



Tesis de Maestría en Biología, Subárea Biología Celular y Molecular, PEDECIBA

# Diapausa I en peces anuales: un análisis transcriptómico comparativo entre embriones del campo y del laboratorio.

Lic. Daniel Blanco

Laboratorio de Biología Molecular de Organismos Acuáticos, Sección Biología Celular, Facultad de Ciencias, UDELAR

Orientadora:

Dra. María José Arezo

Laboratorio de Biología Molecular de Organismos Acuáticos, Sección Biología Celular, Facultad de Ciencias, UDELAR.

Co-orientadores:

Dra. Verónica Gutiérrez<sup>1</sup>; Dr. José Sotelo-Silveira<sup>2,3</sup>

1 Departamento de Biodiversidad y Genética, Instituto de Investigaciones Biológicas Clemente Estable, MEC.

2 Departamento de Genómica, Instituto de Investigaciones Biológicas Clemente Estable, MEC.

3 Sección Biología Celular, Facultad de Ciencias, UDELAR.

TRIBUNAL:

Presidenta: Dra. Ivanna Tomasco.

Vocales: Dres. Inés Carrera y Santiago Fontenla.

## ÍNDICE

<b>Resumen</b> .....	4
<b>Introducción</b> .....	5
Peces anuales sudamericanos.....	5
Características particulares del desarrollo embrionario.....	6
Desarrollo embrionario de los peces anuales en su hábitat natural.....	11
Genes candidatos y vías regulatorias implicadas en las diapausas.....	12
<b>Hipótesis de trabajo</b> .....	14
<b>Objetivo general</b> .....	14
Objetivos específicos. ....	14
<b>Materiales y Métodos</b> .....	15
1. Colecta y mantenimiento de adultos en el laboratorio. ....	15
2. Diseño experimental para la obtención de los estadios embrionarios en laboratorio .....	15
3. Colecta de embriones en el campo.....	16
4. Flujo de trabajo para la secuenciación de los transcriptomas embrionarios.....	16
4.1 Extracción de ARN total.....	17
4.2 Secuenciación masiva del transcriptoma.....	18
4.3 Control de calidad y edición de las secuencias.....	18
4.4 Ensamblado <i>de novo</i> para la obtención del transcriptoma de referencia.....	18
4.5 Anotación funcional del transcriptoma.....	21
4.6 Búsqueda de genes con expresión diferencial .....	22
4.7 Construcción y análisis de redes de co-expresión génica ponderadas.....	23
4.8 Ontología génica y análisis de enriquecimiento de genes.....	23
5. Análisis filogenético y de selección.....	24
<b>Resultados</b> .....	27
Ensamblado <i>de novo</i> del transcriptoma embrionario de <i>Garcialebias charrua</i> .....	27

Análisis preliminares de la expresión génica en diapausa I.....	30
Anotación del transcriptoma mediante el programa Trinotate.....	32
Genes con expresión diferencial entre los fenotipos de diapausa I y los ambientes.....	32
Análisis de Redes de Co-expresión Génica Ponderadas (WGCNA).....	38
Estudios de Ontología génica (GO) y enriquecimiento.....	41
Genes seleccionados de interés.....	43
Análisis de redes vinculadas con los genes de interés.....	46
Análisis filogenético de FoxH1.....	49
Análisis de selección.....	51
Análisis de codón 102.....	54
<b>Discusión.</b> ....	56
Análisis de los genes con expresión diferencial entre campo y laboratorio.....	56
Estudios de Ontología Génica.....	58
Redes de enriquecimiento de FoxH1.....	59
Análisis filogenético de FoxH1.....	61
Mutaciones en los codones que tienen selección purificadora.....	63
<b>Conclusiones</b> .....	64
<b>Perspectivas</b> .....	65
<b>Agradecimientos.</b> ....	67
<b>Referencias</b> .....	68
<b>Anexos</b> .....	82

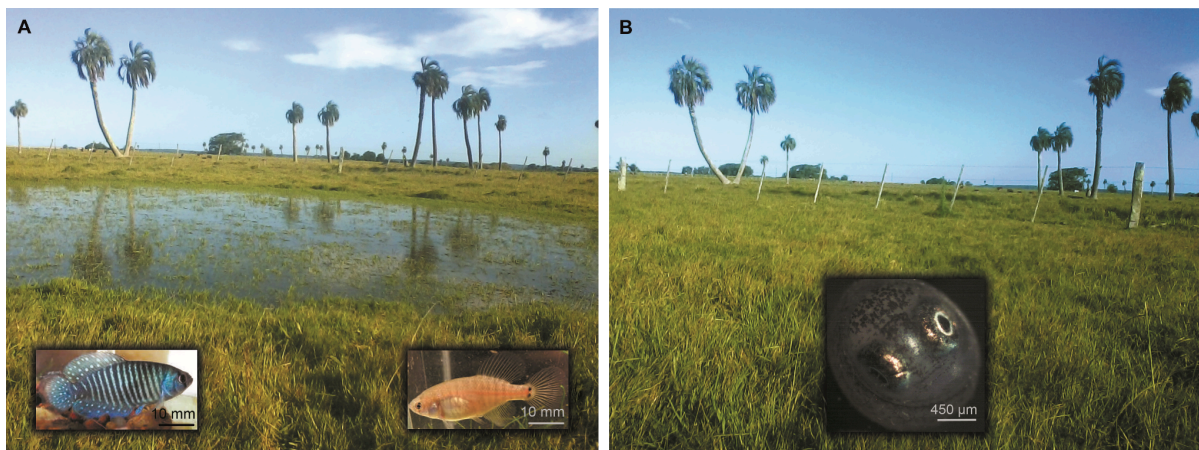
## Resumen

Los estados de dormancia permitieron a los seres vivos conquistar regiones extremas del planeta mediante adaptaciones a ambientes hostiles. Las diapausas de peces anuales son ejemplos de dormancia funcionado como estrategia de supervivencia a la desecación de los charcos estacionales donde habitan. Durante estas tres detenciones, las actividades vitales disminuyen hasta que las señales ambientales determinan su reactivación. A pesar del papel crucial de las diapausas en el ciclo de vida de los peces anuales, los mecanismos moleculares implicados principalmente en la diapausa I, son desconocidos. Nuestro grupo de investigación se centra en comprender los mecanismos involucrados en la regulación de las diapausas con un enfoque interdisciplinario. En este contexto, se inició un estudio de la expresión génica diferencial mediante un análisis comparativo entre el repertorio de ARNm de embriones en diapausa I del pez anual sudamericano *Garcialebias charrua*, obtenidos en dos condiciones ambientales diferentes (laboratorio y naturaleza). A partir de estos ARNs, se obtuvo un ensamblado *de novo* con 10069 transcriptos totales. El análisis de expresión génica diferencial detectó 5760 transcriptos que cambian su expresión entre las diapausas I obtenidas en condiciones de laboratorio y en el campo. Entre los fenotipos de diapausa I (A y B) de los embriones inducidos en el laboratorio se detectaron 60 transcriptos expresados diferencialmente mientras que en los mismos fenotipos obtenidos en la naturaleza, se detectaron 62 transcriptos con expresión diferencial. Los análisis de componentes principales y de correlación de Pearson evidenciaron claras diferencias entre los embriones en diapausa I colectados en el campo y aquellos inducidos en el laboratorio, indicando que las condiciones ambientales influyen fuertemente en la expresión génica durante esta detención. Mediante el análisis de coexpresión se identificaron grupos de genes con expresión específica en embriones provenientes del campo enriquecidos en funciones moleculares asociadas a la regulación transcripcional, unión al ADN, protección frente al daño genómico y regulación enzimática. Los análisis de ontología génica y de enriquecimiento funcional mostraron la participación de vías de señalización como la IGF, la TGF- $\beta$ /SMAD, y la Nodal. Entre los genes de interés identificados se destacaron los factores de transcripción FoxH1 y Homeobox-like 4, una proteína de choque térmico frío, el elemento transponible LINE1 y un integrante de la vía IGF (IGF2BP) cuya máxima expresión se observó en las diapausas I del campo. Estos genes representan potenciales reguladores de la dormancia embrionaria en respuesta a las señales ambientales. El análisis filogenético del gen FoxH1 mostró una fuerte conservación funcional entre especies de peces anuales. Los estudios de selección molecular revelaron regiones bajo presión selectiva purificadora y codones bajo selección diversificadora episódica. Estos resultados demuestran que la regulación de la diapausa I en *G. charrua* depende de una interacción compleja entre factores ambientales y redes genéticas altamente conservadas, pero modulables, lo que otorga a esta especie una notable plasticidad adaptativa. Este trabajo aporta evidencia sobre el efecto que ejercen las señales ambientales en el campo y en el laboratorio. El estudio de las diapausas en peces anuales permite plantear preguntas que entrelazan las dimensiones ecológica, evolutiva y de biología del desarrollo.

## Introducción

Los organismos extremófilos habitan regiones con condiciones adversas y presentan mecanismos de adaptación (dormancia) que les permiten sobrevivir en esos ambientes. Representantes de diferentes niveles evolutivos experimentan estos estados tan especiales. Son ejemplos de dormancia la criptobiosis en tardígrados y rotíferos, la hibernación en un insecto del Ártico, la esporulación en las levaduras, el enquistamiento del crustáceo *Artemia*, la dormancia en las semillas de las plantas y las detenciones reversibles en el desarrollo embrionario o diapausas en más de 130 especies de mamíferos y en los peces anuales (Lubzens y col. 2010; Fenelon y col. 2014; Hand y col. 2016).

Los peces anuales (Cyprinodontiformes) son teleósteos dulceacuícolas de América y África. Habitan charcos temporales con marcadas fluctuaciones en factores abióticos (pH, turbidez, oxígeno disuelto y temperatura) que implican para los peces una gran adaptación y tolerancia al estrés ambiental (Podrabsky y col. 2016) y una historia de vida que consta de dos fases: seca y húmeda. Durante la fase húmeda (época de lluvias), los adultos se reproducen en los charcos generando embriones que permanecen en el fondo de los mismos. Al iniciarse la estación seca, las masas de agua desaparecen causando la muerte masiva de la población adulta (Figura 1). Por lo tanto, la supervivencia de estas especies depende de los embriones enterrados en el sustrato (Wourms, 1972a).



**Figura 1.** Ciclo de vida de los peces anuales en las masas de agua dulce temporales que habitan. **A:** charco inundado durante la estación lluviosa. Se muestra un macho (izquierda) y una hembra (derecha) de la especie *Garcialebias charrua*. **B:** el mismo charco durante la estación seca (verano), momento en el que los adultos han desaparecido y los embriones permanecen en una etapa de detención del desarrollo previa a la eclosión. Tomada de Berois y col. (2024, p. 51).

## Peces anuales sudamericanos

Dentro de los peces anuales sudamericanos se destaca el grupo de géneros *Austrolebias* (Cyprinodontiformes, Rivulidae) (Costa 2006, Alonso y col. 2023), un modelo emergente de peces anuales (Berois y col. 2014). Este grupo incluye 11 géneros y 61 especies distribuidas en la cuenca de los ríos Paraná-La Plata, en el sistema de lagunas costeras Patos-Merín y en la cuenca amazónica del sudoeste de América (Alonso y col. 2023). Es uno de los grupos más extensos de peces anuales neotropicales y para Uruguay están citadas más de 20 especies, algunas de las cuales son endémicas y otras son compartidas con Brasil y Argentina (Loureiro y col. 2018, Alonso y col. 2023). Al igual que

otros teleósteos, estos peces presentan características ventajosas para la investigación del desarrollo embrionario tales como la oviparidad, la transparencia de los embriones, largos períodos de reproducción y la facilidad de mantenimiento en condiciones de laboratorio. Asimismo, presentan algunos atributos particulares relacionados con su singular ciclo de vida. Han desarrollado un ciclo de vida anual (menor a un año) que les permite sobrevivir y reproducirse en ese entorno efímero (Wourms, 1967) y son considerados uno de los grupos de vertebrados con embriones extremófilos más peculiares (Berois y col. 2012).

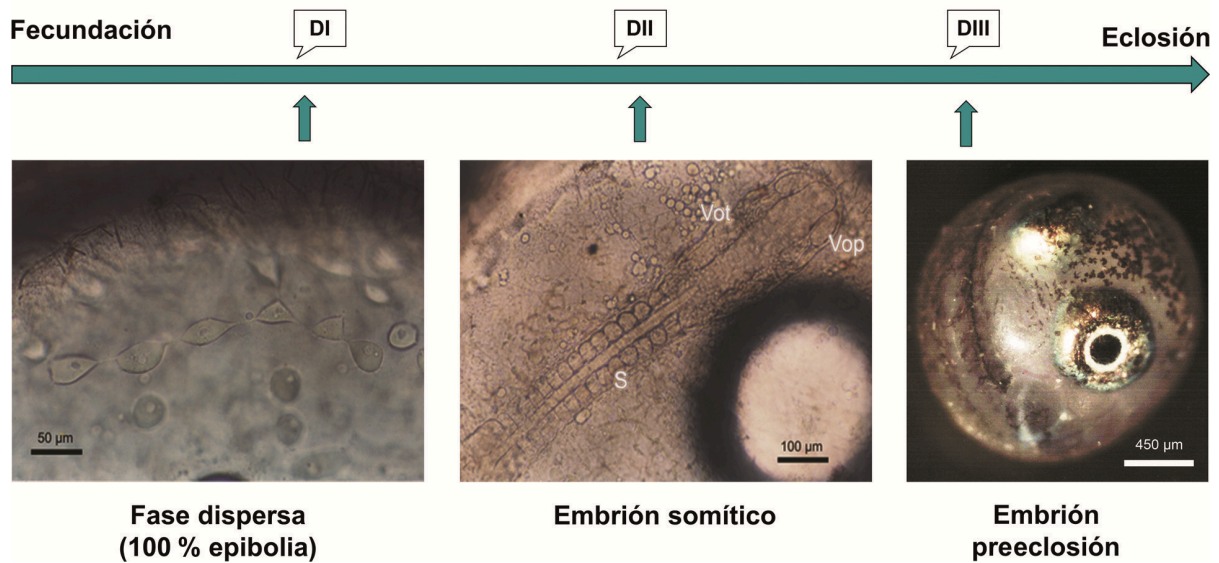
Las especies del grupo de géneros *Austrolebias* presentan un tamaño genómico de aproximadamente  $5,95 \pm 0,45$  pg/célula diploide, similar al del genoma humano, mientras que *Cynopoecilus melanotaenia*, una especie anual con una distribución geográfica restringida al sureste de Uruguay y sur de Brasil, presenta un tamaño de  $2,72 \pm 0,06$  pg/célula diploide (García y col. 2014). El drástico aumento de ADN nuclear encontrado en especies de este grupo estaría asociado con un aumento considerable en la proporción de elementos transponibles (ETs) (García y col. 2015, Gajardo-Escobar y col. 2026). La mayoría de estas especies son endémicas de áreas geográficas muy restringidas. Su ubicación y distribución limitada, junto con la grave transformación del uso del suelo que se está produciendo en la región (deforestación, cultivos de arroz y soja, minería a cielo abierto y turismo), sitúan a muchas de ellas en la Lista Roja de la UICN (Loureiro y col. 2018; Garcia y col. 2025).

Dentro de este grupo, *Garcialebias charrua*, anteriormente conocida como *Austrolebias charrua* (Costa y Cheffe, 2001; Alonso y col. 2023), se distribuye en tierras bajas que incluyen los Bañados del Este, una Reserva de la Biosfera y Sitio Ramsar del este de Uruguay (Probides, 1999) y el sur de Brasil. Esta especie ha mostrado una adaptación óptima a las condiciones de laboratorio, por lo tanto, ha sido posible analizarla mediante diferentes enfoques multidisciplinarios como la proteómica (Chalar y col., 2021), la transcriptómica (Papa y col. 2024) y la genómica (Gajardo-Escobar y col. 2026)

### **Características particulares del desarrollo embrionario**

Los embriones de los peces anuales son resistentes al estrés ambiental (deseccación, temperaturas variables e hipoxia severa) (Podrabsky y col. 2016). Durante el desarrollo temprano, ocurre una fase de dispersión y reagregación (D/R) de blastómeras que desacopla la epibolia (primer movimiento de gastrulación en peces) del ingreso del endomesodermo en las dimensiones temporal y espacial, a diferencia de lo que ocurre en los teleósteos no anuales. Aunque este proceso fue descrito hace varias décadas (Wourms, 1972a,b,c), el mecanismo molecular que lo controla aún no ha sido dilucidado. Se ha evidenciado que el ARNm del gen *cordina* (asociado con la formación del eje dorso-ventral en vertebrados y muy conservado en la evolución), presenta un patrón de expresión único en el pez anual sudamericano *Austrofundulus limnaeus* comparado con otros teleósteos. Este hecho indicaría un posible rol de este gen durante la fase D/R específica del desarrollo de los peces anuales (Wagner y Podrabsky, 2015). Otra particularidad de este grupo de peces es que los embriones son capaces de detener el desarrollo en tres estadios diferentes: diapausa I (DI), diapausa II (DII) y diapausa III (DIII) (Figura 2) (Wourms, 1972c).

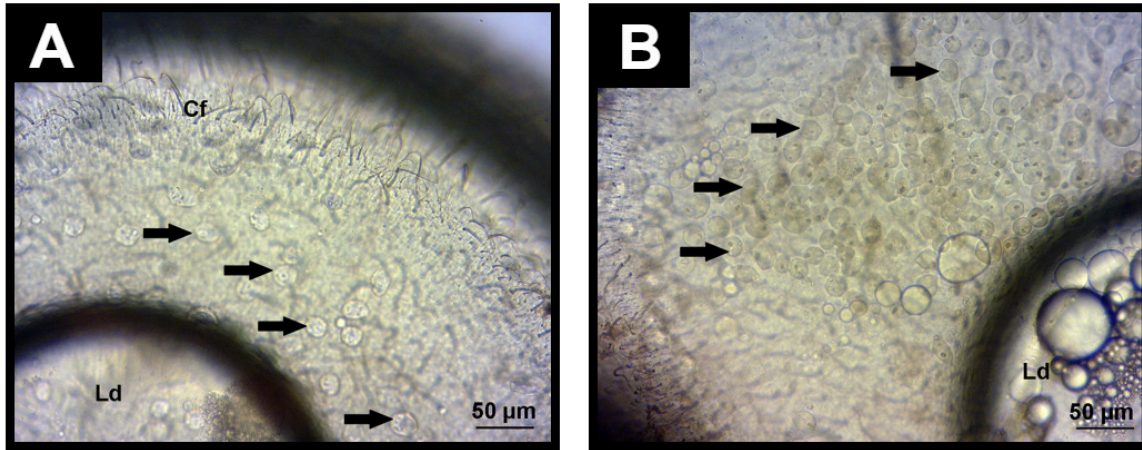
## Diapausas en peces anuales



**Figura 2.** Ubicación de las diapasas durante el desarrollo de un pez anual, desde la fecundación hasta la eclosión. Las imágenes corresponden a embriones de *Garcialebias charrua* y representan las etapas de fase dispersa (100 % de epibolia; DI), embrión somítico (DII) y embrión preeclosión (DIII). S: somites; Vop: vesículas ópticas; Vot: vesículas óticas. Tomada de Berois y col. (2024, p. 57).

Cada una de las diapasas presenta una duración variable y originalmente se las clasificó en facultativas u obligatorias dependiendo del género (Wourms 1972c). Sin embargo, Podrabsky y Wagner, 2016 han sugerido que las diapasas son siempre facultativas en peces anuales, pero se vuelven obligatorias en ciertas circunstancias ambientales.

La DI ocurre durante la fase de dispersión total de blastómeras (100% epibolia), es inducible por bajas temperaturas e hipoxia y ha sido interpretada como un control temprano ejercido por el ambiente sobre el desarrollo (Wourms, 1972b; Arezo y col. 2017). En los peces anuales africanos del género *Nothobranchius*, esta detención fue descrita como consecuencia de una señal química secretada por los peces adultos (Inglima y col. 1981; Denucé 1989) hasta ahora no identificada. Asimismo, se ha demostrado, mediante experimentos *in vivo* en embriones de *N. furzeri*, que las células dispersas en DI no se dividen (Dolfi y col. 2019). En *G. charrua* se ha evidenciado la inducción de esta diapausa en embriones cultivados en presencia de adultos, bajas temperaturas e hipoxia. Además, se ha observado el ingreso a la DI de manera espontánea (en ausencia de señales de inducción) y la existencia de dos fenotipos alternativos de la misma: DIA y DIB. Estos fenotipos alternativos (Figura 3) aportarían una mayor diversidad entre la población de embriones, los cuales responderían de forma diferencial a las señales ambientales (Arezo y col., 2017).



**Figura 3.** Fenotipos de la DI identificados en *G. charrua*. **A)** DIA: las blastómeras profundas (flechas) se encuentran separadas unas de otras, **B)** DIB: las blastómeras profundas (flechas) se encuentran formando agrupaciones. Modificado de Arezo y col. 2017 (Ld: gota lipídica, Cf: filamentos coriónicos).

Recientemente, se han identificado 695 transcritos cuya expresión varía en embriones de *G. charrua* inducidos a DI en condiciones de laboratorio, respecto al estadio más temprano del desarrollo embrionario (estadio materno: 1 a 16 células). Ese análisis transcriptómico reveló que las vías de señalización asociadas a la diapausa en otras especies (por ejemplo: insulina/IGF, vitamina D, Wnt y polycomb), así como las proteínas de choque térmico, también están involucradas en la inducción de la DI en esta especie (Papa y col. 2024). Asimismo, se ha documentado una expresión diferencial de genes relacionados con la replicación del ADN, el ciclo celular y otros procesos biológicos al comparar embriones en DI y DIII de *N. guentheri* (Gao y col. 2022).

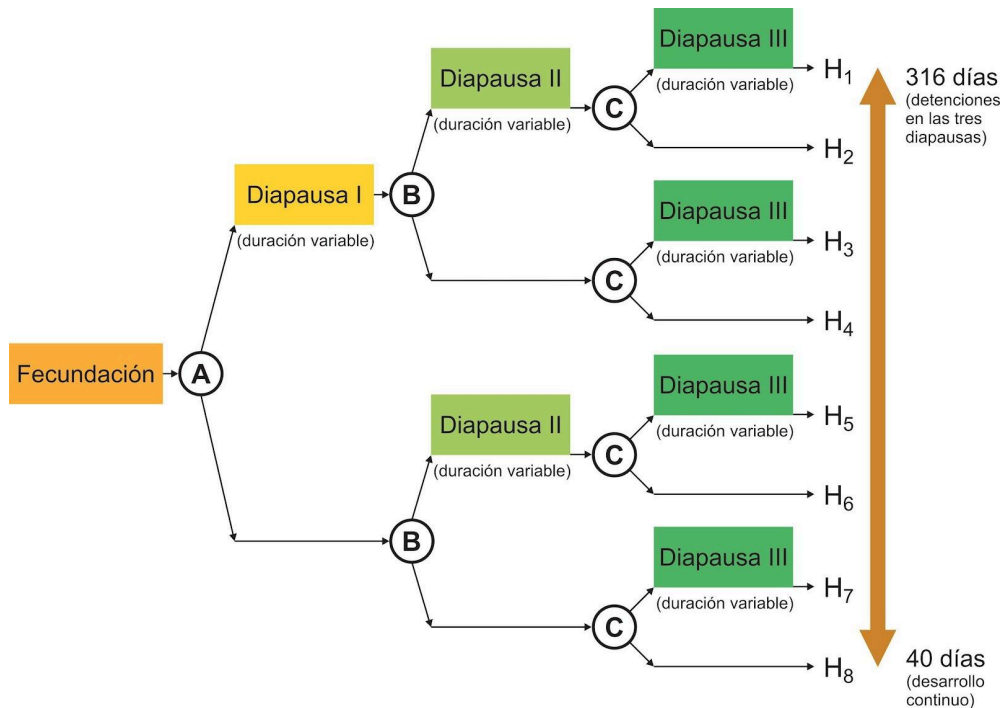
La DII ocurre durante la organogénesis (estadio de 35-40 somites). Los datos más significativos sobre esta etapa, que es la más estudiada, se han obtenido en *A. limnaeus* (Podrabsky y Hand, 2015; Clouser y col. 2025). Los embriones en DII de *A. limnaeus* presentan el metabolismo deprimido en un 90% y un alto nivel de tolerancia a la anoxia (Podrabsky y Hand, 1999; Podrabsky y col. 2007). La vía de señalización del factor de crecimiento similar a la insulina (IGF), el procesamiento alternativo de ARNm y pequeños ARNnc (no codificantes) estarían regulando esta etapa (Romney y Podrabsky, 2017; Woll y Podrabsky, 2017). Asimismo, Romney y col. (2018) demostraron que la vía de señalización de la vitamina D es capaz de integrar la información ambiental a los programas de desarrollo asociados con las trayectorias alternativas de estos embriones, de manera que la síntesis de vitamina D3 promueve el desarrollo activo, mientras que su inhibición conduce a la entrada en DII. Recientemente, Clouser y col. (2025) demostraron que la condición de anoxia experimentada durante esta detención en *A. limnaeus* despliega una respuesta transcriptómica al estrés consistente con los desafíos para la homeostasis celular en la anoxia. Otros análisis transcriptómicos de la DII en *N. furzeri* revelaron que, además de ser mantenida activamente por reguladores específicos de la cromatina, involucra cambios fisiológicos que protegen al organismo del envejecimiento, preservándolo durante períodos prolongados (equivalentes a la vida media de los adultos), sin consecuencias en el crecimiento, fertilidad y expectativa de vida del organismo (Hu y col. 2020). Estudios integrativos multiómicos mostraron que la DII evolucionó mediante una remodelación

reciente de elementos reguladores en duplicados de genes muy antiguos (parálogos) presentes en todos los vertebrados. Utilizando técnicas de edición génica basadas en CRISPR-Cas9, identificaron los factores de transcripción REST/NRSF y FOXO como elementos cruciales para el programa de expresión génica en la DII, incluyendo genes involucrados en el metabolismo lipídico (la diapausa muestra un perfil lipídico distintivo, con un aumento de triglicéridos con ácidos grasos de cadena muy larga) (Singh y col. 2024).

La DIII ocurre en la etapa pre-eclosión (Wourms, 1972c). En esta diapausa el metabolismo decrece progresivamente comparado con los embriones inducidos a eclosionar (Levels y col., 1986; Podrabsky y Hand, 1999). Análisis transcriptómicos, han reportado 945 genes expresados diferencialmente entre embriones en DIII y larvas del pez anual sudamericano *Nematolebias whitei* (Thompson y Orti, 2016) así como la importancia del metabolismo de la glucosa entre embriones en etapa de eclosión y en DIII del africano *N. guentheri* (Gao y col. 2022). Un estudio enfocado en la expresión diferencial de diversas proteínas entre embriones en DIII y pre-eclosión de *G. charrua*, reveló un cambio de expresión en varias proteínas involucradas en vías metabólicas, tolerancia al estrés, reparación del ADN, modificaciones epigenéticas y cambios en proteínas del citoesqueleto (Chalar y col. 2021).

Muchas de las vías de señalización implicadas en la regulación de las diapausas en diferentes organismos, también regulan el crecimiento durante el desarrollo activo (Karp, 2021). Se ha propuesto que la regulación negativa de estas vías promotoras del crecimiento es un mecanismo molecular común en diferentes especies de invertebrados y vertebrados que permitió la evolución de las diapausas (Romney y col. 2018; Karp, 2021). La vía de señalización de la insulina/IGF funcionaría universalmente como un mecanismo anti-diapausa que controlaría la fisiología celular mediante modificaciones metabólicas. Asimismo, la señalización por receptores hormonales nucleares (NHR) también regula el ingreso a la diapausa. Sin embargo, el tipo de receptores involucrados no se conservan entre las distintas especies (Karp, 2021).

Las diapausas en peces anuales son una estrategia de supervivencia en un ambiente efímero, cambiante y hostil para un teleósteo. Durante estas etapas se produce una reducción o suspensión de las actividades fisiológicas hasta que las condiciones ambientales se vuelven favorables con la llegada de las lluvias. Los embriones pueden desarrollarse continuamente o atravesar las diapausas con duraciones variables, por lo que embriones de un mismo desove pueden desarrollarse y eclosionar en momentos diferentes. Esta adaptación fisiológica en el control del desarrollo embrionario se denomina “efecto multiplicador” (Figura 4). Desde una perspectiva ecológica, este fenómeno introduce diferentes tiempos de desarrollo en la trayectoria embrionaria y constituye una ventaja adaptativa en un ambiente tan cambiante como es el de los charcos temporales (Wourms, 1972c).



**Figura 4.** Esquema que representa el “efecto multiplicador”, donde se muestran las trayectorias alternativas (A, B y C) que puede seguir un embrión de pez anual. Los tiempos de cada trayectoria son especie-específicos. H1-H8, representan los distintos tiempos totales posibles de desarrollo. Tomado de Berois y col. (2024, p. 60).

La respuesta embrionaria a los factores ambientales (plasticidad fenotípica) genera, en sentido probabilístico, la reacción apropiada a las condiciones ambientales. Embriones de un mismo desove, cultivados en condiciones similares a su ambiente natural pero sin información sobre el momento de la estación anual, muestran variabilidad en la entrada, duración y salida de las diapausas lo que sugiere la influencia de factores intrínsecos como la variabilidad genética. Este fenómeno es consistente con la estrategia de “bet-hedging” (no apostar la eclosión de los embriones a un mismo momento) que genera un amplio rango de fenotipos que maximizan la adaptación a estos ambientes (Furness y col. 2015). Por ejemplo, en *N. furzeri* los embriones son capaces de sobrevivir largos períodos en diapausa obteniéndose eclosiones exitosas luego de tres años de almacenamiento (Cellerino y col., 2015).

Por lo planteado anteriormente, la entrada en las DI y DII o continuación hasta la DIII y eclosión depende de varios factores tales como la edad materna, el fotoperíodo, la temperatura, la hipoxia, la presencia de adultos y la variabilidad genética poblacional (Podrabsky y col. 2016). Los mecanismos involucrados en la transferencia e integración de información de los adultos a los embriones son probablemente epigenéticos (Podrabsky y col. 2010). La producción de embriones diapáusicos dependería de la presencia de determinados tipos de ARNs o proteínas mientras que variantes de procesamiento alternativo de ARNs generarían embriones de escape (desarrollo continuo) (Romney y Podrabsky, 2017). Sin embargo, es importante considerar que, aunque las influencias maternas comprometerían al embrión hacia determinada trayectoria, los embriones responderán a las claves ambientales experimentadas superando a las señales de origen materno (Levels y Denuce, 1988; Podrabsky y col. 2016).

La trayectoria de desarrollo continuo es típica de los peces no anuales del género Cyprinodontiformes y de los teleósteos en general, y posiblemente sea el mecanismo ancestral. La diapausa habría evolucionado por evolución convergente (Furness y col. 2015).

### **Desarrollo embrionario de los peces anuales en su hábitat natural**

Los peces anuales actualmente están atrayendo la atención de los investigadores debido a las particularidades mencionadas de su biología, la posibilidad de mantenerlos en condiciones de laboratorio (Berois y col. 2012), ser un valioso modelo en la investigación sobre las diapausas y la senescencia (Hu y col. 2020; Singh y col. 2024) y su fuerte vínculo con los humedales como un ecosistema en peligro global de extinción, especialmente a la luz del cambio climático en curso (Erwin, 2009, García y col. 2025).

