MC. Host-transposable element coexistence: a matter of resistance, tolerance and trade-off. *EMBO J.* 2026 May 13;45(12):4014–33. doi:10.1038/s44318-026-00795-z
48. Wells JN, Feschotte C. A Field Guide to Eukaryotic Transposable Elements. *Annu Rev Genet.* 2020 Nov 23;54(1):539–61. doi:10.1146/annurev-genet-040620-022145
49. McClintock B. The origin and behavior of mutable loci in maize. *Proc Natl Acad Sci USA.* 1950 Jun;36(6):344–55. doi:10.1073/pnas.36.6.344
50. Jones RN. McClintock's controlling elements: the full story. *Cytogenet Genome Res.* 2005;109(1–3):90–103. doi:10.1159/000082387
51. Ravindran S. Barbara McClintock and the discovery of jumping genes. *Proc Natl Acad Sci USA.* 2012 Dec 11;109(50):20198–9. doi:10.1073/pnas.1219372109
52. Hua-Van A, Le Rouzic A, Maisonhaute C, Capy P. Abundance, distribution and dynamics of retrotransposable elements and transposons: similarities and differences. *Cytogenet Genome Res.* 2005;110(1–4):426–40. doi:10.1159/000084975
53. Lampe DJ, Churchill ME, Robertson HM. A purified mariner transposase is sufficient to mediate transposition in vitro. *EMBO J.* 1996 Oct 1;15(19):5470–9. PubMed PMID: 8895590; PubMed Central PMCID: PMC452289.
54. Miller WJ, McDonald JF, Pinsker W. Molecular domestication of mobile elements. *Genetica.* 1997;100(1–3):261–70. PubMed PMID: 9440279.
55. Sinzelle L, Izsvák Z, Ivics Z. Molecular domestication of transposable elements: From detrimental parasites to useful host genes. *Cell Mol Life Sci.* 2009 Mar;66(6):1073–93. doi:10.1007/s00018-009-8376-3
56. Modzelewski AJ, Gan Chong J, Wang T, He L. Mammalian genome innovation through transposon domestication. *Nat Cell Biol.* 2022 Sep;24(9):1332–40. doi:10.1038/s41556-022-00970-4
57. Volff J. Turning junk into gold: domestication of transposable elements and the creation of new genes in eukaryotes. *BioEssays.* 2006 Sep;28(9):913–22. doi:10.1002/bies.20452
58. Levin HL. A Novel Mechanism of Self-Primed Reverse Transcription Defines a New Family of Retroelements. *Mol Cell Biol.* 1995 Jun 1;15(6):3310–7. doi:10.1128/MCB.15.6.3310
59. McCarthy EM, McDonald JF. Long terminal repeat retrotransposons of *Mus musculus*. *Genome Biol.* 2004 Feb 13;5(3):R14. doi:10.1186/gb-2004-5-3-r14
60. Greally JM. Short interspersed transposable elements (SINEs) are excluded from imprinted regions in the human genome. *Proc Natl Acad Sci USA.* 2002 Jan 8;99(1):327–32. doi:10.1073/pnas.012539199
61. Burke WD. R5 Retrotransposons Insert into a Family of Infrequently Transcribed 28S rRNA Genes of Planaria. *Molecular Biology and Evolution.* 2003 Apr 25;20(8):1260–70. doi:10.1093/molbev/msg141
62. Han JS. Non-long terminal repeat (non-LTR) retrotransposons: mechanisms, recent developments, and unanswered questions. *Mobile DNA.* 2010;1(1):15. doi:10.1186/1759-8753-1-15

63. Denli AM, Narvaiza I, Kerman BE, Pena M, Benner C, Marchetto MCN, et al. Primate-Specific ORF0 Contributes to Retrotransposon-Mediated Diversity. *Cell*. 2015 Oct;163(3):583–93. doi:10.1016/j.cell.2015.09.025
64. Carelli FN, Hayakawa T, Go Y, Imai H, Warnefors M, Kaessmann H. The life history of retrocopies illuminates the evolution of new mammalian genes. *Genome Res*. 2016 Mar;26(3):301–14. doi:10.1101/gr.198473.115
65. Kubiak MR, Makalowska I. Protein-Coding Genes' Retrocopies and Their Functions. *Viruses*. 2017 Apr 13;9(4):80. doi:10.3390/v9040080
66. Platt RN, Vandeweghe MW, Ray DA. Mammalian transposable elements and their impacts on genome evolution. *Chromosome Res*. 2018 Mar;26(1–2):25–43. doi:10.1007/s10577-017-9570-z
67. Sengstag C, Iida S, Hiestand-Nauer R, Arber W. Terminal inverted repeats of prokaryotic transposable element IS186 which can generate duplications of variable length at an identical target sequence. *Gene*. 1986 Jan;49(1):153–6. doi:10.1016/0378-1119(86)90395-1