Como se mencionó anteriormente, estos peces poseen la capacidad de desarrollarse activamente frente a condiciones ambientales muy variables, a menudo extremas, que la mayoría de las especies no soportaría o que provocarían un desarrollo embrionario anormal. Este grupo proporciona una oportunidad incomparable para estudiar la evolución de fenotipos alternativos del desarrollo (desarrollo directo en comparación con la diapausa) y la alta tolerancia a factores de estrés medioambientales. Actualmente, existe un creciente interés en conocer cómo se integran estos factores extrínsecos en los programas de desarrollo embrionario para comprender las consecuencias ecológicas y evolutivas de posibles patrones de desarrollo alterados. En este sentido, es esencial trabajar con organismos en un contexto ambiental y ecológicamente relevante. Los organismos tradicionalmente utilizados como modelo en estudios sobre la biología del desarrollo, no siempre son la mejor opción para responder estas interrogantes, ya que corresponden principalmente a líneas de laboratorio alejadas de su propia historia de vida en la naturaleza. Por lo tanto, hay una creciente necesidad de expandir el enfoque hacia estudios sobre el desarrollo embrionario en organismos que sean adecuados para experimentar en el laboratorio pero que, a su vez, brinden la oportunidad de investigar su historia de vida en un contexto ecológico-ambiental. Esto habilita un marco relevante para abordar preguntas que integran las dimensiones ecológica, evolutiva y de biología del desarrollo (Podrabsky y Arezo, 2017).

Los datos básicos sobre la historia de vida y la ecología de las poblaciones de peces anuales son muy escasos. En América del Sur, los humedales están ubicados en países en desarrollo que juegan un papel esencial en la producción internacional de cultivos y contienen charcos temporales que albergan una gran diversidad de especies de peces anuales en riesgo de extinción (García y col. 2025). A su vez, la principal falta de conocimiento en relación a la historia de vida de los peces anuales en la naturaleza radica principalmente en el desarrollo embrionario. Los primeros datos en esta área provienen del pez anual de Méjico, *Millerichthys robustus*, en el cual se reportó que todos los embriones extraídos del sustrato se encontraban en DI (Domínguez-Castanedo y col. 2013).

Un estudio más exhaustivo sobre el curso del desarrollo en el ambiente natural fue realizado por Polacik y col. (2021) en peces anuales africanos. En ese trabajo de campo se observó que durante la fase de inundación y el pico de la fase seca, el banco de embriones muestra una estricta sincronía en el desarrollo en respuesta a los factores ambientales imperantes. Mientras los charcos están inundados, los embriones permanecen detenidos en DI (señal ambiental confiable: hipoxia). Durante el pico de la estación seca, los embriones

se encuentran en DII (señal ambiental confiable: aumento de oxígeno en el sustrato). Hacia el comienzo y el final de la estación seca, donde las señales ambientales son más ambiguas, los embriones se encuentran distribuidos en DII, post-diapausa II y DIII. En estas fases del ciclo, los embriones podrían experimentar la estrategia de “bet-hedging” (Furness y col. 2015) en respuesta a condiciones ambientales inciertas. Esa asincronía en el desarrollo, durante esas fases del ciclo, sería una consecuencia del cambio en las micro-condiciones de incubación donde el microhábitat de cada embrión experimentaría cambios continuos pero graduales (por ejemplo: el secado progresivo de superficie al fondo del charco) (Polacik y col. 2021).

Si bien los datos obtenidos en el campo no demuestran un vínculo causal directo entre los factores ambientales y el desarrollo embrionario, sí revelaron patrones de desarrollo claros asociados a cambios en determinadas condiciones ambientales (factor ambiental unificador: inundación/sequía). Estos hallazgos son robustos y generalizables ya que son replicables entre poblaciones de distintos charcos y fueron observados en distintas especies africanas de peces anuales y en regiones geográficas distantes. Asimismo, estas observaciones están apoyadas por diseños experimentales en el laboratorio (Polacik y col. 2021). Este mismo patrón fue corroborado en algunas especies anuales sudamericanas, *Austrolebias bellottii*, *Titanolebias elongatus*, *Argolebias nigripinnis*, *G. charrua*, y *Acantholebias luteoflammulatus* que habitan en Uruguay (Polacik y col. 2023).

Estos trabajos mostraron que el desarrollo embrionario observado en el laboratorio difiere del identificado en la naturaleza. En el laboratorio la dinámica es principalmente asincrónica y en la naturaleza es sincrónica prácticamente durante todo el ciclo (excepto en los momentos de inicio o fin de la fase seca) (Polacik y col. 2021, 2023).

Con respecto a la DI, en condiciones de laboratorio, se ha reportado que esta detención es extremadamente rara o ausente en embriones de *Argolebias nigripinnis* (Peters, 1963), *Austrofundulus myersi*, *Nothobranchius furzeri* (Wourms, 1972a), *Aphyosemion*, *Cynopoecilus melanotaenia* (Wourms, 1972c), *Garcialebias viarius* (Arezo y col., 2005) y *A. limnaeus* (Wagner y Podrabsky, 2015) cultivados a 25 °C. Sin embargo, los trabajos de Polacik y col. (2021, 2023) mostraron que, durante la fase de inundación, los embriones de las especies analizadas se encuentran detenidos en DI. Estos hallazgos evidencian la relevancia ecológica que presenta la DI sugerida previamente por Arezo y col. (2017).

## **Genes candidatos y vías regulatorias implicadas en las diapausas**

En diversos trabajos se han estudiado funciones moleculares y procesos biológicos de varios genes involucrados en las diapausas. En *A. limnaeus*, se encontró que la vía de señalización de IGF podría desempeñar un papel relevante en las transiciones metabólicas y en la regulación de los estadios diapáusicos (Woll y Podrabsky, 2017).

En los últimos años, se destacan los factores de transcripción de la familia Forkhead box (FOX), ampliamente conocidos por su papel en la regulación del ciclo celular, la diferenciación, el metabolismo y la respuesta al estrés (Lam y col. 2013). En los peces anuales, los FOX han sido implicados directamente en la regulación de distintos estadios de diapausa. En *A. limnaeus*, se han identificado a miembros de la familia FOX entre los candidatos asociados a la entrada, mantenimiento y salida de la DII (Woll y Podrabsky, 2017). Análisis proteómicos en *G. charrua* también han detectado proteínas de la familia FOX específicamente enriquecidas durante la DIII, lo que sugiere un papel funcional en el mantenimiento o establecimiento de esta diapausa (Chalar y col. 2021). La importancia de

esta familia no se limita a los peces anuales. Revisiones comparadas en distintos organismos que experimentan dormancia (incluyendo insectos, nemátodos y mamíferos), destacan que los factores FOX, en particular los subgrupos FOXO, actúan como nodos centrales en la regulación de vías de supervivencia, control del metabolismo, resistencia al estrés y detención del ciclo celular, todos procesos fundamentales en las diapausas (Webb y Brunet, 2014; Martins y col. 2016). Esta convergencia funcional sugiere que la participación de los FOX en la dormancia es un rasgo evolutivamente conservado en muchos grupos, géneros y especies (Sweet y Hu 2025).

Un trabajo reciente sugiere la existencia de un núcleo de genes conservados entre especies de peces anuales involucrados en la regulación de la DII (Thompson y col. 2024). Entre estos se encuentran genes Homeobox, una familia de factores de transcripción fundamentales en la regulación de programas de desarrollo embrionario (McGinnis y Krumlauf, 1992; Keynes y Krumlauf, 1994). En peces anuales del grupo de géneros *Austrolebias* ya se han reportado genes Hox, destacando su posible relevancia en procesos de desarrollo en estos peces (Gutiérrez y col. 2007). Asimismo, en otros organismos se ha asociado la regulación de genes Homeobox con estados de dormancia y reprogramación metabólica. En insectos en diapausa se han detectado cambios en la expresión génica relacionados con el control del desarrollo y con la regulación metabólica (Bradshaw y Holzapfel, 2010; Bao y Xu, 2011). En mamíferos, además, se ha demostrado la participación de genes Homeobox específicos en la regulación de la diapausa embrionaria durante etapas tempranas del desarrollo (Cha y col. 2020).

En este contexto, el presente trabajo se enfocó en estudiar los genes expresados diferencialmente y los grupos de genes co-expresados entre embriones en DI de *G. charrua* colectados en el campo y obtenidos en el laboratorio.

## **Hipótesis de trabajo:**

Los embriones detenidos en DI (fenotipos A y B) inducidos en condiciones de laboratorio son similares, a nivel transcriptómico, a los colectados en su hábitat natural.

## **Objetivo general**

Aportar al conocimiento de las diapausas de *Garcialebias charrua*, mediante la caracterización de los perfiles de expresión diferencial y genes co-expresados en la DI (A y B) obtenidas en condiciones naturales y en el laboratorio.

## **Objetivos específicos**

1. Obtener ARN total de embriones de *G. charrua* en DI (fenotipos A y B) colectados en la naturaleza e inducidos en el laboratorio, así como de las etapas anteriores a la DI (materno, epibolia) y de la etapa posterior (reagregado) en condiciones de laboratorio.
2. Realizar el ensamblado *de novo* del transcriptoma de referencia para los estadios embrionarios mencionados a partir de las secuencias obtenidas.
3. Estudiar la población de ARN mensajeros de los estadios embrionarios mediante secuenciación masiva y analizar la expresión diferencial de ARN mensajeros de embriones en DI con respecto a los estadios del desarrollo previamente mencionados.
4. Analizar los grupos de genes co-expresados en embriones de *G. charrua* en la DI de campo y laboratorio.
5. Identificar diferentes genes candidatos vinculados a la diapausa (vía IGF, proteínas de choque térmico, Forkhead, Homeobox) en los transcriptomas ensamblados.
6. Evaluar las diferencias y similitudes entre los embriones detenidos en DI colectados en el campo y los inducidos en el laboratorio.
7. Realizar análisis filogenéticos y de selección en alguno de los genes de interés con función relevante en la DI.

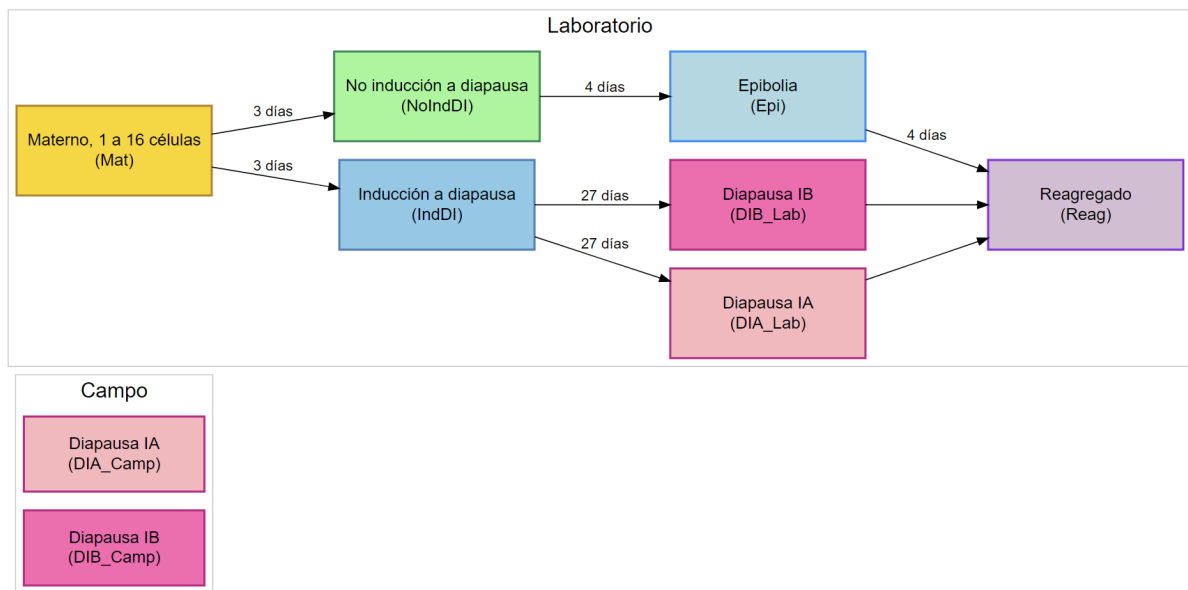
## **Materiales y Métodos**

### **1. Colecta y mantenimiento de los adultos en el laboratorio**

Se colectaron hembras y machos adultos de *G. charrua* en agosto del 2019 en un charco ubicado en La Coronilla (Departamento de Rocha, Uruguay, 33°53'46.5"S 53°30'56.2"W). Luego del período de cuarentena, se colocaron 2 o 3 hembras junto con 1 macho en los tanques de reproducción de 20 litros, dentro del sistema de recirculación, siguiendo las condiciones descritas en Papa y col. (2016) y de acuerdo al protocolo aprobado por la Comisión Honoraria de Experimentación Animal (CHEA) de la Universidad de la República (código de protocolo 240011-002308-14). En el sistema de recirculación, los peces fueron mantenidos en un ciclo de 12 horas de luz y 12 horas de oscuridad y a una temperatura del agua de 19 °C, con alimentación viva ofrecida diariamente (*Daphnia magna* y *Enchytraeus sp.*). El desove se realizó de forma natural en el sustrato ofrecido para el mismo que contenía esferas de vidrio (500 µm de diámetro Thomas Scientific Swedesboro USA) en el fondo (Papa y col. 2016). Los eventos de desove se limitaron a un período entre 5 a 12 horas para asegurar la recolección de embriones entre 1 y 16 células.

### **2. Diseño experimental para la obtención de los estadios embrionarios en laboratorio**

Se recolectaron los huevos recién fecundados y los embriones viables se dividieron en distintos grupos. Un grupo, denominado estadio materno (Mat), contenía embriones entre 1 y 16 células. Otro grupo contenía embriones sin señales de inducción a diapausa (NoInDI), obtenidos a partir de embriones en Mat cultivados en agua de acuario sin adultos. El agua de acuario estaba desionizada, restituida con sal sintética para acuarios marinos Tetra Marine Salt Pro, con una conductividad de 300 µS y pH entre 7 y 7,5 a 19 °C durante 72 horas. Posteriormente se obtuvieron los estadios de epibolia 80-90% (Epi) (Arezo y col. 2005) y reagregado II (Reag) (Pereiro y col. 2017) cultivando los embriones desde Mat hasta alcanzar estos estadios. En paralelo, se colocaron los embriones en Mat en los dispositivos de inducción siguiendo el protocolo descrito por Arezo y col. (2017). A las 72 horas se obtuvo un grupo denominado inducción de diapausa I (IndDI). Transcurridos 30 días de permanecer en los dispositivos de inducción se obtuvieron embriones en DI, en sus dos fenotipos (DIA\_Lab y DIB\_Lab) (Figura 5).



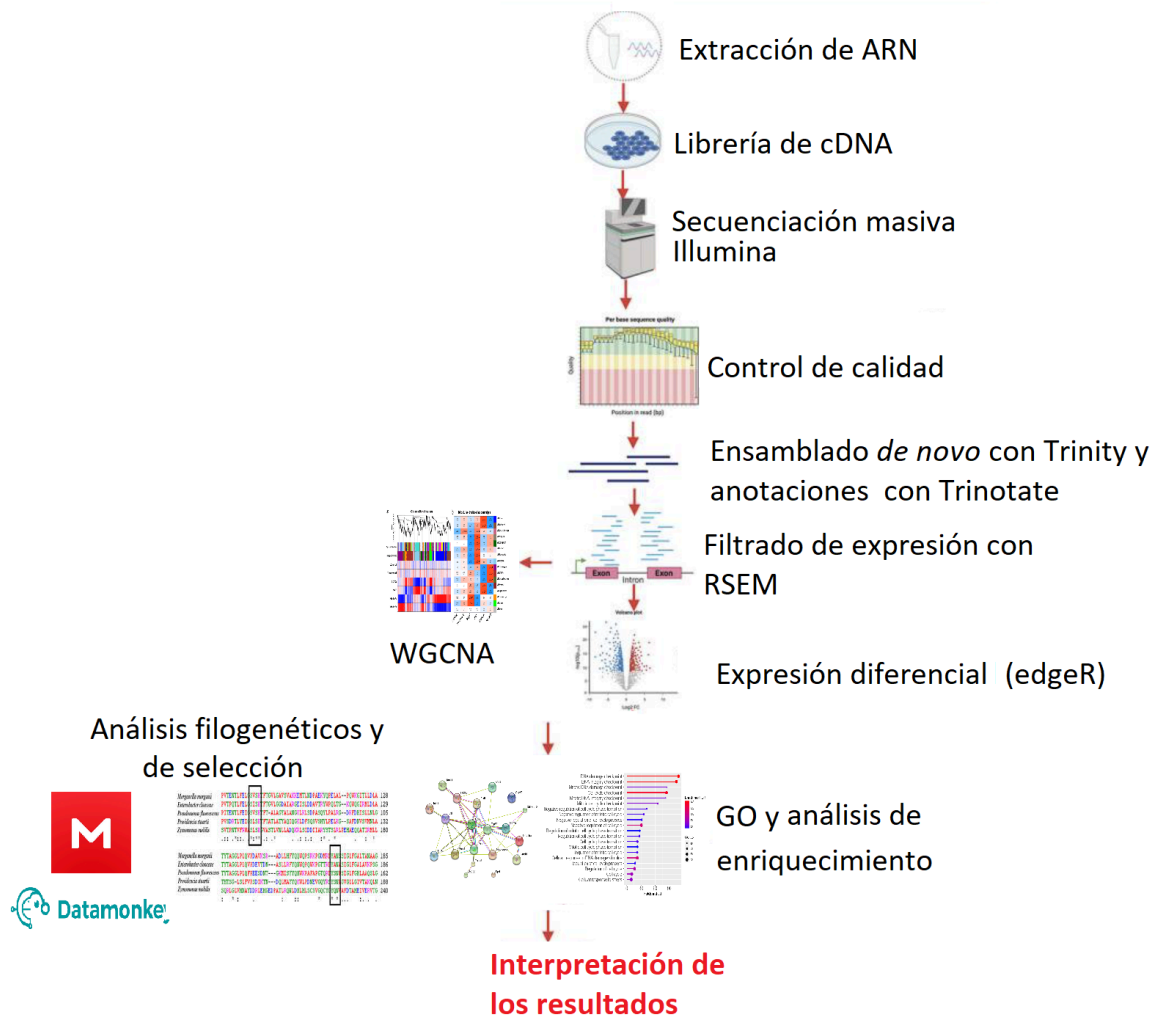
**Figura 5.** Diagrama que ilustra temporalmente los diferentes estadios del desarrollo embrionario de *G. charrua*. Se representan los distintos caminos experimentales realizados en condiciones de laboratorio para obtener embriones en cada estadio. Los tiempos indicados (en días) corresponden a la duración entre cada etapa bajo condiciones experimentales controladas. En el campo (diagrama abajo a la derecha), se obtuvieron embriones en diapausa IA y IB.

### 3. Colecta de embriones en el campo

Se realizaron tres salidas de campo para colectar embriones en DIA y DIB, en condiciones del ambiente natural. Las muestras se obtuvieron del mismo charco donde se capturaron los reproductores que generaron las muestras de laboratorio. Los embriones se obtuvieron colando el barro asociado a las raíces de las plantas acuáticas con un colador de metal, utilizando un método adaptado de Polacik y col. (2023). Estos embriones se diagnosticaron mediante microscopía fotónica en el laboratorio montado en la Colonia de Vacaciones del Ministerio de Transporte y Obras Públicas (MTO) en La Coronilla.

### 4. Flujo de trabajo para la secuenciación de los transcriptomas embrionarios

La secuenciación masiva de los transcriptomas de los distintos estadios embrionarios de *G. charrua* se llevó a cabo mediante una serie de pasos que se describen a continuación (Figura 6).



**Figura 6:** Esquema del flujo de trabajo para la obtención, anotación e interpretación de los transcriptomas obtenidos en los diferentes estadios del desarrollo embrionario (modificado de (Omondi y col. 2023)).

#### 4.1 Extracción de ARN total

Se generaron triplicados biológicos para cada uno de los grupos obtenidos en el laboratorio: Mat, IndDI, NoIndDI, DIA\_Lab, DIB\_Lab, Epi y Reag (Figura 5), utilizando entre 28 y 33 embriones por muestra. En las muestras de campo (DIA\_Camp y DIB\_Camp), cada réplica incluyó entre 11 y 24 embriones. En total, se obtuvieron 27 muestras. Se estabilizaron los ARNs de los embriones con TRIzol™ (Thermo Fisher Scientific, Wilmington, DE) utilizando una relación de volumen de 50 µl por embrión. Se almacenaron a -80°C hasta su procesamiento para el aislamiento de ARN total.

La extracción de ARN total se realizó de acuerdo a las instrucciones detalladas en el prospecto del TRIzol™. La cantidad y calidad del ARN extraído se evaluó mediante

electroforesis capilar en un Bioanalyzer 2100 utilizando el kit Eukaryote Total RNA Nano (Agilent, Santa Clara, CA). Para los análisis posteriores, se seleccionaron las muestras con concentraciones entre 0,5 y 1,2 µg y valores de integridad (RIN, por su sigla en inglés) superiores a 6,8.

#### **4.2 Secuenciación masiva del transcriptoma**

Mediante la técnica de RNA-seq, se realizó la secuenciación del transcriptoma extraído en cada uno de los estadios embrionarios de interés (Mat, IndDI, NoIndDI, DIA\_Lab, DIB\_Lab, DIA\_Camp, DIB\_Camp, Epi y Reag) (Figura 5) y sus triplicados biológicos, utilizando la plataforma Novaseq de 40 millones de lecturas por muestra y fragmentos de 100 nucleótidos pareados. Se construyeron y secuenciaron bibliotecas de ARNm de cadena TruSeq (plataforma 2x150 Illumina HiSeq2500) en la empresa Macrogen (Seúl, Corea).

#### **4.3 Control de calidad y edición de las secuencias**

Se verificó la calidad de los resultados de secuenciación obtenidos con el programa FastQC (Andersson, 2010), donde se buscó que las bases tuvieran una calidad (Q) mayor o igual a 30 (Phred mayor o igual a 30). Esto significa que la probabilidad de que esa base sea incorrecta es de 1 en 1000. Algo similar ocurre con el valor de Q20 que corresponde a una probabilidad de error de 1 en 100 (precisión del 99%). Estos valores permiten estimar la confiabilidad de los datos antes del ensamblaje. Las bases con menor valor de calidad se descartaron. Luego se eliminaron los adaptadores de secuenciación con el programa CLC Genomics Workbench 10.1.1 ([www.qiagenbioinformatics.com](http://www.qiagenbioinformatics.com)).

#### **4.4 Ensamblado *de novo* para la obtención del transcriptoma de referencia**

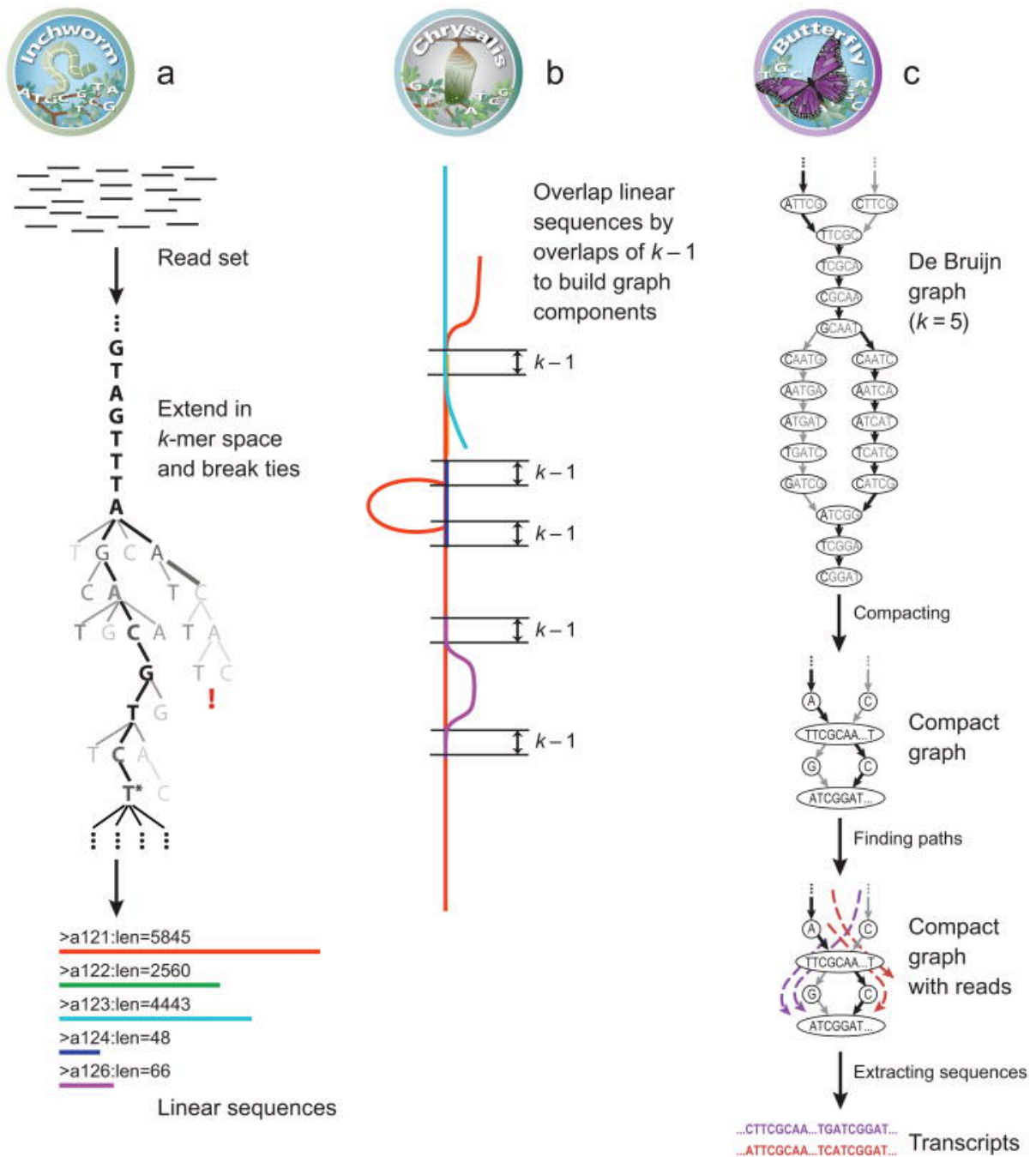
Para realizar el ensamblado *de novo*, se utilizó el servidor del Instituto de Investigaciones Biológicas Clemente Estable (IIBCE), equipado con 20 microprocesadores y 220,37 GB de memoria RAM. El servidor está configurado con un entorno de línea de comandos y configuración de paquetes de software de Bio-Anaconda (Anaconda, Inc. <https://www.anaconda.com/products/distribution>).

Se realizaron análisis de calidad, recorte de longitud (realizado sólo en secuencias adaptadoras) y RNA-seq utilizando CLC Genomics Workbench 10.1.1 ([www.qiagenbioinformatics.com](http://www.qiagenbioinformatics.com)). Las lecturas de la secuenciación de las muestras de los embriones (Mat, IndDI, NoIndDI, DIA, DIB de campo y laboratorio, Epi y Reag) y sus triplicados, se alinearon con el transcriptoma de referencia de *G. charrua*, generado y anotado por Gajardo-Escobar y col. (2026), utilizando un conjunto de procesos (pipeline) de RNA-seq de CLC Genomics Workbench 10.1.1 para la normalización TMM (Robinson y Oshlack 2010). El mapeo, la reproducibilidad y las propiedades estadísticas generales se evaluaron utilizando la correlación de Pearson y un análisis de componentes principales (PCA).

Las lecturas de la secuenciación masiva depuradas en el paso anterior, se trataron mediante un ensamblado *de novo* con el programa Trinity v2.8.5 (Grabherr y col. 2011). Dentro del mismo, se utilizó el módulo "Inchworm" para construir los contigs iniciales a partir de k-mers frecuentes (k-mer es una subsecuencia de longitud fija k a partir de una secuencia de ADN que se usa en un algoritmo para reconstrucción de genes). Luego se

utilizó el módulo “Chrysalis” para agrupar los contigs en componentes (o grupos) basados en k-1-mers compartidos, generar las lecturas que los conectan y construir los grafos de Bruijn (Compeau y col. 2011) para cada grupo. Por último, se usó el módulo “Butterfly” para procesar cada grafo, reconciliar en lecturas individuales y pareadas (paired-end), así como construir isoformas alternativas y genes parálogos. Finalmente, se obtuvo el ensamblado *de novo* del transcriptoma (archivo FASTA con transcritos completos) (Figura 7).

Para filtrar las secuencias correspondientes a bacterias y excluirlas del transcriptoma, se utilizó el programa Kraken (Wood y Salzberg, 2014). El criterio utilizado para el filtrado del ensamblado fue eliminar todo transcritos con un 45% o más de homología con transcritos de origen bacteriano. De esta manera se descartaron esas secuencias bacterianas del ensamblado. Para visualizar los resultados se utilizó el software de Krona (Ondov y col. 2011, Figura A1 en el anexo).



**Figura 7.** Esquema del proceso de ensamblado *de novo* en Trinity. (a) Inchworm ensambla las lecturas depuradas de la secuenciación masiva (Read set) buscando con avidez los caminos en un grafo de  $k$ -mer (medio), generando una colección de contigs lineales (Linear sequences) en la que cada  $k$ -mer está presente sólo una vez. (b) Chrysalis agrupa los contigs que comparten al menos un  $k-1$ -mer y están conectados por lecturas superpuestas, construyendo grafos de Bruijn individuales para cada grupo (líneas de colores). (c) Butterfly toma cada grafo de Bruijn, elimina los bordes espurios y compacta los caminos lineales. Luego reconcilia el grafo con las lecturas (Compact graph

with reads) y produce una secuencia lineal para cada forma de empalme (isoforma) (Transcripts), (tomada de Grabherr y col. 2011).

El programa Trinity incluye opciones adicionales para automatizar varios aspectos del procesamiento de las lecturas de RNA-seq que debieron considerarse antes de ejecutar el ensamblaje *de novo* (Figura 7). Esto incluyó el recorte de calidad de las lecturas mediante Trimmomatic (Bolger y col. 2014), o la normalización *in silico* de las lecturas totales para reducir la cantidad de lecturas que están sujetas al ensamblaje *de novo*, lo que mejora el tiempo de ejecución del ensamblaje. Al correr: TrinityStats.pl, Trinity.fasta, se generó el archivo fasta y luego lanzamos el ensamblado de esta forma: `$ Trinity --seqType fq --max_memory 80G --min_kmer_cov 3 --normalize_max_read_cov 50 --samples_file ../samples9cond.txt --CPU 8 > nohup _9_cond_full_ens2.txt &`. Se ajustaron los requerimientos computacionales del programa, se agregaron las rutas de acceso al archivo que contiene las etiquetas de cada condición en un archivo de texto (esto es 3 réplicas por cada estadio embrionario de campo y de laboratorio), y con respecto al ensamblado subimos el mínimo de cobertura de los k-mer a 3 y bajamos la cobertura máxima a 50. Usando este programa se realizó la normalización con las lecturas obtenidas de una réplica por grupo de embriones.