68. Koga A. Vertebrate DNA Transposon as a Natural Mutator: The Medaka Fish Tol2 Element Contributes to Genetic Variation without Recognizable Traces. *Molecular Biology and Evolution*. 2006 May 5;23(7):1414–9. doi:10.1093/molbev/msl003
69. Thomas J, Pritham EJ. *Helitrons*, the Eukaryotic Rolling-circle Transposable Elements. Chandler M, Craig N, editors. *Microbiol Spectr*. 2015 Jul 2;3(4):3.4.03. doi:10.1128/microbiolspec.MDNA3-0049-2014
70. Feschotte C, Pritham EJ. DNA Transposons and the Evolution of Eukaryotic Genomes. *Annu Rev Genet*. 2007 Dec 1;41(1):331–68. doi:10.1146/annurev.genet.40.110405.090448
71. Shi Z, Xu J, Niu L, Shen W, Yan S, Tan Y, et al. Evolutionarily distinct and sperm-specific supersized chromatin loops are marked by Helitron transposons in *Xenopus tropicalis*. *Cell Reports*. 2023 Mar;42(3):112151. doi:10.1016/j.celrep.2023.112151
72. Wessler S. LTR-retrotransposons and MITEs: important players in the evolution of plant genomes. *Current Opinion in Genetics & Development*. 1995 Dec;5(6):814–21. doi:10.1016/0959-437X(95)80016-X
73. Zhang Q, Arbuckle J, Wessler SR. Recent, extensive, and preferential insertion of members of the miniature inverted-repeat transposable element family *Heartbreaker* into genic regions of maize. *Proc Natl Acad Sci USA*. 2000 Feb;97(3):1160–5. doi:10.1073/pnas.97.3.1160
74. Oosumi T, Garlick B, Belknap WR. Identification of putative nonautonomous transposable elements associated with several transposon families in *Caenorhabditis elegans*. *J Mol Evol*. 1996 Jul;43(1):11–8. doi:10.1007/BF02352294
75. Izsvák Z, Ivics Z, Shimoda N, Mohn D, Okamoto H, Hackett PB. Short Inverted-Repeat Transposable Elements in Teleost Fish and Implications for a Mechanism of Their Amplification. *J Mol Evol*. 1999 Jan;48(1):13–21. doi:10.1007/PL00006440
76. Smit AF, Riggs AD. Tiggers and DNA transposon fossils in the human genome. *Proc Natl Acad Sci USA*. 1996 Feb 20;93(4):1443–8. doi:10.1073/pnas.93.4.1443

77. Zhang X, Jiang N, Feschotte C, Wessler SR. *PIF* - and *Pong* -Like Transposable Elements: Distribution, Evolution and Relationship With *Tourist* -Like Miniature Inverted-Repeat Transposable Elements. *Genetics*. 2004 Feb 1;166(2):971–86. doi:10.1093/genetics/166.2.971
78. Lu C, Chen J, Zhang Y, Hu Q, Su W, Kuang H. Miniature Inverted-Repeat Transposable Elements (MITEs) Have Been Accumulated through Amplification Bursts and Play Important Roles in Gene Expression and Species Diversity in *Oryza sativa*. *Molecular Biology and Evolution*. 2012 Mar 1;29(3):1005–17. doi:10.1093/molbev/msr282
79. Ye C, Ji G, Liang C. detectMITE: A novel approach to detect miniature inverted repeat transposable elements in genomes. *Sci Rep*. 2016 Jan 22;6(1):19688. doi:10.1038/srep19688
80. Hu J, Zheng Y, Shang X. MiteFinderII: a novel tool to identify miniature inverted-repeat transposable elements hidden in eukaryotic genomes. *BMC Med Genomics*. 2018 Nov;11(S5):101. doi:10.1186/s12920-018-0418-y
81. Oki N, Yano K, Okumoto Y, Tsukiyama T, Teraishi M, Tanisaka T. A genome-wide view of miniature inverted-repeat transposable elements (MITEs) in rice, *Oryza sativa* ssp. *japonica*. *Genes Genet Syst*. 2008;83(4):321–9. doi:10.1266/ggs.83.321
82. Kuang H, Padmanabhan C, Li F, Kamei A, Bhaskar PB, Ouyang S, et al. Identification of miniature inverted-repeat transposable elements (MITEs) and biogenesis of their siRNAs in the Solanaceae: New functional implications for MITEs. *Genome Res*. 2009 Jan;19(1):42–56. doi:10.1101/gr.078196.108
83. Kim S, Wysocka J. Deciphering the multi-scale, quantitative cis-regulatory code. *Molecular Cell*. 2023 Feb;83(3):373–92. doi:10.1016/j.molcel.2022.12.032
84. Buenrostro JD, Giresi PG, Zaba LC, Chang HY, Greenleaf WJ. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat Methods*. 2013 Dec;10(12):1213–8. doi:10.1038/nmeth.2688
85. Van Kruijsbergen I, Hontelez S, Elurbe DM, Van Heeringen SJ, Huynen MA, Veenstra GJC. Heterochromatic histone modifications at transposons in *Xenopus tropicalis* embryos. *Developmental Biology*. 2017 Jun;426(2):460–71. doi:10.1016/j.ydbio.2016.08.031
86. Hendrickson DG, Soifer I, Wranik BJ, Botstein D, Scott Mclsaac R. Simultaneous Profiling of DNA Accessibility and Gene Expression Dynamics with ATAC-Seq and RNA-Seq. In: Von Stechow L, Santos Delgado A, editors. *Computational Cell Biology* [Internet]. New York, NY: Springer New York; 2018 [cited 2026 Jun 20]. p. 317–33. (Methods in Molecular Biology). Available from: http://link.springer.com/10.1007/978-1-4939-8618-7_15 doi:10.1007/978-1-4939-8618-7_15
87. Santiago-Algarra D, Dao LTM, Pradel L, España A, Spicuglia S. Recent advances in high-throughput approaches to dissect enhancer function. *F1000Res*. 2017 Jun 19;6:939. doi:10.12688/f1000research.11581.1
88. Siddique A, Buisine N, Chalmers R. The Transposon-Like Correia Elements Encode Numerous Strong Promoters and Provide a Potential New Mechanism for Phase Variation in the *Meningococcus*. Casadesús J, editor. *PLoS Genet*. 2011 Jan 20;7(1):e1001277. doi:10.1371/journal.pgen.1001277
89. Friedli M, Trono D. The Developmental Control of Transposable Elements and the Evolution of Higher Species. *Annu Rev Cell Dev Biol*. 2015 Nov 13;31(1):429–51. doi:10.1146/annurev-cellbio-100814-125514

90. Wang J, Xie G, Singh M, Ghanbarian AT, Raskó T, Szvetnik A, et al. Primate-specific endogenous retrovirus-driven transcription defines naive-like stem cells. *Nature*. 2014 Dec;516(7531):405–9. doi:10.1038/nature13804
91. Gerdes P, Richardson SR, Mager DL, Faulkner GJ. Transposable elements in the mammalian embryo: pioneers surviving through stealth and service. *Genome Biol*. 2016 Dec;17(1):100. doi:10.1186/s13059-016-0965-5
92. Modzelewski AJ, Shao W, Chen J, Lee A, Qi X, Noon M, et al. A mouse-specific retrotransposon drives a conserved Cdk2ap1 isoform essential for development. *Cell*. 2021 Oct;184(22):5541–5558.e22. doi:10.1016/j.cell.2021.09.021
93. Sniezewski L, Janik S, Laszkiewicz A, Majkowski M, Kisielow P, Cebart M. The evolutionary conservation of the bidirectional activity of the NWC gene promoter in jawed vertebrates and the domestication of the RAG transposon. *Developmental & Comparative Immunology*. 2018 Apr;81:105–15. doi:10.1016/j.dci.2017.11.013
94. Villar D, Berthelot C, Aldridge S, Rayner TF, Lukk M, Pignatelli M, et al. Enhancer Evolution across 20 Mammalian Species. *Cell*. 2015 Jan;160(3):554–66. doi:10.1016/j.cell.2015.01.006
95. Lynch VJ, Nnamani MC, Kapusta A, Brayer K, Plaza SL, Mazur EC, et al. Ancient Transposable Elements Transformed the Uterine Regulatory Landscape and Transcriptome during the Evolution of Mammalian Pregnancy. *Cell Reports*. 2015 Feb;10(4):551–61. doi:10.1016/j.celrep.2014.12.052
96. Chuong EB, Rumi MAK, Soares MJ, Baker JC. Endogenous retroviruses function as species-specific enhancer elements in the placenta. *Nat Genet*. 2013 Mar;45(3):325–9. doi:10.1038/ng.2553
97. Pontis J, Pulver C, Playfoot CJ, Planet E, Grun D, Offner S, et al. Primate-specific transposable elements shape transcriptional networks during human development. *Nat Commun*. 2022 Nov 23;13(1):7178. doi:10.1038/s41467-022-34800-w
98. Okhovat M, Nevonen KA, Davis BA, Michener P, Ward S, Milhaven M, et al. Co-option of the lineage-specific LAVA retrotransposon in the gibbon genome. *Proc Natl Acad Sci USA*. 2020 Aug 11;117(32):19328–38. doi:10.1073/pnas.2006038117
99. Ye M, Goudot C, Hoyler T, Lemoine B, Amigorena S, Zueva E. Specific subfamilies of transposable elements contribute to different domains of T lymphocyte enhancers. *Proc Natl Acad Sci USA*. 2020 Apr 7;117(14):7905–16. doi:10.1073/pnas.1912008117
100. Horton I, Kelly CJ, Dziulko A, Simpson DM, Chuong EB. Mouse B2 SINE elements function as IFN-inducible enhancers. *eLife*. 2023 May 9;12:e82617. doi:10.7554/eLife.82617