El ensamblado se filtró por niveles de expresión con RSEM (Li y Dewey 2011), y TransRate (Smith-Unna y col. 2016), para hacer los filtrados correspondientes en TPM (transcriptos por millón de lecturas o reads). Se anotó el transcriptoma resultante con Transdecoder (Haas y col. 2013), herramienta que nos permitió identificar y predecir regiones codificantes de proteínas, es decir marcos abiertos de lecturas (por sus siglas en inglés: ORFs). TransDecoder también pudo integrar información de dominios de proteínas conocidos (usando bases de datos como Pfam) para mejorar la precisión de las predicciones, generando un archivo fasta de nucleótidos (longest\_orfs.cds), un fasta de aminoácidos (longest\_orfs.pep) y un archivo de anotación (gff. longest\_orfs.gff3)

#### **4.5 Anotación funcional del transcriptoma**

La anotación completa de las secuencias ensambladas con Trinity se realizó utilizando los ORFs y el programa Trinotate 4.1 (Bryant y col. 2017, <http://trinotate.github.io/>). Este software está diseñado para realizar una anotación funcional automática de transcriptomas, mediante la búsqueda de homologías de secuencias en las transcripciones generadas por Trinity. Para ello se realizó una búsqueda blast (Altschul. y col., 1990) contra bases de datos de proteínas como SwissProt y Uniref90. La anotación también incluyó la identificación de dominios proteicos mediante HMMER sobre la base de datos Pfam (Finn y col. 2011, Punta y col. 2012), la predicción de péptidos señal y dominios transmembrana a través de las herramientas Signal y tmHMM (Krogh y col. 2001, Petersen y col. 2011), y la comparación con anotaciones disponibles en bases de datos muy utilizadas como Uniprot, eggNOG, y Gene Ontology (GO) (Ashburner y col, 2000, Kanehisa y col., 2012, Powell y col., 2012). La búsqueda en todas las bases de datos se realizó con el algoritmo de Diamond Blast (Buchfink y col. 2021).

Todas las búsquedas se realizaron con el programa SQLite, que integra todos los métodos de anotación como el blast-tx y el blast-n de la base de datos Swiss-prot, n-Prot.

#### 4.6 Búsqueda de genes con expresión diferencial

La búsqueda de genes con expresión diferencial (GED) se realizó con el programa EdgeR (Robinson y col. 2010), que estima el parámetro de dispersión común utilizando un promedio ponderado de las dispersiones entre todos los genes, proporcionando un valor único que representa la dispersión asumida para todos los genes. Utilizando esta dispersión común, EdgeR ajusta un modelo binomial negativo a los datos, por lo que considera la dispersión común estimada y los niveles de expresión promedio de cada gen (Robinson y col. 2010; Chen y col. 2014). Este análisis se realizó para cada una de las réplicas biológicas. Como se describió anteriormente, cada una de ellas está compuesta por un grupo de embriones, y esta estrategia de agrupación reduce eficazmente la variación y con el uso del método EdgeR, redundan en la obtención de resultados robustos (Schurch y col. 2016; Takele Assefa y col. 2020).

Las lecturas pareadas de cada muestra se mapearon utilizando como referencia el ensamblado obtenido en el paso anterior mediante RSEM. Los resultados cuantitativos fueron comparados estadísticamente utilizando el programa EdgeR (Robinson y col. 2010). Se realizaron comparaciones entre pares o entre más condiciones, generando listas de genes a partir de las cuales se obtuvieron aquellos con expresión diferencial. Las expresiones diferenciales surgieron de la comparación entre las diapausas (DIA y DIB) de campo y de laboratorio. Además, estos transcritos se contrastaron con los identificados previamente como específicos de DI (GED entre DI y los estadios activos de desarrollo, Papa, 2025).

Para determinar los GED se aplicaron los parámetros de la tasa de falsos descubrimientos (FDR), una metodología estadística empleada para corregir los resultados de pruebas múltiples y controlar la proporción de falsos positivos en los análisis de expresión génica, estableciendo un umbral  $\leq 0,05$ ; el p-value  $\leq 0,01$  y el Fold Change (FC, cambio en pliegues), que es una medida que describe la variación en la expresión de un gen entre dos condiciones experimentales, se calcula como la razón entre los niveles de expresión en una condición respecto a la otra y se expresa en logaritmo base dos ( $\log_2FC$ ) para facilitar su interpretación y simetrizar los valores de aumento o disminución. Esta medida se utilizó para identificar genes diferencialmente expresados, los cuáles se emplearon posteriormente en los análisis de enriquecimiento funcional para evaluar si determinados conjuntos de genes están asociados a categorías o funciones biológicas específicas.

Finalmente, utilizando el programa web Venny 2.0 (Oliveros y col. 2015) se obtuvieron representaciones gráficas globales en diagramas de Venn a partir del cruzamiento de las listas de los GED entre las DI de campo y las de laboratorio. Se obtuvieron mapas de calor usando el programa Morpheus (<https://software.broadinstitute.org/morpheus>) para los GED y listas de genes derivadas de los diagramas de Venn compartida anteriormente.

En los siguientes análisis de expresión diferencial y de anotación, se seleccionó el transcriptoma obtenido con el filtrado más robusto (10 TPM) que contenía sólo genes correspondientes a *G. charrua*. Para analizar la gran cantidad de transcritos, se eligió utilizar solamente los genes que se expresan, sin considerar sus isoformas. De esta manera, se evitó la redundancia en los análisis de ontología génica y enriquecimiento, así como en el estudio de las redes y vías metabólicas posiblemente involucradas. Además, se

realizaron anotaciones utilizando el filtrado de 5 TPM, con el objetivo de analizar las isoformas de los genes de interés en una segunda etapa de análisis a mayor profundidad.

#### **4.7 Construcción y análisis de redes de co-expresión génica ponderadas**

El método WGCNA (Weighted Gene Co-expression Network Analysis) (Langfelder y col. 2008), es una técnica utilizada para analizar datos de expresión génica dentro del programa R (Rstudio Posit Software, PBC, <http://www.posit.co/>). Su objetivo principal es identificar y caracterizar redes de genes que están co-expresados, es decir, aquellos que muestran patrones similares de expresión bajo diferentes condiciones.

En el análisis WGCNA se construyeron redes de co-expresión génica a partir de datos de expresión de alta dimensión. De forma predeterminada, el método organiza estos datos en grupos que contienen genes con patrones de expresión similares. Esta clasificación permite identificar conjuntos de genes que probablemente participan en procesos biológicos relacionados o están regulados de manera conjunta.

Para este análisis se utilizó la matriz que contenía únicamente genes, a la cual se le realizó un filtrado para eliminar todos los genes cuya expresión fuera igual a cero en al menos 15 de las condiciones analizadas. Otro criterio considerado fue mantener solo los genes cuya expresión fuera mayor o igual a 5 TPM en 4 condiciones de DI (sin importar el fenotipo, o si provenían del campo o del laboratorio).

Las 27 muestras se sometieron a la construcción de esas redes de co-expresión. Para ello, se calculó la correlación entre todos los pares de genes utilizando la correlación de Pearson. Esto generó una matriz de correlación (gen x gen), que reflejaba la similitud en los perfiles de expresión entre pares de genes en todas las muestras. Luego se realizó un análisis de agrupamiento jerárquico sobre la matriz de topología de red (TOM, por sus siglas en inglés que significa matriz de solapamiento topológico) para agrupar los genes en módulos.

Los genes con altos valores de co-expresión (altos TOM) se agruparon más cercanamente. Se utilizaron las funciones `cutreeDynamic` o `cutreeHybrid` para cortar el dendrograma y definir los grupos de genes. Se ajustó el parámetro `minModuleSize` (tamaño mínimo de módulo, por defecto) a un valor de 30 y se activó la opción `deepSplit` para mayor sensibilidad en la detección de módulos pequeños. Por último se ajustó el parámetro `mergeCutHeight`, que indica el umbral de fusión de grupos de interés, seleccionando un valor 0,25.

Los gráficos resultantes se generaron en Rstudio. El resultado principal fue un mapa de calor con los índices de correlación calculados a partir del valor promedio dado por un gen hipotético representativo de ese grupo en cada condición (réplica de cada estadio), y la primera componente del análisis del PCA, empleando la correlación biserial puntual (<https://datatab.net/tutorial/point-biserial-correlation>) como método matemático integrándose con las ya mencionadas anteriormente.

#### **4.8 Ontología génica y análisis de enriquecimiento de genes**

Se examinaron las vías regulatorias relacionadas a los genes diferenciales y co-expresados mediante paquetes especializados en análisis de ontología génica (GO)

como Panther (Thomas y col. 2003) y String enrichment analysis (Szklarczyk y col. 2019), con el fin de comprender la relevancia biológica de los GED. Posteriormente, se seleccionaron las secuencias de algunos transcritos de mayor relevancia biológica, para análisis filogenéticos y de selección.

En la entrada al programa String, se ingresaron las secuencias aminoacídicas de los transcritos en formato FASTA (no se utilizaron los nombres de las anotaciones de Trinotate para evitar errores de nomenclatura). El organismo de referencia para los estudios de ontología genética fue el pez anual *A. limnenus*, lo que permitió incrementar la confiabilidad en la asignación funcional de las proteínas del enriquecimiento. Se usaron parámetros matemáticos de vinculación de genes con procesos biológicos y funciones moleculares, y altos grados de similitud. String genera redes de interacción basadas en evidencia experimental, bases de datos de coexpresión, y predicciones computacionales. Para los análisis se seleccionaron parámetros estándar de alto grado de confianza (score  $\geq 0.7$ ), para garantizar que los enlaces representaran asociaciones biológicas robustas. Se descartaron interacciones de baja confianza ( $< 0.4$ ) para reducir la inclusión de falsos positivos. A partir de las redes obtenidas, se realizó el enriquecimiento funcional en las categorías: procesos biológicos (BP) y funciones moleculares (MF). Se emplearon pruebas estadísticas internas de String (enriquecimiento de tipo Fisher exact test) y corrección por falsos descubrimientos (FDR  $< 0.05$ ) para determinar la significancia.

De los grupos de genes con las características de interés, se seleccionaron 5 genes diferencialmente expresados, de acuerdo a la bibliografía (Karp 2021, Thompson y col. 2024). Uno de los GED seleccionado fue el correspondiente a la proteína de choque térmico frío (csde1) ya que está asociado con respuestas a bajas temperaturas, un factor crucial para especies que experimentan variaciones ambientales extremas. Otros fueron, el Forkhead box H1 (FoxH1), que participa en la regulación de rutas de señalización embrionarias fundamentales para el desarrollo temprano y un Homeobox (Alx4) por ser reguladores maestros del desarrollo corporal y morfogénesis, procesos críticos para la viabilidad de especies con ciclos de vida acelerados como los peces anuales. Se utilizaron criterios estadísticos de enriquecimiento como puntuación GO mínima de interacciones requerida, alto grado de confianza y un número máximo de interacciones por gen de 10.

Se procedió a graficar los análisis de GO con el programa ShinyGo v.0.82 (Ge y col. 2020), a partir de las listas de genes y sus respectivos valores de GO obtenidos en Trinotate. De este modo, se visualizaron las agrupaciones de funciones moleculares más enriquecidas y sus respectivos valores de fold change. Para las muestras de laboratorio, se consideraron los genes resultantes del análisis de expresión diferencial realizado en EdgeR ( $\geq 10$  TPM), el cual fue sometido a un filtrado adicional, eliminando todos los transcritos que exhibieron un valor de cero en al menos 24 de las 27 réplicas biológicas, con el fin de establecer un criterio más riguroso y evitar la sobrerepresentación de genes diferenciales. Finalmente, se repitió el procedimiento antes descrito para elaborar la tabla correspondiente de categorías, familias génicas y estudio de GO.

## 5. Análisis filogenético y de selección

De los genes de interés mencionados anteriormente, se eligió profundizar en el análisis del gen FoxH1, por ser miembro de la familia Forkhead box, un grupo de factores

de transcripción con roles fundamentales en la regulación del desarrollo temprano y el destino celular (Golson y Kaestner, 2016) y por su rol durante el desarrollo temprano en el pez cebra (Pei y col. 2007). De las diferentes isoformas observadas de este gen, se seleccionó aquella cuya expresión aumentaba en los embriones en diapausa obtenidos en el campo (DIA\_Camp y DIB\_camp). Mediante un blast-n, se obtuvieron del GenBank (<https://www.ncbi.nlm.gov/genbank>) 10 secuencias homólogas a esa isoforma en otras especies de Cyprinodontiformes, incluidos algunos peces anuales, seleccionando aquellas con valor E de 0.0. Como grupo externo se incluyó la secuencia homóloga de *Oryzias latipes* (Tabla 1) (Setiamarga y col. 2008).

Se realizó un alineamiento múltiple en el programa ClustalW implementado en el MEGA11 (Tamura y col. 2021). Luego, se editaron las secuencias manualmente y se realizó un análisis de polimorfismos, determinando el número de sitios conservados (C), variables (V) y filogenéticamente informativos según el criterio de parsimonia (Pi). El análisis filogenético de Máxima Verosimilitud (ML) se realizó en el programa MEGA11 utilizando la opción de búsqueda del mejor modelo de sustitución nucleotídica en base al criterio de información de Akaike (AIC). El análisis se llevó a cabo bajo el modelo de sustitución de Tamura 3-parámetros (T92) con dos categorías gamma. El soporte de los nodos se estimó por el método bootstrap y se descartaron aquellos valores inferiores al 75%. El parámetro de forma del modelo gamma obtenido para FoxH1 ( $\alpha$ ), que describe cuánta variación de tasas de evolución hay entre los sitios de la secuencia, se calculó mediante la máxima verosimilitud durante la optimización del árbol filogenético en MEGA11.

**Tabla 1.** Especies de Cyprinodontiformes y número de acceso de las secuencias nucleotídicas seleccionadas del GenBank homólogas a la isoforma de FoxH1, (\*) especies de la familia Rivulidae, (\*\*) especies de peces anuales.

Espece	N° de acceso
<i>Nematolebias whitei</i> (*) (**)	XM_037693925.1
<i>Austrofundulus limnaeus</i> (*) (**)	XM_014018757.1
<i>Kryptolebias marmoratus</i> (*)	XM_017428403.2
<i>Nothobranchius furzeri</i> (**)	XM_054751106.2
<i>Poeciliopsis prolifica</i>	XM_055037648.1
<i>Gambusia affinis</i>	XM_044103936.1
<i>Poecilia formosa</i>	XM_016669359.1
<i>Fundulus heteroclitus</i>	XM_012863178.3
<i>Girardinichthys multiradiatus</i>	XM_047350151.1
<i>Cyprinodon variegatus</i>	XM_015386601.1
<i>Oryzias latipes</i>	NM_001160471.1

Posteriormente, se realizó un análisis de polimorfismos y divergencia genética en el programa DnaSP v.6.12.03 (Rozas y col. 2017). Se estimó la divergencia inter e intraespecífica mediante los valores de  $K_a$  y  $K_s$ , correspondientes a las tasas de sustituciones no sinónimas (provocan cambios en la secuencia de aminoácidos) y sinónimas (no alteran la secuencia aminoacídica) por sitio, respectivamente. Adicionalmente, se calculó  $K_a/K_s$  para inferir el tipo de presión selectiva actuante sobre la secuencia codificante. Asimismo, se estimaron los niveles de diversidad nucleotídica promedio en sitios sinónimos ( $P_i(s)$ ) y aquella en sitios no sinónimos ( $P_i(a)$ ) mediante el mismo programa (Nei y Li 1979).

A continuación, se procedió a identificar los codones bajo selección diversificadora episódica utilizando la plataforma web Datamonkey (<http://www.datamonkey.org/>) (Kosakovsky y col. 2020). Con el objetivo de detectar procesos evolutivos asociados a la selección natural, se analizó la presencia de presión selectiva a nivel de sitios individuales dentro del gen. Para ello, se empleó el modelo MEME (del inglés Mixed Effects Model of Evolution), recomendado por Datamonkey para la detección de selección episódica, aquella que actúa de manera intermitente en ciertos sitios y linajes evolutivos.

Posteriormente, se construyó una matriz con las secuencias aminoacídicas correspondientes a las 4 especies de peces anuales analizadas (Tabla 1), las cuales se alinearon con el programa ClustalW implementado en MEGA11. En los casos en que se detectaron cambios en la secuencia, se consultó la base de datos UniProt (<https://doi.org/10.1093/nar/gkae1010>) con el objetivo de identificar la región específica de la isoforma del gen FoxH1 en la que se encontraba dicho cambio aminoacídico, utilizando la plataforma gráfica InterProScan (<https://www.ebi.ac.uk/interpro/>). Por último, se realizó el modelado tridimensional de FoxH1 con el programa AlphaFold2 (Jumper y col. 2021) en la plataforma de Google Colab, usando la versión modificada ColabFold (Mirdita y col. 2022).

# Resultados

## Ensamblado *de novo* del transcriptoma embrionario de *Garcialebias charrua*

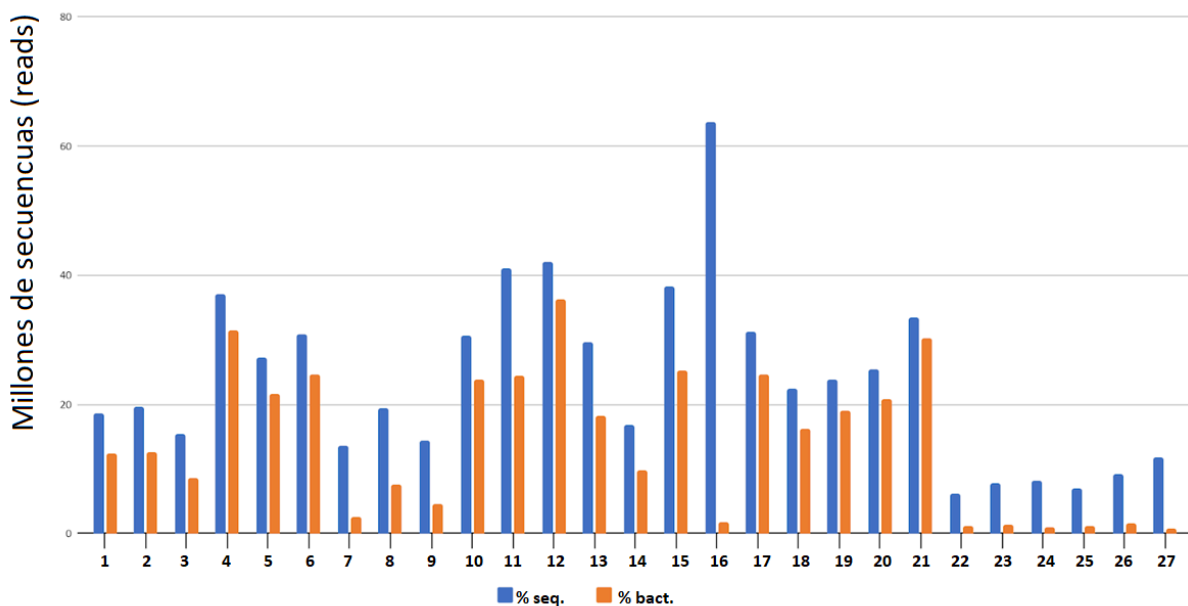
Las muestras de ARN obtenidas de cada uno de los 9 estadios embrionarios y sus triplicados biológicos, presentaron valores de concentración e integridad aceptables para la construcción de bibliotecas y posterior secuenciación. El control de calidad de las lecturas obtenidas mediante RNAseq, evaluado con el programa FastQC, reveló una alta calidad de secuenciación en cada condición y en casi todas sus réplicas biológicas (Figura 8). En cada muestra, se obtuvieron valores  $Q > 30$  en la mayoría de las bases, ausencia de contaminantes, y una distribución uniforme del contenido de nucleótidos. Estos resultados indican que las lecturas obtenidas son confiables para el ensamblado *de novo* del transcriptoma embrionario de *G. charrua*.



**Figura 8.** a. Conteo de lecturas totales en las 27 muestras secuenciadas. En el eje Y cada barra corresponde a una muestra (Sample1, Sample2, etc.) y el eje X corresponde al número total de lecturas (reads) obtenidas después de la secuenciación. b. Relación entre la cantidad de bases con calidad Q20 y las bases con calidad Q30 en un conjunto de datos de secuenciación. En el eje Y se representa cada muestra secuenciada y en el eje X el porcentaje de bases que alcanzan los valores de calidad (azul Q20, anaranjado Q30). c. Diagrama de fastQC de la calidad de las secuencias. En el eje Y, se representa el valor de calidad Phred asignado a cada base (escala Q) y en el eje X se representa la posición dentro de la lectura (desde el primer nucleótido hasta el último). Las áreas del gráfico representan: verde, buena calidad ( $Q \geq 30$ ), amarillo, calidad moderada (Q20–Q30) y rojo, baja calidad ( $< Q20$ ).

En una primera instancia, la anotación del ensamblado (Ensamblado\_2 Comp) con diferentes bases de datos (SwissProt y UniProtKB), reveló que en el transcriptoma había una alta proporción de transcritos correspondientes a bacterias y arqueas en todos los estadios del desarrollo embrionario analizados (Figura 9). Las réplicas de DI obtenidas en condiciones de laboratorio presentaron una mayor proporción de transcritos con alta homología a secuencias bacterianas (Figura 9, números del 1 al 21), en comparación con las réplicas de las DI colectadas en el campo (Figura 9, números del 22 al 27). Por este motivo, se procedió a depurar la base de datos utilizando el programa Kraken (Tabla 2, columnas verdes) a fin de obtener un transcriptoma libre de secuencias contaminantes de bacterias, donde el programa con su propia base de datos, anotó un 24,53% de los transcritos del ensamblado que se detalla en la tabla 1A en anexos.

El ensamblado inicial contenía un total de 522.611 contigs, con una longitud promedio de 757 pb y un GC% de 48 (Tabla 2, columna anaranjada). La depuración de las secuencias bacterianas con Kraken redujo el número a 48.065 secuencias, de las cuales aproximadamente el 60% incluían ORFs, indicando una fracción considerable de transcritos potencialmente codificantes (Tabla 2, columna verde oscuro). Las métricas de longitud (N50 ~1.500–1.800 pb) reflejaron buena continuidad del ensamblado. Los filtrados posteriores por abundancia (1, 2, 5 y 10 TPM) disminuyeron el número de transcritos retenidos, pero aumentaron la longitud promedio, conservando principalmente transcritos más confiables y expresados (Tabla 2, columnas verde claro y azul).



**Figura 9.** Porcentaje de secuencias clasificadas en la base de datos de Kraken. Transcritos eucariotas (% seq. barras azules) y bacterianos (% bacterias, barras naranjas), en las muestras analizadas por triplicado: 1, 2, 3 (Mat), 4, 5, 6, (NoIndDI), 7, 8, 9, (IndDI), 10, 11, 12, (Epi), 13, 14, 15, (DIA\_lab), 16, 17, 18, (DIB\_lab), 19, 20, 21, (Reag), 22, 23, 24 (DIA\_Camp), 25, 26, 27, (DIB\_Camp).

Luego de eliminar todas las secuencias bacterianas según este criterio de filtrado, se creó un nuevo ensamblado. Este procedimiento mejoró la resolución en cuanto a la cantidad total de secuencias que Trinity logró ensamblar (Tabla 2, columnas verdes). Por ejemplo, se obtuvo una mejora en el valor de N50, una mayor cantidad de transcriptos ensamblados con más de 1000 y 10000 pares de bases, y una mayor cantidad de ORFs en las secuencias, entre otros.

Posteriormente, al ensamblado curado (Sin Bacteria) se le realizaron filtrados de 1 TPM, 5 TPM (que contenían varias isoformas de cada gen) y un filtrado más riguroso de 10 TPM que sólo contenía a los genes. De esta forma, se obtuvo un transcriptoma con un total de 30.736 genes, representando el 7% del transcriptoma original sin bacterias (Tabla 2, columnas verdes claro y azul).

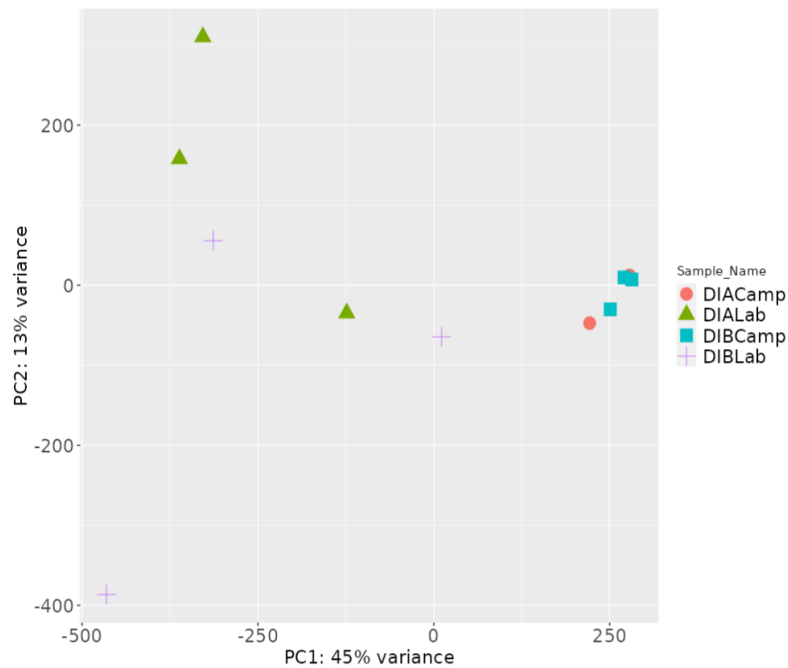
**Tabla 2.** Resultados obtenidos con el programa TransRate de las características del ensamblado y sus respectivos filtrados.

TransRate	Ensamblado_2 Comp	Sin Bacteria-Sin filtro	Sin Bacteria-1TPM	Sin Bacteria-2TPM	Sin Bacteria-5TPM	Sin Bacteria-10TPM
	Contig metrics:	Contig metrics:	Contig metrics:	Contig metrics:	Contig metrics:	Contig metrics:
n seqs	522611	48065	377842	259764	100068	30736
smallest	164	169	169	169	169	169
largest	40832	28202	28202	28202	28202	17734
n bases	395803372	393086067	273009165	183593063	81398910	37501220
mean len	757,26	817,73	722,46	706,66	813,23	899,79
n under 200	271	199	167	151	106	50
n over 1k	98324	101418	65734	44888	22567	11386
n over 10k	350	414	270	186	76	26
n with ORF	126111	115814	83660	59145	28482	13906
mean ORF percent	62,11	59,88	63,5	65,86	65,18	63,6
n90	277	286	261	247	263	300
n70	613	721	561	562	834	1001
n50	1476	1711	1508	1587	1844	1884
n30	2759	2985	2805	2839	2906	2817
n10	5153	5295	5097	5035	4869	4601
GC	0,48	0,47	0,48	0,49	0,49	0,48
Results	522611	Retained: 480659 = 100%	377842/480659 = 78,61%	259764/480659 = 54,04%	100068/480659 = 20,82%	30736/480659 = 7,00%
		TRANSCRIPT				GENE

n seqs: Número de secuencias o tamaño del ensamblado en número total de contigs. smallest: Longitud (en pb) del contig más corto del ensamblado. largest: Longitud (en pb) del contig más largo. n bases: Suma total de bases de todos los contigs (tamaño total del ensamblado). mean len: Longitud media de los contigs (promedio del largo). n under 200: Número de contigs menores a 200 pb. n over 1k: Número de contigs mayores a 1000 pb. n over 10k: Número de contigs mayores a 10 000 pb. n with ORF: Contigs que contienen un marco de lectura abierto (ORF) identificado, indica potencial codificante. mean ORF percent: Porcentaje promedio de longitud del contig ocupado por un ORF, mide cuán "codificantes" son las secuencias. N50: Longitud del fragmento más corto en un conjunto de fragmentos -contigs o scaffolds- que, al sumarse, representan al menos el 50% del total de la longitud del ensamblaje. Similar a lo que sucede con n90 / n70 / n50 / n30 / n10, variando el porcentaje. GC: Contenido de GC en el transcriptoma. Results: indica el porcentaje de transcriptos retenidos respecto al total inicial tras cada filtrado.

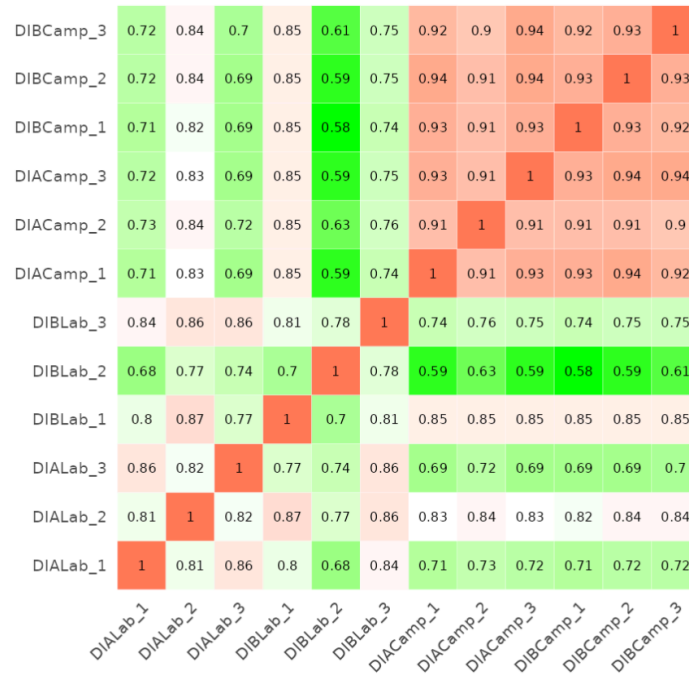
## Análisis preliminares de la expresión génica en la diapausa I

Los resultados obtenidos del PCA muestran un claro agrupamiento de las diapausas del campo que presentan entre sí una similitud muy estrecha en sus niveles de expresión génica. Las diapausas inducidas en el laboratorio exhiben una marcada lejanía en el diagrama, respecto a las muestras del campo, e incluso entre sí, sugiriendo una mayor variabilidad en los niveles de expresión (Figura 10).



**Figura 10.** PCA de los niveles de expresión en las diferentes condiciones de DIA\_Lab, DIB\_Lab, DIA\_Camp y DIB\_Camp. Imagen obtenida con el programa RSEM desde Trinity. Los diferentes colores y formas representan a los diferentes fenotipos y condiciones según la leyenda. Los ejes del gráfico representan dos componentes principales: PC1, con un 45% de varianza, indica cuánta de la variación total explican las diferencias entre campo y laboratorio; PC2, con un 13% de varianza indica las diferencias entre los fenotipos de DI.

El análisis de correlación de Pearson, obtenido a partir de los niveles de expresión de cada muestra, reafirma los resultados del análisis anterior. Las correlaciones entre las muestras de DI del campo son más altas entre sí, mientras que las DI del laboratorio presentan menores valores de correlación (Figura 11).