101. Morata J, Marín F, Payet J, Casacuberta JM. Plant Lineage-Specific Amplification of Transcription Factor Binding Motifs by Miniature Inverted-Repeat Transposable Elements (MITEs). Gonzalez J, editor. *Genome Biology and Evolution*. 2018 May 1;10(5):1210–20. doi:10.1093/gbe/evy073
102. Session AM, Uno Y, Kwon T, Chapman JA, Toyoda A, Takahashi S, et al. Genome evolution in the allotetraploid frog *Xenopus laevis*. *Nature*. 2016 Oct;538(7625):336–43. doi:10.1038/nature19840

103. Kakebeen A, Wills A. Advancing genetic and genomic technologies deepen the pool for discovery in *Xenopus tropicalis*. *Developmental Dynamics*. 2019 Aug;248(8):620–5. doi:10.1002/dvdy.80
104. Zahn N, James-Zorn C, Ponferrada VG, Adams DS, Grzymkowski J, Buchholz DR, et al. Normal Table of *Xenopus* development: a new graphical resource. *Development*. 2022 Jul 15;149(14):dev200356. doi:10.1242/dev.200356
105. Evans BJ, Kelley DB, Tinsley RC, Melnick DJ, Cannatella DC. A mitochondrial DNA phylogeny of African clawed frogs: phylogeography and implications for polyploid evolution. *Molecular Phylogenetics and Evolution*. 2004 Oct;33(1):197–213. doi:10.1016/j.ympev.2004.04.018
106. Kaessmann H. Origins, evolution, and phenotypic impact of new genes. *Genome Res*. 2010 Oct;20(10):1313–26. doi:10.1101/gr.101386.109
107. Van De Peer Y, Maere S, Meyer A. The evolutionary significance of ancient genome duplications. *Nat Rev Genet*. 2009 Oct;10(10):725–32. doi:10.1038/nrg2600
108. Holland PWH, Garcia-Fernández J, Williams NA, Sidow A. Gene duplications and the origins of vertebrate development. *Development*. 1994 Jan 1;1994(Supplement):125–33. doi:10.1242/dev.1994.Supplement.125
109. Zhang J. Evolution by gene duplication: an update. *Trends in Ecology & Evolution*. 2003 Jun;18(6):292–8. doi:10.1016/S0169-5347(03)00033-8
110. Conant GC, Wolfe KH. Turning a hobby into a job: How duplicated genes find new functions. *Nat Rev Genet*. 2008 Dec;9(12):938–50. doi:10.1038/nrg2482
111. Voldoire E, Brunet F, Naville M, Volff JN, Galiana D. Expansion by whole genome duplication and evolution of the sox gene family in teleost fish. Vaudry H, editor. *PLoS ONE*. 2017 Jul 24;12(7):e0180936. doi:10.1371/journal.pone.0180936
112. Wighard SS, Athanasouli M, Witte H, Rödelberger C, Sommer RJ. A New Hope: A Hermaphroditic Nematode Enables Analysis of a Recent Whole Genome Duplication Event. Wolfe K, editor. *Genome Biology and Evolution*. 2022 Dec 7;14(12):evac169. doi:10.1093/gbe/evac169
113. Hellsten U, Harland RM, Gilchrist MJ, Hendrix D, Jurka J, Kapitonov V, et al. The Genome of the Western Clawed Frog *Xenopus tropicalis*. *Science*. 2010 Apr 30;328(5978):633–6. doi:10.1126/science.1183670
114. Grainger RM. *Xenopus tropicalis* as a Model Organism for Genetics and Genomics: Past, Present, and Future. In: Hoppler S, Vize PD, editors. *Xenopus Protocols* [Internet]. Totowa, NJ: Humana Press; 2012 [cited 2026 Jun 23]. p. 3–15. (Methods in Molecular Biology). Available from: http://link.springer.com/10.1007/978-1-61779-992-1_1 doi:10.1007/978-1-61779-992-1_1
115. Buckley RM, Kortschak RD, Raison JM, Adelson DL. Similar Evolutionary Trajectories for Retrotransposon Accumulation in Mammals. *Genome Biology and Evolution*. 2017 Sep;9(9):2336–53. doi:10.1093/gbe/evx179
116. Chang NC, Rovira Q, Wells J, Feschotte C, Vaquerizas JM. Zebrafish transposable elements show extensive diversification in age, genomic distribution, and developmental expression. *Genome Res*. 2022 Jul;32(7):1408–23. doi:10.1101/gr.275655.121

117. Hikosaka A, Kawahara A. A systematic search and classification of T2 family miniature inverted-repeat transposable elements (MITEs) in *Xenopus tropicalis* suggests the existence of recently active MITE subfamilies. *Mol Genet Genomics*. 2010 Jan;283(1):49–62. doi:10.1007/s00438-009-0496-9