**Figura 11.** Matriz de correlación de Pearson basada en los niveles de expresión génica promedio por muestra. Los tonos rojos indican alta correlación positiva entre muestras ( $r \geq 0,9$ ), mientras que los tonos verdes representan baja correlación o ausencia de correlación ( $r \leq 0,4$ ). Los valores intermedios corresponden a correlaciones moderadas ( $0,4 < r < 0,9$ ), con una transición gradual de verde a rojo.

### Anotación del transcriptoma mediante el programa Trinotate

La anotación del transcriptoma de la DI de *G. charrua* mostró que en el filtrado de 10 TPM, aproximadamente la mitad de los genes ensamblados presentaron coincidencias en las bases de datos utilizadas (Tabla 3). En las anotaciones correspondientes al ensamblado del filtrado de 5 TPM, se observó un patrón similar, indicando una consistencia en la representación de los transcritos entre ambos niveles de filtrado (Tabla 3). Estos resultados reflejan una cobertura significativa de la información funcional obtenida a partir del análisis de las secuencias ensambladas.

**Tabla 3.** Comparación entre dos procesos de anotación con Trinotate, de la cantidad de genes (10 TPM genes) o transcritos (5 TPM transcritos) anotados en distintas bases de datos, sobre el total de transcritos.

ANOTACIONES:	BLASTX	infernal	Pfam	egglog	Kegg	gene_ontology_BLASTX	de un total de:
5TPM transcritos	14925	5522	11606	18873	34234	61620	100068
10TPM genes	13681	2257	8704	4571	11726	13462	30735



laboratorio contra el campo; d) Fenotipo B de laboratorio contra el campo. Los puntos rojos representan los genes diferencialmente expresados y los puntos negros representan los genes que no alcanzan valores diferenciales de expresión entre las condiciones analizadas.

Los ajustes aplicados a los valores de FDR y FC ( $\log_2FC$ ) permitieron optimizar la identificación de los GED, reduciendo la probabilidad de falsos positivos y así mejorar la confiabilidad de los resultados obtenidos. En la Tabla 4 se detallan los parámetros empleados en los filtrados y el número de genes retenidos tras la corrección estadística. La referencia a los genes que aumentan su expresión en cada condición indica que el signo del  $\log_2FC$  es positivo para las DI inducidas en el laboratorio sobre las DI obtenidas en el campo y para el fenotipo DIA sobre el DIB.

Es importante destacar que estos resultados muestran una clara similitud entre los fenotipos de DI, considerando el número de transcritos expresados diferencialmente dentro de cada condición, lo que también es coherente con los resultados preliminares de los PCA. Este grupo de transcritos representa el 0,04% del total de genes expresados (30736 genes) en los embriones obtenidos del campo y el 0,16% de los genes expresados en los embriones inducidos en condiciones de laboratorio.

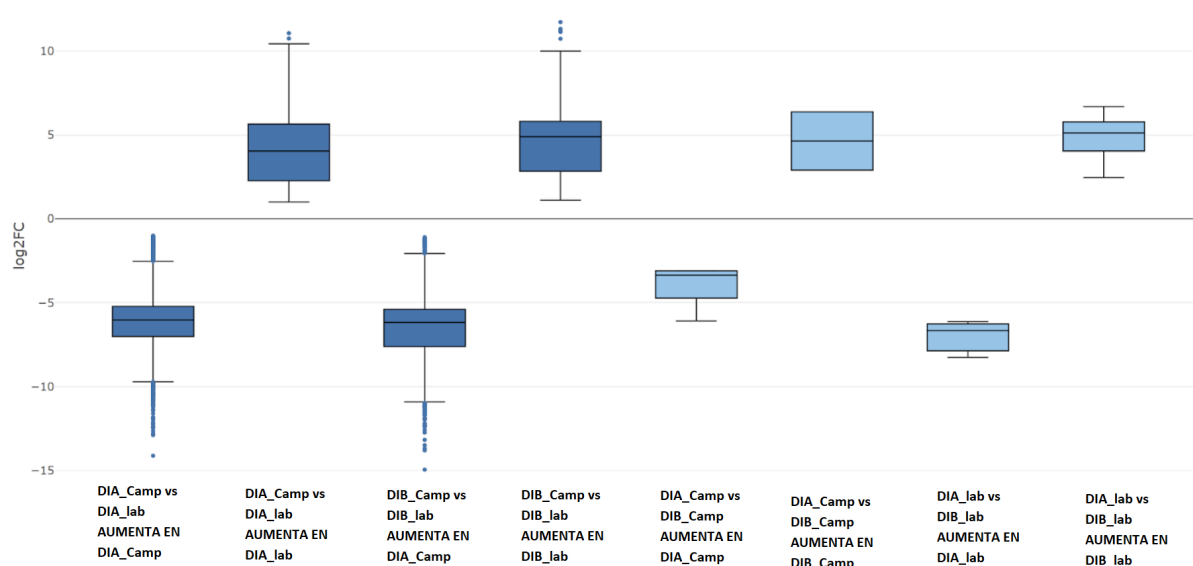
Con respecto a la variable ambiente (campo o laboratorio), el número de genes diferencialmente expresados entre las DI del campo y del laboratorio evidencian que es muy significativo, sin importar el fenotipo. Esto sugiere que las diapausas, en distintas condiciones, responden de manera muy diferente. La cantidad de genes diferencialmente expresados entre DIA\_Camp en relación con DIA\_Lab es del 17,57%, mientras que entre DIB\_Camp y DIB\_Lab es del 12,03% (Tabla 4).

En la tabla 4, donde se encuentran discriminados los genes y transcritos que se expresan al alza en una condición o en la otra, se evidencia que la expresión diferencial no es equilibrada. En DIA\_Lab aumentan su expresión 4854 genes (se expresan a la inversa en la DIA\_Camp) y solo se expresan positivamente en DIA\_Camp 554 genes (se expresan a la inversa en DIA\_Lab). Asimismo, en DIB\_Lab se expresan al alza 3058 genes y en DIB\_Camp aumentan su expresión 640 genes. Esta tendencia refleja que la respuesta transcripcional en el laboratorio fue más amplia y diversa, mientras que en el campo se concentró en un conjunto más reducido de genes.

**Tabla 4.** Datos del análisis de expresión diferencial de los filtrados de 5 TPM (transcritos) y 10 TPM (genes). Ambos análisis se hicieron aplicando parámetros ajustados de  $FDR < 0,05$  y  $FC (\log_2FC)$ .

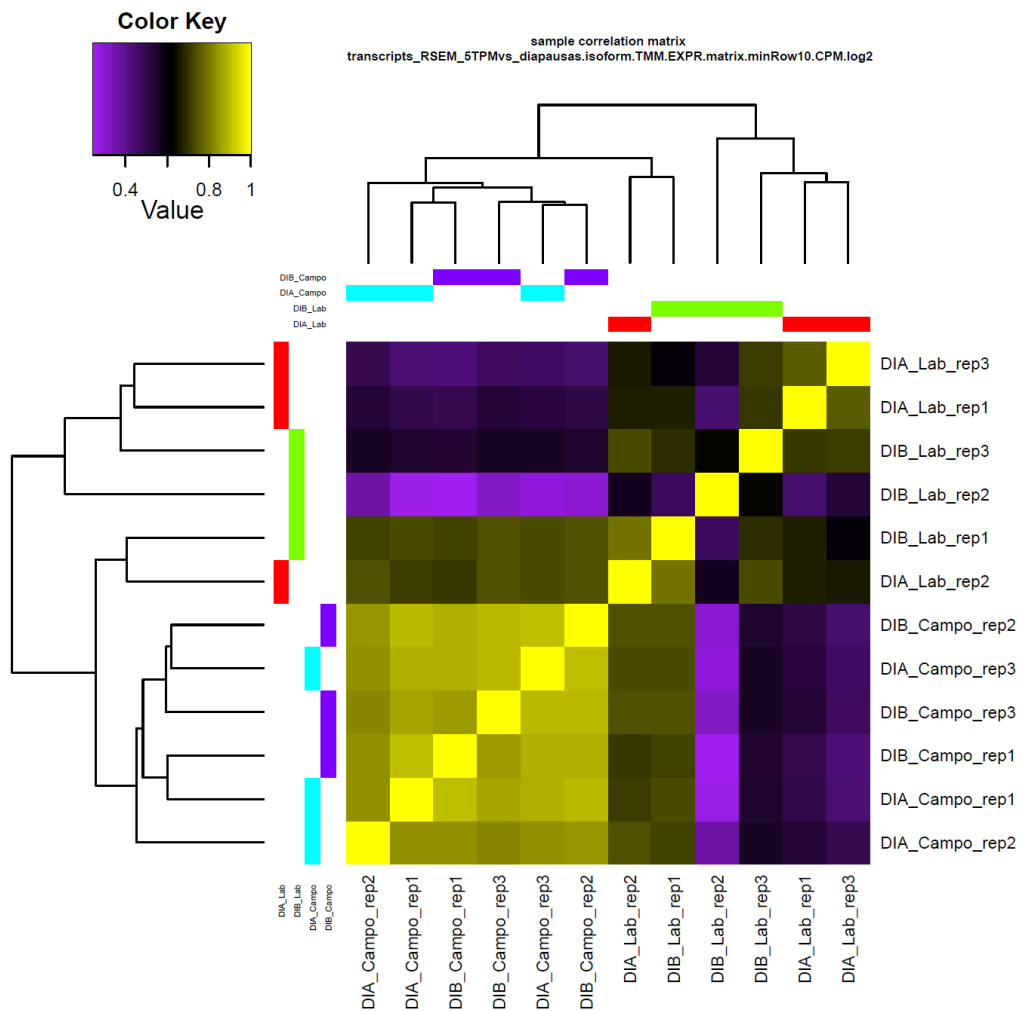
Diferenciales 5 TPM	CONDICIONES	umentan en DIA	umentan en DIB	TOTAL	% DEL TOTAL	GENES TOTALES
con log2FC y FDR<0.05	DIA camp vs DIB camp	30	32	62	0.06	100068
	DIA Lab vs DIBLab	35	25	60	0.06	100068
		umentan en DIAlab	umentan en DIAcamp			100068
	DIA lab vs DIAcampo	4926	804	5730	5.73	100068
		umentan en DIBlab	umentan en DIBcamp			100068
	DIB lab vs DIB campo	5731	1008	6739	6.73	100068
DIFERENCIALES 10 TPM	CONDICIONES	umentan en DIA	umentan en DIB	TOTAL	% DEL TOTAL	GENES TOTALES
con log2FC y FDR<0.05	DIAcampo vs DIB	3	10	13	0.04	30736
	DIA lab vs DIB lab	35	13	48	0.16	30736
		umentan en DIAlab	umentan en DIAcamp			
	DIA lab vs DIA campo	4845	554	5399	17.57	30736
		umentan en DIBlab	umentan en DIBcamp			
	DIB lab vs DIB campo	3058	640	3698	12.03	30736

En la Figura 13 se muestran los cambios en los patrones de expresión resultantes de la aplicación de los filtrados más estrictos en la categoría de genes (Tabla 4, filas azules). El gráfico de cajas muestra la distribución de los genes que aumentan su expresión en cada condición analizada. En términos proporcionales, hay más genes que aumentan su expresión en las DI inducidas en el laboratorio en comparación con las DI del campo. En cuanto al fenotipo, hay muy pocos genes expresados diferencialmente en general, siendo DIB\_Camp y DIA\_Lab las condiciones con mayor el número de GED.



**Figura 13.** Gráficos de Fold Change ( $\log_2FC$ ) del análisis de expresión diferencial realizado en EdgeR sobre el transcriptoma filtrado por 10 TPM, aplicando el filtro de  $FDR < 0,05$ . Se comparan las diapausas de campo y laboratorio, así como los fenotipos DIA y DIB. En azul oscuro se representan las comparaciones entre DI de campo y laboratorio. En celeste se representan los genes expresados diferencialmente entre los fenotipos de DI.

Otra manera de visualizar de forma gráfica estos resultados fue con un análisis de correlación utilizando los mismos parámetros de FC y de FDR que figuran en la tabla 4 (Figura 14). Este análisis confirmó la alta correlación entre las DI del campo, y la gran diferencia que existe entre las DI del campo y del laboratorio, independientemente del fenotipo.



**Figura 14.** Diagrama de correlación de los transcritos filtrados en 10 TPM y normalizados con el método TMM. Los transcritos fueron filtrados aplicando  $FDR \leq 0,05$  y  $\log_2FC$ . Se muestra la similitud y la divergencia en la expresión génica entre las diferentes condiciones y fenotipos de DI. La clusterización euclidiana a la izquierda y en la parte superior del mapa de calor, representan el agrupamiento jerárquico de las muestras según la similitud de sus perfiles de expresión génica global (basada en correlaciones de Pearson). Las barras de colores (celeste, violeta, rojo, verde) que aparecen en esos árboles son etiquetas visuales de grupos experimentales, que representan: en

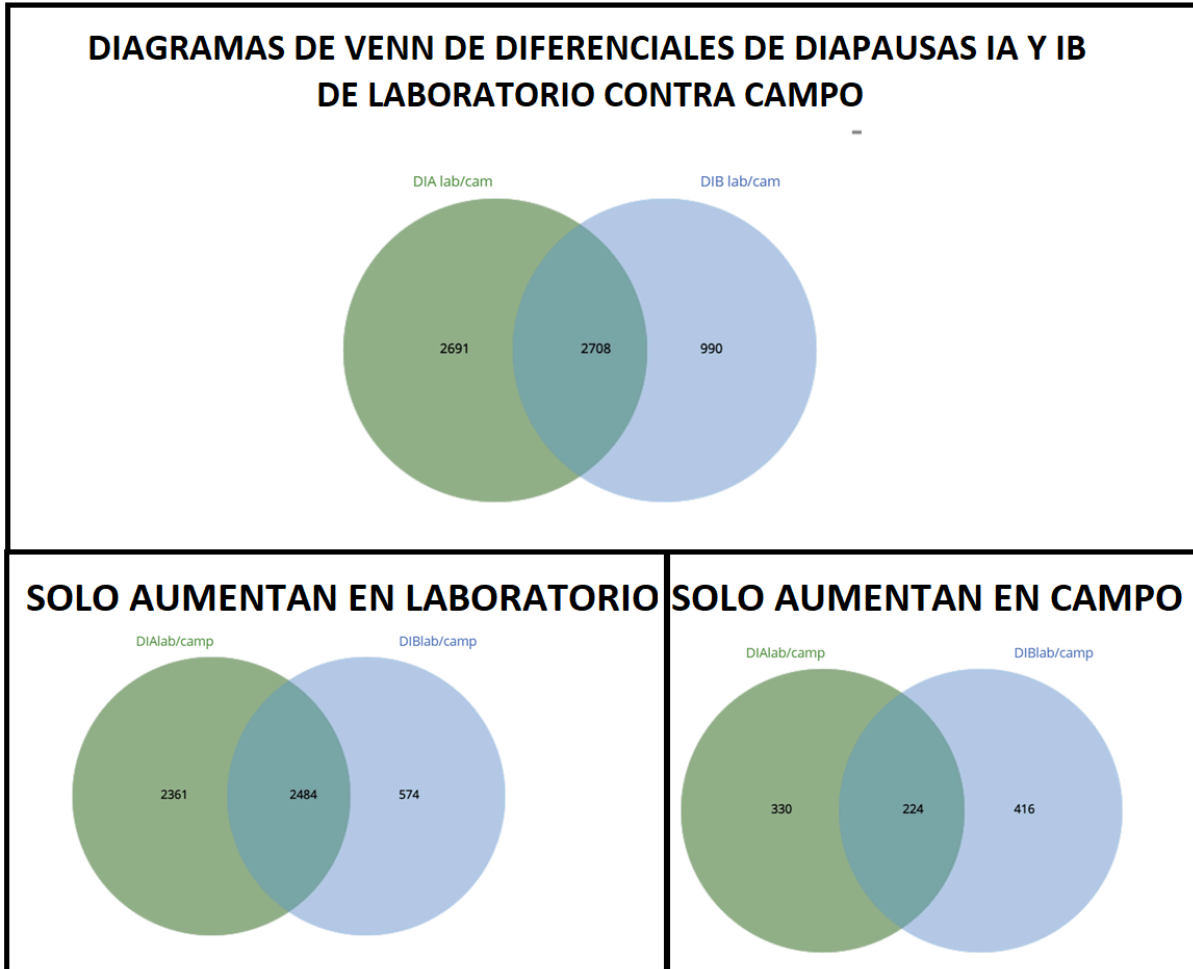
celeste, campo, en violeta, laboratorio, en rojo DIA y en verde DIB. Abajo y a la derecha del mapa de calor, están los nombres de las diferentes muestras analizadas por triplicado. Arriba a la izquierda se encuentra la escala de colores, desde el violeta (menor correlación) hasta el amarillo intenso (máxima correlación).

Los diagramas de Venn evidenciaron la cantidad de genes en común existentes entre los fenotipos de DI con respecto a los genes diferenciales entre las DI del campo y del laboratorio (Figura 15). Esto sugiere que las condiciones ambientales generan grandes diferencias de expresión entre el campo y el laboratorio, mientras que entre fenotipos dentro de un mismo ambiente los perfiles de expresión son muy similares.

En el primer diagrama de Venn, el conjunto de 2708 GED es común entre los dos fenotipos de DI, con un valor de  $FDR < 0.05$  y un  $\log_2FC$  significativo (Figura 15, diagrama superior). Al discriminar el sentido de la regulación génica, se observa que un total de 5419 genes presentaron un aumento en la expresión en las DI del laboratorio, mientras que 970 genes mostraron un incremento en las DI del campo (Figura 15, diagramas inferiores). Este resultado permite definir un primer grupo de genes candidatos que caracterizan los perfiles de expresión génica asociados a los ambientes (laboratorio y campo) durante la DI, sin distinción del fenotipo.

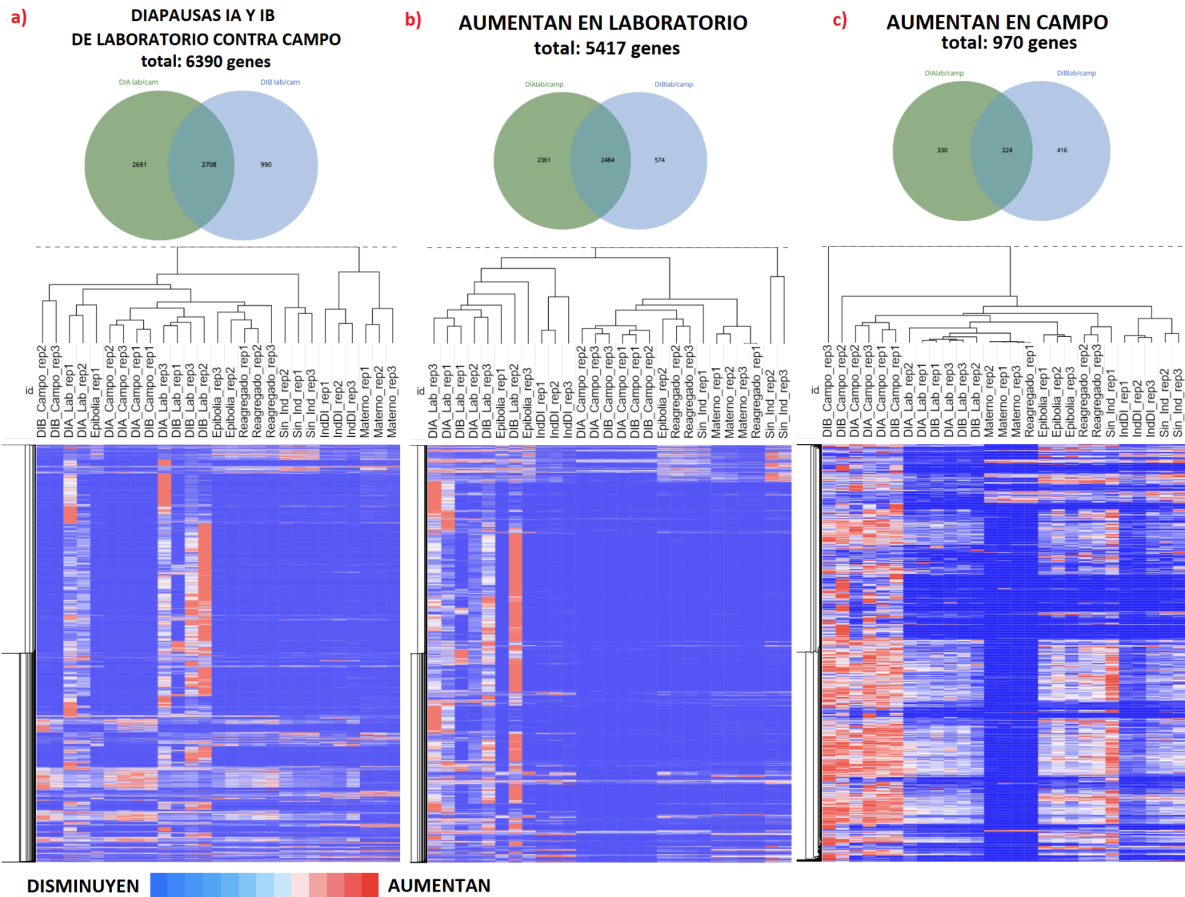
Además de la identificación de los genes diferenciales, los diagramas de Venn permiten visualizar el grado de superposición entre los transcritos regulados en las DI del laboratorio y del campo. El solapamiento observado entre las DIA y DIB confirma que una proporción importante de genes es compartida en ambos fenotipos, mientras que un número menor de genes se regula de manera exclusiva en cada uno.

Estos patrones permiten establecer listas específicas de genes candidatos, diferenciando entre aquellos compartidos y los que son exclusivos de un único fenotipo.



**Figura 15.** Diagramas de Venn de los genes con expresión diferencial mediante EdgeR de las DI del laboratorio contra las DI del campo de ambos fenotipos. Diagrama superior: total de los genes diferenciales con  $FDR < 0.05$  y  $\log_2FC$ . En los diagramas inferiores se discriminan los genes expresados positivamente (aumentan) en cada ambiente (a la izquierda: laboratorio, y a la derecha: campo).

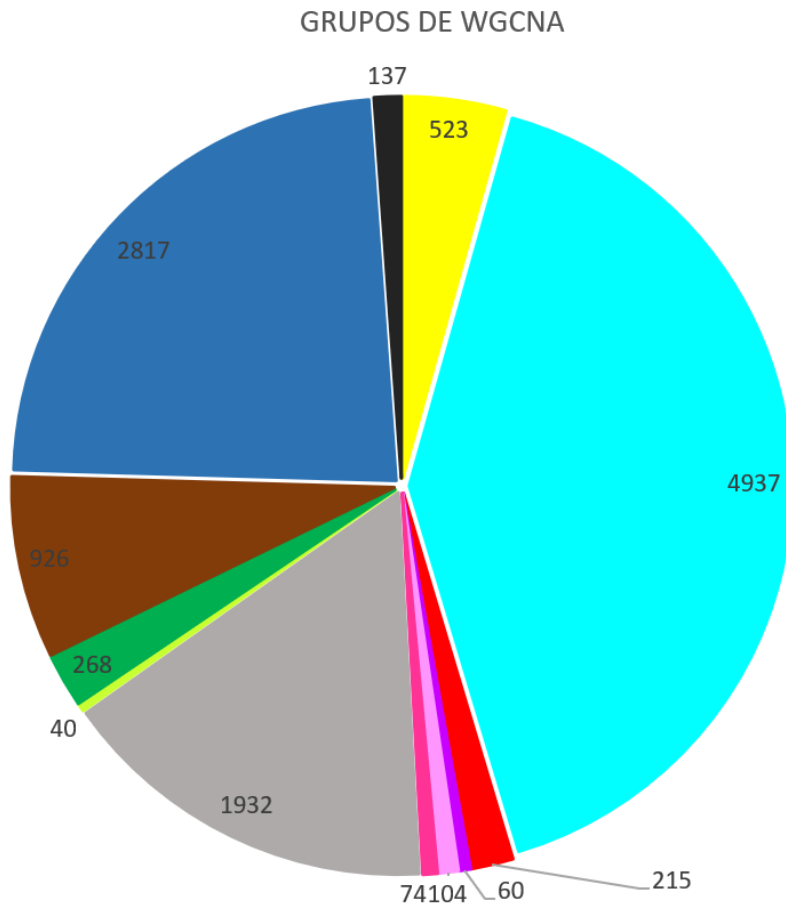
Finalmente, se realizaron mapas de calor con los genes expresados diferencialmente considerando la expresión en todos los estadios secuenciados (Figura 16). En rasgos generales, los mapas de calor (Figura 16) muestran que los GED presentan una alta expresión en las DI y una baja o nula expresión en los otros estadios del desarrollo. Los genes que aumentan su expresión en las DI del laboratorio muestran una fuerte especificidad por esa condición (Figura 16b). En cuanto a los genes expresados al alza en el campo, muestran cierta expresión en otros estadios del desarrollo, incluso en las DI del laboratorio (Figura 16c). Estos resultados permitieron definir conjuntos de genes candidatos asociados diferencialmente a las condiciones ambientales (campo o laboratorio), proporcionando información clave sobre la regulación génica durante la DI de *G. charrua*.



**Figura 16.** Mapas de calor de los genes expresados diferencialmente entre todos los estadios embrionarios secuenciados. Se aplicó clusterización euclidiana para correlacionar los estadios y los genes. En la parte superior de la figura se muestran los diagramas de Venn obtenidos en el paso anterior y en la parte inferior los diagramas de calor obtenidos para cada caso: **a)** Comparación de las DIA y DIB del laboratorio y del campo. **b)** Genes que aumentan su expresión en las DI del laboratorio. **c)** Genes que aumentan su expresión en las DI del campo. La escala corresponde a un gradiente de colores que indica los niveles de expresión.

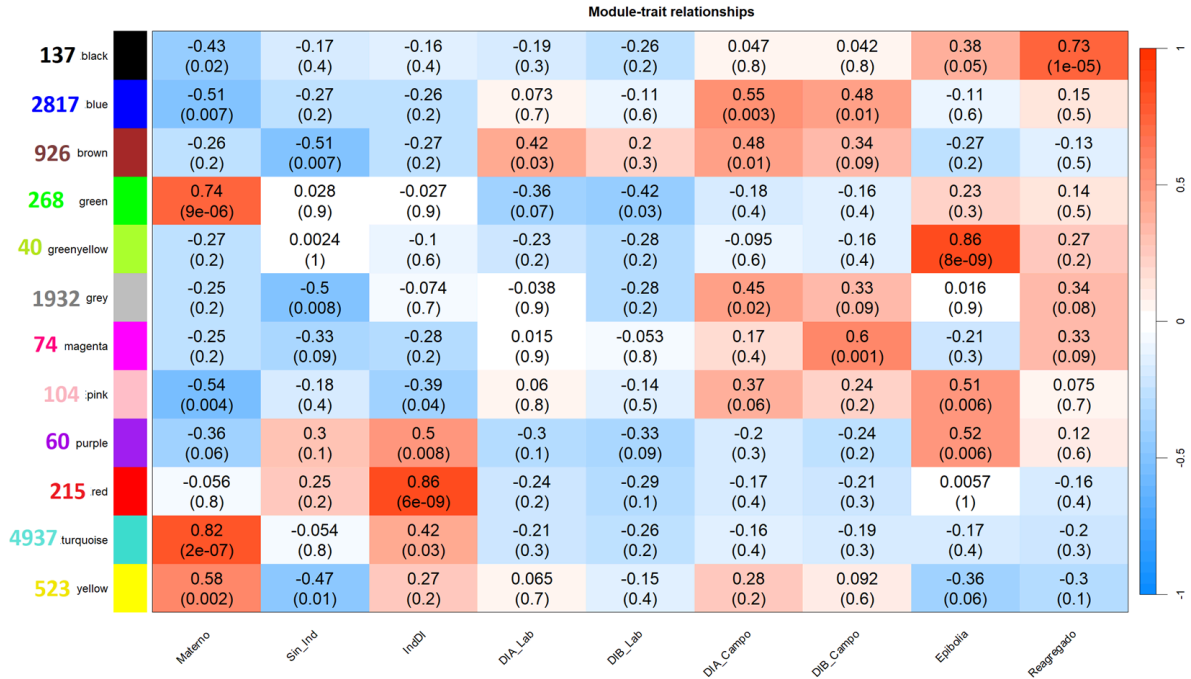
### Análisis de redes de co-expresión génica ponderadas (WGCNA)

La aplicación de los criterios de filtrado (eliminación de los genes con valores de expresión iguales a cero en al menos 15 condiciones y mantenimiento de aquellos con expresión mayor o igual a 5 TPM en 4 condiciones de diapausa sin importar el fenotipo y el ambiente) permitió disminuir de 30735 a 12031 la cantidad de genes a analizar. Con este grupo de genes filtrados se realizó el agrupamiento mediante el método WGCNA, donde se asociaron los genes que presentan similitud de expresión, arrojando como resultado 12 grupos con diferentes genes cada uno (Figura 17).



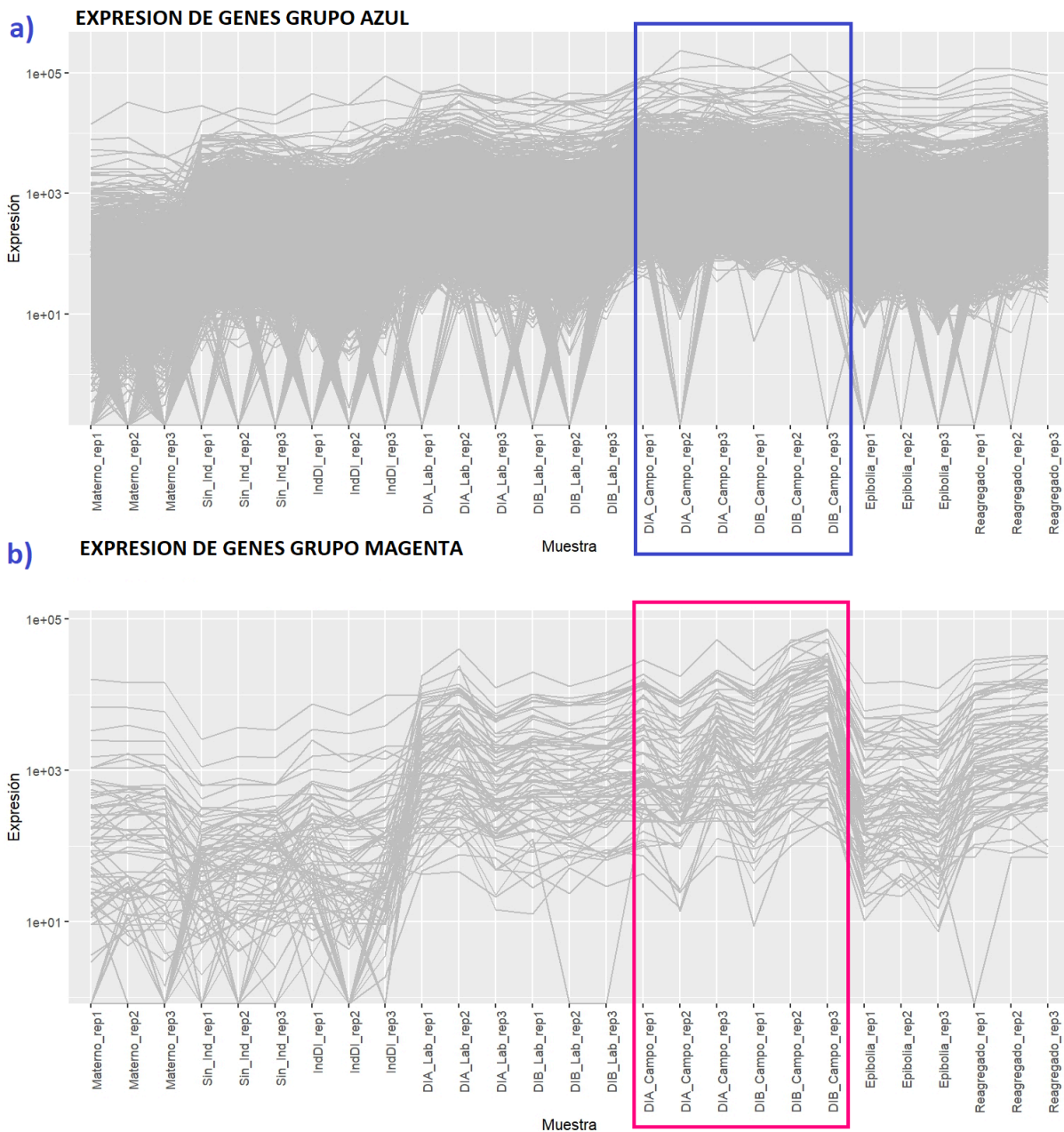
**Figura 17.** Diagrama de torta que grafica el número de genes en cada grupo del WGCNA. El grupo gris representa a los genes que no se agrupan (no exhiben correlación). Dentro de cada grupo está escrito el número de genes correspondiente.