118. Shen JJ, Dushoff J, Bewick AJ, Chain FJJ, Evans BJ. Genomic Dynamics of Transposable Elements in the Western Clawed Frog (*Silurana tropicalis*). *Genome Biology and Evolution*. 2013 May;5(5):998–1009. doi:10.1093/gbe/evt065
119. Tang W, Liang P. Comparative Genomics Analysis Reveals High Levels of Differential Retrotransposition among Primates from the Hominidae and the Cercopithecidae Families. Cordaux R, editor. *Genome Biology and Evolution*. 2019 Nov 1;11(11):3309–25. doi:10.1093/gbe/evz234
120. Nurk S, Koren S, Rhie A, Rautiainen M, Bizikadze AV, Mikheenko A, et al. The complete sequence of a human genome. *Science*. 2022 Apr;376(6588):44–53. doi:10.1126/science.abj6987
121. Mouse Genome Sequencing Consortium. Initial sequencing and comparative analysis of the mouse genome. *Nature*. 2002 Dec;420(6915):520–62. doi:10.1038/nature01262
122. The FANTOM Consortium, Andersson R, Gebhard C, Miguel-Escalada I, Hoof I, Bornholdt J, et al. An atlas of active enhancers across human cell types and tissues. *Nature*. 2014 Mar;507(7493):455–61. doi:10.1038/nature12787
123. Ogino H, McConnell WB, Grainger RM. Highly efficient transgenesis in *Xenopus tropicalis* using I-SceI meganuclease. *Mechanisms of Development*. 2006 Feb;123(2):103–13. doi:10.1016/j.mod.2005.11.006
124. Kroll KL, Amaya E. Transgenic *Xenopus* embryos from sperm nuclear transplantations reveal FGF signaling requirements during gastrulation. *Development*. 1996 Oct 1;122(10):3173–83. doi:10.1242/dev.122.10.3173
125. Ishibashi S, Kroll KL, Amaya E. Generating Transgenic Frog Embryos by Restriction Enzyme Mediated Integration (REMI). In: Hoppler S, Vize PD, editors. *Xenopus Protocols* [Internet]. Totowa, NJ: Humana Press; 2012 [cited 2026 Jun 26]. p. 185–203. (Methods in Molecular Biology). Available from: http://link.springer.com/10.1007/978-1-61779-992-1_11 doi:10.1007/978-1-61779-992-1_11
126. Rojas J, Castillo H, Mundaca M, Fraga J, Gilbert C, Valdebenito-Maturana B, et al. Liver-specific enhancers evolved from independent episodes of MITE domestication in *Xenopus tropicalis*. *Mobile DNA*. 2025 Dec 4;16(1):46. doi:10.1186/s13100-025-00384-7
127. Miao B, Fu S, Lyu C, Gontarz P, Wang T, Zhang B. Tissue-specific usage of transposable element-derived promoters in mouse development. *Genome Biol*. 2020 Dec;21(1):255. doi:10.1186/s13059-020-02164-3
128. Lee HJ, Hou Y, Maeng JH, Shah NM, Chen Y, Lawson HA, et al. Epigenomic analysis reveals prevalent contribution of transposable elements to *cis*-regulatory elements, tissue-specific expression, and alternative promoters in zebrafish. *Genome Res*. 2022 Jul;32(7):1424–36. doi:10.1101/gr.276052.121
129. Monsen Ø, Grønvold L, Datsomor A, Harvey T, Kijas J, Suh A, et al. The role of transposon activity in shaping *cis*-regulatory element evolution after whole-genome duplication. *Genome Res*. 2025 Mar;35(3):475–88. doi:10.1101/gr.278931.124

130. Hadji-Azimi I, Coosemans V, Canicatti C. B-lymphocyte populations in *Xenopus laevis*. *Developmental & Comparative Immunology*. 1990 Dec;14(1):69–84. doi:10.1016/0145-305X(90)90009-4
131. Nogawa-Kosaka N, Sugai T, Nagasawa K, Tanizaki Y, Meguro M, Aizawa Y, et al. Identification of erythroid progenitors induced by erythropoietic activity in *Xenopus laevis*. *Journal of Experimental Biology*. 2011 Mar 15;214(6):921–7. doi:10.1242/jeb.050286
132. Omata K, Nomura I, Hirata A, Yonezuka Y, Muto H, Kuriki R, et al. Isolation and evaluation of erythroid progenitors in the livers of larval, froglet, and adult *Xenopus tropicalis*. *Biology Open*. 2023 Aug 15;12(8):bio059862. doi:10.1242/bio.059862
133. Imai Y, Ishida K, Nemoto M, Nakata K, Kato T, Maéno M. Multiple origins of embryonic and tadpole myeloid cells in *Xenopus laevis*. *Cell Tissue Res*. 2017 Aug;369(2):341–52. doi:10.1007/s00441-017-2601-4
134. Goodwin TJD, Poulter RTM. The DIRS1 Group of Retrotransposons. *Molecular Biology and Evolution*. 2001 Nov 1;18(11):2067–82. doi:10.1093/oxfordjournals.molbev.a003748
135. Wicker T, Sabot F, Hua-Van A, Bennetzen JL, Capy P, Chalhoub B, et al. A unified classification system for eukaryotic transposable elements. *Nat Rev Genet*. 2007 Dec;8(12):973–82. doi:10.1038/nrg2165
136. Gazolla CB, Ludwig A, De Moura Gama J, Bruschi DP. Evolutionary dynamics of *DIRS-like* and *Ngaro-like* retrotransposons in *Xenopus laevis* and *Xenopus tropicalis* genomes. Birchler J, editor. *G3 Genes|Genomes|Genetics*. 2022 Feb 4;12(2):jkab391. doi:10.1093/g3journal/jkab391
137. Kojima KK, Bao W. A unique eukaryotic lineage of composite-like DNA transposons encoding a DDD/E transposase and a His-Me finger homing endonuclease. *Mobile DNA*. 2022 Oct 22;13(1):24. doi:10.1186/s13100-022-00281-3
138. Hikosaka A, Konishi S. Multiple massive domestication and recent amplification of Kolobok superfamily transposons in the clawed frog *Xenopus*. *Zoological Lett*. 2018 Dec;4(1):17. doi:10.1186/s40851-018-0100-4
139. Roelants K, Gower DJ, Wilkinson M, Loader SP, Biju SD, Guillaume K, et al. Global patterns of diversification in the history of modern amphibians. *Proc Natl Acad Sci USA*. 2007 Jan 16;104(3):887–92. doi:10.1073/pnas.0608378104
140. Ducret V, Richards AJ, Videlier M, Scalvenzi T, Moore KA, Paszkiewicz K, et al. Transcriptomic analysis of the trade-off between endurance and burst-performance in the frog *Xenopus allofraseri*. *BMC Genomics*. 2021 Dec;22(1):204. doi:10.1186/s12864-021-07517-1
141. Evans BJ, Carter TF, Greenbaum E, Gvoždík V, Kelley DB, McLaughlin PJ, et al. Genetics, Morphology, Advertisement Calls, and Historical Records Distinguish Six New Polyploid Species of African Clawed Frog (*Xenopus*, Pipidae) from West and Central Africa. Veenstra GJC, editor. *PLoS ONE*. 2015 Dec 16;10(12):e0142823. doi:10.1371/journal.pone.0142823
142. Sun YB, Xiong ZJ, Xiang XY, Liu SP, Zhou WW, Tu XL, et al. Whole-genome sequence of the Tibetan frog *Nanorana parkeri* and the comparative evolution of tetrapod genomes. *Proc Natl Acad Sci USA*. 2015 Mar 17;112(11). doi:10.1073/pnas.1501764112
143. Ge X, Ren J, Gu K, Gong W, Shen K, Feng W. The structure and assembly of the hetero-octameric BLOC-one-related complex. *Structure*. 2025 Feb;33(2):234-246.e6. doi:10.1016/j.str.2024.12.001

144. Wang B, Fu J, Lin Y, Lou Y, Lu A, Yang J. ATG101-related signature predicts prognosis and therapeutic option in hepatocellular carcinoma. *Sci Rep.* 2022 Oct 27;12(1):18066. doi:10.1038/s41598-022-22505-5
145. Kim HK, Lee GH, Bhattarai KR, Lee MS, Back SH, Kim HR, et al. TMBIM6 (transmembrane BAX inhibitor motif containing 6) enhances autophagy through regulation of lysosomal calcium. *Autophagy.* 2021 Mar 4;17(3):761–78. doi:10.1080/15548627.2020.1732161
146. Zhang J, Zuo T, Wang D, Peterson T. Transposition-mediated DNA re-replication in maize. *eLife.* 2014 Nov 18;3:e03724. doi:10.7554/eLife.03724
147. Reams AB, Roth JR. Mechanisms of Gene Duplication and Amplification. *Cold Spring Harb Perspect Biol.* 2015 Feb;7(2):a016592. doi:10.1101/cshperspect.a016592
148. Dunn CA, Romanish MT, Gutierrez LE, Van De Lagemaat LN, Mager DL. Transcription of two human genes from a bidirectional endogenous retrovirus promoter. *Gene.* 2006 Feb;366(2):335–42. doi:10.1016/j.gene.2005.09.003
149. Hollenhorst PC, McIntosh LP, Graves BJ. Genomic and Biochemical Insights into the Specificity of ETS Transcription Factors. *Annu Rev Biochem.* 2011 Jul 7;80(1):437–71. doi:10.1146/annurev.biochem.79.081507.103945
150. Meyer D, Durlat M, Senan F, Wolff M, Andre M, Houdry J, et al. Ets-1 and Ets-2 proto-oncogenes exhibit differential and restricted expression patterns during *Xenopus laevis* oogenesis and embryogenesis. *Int J Dev Biol.* 1997 Aug;41(4):607–20. PubMed PMID: 9303349.