En la figura 18 se observan los 12 grupos de genes que se coexpresan en los distintos estadios. Se seleccionaron aquellos que poseen expresiones al alza o a la baja con un p-valor menor a 0,05 en los estadios de DI del campo. El primer grupo que reúne estas características es el grupo azul, el cual contiene un grupo de genes de interés de acuerdo con el objetivo de este trabajo. En este grupo, el índice de correlación entre las diapausas de campo es positivo, agrupando genes que se expresan al alza en las DI (A y B). El grupo marrón también presenta expresiones positivas y con alta correlación en todos los estadios de DI (A y B), tanto en los embriones obtenidos en el campo como en los inducidos en el laboratorio. En sentido opuesto, el grupo rojo presenta correlaciones negativas, o genes que se expresan a la baja en todos los estadios de DI, pero con un valor de correlación mayor en las DI del laboratorio. Esto podría indicar un perfil de expresión génica exclusivo de las condiciones de laboratorio, por lo cual los grupos marrón y rojo se descartaron para los análisis posteriores.



**Figura 18.** Mapa de calor obtenido del programa WGCNA llamado: Module-trait relationships. A la izquierda del mapa se encuentran los grupos discriminados con colores y la cantidad de genes presentes en cada uno. En el eje x se encuentran los estadios del desarrollo embrionario estudiados en este trabajo. A la derecha, la escala de color indica el gradiente de niveles de expresión génica: los genes que aumentan su expresión en un determinado estadio se representan en tonos anaranjados y los que se expresan a la baja se muestran en tonos azules. En cada estadio se consideraron las tres réplicas biológicas para construir la correlación.

Considerando lo expuesto anteriormente, el análisis de ontología génica se centró en los grupos azul y magenta, ya que ambos presentan correlaciones positivas en las diapausas del campo, aunque el magenta contiene un número menor de genes (Figura 19).



**Figura 19.** Gráficos de expresión génica obtenidos en el WGCNA. En el eje y: nivel de expresión génica; eje x: muestras analizadas. Los recuadros azul y magenta indican las muestras correspondientes a las DI de campo en cada grupo.

En ambos grupos, es posible apreciar una máxima expresión génica en ambos fenotipos de DI del campo en comparación con las DI del laboratorio y con los demás estadios del desarrollo.

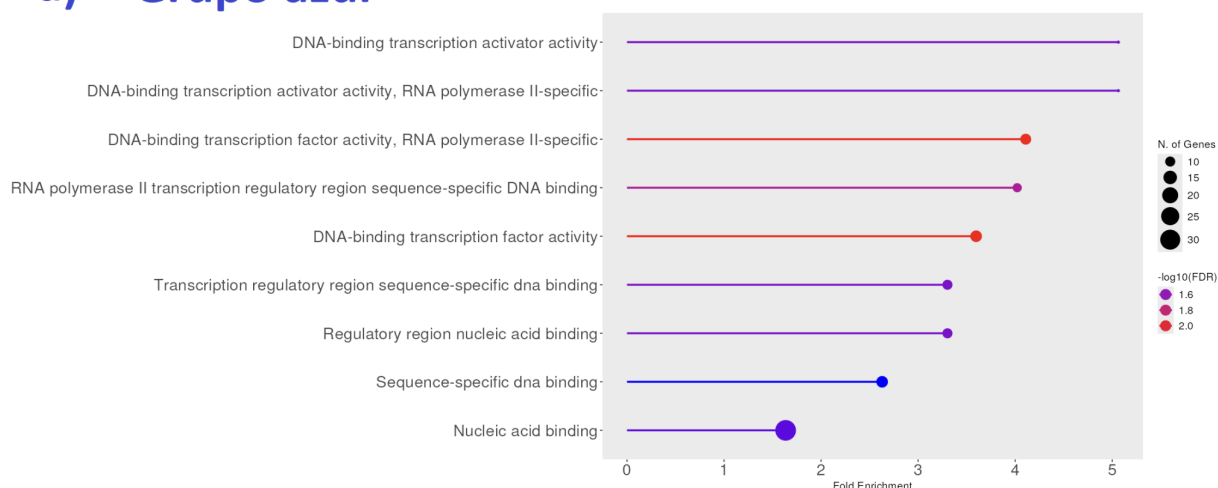
### Estudios de ontología génica (GO) y enriquecimiento

Con el objetivo de conocer las implicancias biológicas de los genes que aumentan su expresión en las DI del campo, se realizó el análisis de GO, utilizando los programas ShinyGO y String, de los grupos de genes azul y magenta obtenidos en el WGCNA.

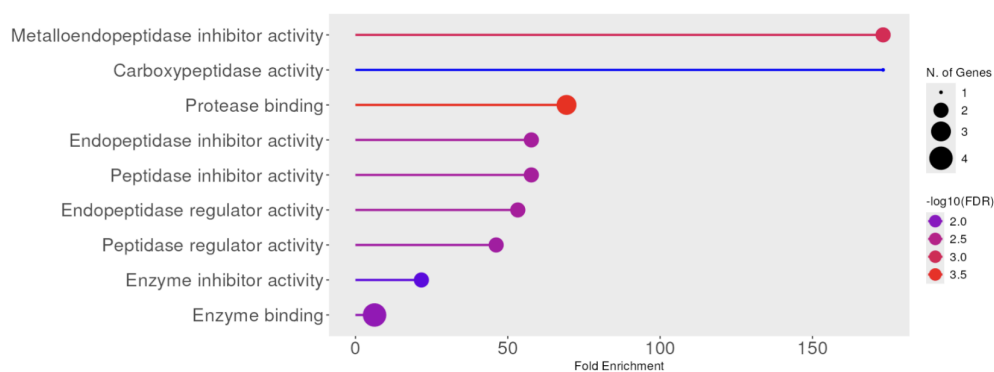
En el grupo azul, de los 2817 genes, solo 329 están anotados en la base de datos de Trinotate. A partir de los genes anotados, se realizó el análisis de GO y enriquecimiento arrojando como resultado los procesos y funciones biológicas de estos grupos de genes que se expresan al alza en las DI del campo (Figura 20; tabla A2 en el anexo). Este grupo está compuesto por categorías de funciones moleculares que, en su mayoría, están vinculadas al núcleo celular, por ejemplo en funciones enzimáticas relacionadas con la polimerasa II. Se identificaron además categorías muy relevantes para este estudio relacionadas con la actividad de factores de transcripción y reguladores transcripcionales entre otros. Dentro de la vía más enriquecida en este estudio (factores de transcripción con motivos de unión al ADN), se encontraron genes de las familias Homeobox y Forkhead box (Figura 20a, primera vía enriquecida). En una categoría, con un bajo enriquecimiento, denominada proteínas citosólicas de unión a enzimas relacionadas con la protección frente al estrés ambiental, se detectaron dos genes de interés. Uno de ellos es el gen que codifica la proteína de choque térmico frío y el otro gen perteneciente a la vía de IGF (Tabla A3 en el anexo).

En el grupo magenta, que contiene 74 genes, solo 42 están anotados. Las funciones moleculares asignadas a éstos muestran un vínculo con la actividad de proteínas reguladoras del metabolismo en general (Figura 20b). Además, se encontraron retrotransposones de la familia LINE1 (elemento nuclear largo disperso-1), que está vinculado con la regulación de respuesta al estrés celular a través del vínculo con brca1. Sin embargo en el gráfico de enriquecimiento no figura ninguna categoría en la que participe este elemento transponible (Tabla A3 en el anexo).

## a) Grupo azul



## b) Grupo magenta

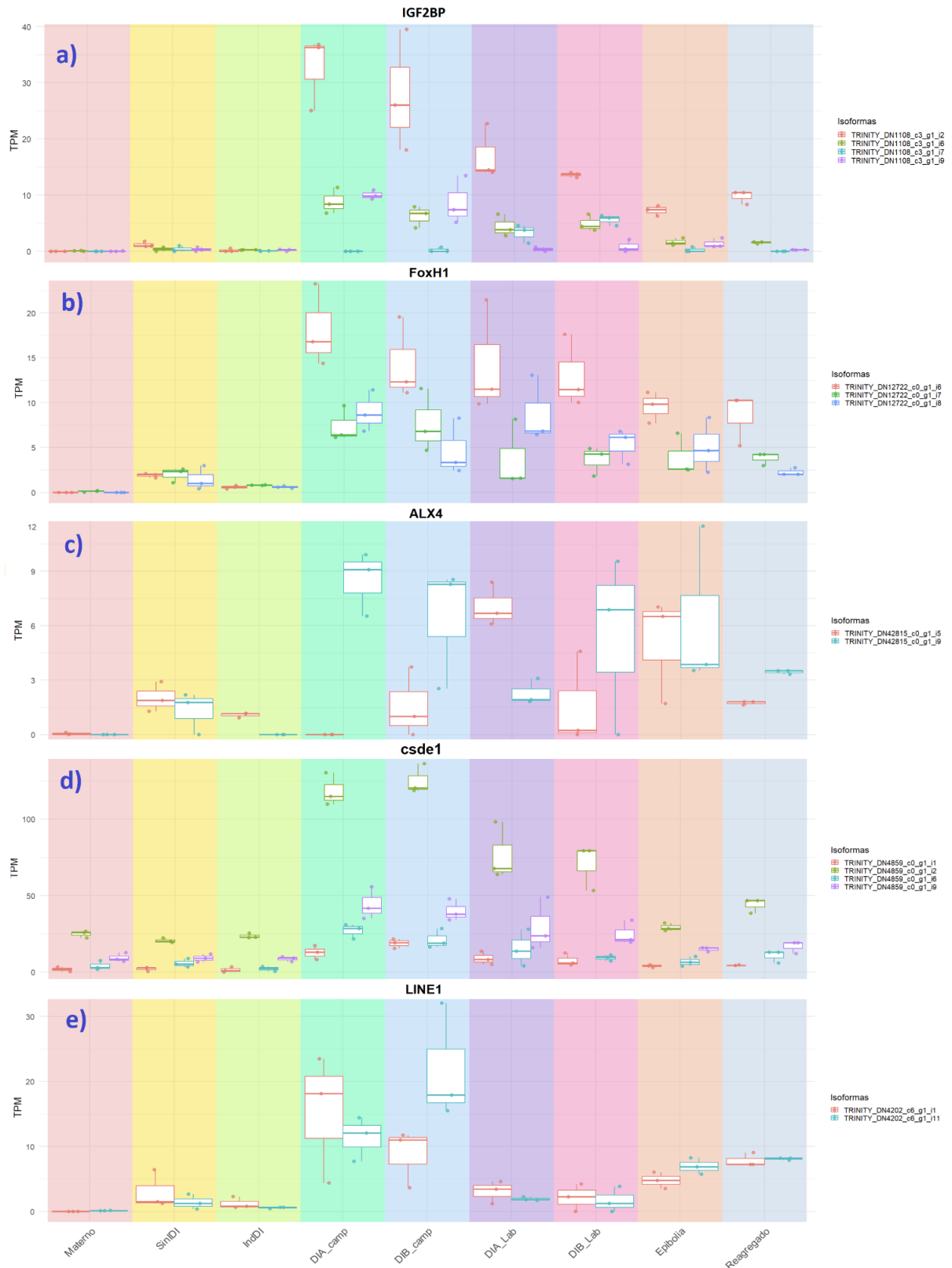


**Figura 20.** Gráficos de enriquecimiento de genes involucrados en algunas funciones moleculares. a) Grupo azul; b) Grupo magenta. En el eje y se encuentran las funciones moleculares a las que pertenecen los distintos genes. En el eje x, el “fold enrichment” corresponde a la comparación de la frecuencia observada de un conjunto de genes con la frecuencia esperada, tomando como referencia el conjunto base (el subconjunto de genes de este grupo). A la derecha de la figura se encuentran las escalas empleadas: arriba (el tamaño de los puntos al final de cada línea, es proporcional al número de genes con esa función), abajo (los colores del gráfico varían según el logaritmo en base 10 de FDR (False discovery rate) que indica, en escala numérica, la tasa de falsos positivos). Gráfico realizado en el programa ShinyGO v0.741 (valores del análisis en la tabla A3 en el anexo).

### Genes seleccionados de interés

De las listas de genes que componen los grupos azul y magenta, surgieron algunos candidatos que han sido reportados en otros trabajos como genes involucrados en los procesos biológicos que regulan a las diapausas en peces anuales y otros organismos según la bibliografía mencionada en materiales y métodos. Este grupo de 5 genes son: la proteína de unión al ARNm del factor de crecimiento similar a la insulina tipo 2 (IGF2BP) de la vía de IGF, los factores de transcripción Forkhead H1 (FoxH1), el Homeobox tipo 4 (ALX4), la proteína de choque térmico frío (csde1) y el retrotransposón LINE1. El patrón de expresión de estos genes en los distintos estadios embrionarios muestra que se encuentran mayoritariamente en las diapausas, principalmente en las DI del campo (Figura 21).

Para los 5 genes seleccionados, el ensamblado generó varias isoformas que muestran distintos perfiles de expresión en los diferentes estadios del desarrollo y en las DI del campo y del laboratorio. El gen IGF2BP, presenta cuatro isoformas con perfiles similares entre ellas, pero con una variación en sus niveles de expresión (Figura 21a). El gen FoxH1, contiene tres isoformas que varían en sus niveles de expresión pero con similares características (Figura 21b). El gen ALX4, cuenta con dos isoformas con perfiles de expresión antagónicos, principalmente en las DI del campo y del laboratorio (Figura 21c). Con respecto al gen csde1, se identificaron cuatro isoformas que mantienen perfiles similares de expresión en los diferentes estadios (Figura 21d). Por último, el LINE1 posee dos isoformas con perfiles similares pero con pequeñas variaciones en las DI del laboratorio (Figura 21e).



**Figura 21.** Gráficos de cajas de los niveles de expresión de las isoformas de los 5 genes de interés por triplicado en cada estadio del desarrollo. El eje y corresponde a TPM y el eje x a los diferentes estadios embrionarios (columnas coloreadas). A la derecha de la imagen se encuentra la leyenda correspondiente a los colores de las isoformas de cada gen: a) IGF2BP; b) FoxH1; c) ALX4; d) csde1 y e) LINE1.

## **Análisis de redes vinculadas con los genes de interés**

Con el objetivo de indagar, según la bibliografía, en qué procesos biológicos, funciones moleculares y vías de señalización están vinculados los 5 genes de interés, se realizó un análisis de enriquecimiento. Los resultados obtenidos evidencian la participación de estos genes en procesos de desarrollo embrionario de vertebrados, tales como la regulación del desarrollo de los ovocitos, vías de señalización nodal y otros procesos involucrados en el desarrollo de la placa neural, según un orden jerárquico. En cuanto a las funciones moleculares, los genes de interés y las vías asociadas a éstos presentan funciones de unión a proteínas SMAD y al ADN principalmente (Tabla 5).

**Tabla 5.** Análisis de enriquecimiento funcional para los 5 genes de interés seleccionados. Tabla superior: procesos biológicos asociados; tabla inferior: funciones moleculares vinculadas. A la derecha de la tabla se muestra una escala con un gradiente de color (rojo a blanco) de acuerdo con el nivel de asociación de mayor a menor.

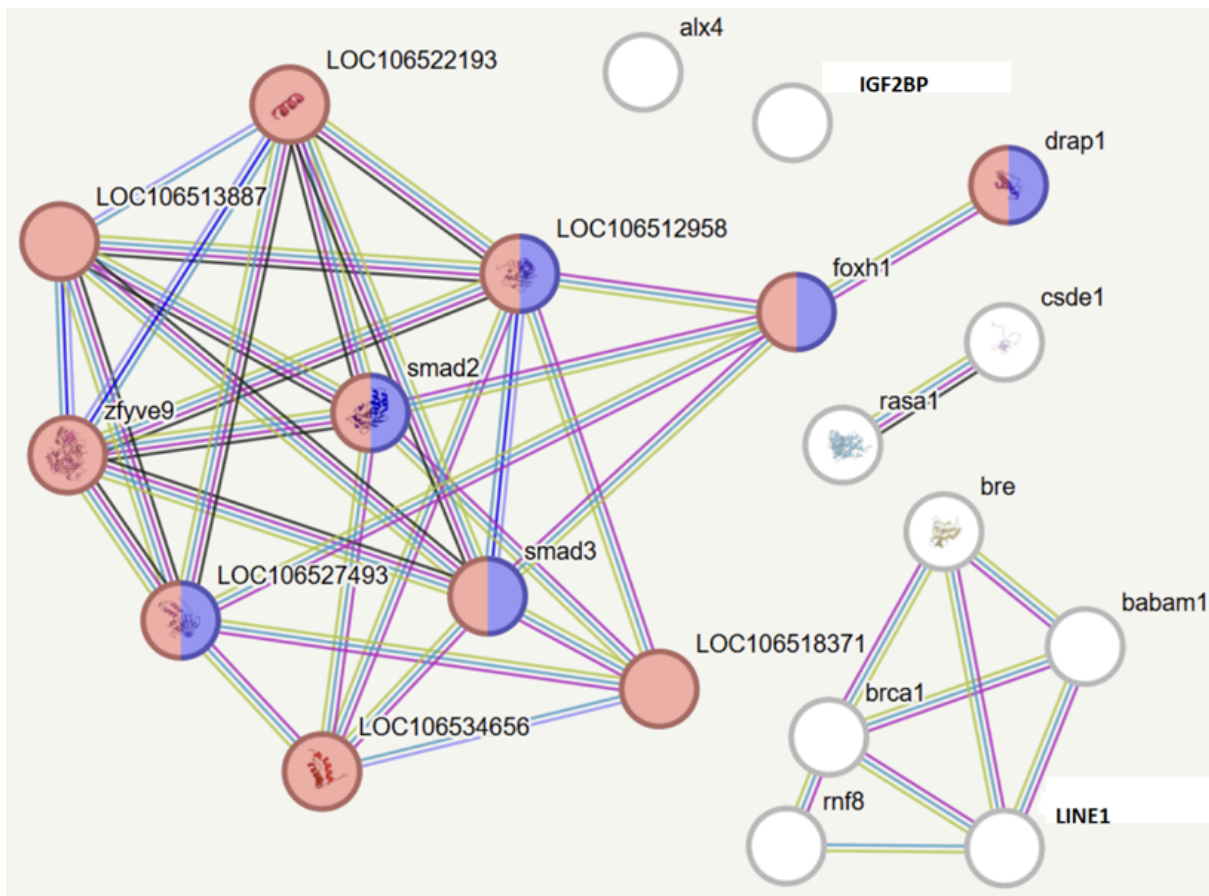
## Enriquecimientos funcionales en su red

Proceso biológico (Ontología genética)	
<i>GO-term</i>	<i>descripción</i>
GO:0060282	Regulación positiva del desarrollo de los ovocitos
GO:0038092	Vía de señalización nodal
GO:0021999	Regionalización anterior/posterior de la placa neural
GO:0009880	Especificación de patrones embrionarios
GO:0010212	Respuesta a la radiación ionizante
GO:0045739	Regulación positiva de la reparación del ADN
GO:0007095	Señalización del punto de control de daño del ADN mitóti...
GO:0007179	Vía de señalización del receptor del factor de crecimient...
GO:0060395	Transducción de señales de la proteína SMAD
GO:0030509	Vía de señalización de BMP
GO:0006302	Reparación de rotura de doble hebra
GO:0009952	Especificación del patrón anterior/posterior
GO:0051052	Regulación del proceso metabólico del ADN
GO:0045934	Regulación negativa del proceso metabólico de compue...
GO:0045892	Regulación negativa de la transcripción, con plantilla de ...
GO:0000122	Regulación negativa de la transcripción por la ARN polim...
GO:0035556	Transducción de señales intracelulares
GO:1903047	Proceso del ciclo celular mitótico
GO:0006325	Organización de la cromatina

Función molecular (ontología genética)	
<i>Término GO</i>	<i>descripción</i>
GO:0070411	Unión I-SMAD
GO:0046332	Unión SMAD
GO:0005545	Unión de 1-fosfatidilinositol
GO:0000976	Unión a la región cis-reguladora de la transcripción
GO:0140110	Actividad reguladora de la transcripción
GO:0000977	Unión específica de ADN a la secuencia de la región ...
GO:0000978	Unión específica de ADN a la secuencia de la región ...
GO:0000981	Actividad del factor de transcripción de unión al AD...

El análisis de las interacciones de los genes de interés (Figura 22), permite apreciar en el diagrama las redes formadas por cada gen seleccionado. El gen que presenta la mayor cantidad de conexiones es el FoxH1. Éste se conecta con genes relacionados con la regulación del desarrollo de ovocitos, la vía de señalización nodal y la unión a proteínas SMAD, entre otras, que se encuentran dentro de las vías más enriquecidas de este análisis (Tabla 5). El segundo grupo con mayor enriquecimiento es el grupo al que pertenece el retrotransposón LINE1 que se relaciona con proteína brca1 presente en las categorías de respuesta al daño del ADN y el estrés celular. Finalmente, el csde1 se interrelaciona con el

gen *rasa1* que está vinculado con procesos de transducción de señales intracelulares y el ciclo celular.



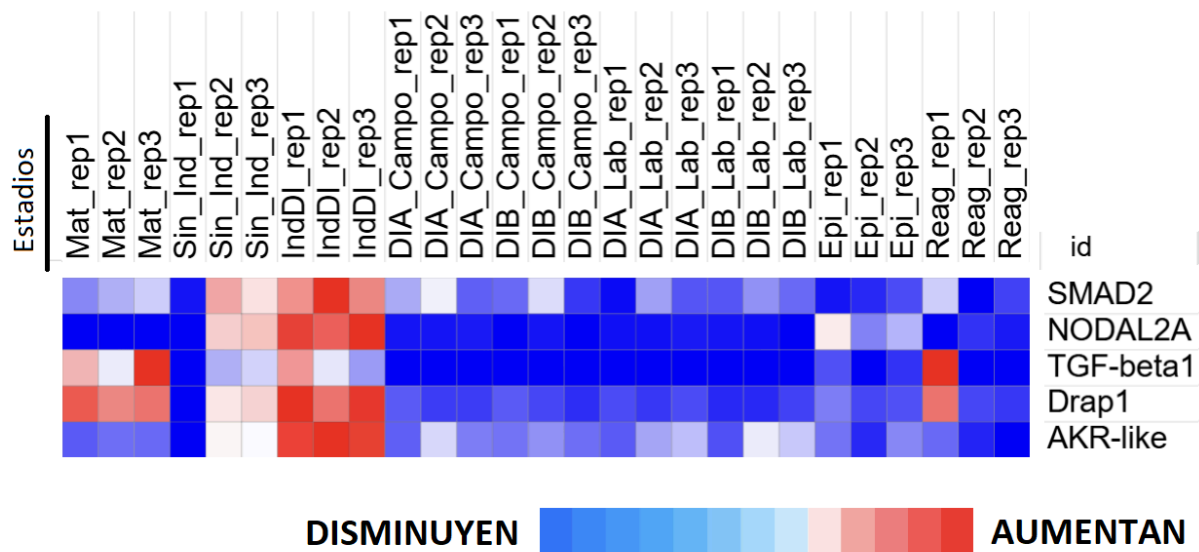
**Figura 22.** Análisis de redes de enriquecimiento utilizando el programa String. Cada círculo corresponde a la función del respectivo gen. En rojo se indican los genes de la vía más enriquecida del estudio GO que corresponde a la regulación positiva del desarrollo de ovocitos que comprende genes de la familia TGF- $\beta$ . Los círculos coloreados mitad rojo y mitad violeta, involucran a la segunda vía más enriquecida que corresponde a la vía de señalización nodal. Los círculos blancos corresponden al resto de los genes. Las líneas que conectan los círculos corresponden a la relación y al rol de cada gen en la vía, mientras más líneas coloreadas, mayor cantidad de funciones se le atribuye a esas conexiones. Los diferentes colores de las líneas indican diferentes artículos bibliográficos que relacionan y forman las redes de enriquecimiento.

Posteriormente, se realizó un mapa de calor para visualizar los niveles de expresión de los genes involucrados en la vía más enriquecida que vincula a FoxH1 (Figura 23). El objetivo fue conocer cuándo se expresan en los estadios estudiados los genes; SMAD2 (coactivador transcripcional de la vía de señalización TGF- $\beta$ ), Drap1 (factor nuclear que se une al complejo DR1/NC2, actuando como represor general de la transcripción) y Homeobox AKR-like (proteína de la familia de las Homeobox).

Paralelamente se realizó una búsqueda bibliográfica de aquellos genes que tienen relevancia en la vía de TGF- $\beta$  y Nodal, y de esta búsqueda quienes están expresados en el transcriptoma de *G. charrua*. De esta búsqueda surgen dos genes de importancia para este proceso que son: NODAL2A (es un gen parálogo del gen nodal, es un ligando que activa la

vía Nodal-SMAD2/3) (Schier y Talbot 2005) y TGF- $\beta$ 1 (factor de crecimiento transformante  $\beta$ 1) (Massagué, 2012).

El mapa de calor revela que estos genes presentan una alta expresión en el estadio de inducción a la DI (IndDI). Sin embargo, en las DI del campo y del laboratorio, su expresión es prácticamente nula y en los demás estadios del desarrollo, es baja en la mayoría de las réplicas (Figura 23).



**Figura 23.** Mapa de calor de la expresión de los 5 genes (a la derecha) relacionados a FoxH1 en los diferentes estadios del desarrollo embrionario (parte superior del gráfico). En la parte inferior de la figura se muestra la escala de colores, como se menciona en las figuras anteriores.

### Análisis filogenético de FoxH1

Considerando el perfil de expresión y la relevancia documentada en diferentes organismos que experimentan diapausa (Karp, 2021; Pei y col. 2007), se realizó un análisis filogenético de la isoforma 7 (i7) de FoxH1 ya que es la más expresada en las DI de los embriones del campo (Figura 21b).

Para el análisis, se obtuvieron del GenBank secuencias nucleotídicas homólogas a FoxH1\_i7 de *G. charrua* correspondientes a otros Cyprinodontiformes y al grupo externo *O. latipes*. El alineamiento múltiple resultante presentó una longitud total de 1209 pb (Anexo A2).

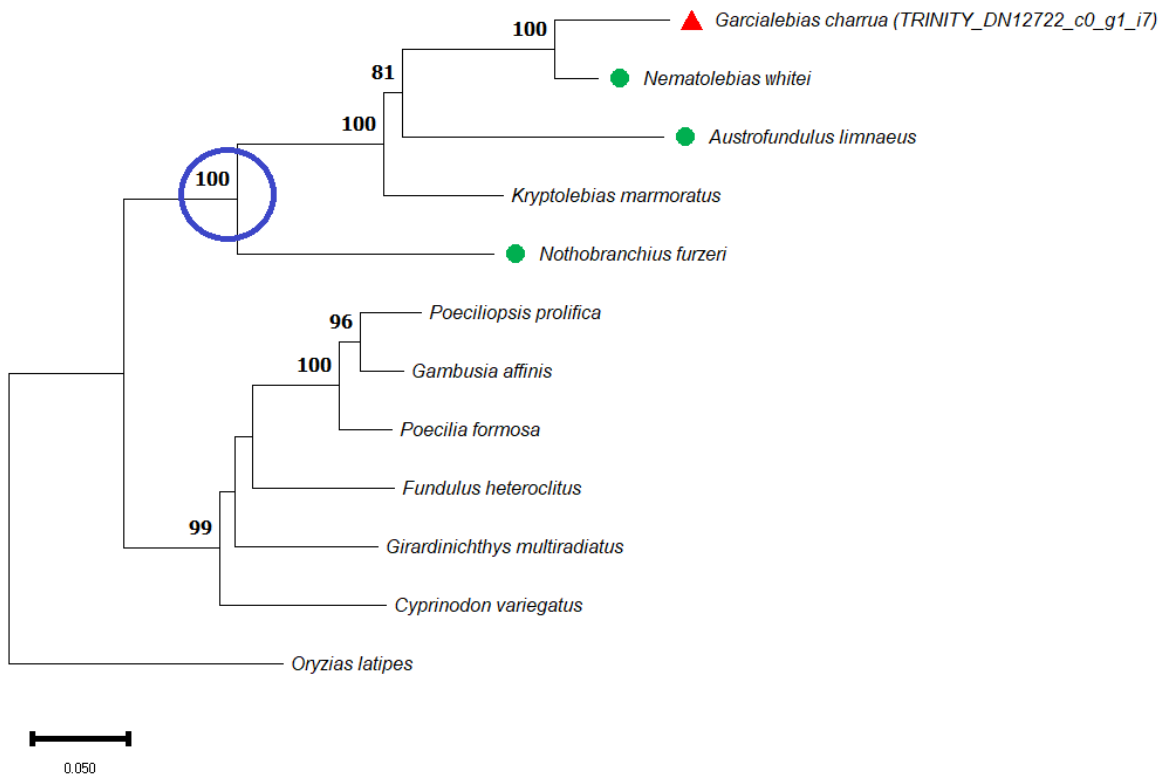
El análisis de polimorfismos efectuado en las 11 secuencias nucleotídicas de los Cyprinodontiformes (Tabla 1), mostró variación en los 1209 pb (Tabla 6). El análisis de los cuatro rivúlidos (*G. charrua*, *N. whitei*, *A. limnaeus* y *K. marmoratus*), así como de los peces anuales (*G. charrua*, *N. whitei*, *A. limnaeus* y *N. furzeri*) revelaron que la secuencia del gen

FoxH1 se encuentra altamente conservada dentro de estos grupos, posiblemente debido a su función biológica, particularmente en la diapausa.