151. Remy P, Baltzinger M. The Ets-transcription factor family in embryonic development: lessons from the amphibian and bird. *Oncogene.* 2000 Dec 18;19(55):6417–31. doi:10.1038/sj.onc.1204044
152. Wang C, Kam RKT, Shi W, Xia Y, Chen X, Cao Y, et al. The Proto-oncogene Transcription Factor Ets1 Regulates Neural Crest Development through Histone Deacetylase 1 to Mediate Output of Bone Morphogenetic Protein Signaling. *Journal of Biological Chemistry.* 2015 Sep;290(36):21925–38. doi:10.1074/jbc.M115.644864
153. Kaczynski J, Cook T, Urrutia R. Sp1- and Kruppel-like transcription factors. *Genome Biol.* 2003;4(2):206. doi:10.1186/gb-2003-4-2-206
154. Hasegawa Y, Struhl K. Different SP1 binding dynamics at individual genomic loci in human cells. *Proc Natl Acad Sci USA.* 2021 Nov 16;118(46):e2113579118. doi:10.1073/pnas.2113579118
155. Ossipova O, Stick R, Pieler T. XSPR-1 and XSPR-2, novel Sp1 related zinc finger containing genes, are dynamically expressed during *Xenopus* embryogenesis. *Mechanisms of Development.* 2002 Jul;115(1–2):117–22. doi:10.1016/S0925-4773(02)00086-2
156. Gao Y, Cao Q, Lu L, Zhang X, Zhang Z, Dong X, et al. Kruppel-like factor family genes are expressed during *Xenopus* embryogenesis and involved in germ layer formation and body axis patterning. *Developmental Dynamics.* 2015 Oct;244(10):1328–46. doi:10.1002/dvdy.24310
157. Bürglin TR, Affolter M. Homeodomain proteins: an update. *Chromosoma.* 2016 Jun;125(3):497–521. doi:10.1007/s00412-015-0543-8

158. Maeda R, Ishimura A, Mood K, Park EK, Buchberg AM, Daar IO. *Xpbox1b* and *Xmeis1b* play a collaborative role in hindbrain and neural crest gene expression in *Xenopus* embryos. *Proc Natl Acad Sci USA*. 2002 Apr 16;99(8):5448–53. doi:10.1073/pnas.082654899
159. Mount SM. A catalogue of splice junction sequences. *Nucl Acids Res*. 1982;10(2):459–72. doi:10.1093/nar/10.2.459
160. Zhang C, Hastings ML, Krainer AR, Zhang MQ. Dual-specificity splice sites function alternatively as 5' and 3' splice sites. *Proc Natl Acad Sci USA*. 2007 Sep 18;104(38):15028–33. doi:10.1073/pnas.0703773104
161. Duan C, Mooney T, Buerer L, Bowers C, Rong S, Kim SW, et al. The unusual gene architecture of polyubiquitin is created by dual-specific splice sites. *Genome Biol*. 2024 Jan 24;25(1):33. doi:10.1186/s13059-023-03157-8
162. Sela N, Mersch B, Gal-Mark N, Lev-Maor G, Hotz-Wagenblatt A, Ast G. Comparative analysis of transposed element insertion within human and mouse genomes reveals Alu's unique role in shaping the human transcriptome. *Genome Biol*. 2007 Jun 27;8(6):R127. doi:10.1186/gb-2007-8-6-r127
163. Sela N, Kim E, Ast G. The role of transposable elements in the evolution of non-mammalian vertebrates and invertebrates. *Genome Biol*. 2010 Jun 2;11(6):R59. doi:10.1186/gb-2010-11-6-r59
164. Smalheiser N, Torvik V. Alu elements within human mRNAs are probable microRNA targets. *Trends in Genetics*. 2006 Oct;22(10):532–6. doi:10.1016/j.tig.2006.08.007
165. Spengler RM, Oakley CK, Davidson BL. Functional microRNAs and target sites are created by lineage-specific transposition. *Human Molecular Genetics*. 2014 Apr 1;23(7):1783–93. doi:10.1093/hmg/ddt569
166. Bastide A, Peretti D, Knight JRP, Grosso S, Spriggs RV, Pichon X, et al. RTN3 Is a Novel Cold-Induced Protein and Mediates Neuroprotective Effects of RBM3. *Current Biology*. 2017 Mar;27(5):638–50. doi:10.1016/j.cub.2017.01.047
167. Palko SI, Benoit MR, Yao AY, Mohan R, Yan R. ER -stress response in retinal Müller glia occurs significantly earlier than amyloid pathology in the Alzheimer's mouse brain and retina. *Glia*. 2024 Jun;72(6):1067–81. doi:10.1002/glia.24514