**Tabla 6.** Análisis de polimorfismos del gen FoxH1, isoforma 7, en todos los grupos analizados.

Grupo	C (Conservados)	V (Variables)	Pi (Sitios filogenéticamente informativos)
<b>Cyprinodontiformes</b>	594	615	356
<b>Rivulidos</b>	874	290	70
<b>No anuales</b>	771	396	106
<b>Anuales</b>	801	366	58

La topología del árbol filogenético resultante (Figura 24), se obtuvo usando el modelo de sustitución nucleotídica T92 con un valor de verosimilitud logarítmica  $-\log L = -6951.53$  y una distribución gamma discreta con dos categorías (+G), con un parámetro de forma  $\alpha = 0.5291$ . Esto indica una alta heterogeneidad, lo que sugiere que algunos sitios están fuertemente conservados mientras que otros evolucionan más rápidamente. En el árbol se observan dos grandes clados. Uno de ellos, con un soporte de bootstrap de 100% (Figura 24, círculo azul), incluye las 4 especies de peces anuales y el rivúlido no anual *K. marmoratus*. Dentro de este grupo, *G. charrua* se ubica como especie hermana de *N. withei* con un soporte de bootstrap del 100%. Este subclado se agrupa con *A. limnaeus* con un soporte del 81% de bootstrap y éste con *K. marmoratus* con un soporte del 100%. En una posición más basal, se encuentra el pez anual africano *N. furzeri*. El otro clado, con un soporte de bootstrap del 99%, está integrado por otros Cyprinodontiformes no anuales, donde las especies de la familia Poeciliidae forman un grupo monofilético con 100% de soporte (Figura 24).



**Figura 24.** Árbol filogenético de MV (bajo el modelo T92) de las secuencias nucleotídicas del gen FoxH1 para las especies de Cyprinodontiformes analizadas. *Oryzias latipes* se empleó como grupo externo. Los valores de bootstrap mayores al 75% están indicados sobre las ramas. Se muestran las longitudes de rama a escala, expresadas como número de sustituciones por sitio. Triángulo rojo: indica a la especie de estudio *G. charrua*. Círculos verdes: indican otras especies de peces anuales. El círculo azul señala el nodo que abarca las especies de interés.

### Análisis de selección

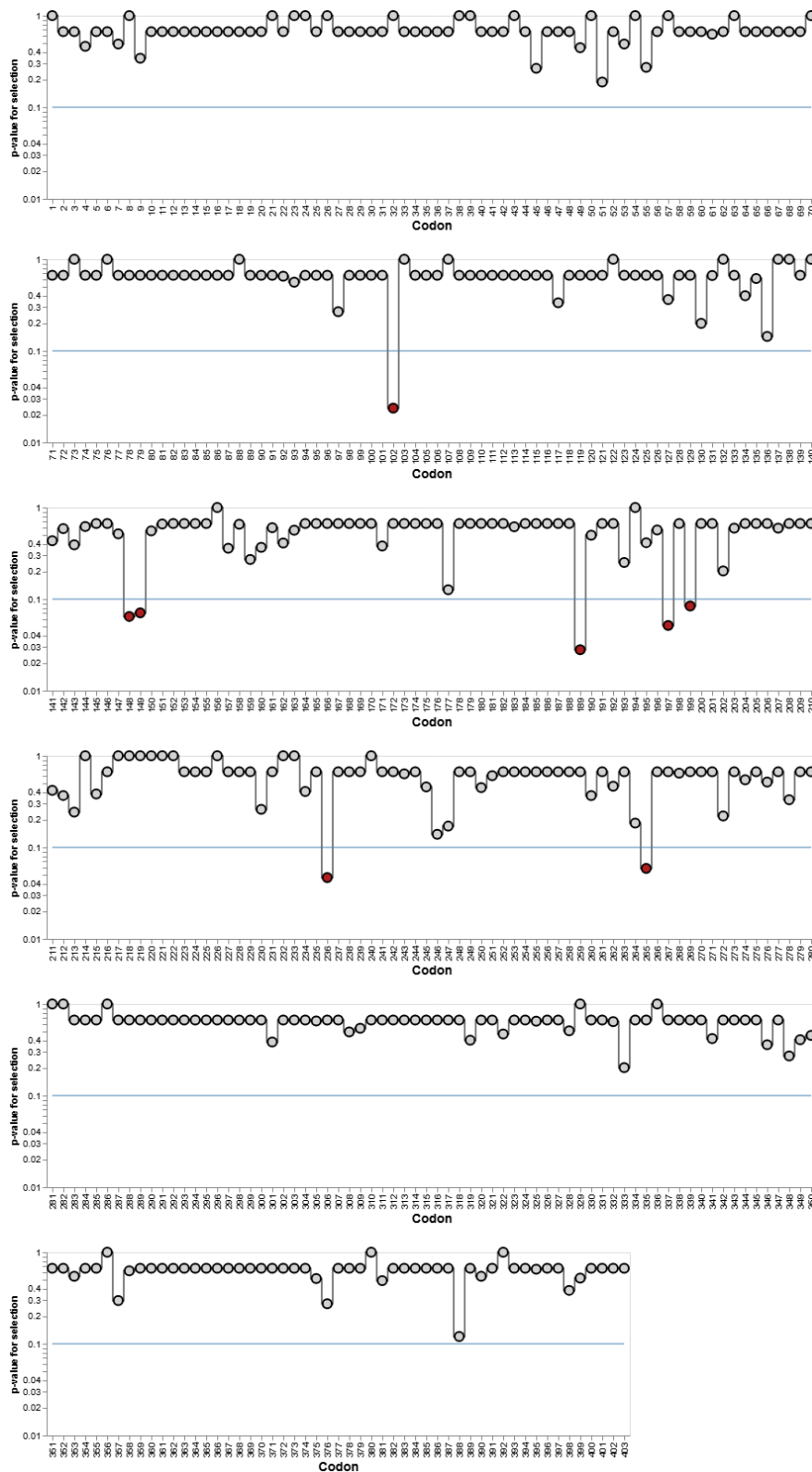
Los datos de las secuencias nucleotídicas de FoxH1 se analizaron evaluando la diversidad nucleotídica y las tasas de divergencia sinónima y no sinónima (Tabla 7).

El análisis de las sustituciones sinónimas ( $K_s$ ) y no sinónimas ( $K_a$ ) mostró consistentemente que la relación  $K_a/K_s$  es menor a 1 en todos los grupos estudiados, lo que indica que el gen FoxH1 está bajo selección purificadora (Tabla 7). Esto sugiere que las sustituciones no sinónimas son eliminadas evolutivamente, preservando la función proteica esencial de FoxH1 dentro de los Cyprinodontiformes. Aunque los valores de  $K_a/K_s$  varían levemente entre grupos, la tendencia general indica una fuerte conservación funcional. Los patrones de diversidad nucleotídica muestran que en todos los grupos la diversidad sinónima ( $P_i(s)$ ) es mayor que la no sinónima ( $P_i(a)$ ), reflejando un predominio de cambios nucleotídicos que no afectan la secuencia aminoacídica. Las relaciones  $P_i(a)/P_i(s)$  confirman esta tendencia, consistente con un escenario en el que FoxH1 mantiene regiones funcionalmente importantes en las diferentes especies de Cyprinodontiformes.

**Tabla 7.** Análisis de selección en el gen FoxH1, isoforma 7, en todos los grupos analizados.

Grupo	Número de secuencias intraspecíficas	Número de sitios segregantes (intraspecíficos)	Pi(s) (sinónima)	Pi(a) (no sinónima)	Relación Pi(a)/Pi(s)	Ks (sinónima)	Ka (no sinónima)	Relación Ka/Ks
Cyprinodontiformes	11	516	0.423	0.100	0.172	0.543	0.129	0.147
Rivulidos	4	280	0.293	0.092	0.263	0.566	0.166	0.178
Anuales	4	337	0.355	0.116	0.261	0.565	0.158	0.169
No anuales	7	398	0.366	0.065	0.136	0.537	0.115	0.132

Profundizando en este análisis, el programa Datamonkey reveló que 8 codones se encuentran bajo selección diversificadora episódica en los Cyprinodontiformes (con  $p \leq 0.1$ ), en las posiciones: 102, 148, 149, 189, 197, 199, 236 y 266 (Figura 25).



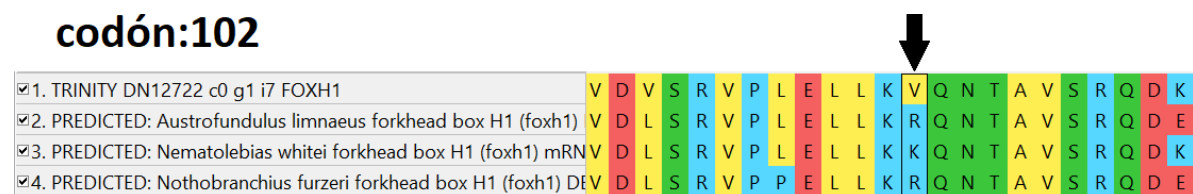
**Figura 25.** Gráfica del análisis de selección episódica en el programa web de Datamonkey. Se muestran todas las comparaciones del gen FoxH1 en los Cyprinodontiformes seleccionados. En el eje X se representan los codones a lo largo de la secuencia del gen analizado (1–420), mientras que el eje Y se indica el valor de p asociado a la probabilidad de selección diversificadora en cada codón. Los círculos rojos del gráfico representan los codones bajo selección diversificadora episódica, y los círculos grises, son codones donde predomina la selección purificadora o la neutralidad. Las líneas que conectan los puntos muestran cómo varían los valores de p entre codones consecutivos,

permitiendo visualizar con claridad los descensos abruptos hacia valores bajos de p que indican evidencia de selección. La línea horizontal gris representa el umbral de significancia estadística de qué se considera selección purificadora (con  $p \leq 0.1$ ).

## Análisis del codón 102

En la estructura primaria de la proteína FoxH1 de *G. charrua*, de los 8 codones con selección diversificadora episódica, se analizó con mayor profundidad el codón ubicado en la posición 102 donde se observó un cambio respecto al aminoácido que posee esta proteína en otros peces anuales (Figura 26). El resto de los codones encontrados con selección diversificadora episódica no revisten gran relevancia respecto a cambios de aminoácidos con propiedades fisicoquímicas diferentes (Figura A2 en el anexo).

En las especies comparadas, los aminoácidos presentes en esta posición son arginina (R) en *A. limnaeus* y *N. furzeri*, y lisina (K) en *N. whitei*, mientras que en *G. charrua* se encuentra una valina (V). Los primeros aminoácidos contienen carga positiva (básicos) y son hidrofílicos (polares), mientras que la valina (V) es neutro (no polar) e hidrofóbico. Esto implica que se está sustituyendo un aminoácido cargado y polar por uno no cargado y no polar, lo que podría modificar propiedades estructurales y funcionales de la proteína.



**Figura 26.** Alineamiento de las secuencias aminoacídicas de FoxH1 de las cuatro especies de peces anuales. TRINITY DN12722 C0 G1 I7 corresponde a FoxH1 de *G. charrua*. La flecha negra marca la posición del aminoácido donde se encuentra el codón del codón 102.

Para analizar en qué región de la proteína FoxH1 se ubica este aminoácido se utilizó la base de datos Uniprot (<https://doi.org/10.1093/nar/gkae1010>), donde se detectaron diferencias entre los distintos organismos analizados. En FoxH1 de *G. charrua* se encuentra fuera del dominio de unión al ADN, pero dentro de un dominio denominado Fork Head 3, también conocido como dominio de hélice alada (Clark y col. 1993). En *N. furzeri* el codón 102 se ubica en la región de unión al ADN. En *A. limnaeus* está posicionado de forma muy similar a *G. charrua* dentro del dominio Fork Head 3 (Figura 27a). En especies de peces no anuales y otros organismos, se encuentra en la región de unión al ADN (datos no mostrados).

Por último, se realizó un modelado tridimensional de FoxH1 marcando el codón 102 donde se presenta la estructura tridimensional de la región de la proteína en la que se encuentra la valina (Figura 27b).

El modelo estructural con alto nivel de confianza en el modelado (Figura 3A de anexos) muestra un dominio central altamente ordenado, representado en color azul oscuro ( $pLDDT > 90$ ), correspondiente a la región con mayor confianza en la predicción. Dentro de este núcleo estructurado se localiza la valina, situada en un segmento compacto que forma parte del pliegue interno del dominio Fork Head 3. La posición de la valina es la 161 en la

estructura primaria de la proteína y está en la zona del núcleo hidrofóbico del dominio, probablemente contribuyendo a la estabilidad estructural de la proteína. Además, su ubicación en una región conservada sugiere que cambios en este residuo podrían afectar el empaquetamiento del dominio.

a)

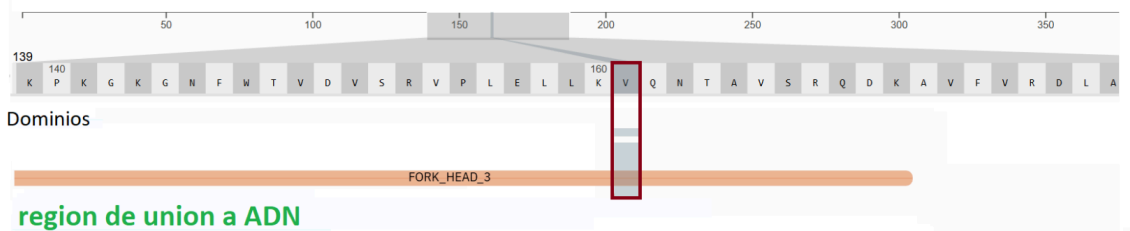
## *Nothobranchius furzeri*



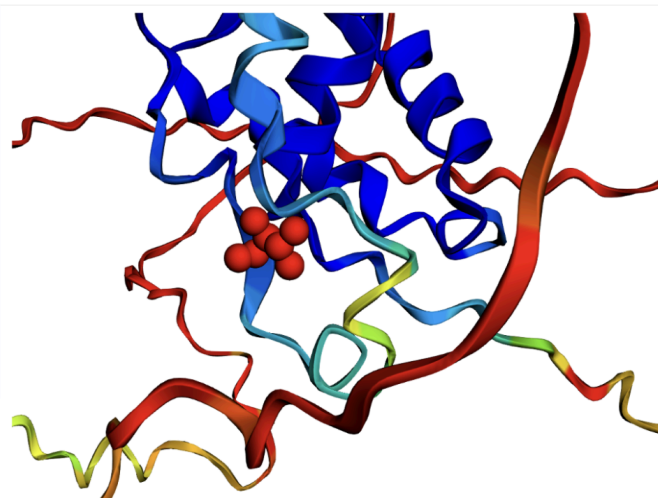
## *Austrofundulus limnaeus*



## *Garcialebias charrua*



b)



pDDT:

- Very low (<50)
- Low (60)
- OK (70)
- Confident (80)
- Very high (>90)

**Figura 27. a)** Secuencia aminoacídica de FoxH1 en cuatro especies de peces anuales. Resaltado en amarillo se encuentra el aminoácido del codon 102 en *N. furzeri* y en *A. limnaeus*. En *G. charrua* es la valina 161 y está marcada con un recuadro rojo. La barra verde indica la región de unión al ADN, la barra anaranjada corresponde al dominio Fork Head 3, que se conoce como dominio de hélice alada. La escala en la parte superior de cada secuencia corresponde a la posición del aminoácido. **b)**

Modelado de la estructura tridimensional de la proteína FoxH1 de *G. charrua* generada en el programa AlphaFold2. En formato de bolas rojas se encuentran en el centro de la figura la valina 161 correspondiente al codón 102. A la derecha de la figura se encuentra una escala que corresponde a la medida de confianza en la predicción de cada parte del modelo (pI-DDT, por sus siglas en inglés predicted Local Distance Difference Test).

## Discusión

En este trabajo se realizó un análisis transcriptómico comparativo entre embriones en DI en dos condiciones ambientales diferentes (campo y laboratorio) con el objetivo de explorar la existencia de una expresión génica diferencial entre ambas condiciones. Asimismo, se analizó la coexpresión de genes candidatos y de las vías relevantes mencionadas que surgen de la bibliografía, las cuales están involucradas en la diapausa II en peces anuales y en otras diapausas caracterizadas.

### **Análisis de los genes con expresión diferencial entre las diapausas I del campo y del laboratorio**

Con respecto a los GED entre los fenotipos de DI (A y B), se observa que el número detectado de genes es muy bajo (Tabla 3). Este resultado confirma que ambos fenotipos son estados de DI en concordancia con la hipótesis sugerida por Arezo y col. (2017) donde se describió al fenotipo B morfológicamente.

En relación a los ambientes, se obtuvo una gran cantidad de GED en los embriones en DI obtenidos en el campo con respecto a los inducidos en condiciones de laboratorio en ambos fenotipos. Considerando que estos embriones se encuentran en el mismo estadio dormante (DI) del desarrollo embrionario, las diferencias de expresión, según las condiciones de cultivo, podrían explicarse por señales ambientales que inciden en los embriones y generan una respuesta celular diferente a factores como la temperatura, niveles de oxígeno u otros factores bióticos y abióticos que varían dependiendo del entorno.

Una posible hipótesis es que el factor principal que podría explicar estas diferencias sea la temperatura. En el diseño experimental de inducción de la DI en el laboratorio se utilizaron tres señales inductoras: hipoxia, presencia de adultos y una temperatura constante de aproximadamente 19°C (Arezo y col. 2017). En condiciones naturales, los embriones también experimentan hipoxia y se encuentran en presencia de adultos, pero los rangos de temperatura experimentados varían entre 6.0 y 28.8° C en el ciclo anual (Errea y Danulat, 2001). Es importante destacar que la temperatura de 19° C, no es informativa para el embrión en relación a la estación del año ya que ésta se registra en dos fases: al comienzo del ciclo (otoño) y al final del mismo (primavera). Considerando que los embriones de peces anuales en diapausa son capaces de tolerar ambientes adversos pero a su vez retienen la capacidad de sentir y responder a claves ambientales (Podrabsky y col. 2016), una temperatura de 19°C, conduciría a los embriones a expresar una mayor diversidad de vías de señalización con respecto a las inducidas en la naturaleza ya que ese factor ambiental (temperatura de 19°C) no es informativo sobre el momento del ciclo anual que está atravesando. Esta respuesta a dicha incertidumbre es consistente con la estrategia “bet-hedging” anteriormente mencionada.

Con el objetivo de encontrar alguna evidencia para esta hipótesis, se realizó una exploración de los genes expresados diferencialmente al alza en las DI obtenidas en el laboratorio y a la baja en las del campo. El análisis de GO (Tabla 8), reveló las vías bioquímicas y funciones biológicas de estos genes. Se encontró que los embriones en DI en condiciones de laboratorio presentan una elevada expresión de genes relacionados con procesos metabólicos, como la fosforilación oxidativa, así como una gran actividad de ARN polimerasas y la activación de vías que favorecen un alto nivel de actividad celular de los embriones en DI (Tabla 8). En el campo, estos mismos genes, presentan una regulación a la baja o ausencia de expresión, reflejando un estado de actividad celular reducida.

**Tabla 8.** Tabla de enriquecimiento de genes que se expresan positivamente en los dos fenotipos de la DI obtenidos en el laboratorio.

Enriquecimiento FDR	nGenes	Genes de la vía	Enriquecimiento en categorías	Funciones biológicas
8.50E-25	35	273	10.8	Citoplasma
1.40E-25	31	175	14.9	Mixta, incluye ribonucleoproteína y biosíntesis de proteínas.
1.10E-10	28	483	4.9	Unión de nucleótidos
7.30E-09	23	395	4.9	Encuadernación de metal
1.90E-16	22	146	12.6	Proteína ribosómica y biosíntesis de proteínas
7.70E-09	21	326	5.4	Membrana celular
9.40E-11	16	123	10.9	Magnesio
4.40E-08	16	200	6.7	Mixto, incluye metabolismo del carbono y fosforilación oxidativa.
4.40E-09	14	117	10	Fosforilación oxidativa y ciclo del citrato (ciclo TCA)
8.30E-09	14	126	9.3	Ciclo mixto, que incluye fosforilación oxidativa y ciclo del citrato (ciclo del TCA)
4.70E-09	13	97	11.2	Proteína ribosómica y biosíntesis de proteínas
1.10E-07	10	65	12.9	NAD
3.10E-10	8	16	41.9	Quinona y ferredoxina similar a la tioredoxina [2Fe-2S]
6.20E-09	8	23	29.2	Translocasa
6.20E-10	7	11	53.4	Mixto, incluye ARN polimerasa y factor sigma-70, región 1.2
6.20E-10	7	11	53.4	Quinona
8.10E-09	6	9	55.9	Síntesis de ATP
1.60E-08	6	10	50.3	Transporte de iones de hidrógeno
1.60E-08	6	10	50.3	Transporte de proteínas
1.60E-08	6	10	50.3	Translocación
3.30E-08	6	11	45.8	Quinona y transportador de membrana conductor de protones
8.30E-09	5	5	83.9	ARN polimerasa

Las condiciones ambientales juegan un papel clave en la regulación molecular de la diapausa. En los embriones en DI obtenidos en el campo se observó una mayor expresión de genes relacionados con la regulación transcripcional, la protección del ADN y la disminución de vías metabólicas (Figura 20 y Tabla A3, anexos), coincidiendo con un estado más profundo de dormancia adaptado a un conjunto de señales específicas y diferentes a las recibidas en el laboratorio (Tabla 8). Por consiguiente, se propone que las señales en la naturaleza configuran molecularmente a los embriones permitiéndoles enfrentar condiciones más extremas del hábitat, mientras que en el laboratorio la ausencia de tales señales conduce a un programa de expresión más activo que presenta mayor diversidad de vías

disponibles, probablemente debido a la falta de certeza sobre el momento del ciclo anual (temperatura de 19 °C).

Otra posible hipótesis es que la composición del ambiente microbiano sea un componente potencialmente relevante. Mushegian y col. (2019) plantearon que las interacciones entre animales y microorganismos durante las diapausas pueden influir en la modulación de la inmunidad, el metabolismo y la reactivación post-diapausa, sugiriendo que los microorganismos ambientales actúan como señales externas capaces de afectar el estado fisiológico del embrión. En *A. limnaeus*, se planteó la hipótesis de que los microorganismos simbióticos contribuirían a sustentar a los embriones durante la DII al proporcionarles rutas metabólicas que favorecen la supervivencia y que no están codificadas en los genomas de los vertebrados (Camara-Lavadores y col. 2022). Así, la exposición diferencial de los embriones a comunidades microbianas naturales frente a las detectadas en el laboratorio (Figura 9) podría contribuir a las diferencias transcriptómicas observadas, afectando la percepción de las señales ambientales.

Este enfoque integrado, que combina la transcriptómica con un contexto ecológico-ambiental, refuerza la necesidad de interpretar la diapausa no sólo como un fenómeno fisiológico en sí mismo, sino también como una estrategia de supervivencia moldeada por la interacción entre los componentes físicos, químicos y biológicos del ambiente, incluida su microbiota asociada.

En suma, los resultados de este estudio mostraron diferencias claras en la cantidad y en los perfiles de los genes expresados diferencialmente entre los embriones en DI obtenidos en el laboratorio y los colectados en el ambiente natural. Este hallazgo evidencia que, a diferencia de lo planteado en la hipótesis inicial y a pesar de la similitud morfológica en las DI de ambos ambientes, a nivel transcriptómico existen diferencias significativas. Esto refuerza la importancia de no extrapolar los resultados obtenidos en condiciones de laboratorio a lo que sucede en el ambiente natural.

## **Estudios de ontología génica**

El análisis del grupo azul (Figura 20a), muestra genes involucrados en la regulación de la expresión génica, factores de transcripción y otras proteínas de unión al ADN. Esto sugiere que, aunque la actividad celular general está disminuida, se mantiene una regulación específica de ciertos genes clave. La presencia de factores relacionados con la ARN polimerasa II implicaría que los embriones presentan componentes listos para activarse rápidamente en la salida de la DI. Esto puede deberse a que las condiciones en el campo activan este perfil de expresión en los embriones de *G. charrua*. Contrariamente, cuando las condiciones externas no son informativas en relación al momento del ciclo anual (temperatura de 19 °C en el laboratorio) estos reguladores transcripcionales no se expresan en el mismo nivel que en el campo (Tabla 5). Asimismo, se observan factores de transcripción que pueden estar implicados en un silenciamiento transcripcional selectivo, sustentando la idea de que las diferencias entre las condiciones experimentadas por los embriones en el campo y en el laboratorio son clave para generar esta diferencia de expresión. En esta categoría, se detectaron genes de gran interés identificados en otras diapausas como los genes de la familia Homeobox y los Forkhead (Podrabsky y col. 2015, Chalar y col. 2021, Sweet y Hu 2025). Además, se encontró un grupo de genes que están involucrados en la protección del ADN frente a daños durante el estrés ambiental, de mayor

relevancia en el ambiente natural donde se exponen a la presencia de radiación ultravioleta y otras posibles toxicidades del ambiente. Las proteínas de choque térmico frío, se ha sugerido que forman parte de un grupo selecto de genes involucrados en el mantenimiento de la diapausa (Thompson y col. 2024)

En el grupo magenta se encuentra el retrotransposon LINE1. Su expresión ha sido asociada con procesos de estrés celular, silenciamiento epigenético y cambios en la regulación génica (Cappucci y col. 2018, Li y col. 2025). Posiblemente la activación de elementos transponibles, como LINE1, en condiciones naturales represente un mecanismo de plasticidad adaptativa frente al estrés ambiental (Teefy y col. 2023), una hipótesis que merece ser explorada profundamente. La regulación de LINE1 podría estar mediada por las proteínas inhibidoras de peptidasas detectadas en el análisis GO, sugiriendo una posible interacción funcional entre la maquinaria epigenética y estas proteínas inhibidoras. Esto concuerda con lo propuesto por Singh y col. (2024), sobre la inserción de los elementos transponibles como posible origen de nuevas isoformas génicas que podrían contribuir a la plasticidad transcriptómica adaptativa de varios genes esenciales en la DI.

Además, es importante destacar que en un trabajo reciente (Gajardo-Escobar y col. 2026), se ha demostrado que este tipo de retrotransposones son los principales contribuyentes al aumento del tamaño genómico en *G. charrua*. La expresión de LINE1 podría estar relacionada con la modulación de la expresión génica a través de interacciones como por ejemplo la alteración de regiones cromatínicas en la diapausa, como fue sugerido por Gajardo-Escobar y col. (2026).

## Redes de enriquecimiento de FoxH1

El factor de transcripción FoxH1 fue identificado originalmente como un mediador crítico en la vía de señalización Nodal a través de su interacción con las proteínas Smad2/3 y Smad4 y desempeña un papel central en la diferenciación de los linajes celulares mesodérmicos y endodérmicos durante la gastrulación (Chen y col, 1997; Chiu y col., 2014).

Zhou y col. (2022) sugieren que, en ratones, FoxH1 participa en diversos procesos epigenéticos que pueden activar o reprimir la transcripción dependiendo de los factores co-reguladores reclutados de acuerdo con el contexto fisiológico. En *Xenopus*, FoxH1, actúa como un factor de transcripción materno pionero que se une al genoma embrionario durante el clivaje (32 a 64 células), etapa previa a la activación del genoma cigótico, participando en esta activación (Pluta y col. 2022; Charney y col. 2017). En asociación con los factores de transcripción Otx1 y Vegt expresados en el endodermo, regula las respuestas transcripcionales de los genes del desarrollo y desempeña un papel importante en la coordinación de las modificaciones epigenéticas de las histonas H3K4me1 (Paraiso y col. 2025). Su y col. (2025) demostraron que FoxH1 es un factor esencial para mantener la pluripotencia en las células madre embrionarias porcinas desempeñando un papel crucial en la inhibición de la expresión prematura de genes relacionados con la diferenciación.

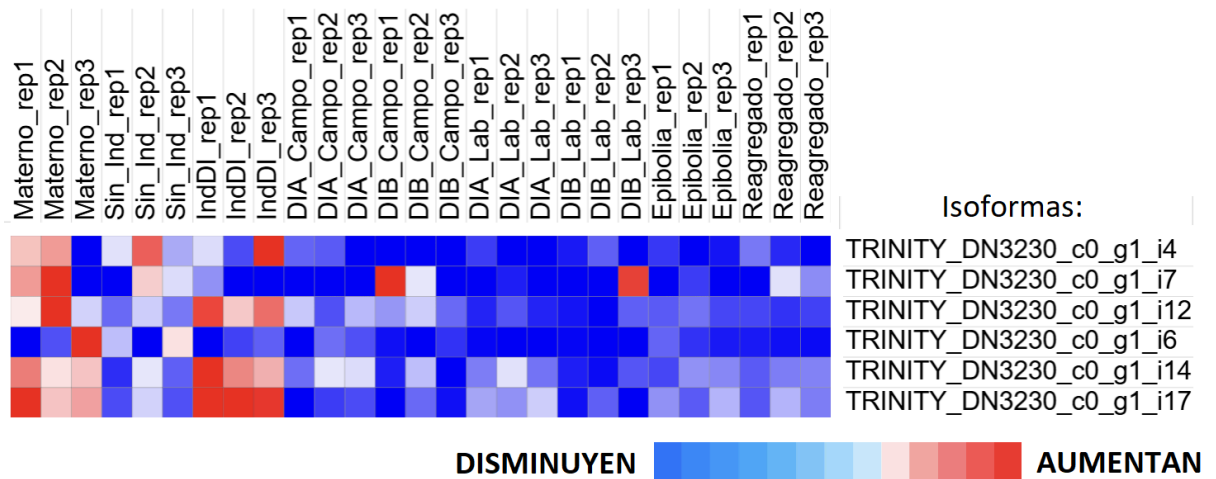
En *D. rerio*, al igual que en *Xenopus*, FoxH1 presenta expresión materna y cigótica. Los transcriptos maternos se localizan en el polo animal durante la ovogénesis. Al inicio de la gastrulación, su expresión se detecta como un gradiente dorsoventral en dominios específicos del mesodermo del escudo embrionario (región del embrión por donde se inicia el ingreso del endomesodermo en la gastrulación de los peces), la notocorda y la placa lateral del mesodermo (Pogoda y col. 2000). Además, Pei y col. (2007) demostraron que en los embriones inyectados con oligonucleótidos antisentido (morfolinos) diseñados para

impedir la traducción de FoxH1, se interrumpen los movimientos de epibolia, internalización y convergencia durante la gastrulación generando un embrión no viable en el primer día de desarrollo.

En *G. charrua* los procesos biológicos identificados incluyen el desarrollo de la placa neural, la regulación de ovocitos y la señalización de la vía Nodal (Tabla 5). En la misma tabla se menciona a las proteínas SMAD, claves por su relación con FoxH1. Estas proteínas están involucradas en la regulación del ciclo celular, la diferenciación celular y la organización del eje embrionario como se ha mencionado (Schier y col. 2003).

En peces anuales, donde la DI ocurre en la fase de dispersión total de blastómeras (100% de epibolia), estos procesos podrían no sólo ralentizarse, sino también reprogramarse transcripcionalmente. La diapausa implica un control activo de la expresión génica, lo que podría permitir al embrión mantener su viabilidad en un estado de adaptación metabólica que conserva las fuentes de energía durante la dormancia. Los mismos genes han sido identificados como reguladores y efectores de la diapausa en diferentes estadios del desarrollo en invertebrados (por ejemplo los genes *daf*, *dauer formation*, en *Caenorhabditis elegans*) y en vertebrados, sugiriendo la presencia de un programa molecular conservado que involucra a las vías como TGF- $\beta$ /Nodal entre otras (revisado por Sweet y Hu, 2025). Estas vías podrían ejercer un papel regulador en la entrada o salida de la diapausa, actuando como interruptores moleculares que permiten detener o reactivar el desarrollo en función de las señales ambientales y/o internas (Podrabsky y col. 2015; Wagner y col. 2018). Concretamente, el análisis de las redes de enriquecimiento (Figura 22), relaciona fuertemente a FoxH1 con estas vías, lo sugiere que este conjunto de genes también estarían directamente involucrados en la modulación de la expresión génica, posiblemente reprimiendo o activando a otros genes según la trayectoria de desarrollo (ingreso a la DI o continuación del desarrollo activo). En conjunto, estos resultados indican que los genes estudiados podrían estar actuando como nodos clave en la red reguladora que controla la DI, ya sea manteniendo el estado dormante o preparándose para reiniciar el desarrollo cuando las condiciones sean favorables, al igual que en otras diapausas estudiadas.

Con respecto al ambiente, un estudio reciente destaca un fuerte acoplamiento de las rutas de respuesta al estrés anóxico con algunos factores de transcripción como SMAD1 en la DII de *A. limnaeus* (Clouser y col. 2025). Este estudio también sugiere que en las diapausas, las vías de estrés metabólico se modulan por factores de transcripción, como por ejemplo PPARA (factor activador del receptor alfa de proliferación de peroxisomas). Este factor presenta múltiples isoformas en el transcriptoma de *G. charrua* expresados en el estadio de IndDI (Figura 28). Cabe destacar que PPARA se expresa al ingreso de la DII en *A. limnaeus* como respuesta celular a la anoxia, sugiriendo la importancia de este factor en las diapausas en los peces anuales.



**Figura 28.** Mapa de calor de las seis isoformas de PPARA obtenidas en el transcriptoma de *G. charrua*. A la derecha, cada una de las isoformas identificadas según el código de Trinity. Arriba, los diferentes estadios del desarrollo embrionario. En la parte inferior de la figura se encuentra la escala de expresión.

En síntesis, los embriones en DI del campo están sometidos a un mayor estrés ambiental con respecto al que experimentan en el laboratorio reflejando diferencias en la expresión génica entre ambas condiciones. En este trabajo, al igual que en otros, se observan similitudes en la expresión de determinados genes y procesos biológicos entre las DI y DII de peces anuales (Wagner y col. 2018, Podrabsky y col. 2015, Sweet y Hu 2025) así como en otros organismos que experimentan diapausa (Karp, 2021; Sweet y Hu, 2025) y en células tumorales dormantes resistentes a la quimioterapia que comparten el mismo propósito de preservar la vida en condiciones adversas (Sweet y Hu, 2025).

La participación de las diferentes vías como la de la IGF/insulina, TGF- $\beta$ , SMAD, PPARA y FOXO, entre otras, en la regulación metabólica y en la detención del ciclo celular durante la diapausa, son esenciales en los procesos moleculares de la dormancia. Esto sugiere la presencia de un programa molecular conservado, compuesto por el mismo conjunto de genes clave, detectados repetidamente en el perfil de expresión de las diapausas descritas (Sweet y Hu 2025), así como en la DI de *G. charrua*.

## Análisis filogenético de FoxH1

El gen FoxH1, fue detectado en las listas de los genes expresados diferencialmente en la DI según el ambiente (campo o laboratorio). El estudio filogenético de FoxH1 podría contribuir en la interpretación de su rol adaptativo en la DI de *G. charrua* y posiblemente a la comprensión de su papel en otros peces anuales.

En concordancia con lo propuesto por Sweet y Hu (2025), la diapausa puede considerarse una plataforma para comprender la dormancia biológica, en un conjunto de rutas moleculares conservadas que regulan la detención del ciclo celular, la regulación metabólica y la protección del genoma. En este contexto, la expresión diferencial de FoxH1 observada en *G. charrua* podría reflejar la activación de componentes de este programa

molecular conservado, reforzando su papel potencial en la regulación del desarrollo y la adaptación a condiciones ambientales variables.

Quienes destacan que múltiples taxones comparten un conjunto conservado de rutas de dormancia que incluyen factores de señalización de TGF- $\beta$ , SMAD y FOXO, indican que la expresión diferencial de FoxH1 como un mediador clave de la vía SMAD (Eustice y col. 2022), en *G. charrua* podría representar la activación de un mecanismo molecular ancestral asociado a la entrada o mantenimiento de la diapausa. Este resultado refuerza la hipótesis de que la regulación de la diapausa comparte elementos con otros tipos de dormancia observados en vertebrados e invertebrados (Martin y Podrabsky, 2017).

El análisis filogenético presentado en la figura 24 evidencia una estrecha relación evolutiva entre FoxH1 de *G. charrua*, otros peces anuales y *K. marmoratus* con un valor de bootstrap del 100%, lo que indica un alto nivel de confianza en esta agrupación. Este hallazgo es coherente con estudios previos que han sugerido una proximidad filogenética entre estos taxones, posiblemente derivada de adaptaciones similares a ambientes temporales y características reproductivas comunes (Gutiérrez y col. 2015). Asimismo, se ha señalado que *K. marmoratus*, aunque no es un pez anual, comparte con estos cierta plasticidad durante el desarrollo embrionario. Por ejemplo, frente a determinadas condiciones ambientales, es capaz de retrasar la eclosión de los embriones (revisado por Martin y Podrabsky, 2017).

El análisis de polimorfismos y divergencia genética en especies de *Cyprinodontiformes* y en particular de los Rivulidae (Tabla 6), revela patrones de diversidad nucleotídica y tasas de sustitución sinónima y no sinónima que ofrecen información sobre las presiones selectivas que actúan sobre el gen FoxH1 en diferentes contextos ecológicos y evolutivos. En general, se observa que la diversidad nucleotídica sinónima ( $P_i(s)$ ) es mayor que la no sinónima ( $P_i(a)$ ) en las especies analizadas. Este patrón sugiere que la mayoría de los cambios en la secuencia estudiada son neutros, indicando una conservación funcional del gen FoxH1 en estas especies. Asimismo, la relación  $K_a/K_s$ , es menor a 1 en todas las categorías analizadas, indicando una fuerte presión selectiva conservadora que mantiene la integridad funcional del gen FoxH1. Al analizar las especies anuales y no anuales, se observa que en las especies anuales, la relación  $P_i(a)/P_i(s)$  es baja, lo que indica que la mayoría de las mutaciones son no sinónimas. Además, la relación  $K_a/K_s$  sugiere una moderada presión selectiva que limita las mutaciones no sinónimas, lo que podría estar relacionado con las adaptaciones específicas de estas especies a ambientes temporales y condiciones de vida extremas. En las especies no anuales, la baja relación  $P_i(a)/P_i(s)$  sugiere que la mayoría de las mutaciones son no sinónimas. La baja relación  $K_a/K_s$  refuerza la idea de que las secuencias están siendo conservadas evolutivamente, lo que podría reflejar una menor presión selectiva en estos linajes. Estos resultados son consistentes con estudios previos que han documentado la alta diversidad genética y la adaptación a ambientes temporales en especies de Rivulidae (García y col. 2004).

El análisis de selección episódica realizado con el programa DataMonkey ha identificado ocho codones que han experimentado presión selectiva diversificadora, con valores de  $p \leq 0.1$ . La selección diversificadora episódica se refiere a eventos de selección positiva que ocurren en momentos específicos de la historia evolutiva, favoreciendo la variabilidad genética en lugar de la conservación de características funcionales (Murrell y col. 2012). La identificación de estos sitios sugiere que las mutaciones en estos codones podrían haber sido favorecidas en determinados linajes o períodos evolutivos, posiblemente en respuesta a cambios ambientales o ecológicos que requirieron adaptaciones específicas. Es importante destacar que el valor de  $p \leq 0.1$  indica una evidencia moderada de selección

diversificadora. Aunque este umbral es más laxo que el convencional  $p \leq 0.05$ , sigue siendo relevante para identificar posibles sitios bajo selección. Sin embargo, es necesario considerar que valores de  $p$  más altos pueden aumentar la probabilidad de falsos positivos, por lo que se recomienda interpretar estos resultados con cautela y, de ser posible, complementarlos con otros métodos o datos adicionales.

### **Mutaciones en los codones que presentan selección purificadora**

La sustitución de una arginina o lisina (residuos básicos y cargados positivamente) por una valina (residuo apolar e hidrofóbico) en el codón 102 de FoxH1 en *G. charrua* (Figura 22), representa un cambio drástico en propiedades fisicoquímicas de la proteína.. Esta mutación podría afectar la estabilidad, la estructura terciaria, o su función (Betts y Russell 2003, Teng y col. 2011). La afinidad de unión al ADN se puede reducir drásticamente si la interacción proteína-ADN depende de cargas positivas que estabilizan la unión. Reemplazar una arginina o lisina por una valina podría alterar esa interacción conduciendo a la inactivación de los factores de transcripción, represores o activadores, según el rol de la proteína. Las consecuencias funcionales podrían incluir la pérdida de su función, un cambio en la especificidad por genes blanco, o también modificar la interacción con las proteínas SMAD involucradas en la alteración de la transcripción de genes específicos (Fisher y col. 2019). En síntesis, la sustitución de residuos básicos por hidrofóbicos en la proteína FoxH1 de *G. charrua* podría ejercer cambios en su capacidad para unirse al ADN y regular la transcripción de genes esenciales para el desarrollo. Esto resalta la importancia de la conservación de la composición de aminoácidos en regiones funcionales importantes de las proteínas (Dai y col. 2021).

El codón 102 se encuentra en la región denominada Fork Head 3, un dominio de unión al ADN conservado, también conocido como hélice alada (Clark y col. 1993). Esa estructura proteica de hélice interactúa con el surco mayor del ADN, permitiendo el acople del sitio de unión específicamente (Littler y col. 2010). Además, estudios recientes muestran que la dinámica, conformación y modo de unión al ADN de los dominios Fork Head pueden variar entre subfamilias de FOX (por ejemplo entre FOXO1, FOXO3, FOXO4), lo que sugiere una flexibilidad estructural y por ende de reconocimiento de secuencias blanco del ADN (Li y col. 2021).

En conjunto, la evidencia obtenida para *G. charrua* indica que la sustitución detectada en el codón 102 de FoxH1 en embriones en DI del campo, localizado dentro del dominio Fork Head 3, una región altamente conservada, responsable del reconocimiento del ADN, constituye un cambio potencialmente importante desde la perspectiva funcional. Este hallazgo es muy interesante ya que, de acuerdo a los resultados obtenidos en este trabajo, este gen se posiciona como un gen muy relevante para la DI y contiene un evento de selección diversificadora episódica en un codón que se encuentra en una región de reconocimiento del ADN. Esto permite sugerir que la variante ejercería un papel esencial en la función de FoxH1 de *G. charrua*.

## Conclusiones

El ensamblado del transcriptoma de *Garcialebias charrua*, generado a partir de la obtención de secuencias de alta calidad, es un aporte esencial para comprender los mecanismos moleculares involucrados en las diapausas de los peces anuales.

El análisis transcriptómico comparativo entre los embriones de *G. charrua* en DI obtenidos en la naturaleza y los inducidos en el laboratorio revela que, aunque ambos correspondan al mismo estadio de dormancia, las condiciones ambientales ejercen una influencia determinante sobre los perfiles de expresión génica. Los embriones del campo muestran un estado de mayor represión metabólica y activación selectiva de vías reguladoras sugiriendo una adaptación específica a las señales ambientales experimentadas, mientras que los embriones del laboratorio mantienen una mayor diversidad de vías activas, probablemente debido a la ausencia de señales ambientales confiables existentes en su entorno natural.

El uso de análisis de correlación y enriquecimiento funcional permitió identificar grupos de genes esenciales, entre los que se destacan los factores de transcripción de las familias Homeobox y Forkhead, así como proteínas relacionadas con la protección frente al estrés ambiental (choque térmico frío) y factores reguladores de las vías de crecimiento y diferenciación celular, como IGF2BP. Estos hallazgos refuerzan la idea de que la diapausa no constituye un estado de inactividad pasiva, sino un proceso regulado dinámicamente por redes moleculares que aseguran las trayectorias de la dormancia y la preparación para la reactivación del desarrollo.

El análisis filogenético y de selección del gen FoxH1 evidenció una fuerte conservación funcional bajo presión selectiva purificadora, aunque con señales de selección diversificadora episódica en sitios específicos. Esto sugiere que, si bien la función básica del gen se preserva, ciertas variaciones podrían haber sido favorecidas en contextos ambientales particulares, contribuyendo a la adaptación evolutiva de los peces anuales a ecosistemas temporales.

En conjunto, estos resultados demuestran que la regulación de la DI en *G. charrua* depende de una interacción compleja entre factores ambientales y redes genéticas altamente conservadas, pero modulables, lo que otorga a esta especie una notable plasticidad adaptativa. Este trabajo no solo aporta evidencia sobre el efecto que ejercen las señales ambientales en los diferentes contextos (campo y laboratorio), sino que también posiciona a genes como FoxH1, IGF2B2, ALX4, csde1 y LINE1 como candidatos centrales en la comprensión de los mecanismos moleculares que sustentan a la DI en *G. charrua* y probablemente en otros peces anuales.

## Perspectivas

### Análisis bioinformáticos

Profundizar en el análisis de las anotaciones del transcriptoma. En este estudio contamos con datos muy significativos en cuanto a la cantidad de transcritos identificados como ARNs no codificantes con los datos que ya están disponibles en la anotación de Trinotate con la base de datos Infernal. Buscar reguladores no codificantes que modulen los genes de DI. Complementar este estudio con programas como miRDeep2 (Friedländer y col. 2012), IntaRNA (Mann y col. 2017) o miRTarBase (Cui y col. 2025). Es necesario realizar consultas en las bases de datos más actualizadas para verificar su clasificación y profundizar sobre la posible relación de estos ARNs con la regulación de la diapausa en condiciones de inducción en el laboratorio y en el campo.

Integrar los datos obtenidos con el programa Trinotate 4.1, que nuclea una gran cantidad de anotaciones consultadas en diversas bases de datos tales como: EggNogg con anotaciones de ontología génica (Hernández-Plaza y col. 2023), y otras bases de datos como Pfam (Finn y col. 2014).

Continuar con los estudios de GO en plataformas con bases de datos como la enciclopedia de Kioto de genes y genomas (KEGG) y la base de datos de transcripción y regulación (TRRUST), para identificar posibles genes reguladores importantes, que posiblemente aportarán gran cantidad de información valiosa para ampliar el conocimiento de este organismo y su respuesta al ambiente.

Reconstruir las redes de co-expresión, usando el programa de WGCNA expandido (mayor cantidad de grupos) o análisis de red de co-expresión génica integrada multiescala (MEGENA) (Song y col. 2015), para inferir redes más completas de co-expresión.

Realizar análisis de secuencias promotoras, buscando elementos de unión de FoxH1, SMAD, Homeobox o PPARA en promotores de genes diferencialmente expresados, por ejemplo con MEME Suite (Timothy y col. 2015) o HOMER (Heinz y col. 2010).

Estudiar la predicción de interacción proteína-proteína en los análisis de enriquecimiento ampliando las redes con STRING, IntAct o BioGRID para visualizar la interconexión de los genes detectados en DI, principalmente en aquellos evidenciados en este trabajo (IGF2BP, Alx4, CS).

Realizar estudios de epigenética y regulación no codificante. Predicción de metilación de ADN o histonas. Explorar si los genes con expresión diferencial presentan motivos CpG o regiones ricas en histonas modificadas (H3K4me3, H3K27me3).

Realizar una integración multiómica (si hay datos disponibles) combinando las aproximaciones de la transcriptómica con la metabolómica y/o la proteómica para validar las vías metabólicas activas (fosforilación oxidativa, inhibición enzimática, etc.). Este análisis puede efectuarse mediante herramientas como MixOmics, MOFA+ o OmicsNet.

Extender análisis filogenéticos y de selección a otros genes, complementar los análisis de selección incorporando metodologías adicionales, como aBSREL, que permitan una detección más robusta de selección a nivel de ramas, así como evaluar el impacto del uso de umbrales de significancia más estrictos y ampliar el análisis a conjuntos más extensos de genes ortólogos, con el fin de capturar patrones evolutivos que podrían no ser evidentes a nivel de secuencias codificantes.

Estudiar la evolución de las isoformas disponibles en el ensamblado y analizar si presentan un perfil de expresión diferencial entre ellas, bajo condiciones de inducción a la diapausa y dentro de la diapausa. Luego ver si son identificables en otras especies de peces anuales en las diapausas del campo y del laboratorio, ya que esto permitirá distinguir si dichos patrones representan mecanismos conservados o innovaciones recientes propias de *G. charrua*.

### **Ensayos experimentales en el laboratorio**

Realizar la validación funcional y de expresión con técnicas como la transcripción reversa y cuantificación en tiempo real (RT-qPCR), de los genes estudiados (FoxH1, IF2B2, csde1, Alx4 y LINE1) en las vías enriquecidas de la DI en el campo para consolidar la evidencia transcriptómica.

Determinar los patrones espacio-temporales de expresión, por ejemplo de FoxH1, en el embrión mediante hibridación *in situ*.

Realizar ensayos de exposición ambiental controlada induciendo la DI con distintos gradientes de temperatura, oxígeno o radiación UV y evaluar los cambios de expresión mediante RNAseq con el objetivo de validar la hipótesis de “plasticidad fenotípica” frente a las señales ambientales.

Realizar experimentos de pérdida (CRISPR/Cas9) y ganancia de función en genes como FoxH1 o IF2B2 para evaluar si afectan la inducción, el mantenimiento y/o la salida de la DI.

## **Agradecimientos**

En primer lugar quiero agradecer a mi familia, especialmente a mi esposa Srtephanie y a mi pequeño hijo Lucio, quienes me dan todo su apoyo y paciencia para poder emplear tanto tiempo y dedicación en estos años al trabajo en la maestría. Un infinito agradecimiento a mis orientadoras, las Dras. Maria Jose Arezo y Veronica Gutierrez que, con una paciencia descomunal, enseñan con dedicación, mucho empeño y esfuerzo para transmitir todo su conocimiento y experiencia a lo largo de estos años. Quisiera además enfatizar que son quienes motivaron mi pasión por el modelo de estudio de los peces anuales con un contagioso amor por la investigación y la profesión de ser investigadores en el área. Asimismo, estoy muy agradecido con mi co-orientador, el Dr. Jose Sotelo, quien me apuntaló de manera fundamental en un área tan compleja como la bioinformática. Me brindó todas las herramientas para emprender esta tesis en este tan desafiante campo de investigación, poniendo a mi disposición su laboratorio y a los integrantes de su equipo para entrenarme en cada uno de los pasos a seguir en el desarrollo de este trabajo. Un agradecimiento por la colaboración y acompañamiento muy de cerca al Dr. Nicolas Papa, que siempre estuvo en contacto conmigo para repasar una y otra vez, las técnicas y protocolos de diferentes estudios o lenguajes de programación sin importar el tiempo que le consumiese, siempre estaba disponible para esas dudas. También estoy muy agradecido con el Dr Santiago Chavez quien nos acompañó, a todo el grupo, con el ensamblaje del transcriptoma, con quien iniciamos la práctica de la gran mayoría de las técnicas y programas de bioinformática y, sin él y su dedicación a estar presente en el uso de estos programas, todo esto no sería posible. Quiero agradecer al Dr. Lucas Inchausti que fue fundamental para el uso de herramientas bioinformáticas al término de mis estudios de este posgrado. Por último, no quiero dejar de mencionar a mi familia en su conjunto, mis padres y mis suegros, quienes acudieron a ayudar a los cuidados de mi bebe Lucio para que pudiera tener reuniones o idas al laboratorio que sin ellos tampoco podría haber realizado este posgrado.

Simplemente, gracias a todos y espero estar a la altura de este esfuerzo en conjunto que originó este lindo trabajo que tengo el honor de presentar y así aportar mi pequeño grano de arena a la ciencia.

## Referencias

### A

Alonso, F., Terán G., Serra Alanís, W., Calviño, P., Montes, M., García, I., Barneche J., Almirón A., Ciotek L., Giorgis P., Casciotta, J. (2023). From the mud to the tree: phylogeny of *Austrolebias* killifishes, new generic structure and description of a new species (Cyprinodontiformes: Rivulidae). *Zoological Journal of the Linnean Society*, 199(1), 280-309. <https://doi.org/10.1093/zoolinlean/zlad032>

Altschul, S., Gish, W., Miller, W., Myers, E. W., Lipman, D. (1990). Basic local alignment search tool. *Journal Of Molecular Biology*, 215(3): 403-410. [https://doi.org/10.1016/S0022-2836\(05\)80360-2](https://doi.org/10.1016/S0022-2836(05)80360-2)

Andersson, R., Gebhard, C., Miguel-Escalada, I., Hoof, I., Bornholdt, J., Boyd, M., Andrews, S. (2010). FastQC: A quality control tool for high throughput sequence data. Version 0.11, 2. <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>

Arezo M. J., Pereiro L, Berois N. 2005. Early development in the annual fish *Cynolebias viarius* (Cyprinodontiformes, Rivulidae). *Journal of Fish Biology*, 66:1357–1370. <https://doi.org/10.1111/j.0022-1112.2005.00688.x>

Arezo, M. J., Papa Rodríguez, N. G., Clivio, G., Montagne, J., De la Piedra, S., Berois Domenech, N. (2017). Annual killifish adaptations to ephemeral environments: diapause I in two *Austrolebias* species. *Developmental Dynamics*, 246(11):848-857 <https://doi.org/10.1002/dvdy.24580>

Ashburner, M., Ball, C., Blake, J, Botstein, D., Butler, H., Cherry, J., Davis, A., Dolinski, K., Dwight, S., Eppig, J., Harris, M., Hill, D., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J. C., Richardson, J., Ringwald, M., Rubin, G., Sherlock, G. (2000). Gene ontology: tool for the unification of biology. *Nature genetics*, 25(1), 25-29. <https://doi.org/10.1038/75556>

Axlid, E., Lewis, P., Carroll, T., & Minicozzi, M. (2023). Functional and behavioral trade-offs between terrestrial and aquatic locomotion in the amphibious fish *Kryptolebias marmoratus*. *Integrative and Comparative Biology*, 63(1), 10–22. <https://doi.org/10.1093/icb/icad003>

---

### B

Bao, B., & Xu, W. H. (2011). Identification of gene expression changes associated with the initiation of diapause in the brain of the cotton bollworm, *Helicoverpa armigera*. *BMC genomics*, 12(1), 224. <https://doi.org/10.1186/1471-2164-12-224>

Berois, N., Arezo, M. J., de Sà, R. (2014) The Neotropical Genus *Austrolebias*: An Emerging Model of Annual Killifishes. *Cell & Developmental Biology* 03, no. 02: 1000136 3: 136.

<https://scholarship.richmond.edu/biology-faculty-publications/115/#:~:text=DOI-,10.4172/2168%2D9296.1000136,-Abstract>

Berois, N., Arezo, M. J., Papa, N., & Clivio, G. (2012). *Annual fish: Developmental adaptation for an extreme environment*. WIREs Developmental Biology, 1, 595–602. <https://doi.org/10.1002/wdev.39>

Berois N., Arezo M. J., Papa N., Schlueb H., Blanco D. & Chalar C. (2024): Ciclo de vida y desarrollo de los peces anuales. Cap. 4, pp. 49-64 en: Arezo MJ, Berois N & García G (eds.): Los peces que vienen de las nubes. Montevideo: DIRAC - Facultad de Ciencias - Udelar, 160 p. ISBN: 978-9974-0-2129-7.

Betts, M. J., Russell, R. (2003). Amino acid properties and consequences of substitutions. En M. R. Barnes & I. C. Gray (Eds.), *Bioinformatics for Geneticists* (pp. 289–316). Wiley. <https://onlinelibrary.wiley.com/doi/book/10.1002/0470867302#page=297:~:text=DOI%3A10.1002/0470867302>

Blažek, R., Polačik, M., Reichard, M. (2013). Rapid growth, early maturation and short generation time in African annual fishes. *EvoDevo*, 4, 24. <https://doi.org/10.1186/2041-9139-4-24>

Bolger, A., Lohse, M., Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, 30(15), 2114–2120. <https://doi.org/10.1093/bioinformatics/btu170>

Bradshaw, W., Holzapfel, C. (2010). Circadian clock genes, ovarian development and diapause. *BMC biology*, 8(1), 115. <https://doi.org/10.1186/1741-7007-8-115>

Bryant, D., Johnson, K., DiTommaso, T., Tickle, T., Couger, M., Payzin-Dogru, D., Lee, T., Leigh, N., Kuo, T., Davis, F., Bateman, J., Bryant, S., Guzikowski, A., Tsai, S., Coyne, S., Ye, W., Freeman, R., Jr., Peshkin, L., Tabin, C., Regev, A., Haas, B., Whited, J. (2017). A tissue-mapped axolotl de novo transcriptome enables identification of limb regeneration factors. *Cell reports*, 18(3), 762-776. <https://doi.org/10.1016/j.celrep.2016.12.063>

Buchfink, B., Reuter, K., Drost, H. (2021). Sensitive protein alignments at tree-of-life scale using DIAMOND. *Nature Methods*, 18 (4), 366–373. <https://doi.org/10.1038/s41592-021-01101-x>

---

## C

Camara-Lavadores, J., Romney, A., Podrabsky, J. (2022). An investigation of microbial symbiosis in vertebrate embryonic diapause using *Austrofundulus limnaeus* as a model. Student Research Symposium, Portland State University. <https://pdxscholar.library.pdx.edu/cgi/viewcontent.cgi?article=1325&context=studentsymposium>

Cappucci, U., Torromino, G., Casale, A. M., Camon, J., Capitano, F., Berloco, M., Mele, A., Pimpinelli, S., Rinaldi, A., Piacentini, L. (2018). Stress-induced strain and brain region-specific activation of LINE-1 transposons in adult mice. *Stress*, 21(6), 575-579. <https://doi.org/10.1080/10253890.2018.1485647>

Cellerino, A., Valenzano, D., Reichwald, K. (2015). From the bush to the bench: the annual fish *Nothobranchius furzeri* as a new model system in biology. *Biological Reviews*, 91(2), 511–533. <https://doi.org/10.1111/brv.12183>

Cha J., Fenelon J., Murphy B., Shaw G., Renfree M., Dey S. (2020) A role for Msx genes in mammalian embryonic diapause. *Bioscience Proceedings*. 10, 44–51. <https://doi.org/10.1530/biosciprocs.10.002>

Chalar, C., Clivio, G., Montagne, J., Costábile, A., Lima, A., Papa, N.G. Berois, N., Arezo, M. J. (2021). Embryonic developmental arrest in the annual killifish *Austrolebias charrua*: A proteomic approach to diapause III. *PLoS One*, 16(6), e0251820. <https://doi.org/10.1371/journal.pone.0251820>

Charney, R., Forouzmand, E., Cho, J., Cheung, J., Paraiso, K., Yasuoka, Y., Takahashi, S., Taira, M., Blitz, I., Xie, X., Cho, K. (2017). Foxh1 occupies cis-regulatory modules prior to dynamic transcription factor interactions controlling the mesendoderm gene program. *Developmental cell*, 40(6), 595-607. <https://doi.org/10.1016/j.devcel.2017.02.017>

Chen, Y., Lun, A., Smyth, G. (2014). Differential expression analysis of complex RNA-seq experiments using edgeR. *Statistical analysis of next generation sequencing data*, 51-74. [https://doi.org/10.1007/978-3-319-07212-8\\_3](https://doi.org/10.1007/978-3-319-07212-8_3)

Clark, K., Halay, E., Lai, E., Burley, S. (1993). Co-crystal structure of the HNF-3/fork head DNA-recognition motif resembles histone H5. *Nature*, 364(6436), 412-420. <https://doi.org/10.1038/364412a0>

Clouser, P., Riggs, C., Romney, A., Podrabsky, J. (2025). Diapause and Anoxia-Induced Quiescence Are Unique States in Embryos of the Annual Killifish, *Austrofundulus limnaeus*. *Biomolecules*, 15 (4), 515. <https://doi.org/10.3390/biom15040515>

Compeau, P., Pevzner, P., Tesler, G. (2011). How to apply de Bruijn graphs to genome assembly. *Nature biotechnology*, 29(11), 987-991. <https://doi.org/10.1038/nbt.2023>

Costa W., Cheffe M. (2001) Three new annual fishes of the genus *Austrolebias* from the Laguna dos Patos system, southern Brazil, and a redescription of *A. adloffii* (Ahl) (Cyprinodontiformes: Rivulidae). *Comun do Mus Ciências e Tecnol da PUCRS. Série Zool* 14:179–200 <https://doi.org/10.5281/zenodo.10138519>

Costa, W., Amorim, P., Mattos, J. (2017). Molecular phylogeny and timing of diversification in South American *Cynolebiini* seasonal killifishes. *Molecular Phylogenetics and Evolution*, 116, 61–68. <https://doi.org/10.1016/j.ympev.2017.07.020>

Costa, W. (2006). The South American annual killifish genus *Austrolebias* (Teleostei: Cyprinodontiformes: Rivulidae): phylogenetic relationships, descriptive morphology and taxonomic revision. *Zootaxa*, 1213, 1–64. <https://doi.org/10.11646/zootaxa.1213.1.1>

Cui, S., Yu, S., Huang, H., Lin, Y., Huang, Y., Zhang, B., Xiao, J., Zuo, H., Wang, J., Li, Z., Li, G., Ma, J., Chen, B., Zhang, H., Fu, J., Wang, L., Huang, H. (2025). miRTarBase 2025: updates to the collection of experimentally validated microRNA–target interactions. *Nucleic acids research*, 53(D1), D147-D156. <https://doi.org/10.1093/nar/gkae1072>

---

## D – F

Dai, S., Qu, L., Li, J., Chen, Y. (2021). Toward a mechanistic understanding of DNA binding by forkhead transcription factors and its perturbation by pathogenic mutations. *Nucleic acids research*, 49(18), 10235-10249. <https://doi.org/10.1093/nar/gkab807>

Denucé, J. (1989). Developmental changes in annual fish embryos kept in the proximity of adult fishes. *Acta embryologiae et morphologiae experimentalis* (1980), 10(1), 87-94.

Dolfi, L., Ripa, R., Antebi, A., Valenzano, D. R., Cellerino, A. (2019). Cell cycle dynamics during diapause entry and exit in an annual killifish revealed by FUCCI technology. *EvoDevo*, 10(1), 29. <https://doi.org/10.1186/s13227-019-0142-5>

Dominguez-Castanedo O, Mosqueda-Cabrera MA, Valdesalici S. (2013). First observations of annualism in *Millerichthys robustus* (Cyprinodontiformes: Rivulidae). *Ichthyological Exploration of Freshwaters*, 24:15–20.