168. Doycheva D, Kaur H, Tang J, Zhang JH. The characteristics of the ancient cell death suppressor, TMBIM6, and its related signaling pathways after endoplasmic reticulum stress. *J of Neuroscience Research*. 2020 Jan;98(1):77–86. doi:10.1002/jnr.24434
169. Kim HK, Lee GH, Bhattarai KR, Lee MS, Back SH, Kim HR, et al. TMBIM6 (transmembrane BAX inhibitor motif containing 6) enhances autophagy through regulation of lysosomal calcium. *Autophagy*. 2021 Mar 4;17(3):761–78. doi:10.1080/15548627.2020.1732161
170. Krajewska M, Xu L, Xu W, Krajewski S, Kress CL, Cui J, et al. Endoplasmic reticulum protein BI-1 modulates unfolded protein response signaling and protects against stroke and traumatic brain injury. *Brain Research*. 2011 Jan;1370:227–37. doi:10.1016/j.brainres.2010.11.015
171. Jeon K, Lim H, Kim JH, Han D, Lee ER, Yang GM, et al. Bax inhibitor-1 enhances survival and neuronal differentiation of embryonic stem cells via differential regulation of mitogen-

- activated protein kinases activities. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research*. 2012 Dec;1823(12):2190–200. doi:10.1016/j.bbamcr.2012.08.005
172. Derlig K, Ehrhardt T, Giehl A, Brandstätter JH, Enz R, Dahlhaus R. Simiate is an Actin binding protein involved in filopodia dynamics and arborization of neurons. *Front Cell Neurosci*. 2014 Apr 8;8. doi:10.3389/fncel.2014.00099
 173. Sirrs S, Van Karnebeek CD, Peng X, Shyr C, Tarailo-Graovac M, Mandal R, et al. Defects in fatty acid amide hydrolase 2 in a male with neurologic and psychiatric symptoms. *Orphanet J Rare Dis*. 2015 Dec;10(1):38. doi:10.1186/s13023-015-0248-3
 174. Krug RG, Lee HB, El Khoury LY, Sigafos AN, Petersen MO, Clark KJ. The endocannabinoid gene *faah2a* modulates stress-associated behavior in zebrafish. Brennan CH, editor. *PLoS ONE*. 2018 Jan 5;13(1):e0190897. doi:10.1371/journal.pone.0190897
 175. Mencacci NE, Brockmann MM, Dai J, Pajusalu S, Atasu B, Campos J, et al. Biallelic variants in *TSPOAP1*, encoding the active-zone protein RIMBP1, cause autosomal recessive dystonia. *Journal of Clinical Investigation*. 2021 Apr 1;131(7):e140625. doi:10.1172/JCI140625
 176. Taylor EM, Copsey AC, Hudson JJR, Vidot S, Lehmann AR. Identification of the Proteins, Including MAGEG1, That Make Up the Human SMC5-6 Protein Complex. *Molecular and Cellular Biology*. 2008 Feb 1;28(4):1197–206. doi:10.1128/MCB.00767-07
 177. Lorite NP, Apostolova S, Guasch-Vallés M, Pryer A, Unzueta F, Freire R, et al. Crucial role of the NSE1 RING domain in Smc5/6 stability and FANCM-independent fork progression. *Cell Mol Life Sci*. 2024 Dec;81(1):251. doi:10.1007/s00018-024-05275-3
 178. Lerchner W, Latinkic BV, Remacle JE, Huylebroeck D, Smith JC. Region-specific activation of the *Xenopus Brachyury* promoter involves active repression in ectoderm and endoderm: a study using transgenic frog embryos. *Development*. 2000 Jun 15;127(12):2729–39. doi:10.1242/dev.127.12.2729
 179. Ryffel GU, Lingott A. Distinct promoter elements mediate endodermal and mesodermal expression of the HNF1 α promoter in transgenic *Xenopus*. *Mechanisms of Development*. 2000 Jan;90(1):65–75. doi:10.1016/S0925-4773(99)00230-0
 180. Liu X, Green CB. A Novel Promoter Element, Photoreceptor Conserved Element II, Directs Photoreceptor-specific Expression of Nocturnin in *Xenopus laevis*. *Journal of Biological Chemistry*. 2001 Jan;276(18):15146–54. doi:10.1074/jbc.M009970200
 181. Sipe CW, Gruber EJ, Saha MS. Short upstream region drives dynamic expression of hypoxia-inducible factor 1 α during *Xenopus* development. *Developmental Dynamics*. 2004 Jun;230(2):229–38. doi:10.1002/dvdy.20049