Erwin, K. (2009). Wetlands and global climate change: the role of wetland restoration in a changing world. *Wetlands Ecology and Management*, 17(1), 71–84. <https://doi.org/10.1007/s11273-008-9119-1>

Eustice, M., Konzman, D., Reece, J. M., Ghosh, S., Alston, J., Hansen, T., Golden, A., Bond, M., Abramowitz, L., Hanover, J. (2022). Nutrient sensing pathways regulating adult reproductive diapause in *C. elegans*. *PLoS One*, 17(9), e0274076. <https://doi.org/10.1371/journal.pone.0274076>

Errea A., Danulat E. (2001). Growth of the annual fish, *Cynolebias viarius* (Cyprinodontiformes), in the natural habitat compared to laboratory conditions. *Environmental Biology of Fishes*, 61: 261–268. <https://doi.org/10.1023/A:1011047611009>

Fenelon J., Banerjee A., Murphy B. (2014) Embryonic diapause: development on hold. *International Journal of Developmental Biology*, 58:163–174. <https://doi.org/10.1387/ijdb.140074bm>

Finn, R., Bateman, A., Clements, J., Coggill, P., Eberhardt, R., Eddy, S. R., Heger, A., Hetherington, K., Holm, L., Mistry, J., Sonnhammer, E., Tate, J., Punta, M. (2014). Pfam: the protein families database. *Nucleic acids research*, 42(D1), D222-D230. <https://doi.org/10.1093/nar/gkt1223>

Fischer, P., Chen, H., Pacho, F., Rieder, D., Kimmel, R., Meyer, D. (2019). FoxH1 represses miR-430 during early embryonic development of zebrafish via non-canonical regulation. *BMC biology*, 17(1), 61. <https://doi.org/10.1186/s12915-019-0683-z>

Friedländer, M., Mackowiak, S., Li, N., Chen, W., Rajewsky, N. (2012). *miRDeep2* accurately identifies known and hundreds of novel microRNA genes in seven animal clades. *Nucleic Acids Research*, 40(1), 37–52. <https://doi.org/10.1093/nar/gkr688>

Furness, A., Lee, K., Reznick, D. (2015). Adaptation in a variable environment: Phenotypic plasticity and bet-hedging during egg diapause and hatching in an annual killifish. *Evolution*, 69(6), 1461–1475. <https://doi.org/10.1111/evo.12669>

---

## G

Gajardo-Escobar, F., Valdivieso, C., Di Genova, A., Pereiro, L., Arezo, M. J., Nardocci, G., Allende, M. L. (2026). A recent and rapid genome expansion driven by the amplification of transposable elements in the Neotropical annual killifish *Garcialebias charrua*. *Biological Research*, 59(1), 1. <https://doi.org/10.1186/s40659-025-00649-8>

Gao, X., Cai, T., Lin, Y., Zhu, R., Hao, W., Guo, S., Hu, G. (2022). The function of glucose metabolism in embryonic diapause of annual killifish. *Comparative Biochemistry and Physiology Part D: Genomics and Proteomics*, 42, 100965. <https://doi.org/10.1016/j.cbd.2022.100965>

García, G., Claramunt, S., Lalanne, A. (2004). Genetic differentiation among annual fishes of the genus *Cynolebias* in a biosphere reserve site from Uruguay. *Environmental Biology of Fishes*, 70(3), 247–256. <https://doi.org/10.1023/B:EBFI.0000033339.55585.f6>

García, G., Gutiérrez, V., Ríos, N., Turner, B., Santiñaque, F., López-Carro, B., Folle, G. (2014). Burst speciation processes and genomic expansion in the neotropical annual killifish genus *Austrolebias* (Cyprinodontiformes, Rivulidae). *Genetics*, 142(1), 87-98. <https://doi.org/10.1007/s10709-014-9756-7>

García, G., Ríos, N., Gutiérrez, V. (2015). Next-generation sequencing detects repetitive elements expansion in giant genomes of annual killifish genus *Austrolebias* (Cyprinodontiformes, Rivulidae). *Genetics*, 143(3), 353-360. <https://doi.org/10.1007/s10709-015-9834-5>

García, G., Gutiérrez, V., Ríos, N. (2022). Living in temporary ponds loading giant genomes: The Neotropical annual killifish genus *Austrolebias* as new outstanding evolutionary model. *Frontiers in Genetics*, 13, 903683. <https://doi.org/10.3389/fgene.2022.903683>

García, G., Gutiérrez, V., Ríos, N. (2025). Conservation Strategies in the South American Annual Killifish of the *Austrolebias* sensu lato Linked to the Riparian Wetlands Zones. *Aquatic Conservation: Marine and Freshwater Ecosystems*, 35(4), e70141. <https://doi.org/10.1002/aqc.70141>

Ge, S., Jung, D., Yao, R. (2020). *ShinyGO*: A graphical gene-set enrichment tool for animals and plants. *Bioinformatics*, 36(8), 2628–2629. <https://doi.org/10.1093/bioinformatics/btz931>

Golson, M. L., Kaestner, K. (2016). Fox transcription factors: from development to disease. *Development*, 143(24), 4558–4570. <https://doi.org/10.1242/dev.112672>

Grabherr, M., Haas, B., Yassour, M., Levin, J., Thompson, D. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology*, 29(7), 644–652. <https://doi.org/10.1038/nbt.1883>

Gutiérrez, V., Rego, N., Naya, H., García, G. (2015). First complete mitochondrial genome of the South American annual fish *Austrolebias charrua* (Cyprinodontiformes: Rivulidae): peculiar features among cyprinodontiforms mitogenomes. *BMC genomics*, 16(1), 879. <https://doi.org/10.1186/s12864-015-2090-3>

---

## H – K

Haas, B. J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P. D., Bowden, J., Couger, M. B., Eccles, D., Li, B., Lieber, M., MacManes, M. D., Ott, M., Orvis, J., Pochet, N., Strozzi, F., Weeks, N., Westerman, R., William, T., Dewey, C. N., Henschel, R., LeDuc, R. D., Friedman, N., Regev, A. (2013). *De novo* transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nature protocols*, 8(8), 1494-1512. <https://doi.org/10.1038/nprot.2013.084>

Hand, S., Denlinger, D., Podrabsky, J., Roy, R. (2016). Mechanisms of animal diapause: recent developments from nematodes, crustaceans, insects, and fish. *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology*, 310(11), R1193-R1211. <https://doi.org/10.1152/ajpregu.00250.2015>

Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y., Laslo, P., Cheng, J., Murre, C., Singh, H., Glass, C. (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Molecular cell*, 38(4), 576-589. <https://doi.org/10.1016/j.molcel.2010.05.004>

Hernández-Plaza, A., Szklarczyk, D., Botas, J., Cantalapiedra, C., Giner-Lamia, J., Mende, D., Kirsch, R., Rattei, T., Letunic, I., Jensen, L. J., Bork, P., von Mering, C., Huerta-Cepas, J. (2023). eggNOG 6.0: enabling comparative genomics across 12 535 organisms. *Nucleic Acids Research*, 51(D1), D389-D394. <https://doi.org/10.1093/nar/gkac1022>

Hu, C., Wang, W., Brind'Amour, J., Singh, P., Reeves, G., Lorincz, M., Alvarado, A., Brunet, A., (2020). Vertebrate diapause preserves organisms long term through Polycomb complex members. *Science*, 367(6480), 870-874. <https://doi.org/10.1126/science.aaw2601>

Inglima, K., Perlmutter, A., Markofsky, J. (1981). Reversible stage-specific embryonic inhibition mediated by the presence of adults in the annual fish *Nothobranchius guentheri*. *Journal of Experimental Zoology*, 215(1), 23-33.

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S. A. A., Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., Back, T., Petersen, S., Reiman, D., Clancy, E., Zielinski, M., Hassabis, D. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature*, 596(7873), 583-589. <https://doi.org/10.1038/s41586-021-03819-2>

Kanehisa, M., Goto, S., Sato, Y., Furumichi, M., Tanabe, M. (2012). KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic Acids Research*, 40(D1), D109–D114. <https://doi.org/10.1093/nar/gkr988>

Karp, X. (2021) Hormonal regulation of diapause and development in nematodes, insects, and fishes. *Front Ecol Evol* 9:735924. <https://doi.org/10.3389/fevo.2021.735924>

Keynes, R., Krumlauf, R. (1994). Hox genes and regionalization of the nervous system. *Annual review of neuroscience*, 17(1), 109-132.

Kosakovskiy Pond, S., Poon, A., Velazquez, R., Weaver, S., Hepler, N., Murrell, B., Shank, S. D., Rife Magalis, B., Bouvier, D., Nekrutenko, A., Wisotsky, S., Spielman, S. J., Frost, S. D. W., Muse, S. V. (2020). HyPhy 2.5—a customizable platform for evolutionary hypothesis testing using phylogenies. *Molecular biology and evolution*, 37(1), 295-299. <https://doi.org/10.1093/molbev/msz197>

Krogh, A., Larsson, B., Von Heijne, G., Sonnhammer, E. (2001). Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *Journal of molecular biology*, 305(3), 567-580. <https://doi.org/10.1006/jmbi.2000.4315>

---

## L

Lam, E., Brosens, J., Gomes, A., Koo, C. (2013). Forkhead box proteins: Tuning forks for transcriptional harmony. *Nature Reviews Cancer*, 13 (7), 482–495. <https://doi.org/10.1038/nrc3539>

Langfelder, P., Horvath, S. (2008). WGCNA: An R package for weighted correlation network analysis. *BMC Bioinformatics*, 9 (1), 559. doi:10.1186/1471-2105-9-559 <https://doi.org/10.1186/1471-2105-9-559>

Levels, P., Gubbels, R., Denucé, J. (1986). Oxygen consumption during embryonic development of the annual fish *Nothobranchius korthausae* with special reference to diapause. *Comparative biochemistry and physiology. A, Comparative physiology*, 84(4), 767-770. [https://doi.org/10.1016/0300-9629\(86\)90403-2](https://doi.org/10.1016/0300-9629(86)90403-2)

Levels, P., Denucé, J. (1988). Intrinsic variability in the frequency of embryonic diapauses of the annual fish *Nothobranchius korthausae*, regulated by light: dark cycle and temperature. *Environmental Biology of Fishes*, 22(3), 211-224. <https://doi.org/10.1007/BF00005382>

Li, B., Dewey, C. (2011). RSEM: Accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics*, 12 (1), 323. <https://doi.org/10.1186/1471-2105-12-323>

Li, J., Dai, S., Chen, X., Liang, X., Qu, L., Jiang, L., Guo, M., Zhou, Z., Wei, H., Zhang, H., Chen, Z., Chen, L., Chen, Y. (2021). Mechanism of forkhead transcription factors binding to a novel palindromic DNA site. *Nucleic Acids Research*, 49(6), 3573-3583. <https://doi.org/10.1093/nar/gkab086>

Li, X., Liu, N. (2025). Advances in understanding LINE-1 regulation and function in the human genome. *Trends in Genetics*. <https://doi.org/10.1016/j.tig.2025.04.011>

Littler, D., Alvarez-Fernández, M., Stein, A., Hibbert, R., Heidebrecht, T., Aloy, P., Medema, R. H., Perrakis, A. (2010). Structure of the FoxM1 DNA-recognition domain bound to a promoter sequence. *Nucleic acids research*, 38(13), 4527-4538. <https://doi.org/10.1093/nar/gkq194>

Loureiro M., de Sá R., Serra W., Alonso F., Lanés L., Volcan M., Calviño P., Nielsen D., Duarte A., García G. (2018) Review of the family Rivulidae (*Cyprinodontiformes*, *Aplocheiloidei*) and a molecular and morphological phylogeny of the annual fish genus *Austrolebias* Costa 1998. *Neotrop Ichthyol* 16:e180007. <https://doi.org/10.1590/1982-0224-20180007>

Lubzens E., Cerdà J., Clark M. (2010). In *Dormancy and resistance in harsh environments* (pp. 43-67). Berlin, Heidelberg: Springer Berlin Heidelberg. <https://doi.org/10.1007/978-3-642-12422-8>

---

## M

McGinnis, W., Krumlauf, R. (1992). Homeobox genes and axial patterning. *Cell*, 68(2), 283-302. [https://doi.org/10.1016/0092-8674\(92\)90471-N](https://doi.org/10.1016/0092-8674(92)90471-N)

Mann, M., Wright, P., Backofen, R. (2017). *IntaRNA 2.0*: Enhanced and customizable prediction of RNA–RNA interactions. *Nucleic Acids Research*, 45(W1), W435–W439. <https://doi.org/10.1093/nar/gkx279>

Martin, K., Podrabsky, J. (2017). Hit pause: Developmental arrest in annual killifishes and their close relatives. *Developmental Dynamics*, 246(11), 858–866. <https://doi.org/10.1002/dvdy.24507>

Martin, R., & Polačik, M. (2019). The natural history of model organisms: *Nothobranchius furzeri*, an “instant” fish from an ephemeral habitat. *eLife*, 8, e41548. <https://doi.org/10.7554/eLife.41548>

Martins, R., Lithgow, G., Link, W. (2016). Long live FOXO: Unraveling the role of FOXO proteins in aging and longevity. *Aging cell*, 15(2), 196-207. <https://doi.org/10.1111/acer.12427>

Massagué, J. (2012). *TGFβ* signalling in context. *Nature Reviews Molecular Cell Biology*, 13(10), 616–630. <https://doi.org/10.1038/nrm3434>

Mirdita, M., Schütze, K., Moriwaki, Y., Heo, L., Ovchinnikov, S., Steinegger, M. (2022). ColabFold: making protein folding accessible to all. *Nature methods*, 19(6), 679-682. <https://doi.org/10.1038/s41592-022-01488-1>

Murrell, B., Wertheim, J., Moola, S., Weighill, T., Scheffler, K., Kosakovsky Pond, S. (2012). Detecting individual sites subject to episodic diversifying selection. *PLoS genetics*, 8(7), e1002764. <https://doi.org/10.1371/journal.pgen.1002764>

Mushegian, A., Tougeron, K. (2019). Animal–microbe interactions in the context of diapause. *Biological Bulletin*, 237(2), 180–191. <https://doi.org/10.1086/706078>

Nei, M., Li, W. (1979). Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proceedings of the National Academy of Sciences*, 76(10), 5269-5273. <https://doi.org/10.1073/pnas.76.10.5269>

---

## O – P

Oliveros, J. (2007-2015) Venny. An interactive tool for comparing lists with Venn's diagrams. <https://bioinfogp.cnb.csic.es/tools/venny/index.html>

Ondov, B., Bergman, N., Phillippy, A. (2011). Interactive metagenomic visualization in a web browser. *BMC Bioinformatics*, 12 (1), 385. <https://doi.org/10.1186/1471-2105-12-385>

Papa, N., Clivio, G., Montagne, J. (2016). Husbandry, reproduction, and embryo culture of annual fishes under laboratory conditions. *Annual Fishes: Life history strategy, diversity, and evolution*, 99-107. Eds: Nibia Berois, Graciela García, Rafael de Sá. Editorial: CRC Press Taylor and Francis. ISSN/ISBN: 9781482299717 <https://doi.org/10.1201/b19016-9>

Papa, N., Sotelo-Silveira, J., Inchausti, L., & Arezo, M. J. (2024). *Exploring the molecular basis of diapause I induction in the annual killifish *Garcialebias charrua*: A transcriptomic approach*. *Environmental Biology of Fishes*, 107(6), 693-717. <https://doi.org/10.1007/s10641-024-01562-x>

Papa, N. (2025) Caracterización de la diapausa I en peces anuales del género *Garcialebias*: un abordaje transcriptómico, Tesis de Doctorado. PEDECIBA, UDELAR.

Paraiso, K., Blitz, I., Cho, K. (2025). Maternal and zygotic contributions to H3K4me1 chromatin marking during germ layer formation. *Developmental biology*, 518, 8-19. <https://doi.org/10.1016/j.ydbio.2024.11.006>

Pei, W., Noushmehr, H., Costa, J., Ouspenskaia, M., Elkahlon, A, Feldman, B. (2007). An early requirement for maternal FoxH1 during zebrafish gastrulation. *Developmental biology*, 310(1), 10-22. <https://doi.org/10.1016/j.ydbio.2007.07.011>

Peters N. 1963. Embryonale Anpassungen oviparer Zahnkarpfenaus periodisch austrocknenden Gewässern. *Int Rev Ges Hydrobiol*48:257–313.

Petersen, T., Brunak, S., von Heijne, G., Nielsen, H. (2011). *SignalP 4.0*: Discriminating signal peptides from transmembrane regions. *Nature Methods*, 8 (10), 785–786. <https://doi.org/10.1038/nmeth.1701>

Pereiro L, Loosli F, Fernández J, Härtel S, Wittbrodt J, Concha M. (2017). Gastrulation in an annual killifish: Molecular and cellular events during germ layer formation in *Austrolebias*. *Dev Dyn*. Nov;246(11):812-826. Epub 2017 Mar 24. PMID: 28249357. <https://doi.org/10.1002/dvdy.24496>

Pluta, R., Aragón, E., Prescott, N., Ruiz, L., Mees, R., Baginski, B., Flood, J., Martin-Malpartida, P., Massagué, J., David, Y., Macias, M. (2022). Molecular basis for DNA recognition by the maternal pioneer transcription factor FoxH1. *Nature Communications*, 13(1), 7279. <https://doi.org/10.1038/s41467-022-34925-y>

Podrabsky, J., Hand, S. C. (1999). The bioenergetics of embryonic diapause in an annual killifish, *Austrofundulus limnaeus*. *Journal of Experimental Biology*, 202 (19), 2567–2580. <https://doi.org/10.1242/jeb.202.19.2567>

Podrabsky J., Garrett I., Kohl Z. (2010). Alternative developmental pathways associated with diapause regulated by temperature and maternal influences in embryos of the annual killifish *Austrofundulus limnaeus*. *The Journal of Experimental Biology*. 213: 3280–3288. <https://doi.org/10.1242/jeb.045906>

Podrabsky, J., Hand, S. (2015). Physiological strategies during animal diapause: lessons from brine shrimp and annual killifish. *The Journal of Experimental Biology*, 218(12), 1897-1906. <https://doi.org/10.1242/jeb.116194>

Podrabsky, J., Riggs, C., Wagner, J. (2016). Tolerance of environmental stress. Annual fishes. Life history strategy, diversity, and evolution, 159-184. Eds: Nibia Berois, Graciela García, Rafael de Sá. Editorial: CRC Press Taylor and Francis. ISSN/ISBN: 9781482299717

Podrabsky J., Romney A., Culpepper K. (2016). Alternate developmental pathway. Annual fishes: life history strategy, diversity and evolution. p 63–73 Eds: Nibia Berois, Graciela García, Rafael de Sá. Editorial: CRC Press Taylor and Francis. ISSN/ISBN: 9781482299717

Podrabsky J., Lopez, J., Fan, T., Higashi, R., Somero, G. (2007). Extreme anoxia tolerance in embryos of the annual killifish *Austrofundulus limnaeus*: insights from a metabolomics analysis. *Journal of Experimental Biology*, 210(13), 2253–2266. <https://doi.org/10.1242/jeb.005116>

Podrabsky, J., Arezo, M. J. (2017) Annual Killifishes as Model Systems for Advancing Understanding of Evolution and Developmental Biology. *Developmental Dynamics*. 246: 778. Reseña Editorial. <https://doi.org/10.1002/dvdy.24594>

Pogoda, H., Solnica-Krezel, L., Driever, W., Meyer, D. (2000). The zebrafish forkhead transcription factor FoxH1/Fast1 is a modulator of nodal signaling required for organizer

formation. Current Biology, 10(17), 1041-1049.  
[https://doi.org/10.1016/S0960-9822\(00\)00669-2](https://doi.org/10.1016/S0960-9822(00)00669-2)

Poláček, M., Vrtílek, M., Reichard, M., Žák, J., Podrabsky, J. (2021). Embryo ecology: Developmental synchrony and asynchrony in wild annual fish populations. *Ecology and Evolution*, 11(9), 4945–4956. <https://doi.org/10.1002/ece3.7402>

Poláček, M., García, D., Arezo, M. J., Papa, N., Schlueb, H., Blanco, D., Podrabsky, J., Vrtílek, M. (2023). Embryonic development of natural annual killifish populations of the genus *Austrolebias*: Evolutionary parallelism and the role of environment. *Freshwater Biology*, 68, 1726–1738. <https://doi.org/10.1111/fwb.14161>

Powell, S., Szklarczyk, D., Trachana, K., Roth, A., Kuhn, M., Muller, J., Arnold, R., Rattei, T., Letunic, I., Doerks, T., Jensen, L. J., von Mering, C., Bork, P. (2012). eggNOG v3. 0: orthologous groups covering 1133 organisms at 41 different taxonomic ranges. *Nucleic acids research*, 40(D1), D284-D289. <https://doi.org/10.1093/nar/gkr1060>

Probides. 1999. Plan director. Reserva de Biosfera “Bañados del Este”. p 159. Rocha, Uruguay

Punta, M., Coggill, P. C., Eberhardt, R. Y., Mistry, J., Tate, J., Boursnell, C., Pang, N., Forslund, K., Ceric, G., Clements, J., Heger, A., Holm, L., Sonnhammer, E. L., Eddy, S. R., Bateman, A., Finn, R. (2012). The Pfam protein families database. *Nucleic acids research*, 40(D1), D290-D301. <https://doi.org/10.1093/nar/gkr1065>

---

## R – S

Robinson, M., McCarthy, D., Smyth, G. (2010). edgeR: differential expression analysis of digital gene expression data. *Bioinformatics*, 26(1), 139-140. <https://doi.org/10.1093/bioinformatics/btp616>

Robinson, M. D., Oshlack, A. (2010). A scaling normalization method for differential expression analysis of RNA-seq data. *Genome biology*, 11(3), R25. <https://doi.org/10.1186/gb-2010-11-3-r25>

Romney, A., Davis, E., Corona, M., Wagner, J., Podrabsky, J. (2018). Temperature-dependent vitamin D signaling regulates developmental trajectory associated with diapause in an annual killifish. *Proceedings of the National Academy of Sciences*, 115, 12763–12768. <https://doi.org/10.1073/pnas.1804590115>

Romney, A., Podrabsky, J. (2017). Transcriptomic analysis of maternally provisioned cues for phenotypic plasticity in the annual killifish, *Austrofundulus limnaeus*. *Evodevo* 8:6. [doi: 10.1186/s13227-017-0069-7](https://doi.org/10.1186/s13227-017-0069-7)

Rozas, J., Ferrer-Mata, A., Sánchez-DelBarrio, J., Guirao-Rico, S., Librado, P., Ramos-Onsins, S., Sánchez-Gracia, A. (2017). DnaSP 6: DNA sequence polymorphism analysis of large data sets. *Molecular biology and evolution*, 34(12), 3299-3302. <https://doi.org/10.1093/molbev/msx248>

- Schier, A., Talbot, W. (2005). Molecular genetics of axis formation in zebrafish. *Annual Review of Genetics*, 39(1), 561–613. <https://doi.org/10.1146/annurev.genet.37.110801.143752>
- Schier, A. (2003). Nodal signaling in vertebrate development. *Annual Review of Cell and Developmental Biology*, 19, 589–621. <https://doi.org/10.1146/annurev.cellbio.19.041603.094522>
- Schurch, N., Schofield, P., Gierliński, M., Cole, C., Sherstnev, A., Singh, V., Wrobel, N., Gharbi, K., Simpson, G. G., Owen-Hughes, T., Blaxter, M., Barton, G. J. (2016). How many biological replicates are needed in an RNA-seq experiment and which differential expression tool should you use?. *Rna*, 22(6), 839-851. <https://doi.org/10.1261/rna.053959.115>
- Setiamarga, D., Miya, M., Yamanoue, Y., Mabuchi, K., Satoh T., Inoue, J., Nishida, M. (2008). Interrelationships of Atherinomorpha (medakas, flyingfishes, killifishes, silversides, and their relatives): The first evidence based on whole mitogenome sequences. *Mol Phylogenet Evol.* Nov;49(2):598-605. doi: 10.1016/j.ympev.2008.08.008. Epub 2008 Aug 19. PMID: 18771739. <https://doi.org/10.1016/j.ympev.2008.08.008>
- Singh, P., Reeves, G., Contrepolis, K., Papsdorf, K., Miklas, J., Ellenberger, M., Hu, C., Snyder, M., Brunet, A. (2024). Evolution of diapause in the African turquoise killifish by remodeling the ancient gene regulatory landscape. *Cell*, 187(13), 3338-3356. <https://doi:10.1016/j.cell.2024.04.048>
- Smith-Unna, R., Bournsnel, C., Patro, R., Hibberd, J., Kelly, S. (2016). TransRate: reference-free quality assessment of de novo transcriptome assemblies. *Genome research*, 26(8), 1134-1144. <https://doi.org/10.1101/gr.196469.115>
- Song, W. M., & Zhang, B. (2015). Multiscale embedded gene co-expression network analysis. *PLoS Computational Biology*, 11(11), e1004574. <https://doi.org/10.1371/journal.pcbi.1004574>
- Su, P., Wu, L., Li, D., Song, W., Tao, D., Liu, L., Wang, Q., Gao, M., Xu, T., Liu, X., Xie, S., Zhang, X., Zhou, J., Miao, Y. (2025). CRISPR screening reveals a novel role for FOXH1 in regulating pluripotency of Porcine embryonic stem cells. *Advanced Science*, 12(34), e09495. <https://doi.org/10.1002/advs.202509495>
- Sweet, N., Hu, C. (2025). Using diapause as a platform to understand the biology of dormancy. *Open Biology*, 15 (8), 250104. <https://doi.org/10.1098/rsob.250104>
- Szklarczyk, D., Gable, A., Lyon, D., Junge, A., Wyder, S., Huerta-Cepas, J., Simonovic, M., Doncheva, N., Morris, J., Bork, P., Jensen, L., von Mering, C. (2019). STRING v11: protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic acids research*, 47(D1), D607-D613. <https://doi.org/10.1093/nar/gky1131>
-

Takele, A., Vandesompele, J., Thas, O. (2020). On the utility of RNA sample pooling to optimize cost and statistical power in RNA sequencing experiments. *BMC genomics*, 21(1), 312. <https://doi.org/10.1186/s12864-020-6721-y>

Tamura, K., Stecher, G., Kumar, S. (2021). *MEGA11: Molecular Evolutionary Genetics Analysis* version 11. *Molecular Biology and Evolution*, 38(7), 3022–3027. <https://doi.org/10.1093/molbev/msab120>

Teefy, B., Adler, A., Xu, A., Hsu, K., Singh, P., Benayoun, B. (2023). Dynamic regulation of gonadal transposon control across the lifespan of the naturally short-lived African turquoise killifish. *Genome Research*, 33(1), 141-153. <https://doi.org/10.1101/gr.277301.122>

Thomas, P., Kejariwal, A., Campbell, M., Mi, H., Diemer, K., Guo, N., Ladunga, I., Ulitsky-Lazareva, B., Muruganujan, A., Rabkin, S., Vandergriff, J. A., Doremieux, O. (2003). PANTHER: a browsable database of gene products organized by biological function, using curated protein family and subfamily classification. *Nucleic acids research*, 31(1), 334-341. <https://doi.org/10.1093/nar/gkg115>

Thompson, A., Ortí, G. (2016). Annual killifish transcriptomics and candidate genes for metazoan diapause. *Molecular Biology and Evolution*, 33(9), 2391–2395. <https://doi.org/10.1093/molbev/msw110>

Thompson, A. W., Black, A. C., Huang, Y. Shi, Q., Furness, A. I., Braasch, I., Hoffmann, F. G., Ortí, G. (2024). Transcriptomic data support phylogenetic congruence and reveal genomic changes associated with the repeated evolution of annualism in aplocheiloid killifishes (*Cyprinodontiformes*). *Molecular Phylogenetics and Evolution*, 201, 108209. <https://doi.org/10.1016/j.ympev.2024.108209>

Timothy, L., Johnson, J., Grant, C., Noble, W. (2015). The MEME Suite. *Nucleic Acids Research*, 43(W1), W39–W49. <https://doi.org/10.1093/nar/gkv416>

---

## U – Z

UniProt Consortium. (2025). UniProt: The universal protein knowledgebase. *Nucleic Acids Research*, 53(D1), D609–D617.

Wagner, J., Podrabsky, J. (2015). Gene expression patterns that support novel developmental stress buffering in embryos of the annual killifish *Austrofundulus limnaeus*. *EvoDevo*, 6(1), 2. <https://doi.org/10.1186/2041-9139-6-2>

Wagner, J., Singh, P., Romney, A., Riggs, C., Minx, P., Woll, S., Roush, J., Warren, W., Brunet, A., Podrabsky, J. (2018). The genome of *Austrofundulus limnaeus* offers insights into extreme vertebrate stress tolerance and embryonic development. *BMC genomics*, 19(1), 155. <https://doi.org/10.1186/s12864-018-4539-7>

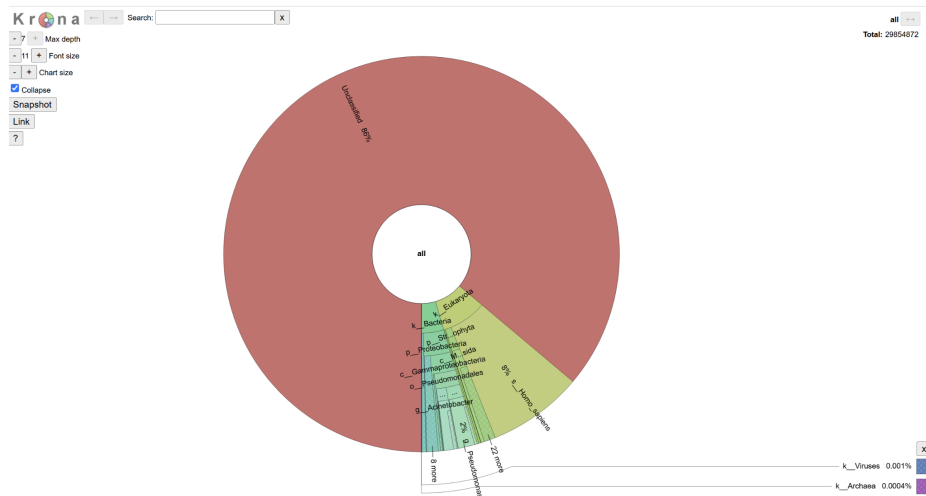
Webb, A., Brunet, A. (2014). *FOXO* transcription factors: Key regulators of cellular quality control. *Trends in Biochemical Sciences*, 39(4), 159–169. <https://doi.org/10.1016/j.tibs.2014.02.003>

- Woll, S., Podrabsky, J. (2017). Insulin-like growth factor signaling regulates developmental trajectory associated with diapause in embryos of the annual killifish *Austrofundulus limnaeus*. *Journal of Experimental Biology*, 220(15), 2777-2786. <https://doi.org/10.1242/jeb.151373>
- Wood, D., Salzberg, S. (2014). *Kraken*: Ultrafast metagenomic sequence classification using exact alignments. *Genome Biology*, 15 (3), R46. <https://doi.org/10.1186/gb-2014-15-3-r46>
- Wourms, J. (1967). Annual fish eggs: diapause and survival in the embryo. *American Zoologist*, 7(3), 363–377.
- Wourms, J.. (1972a). The developmental biology of annual fishes. I. Stages in the normal development of *Austrofundulus myersi*. *Journal of Experimental Zoology*, 182, 143–168. <https://doi.org/10.1002/jez.1401820202>
- Wourms, J. (1972b). The developmental biology of annual fishes. II. Naturally occurring dispersion and reaggregation of blastomeres during the development of *Austrofundulus myersi*. *Journal of Experimental Zoology*, 182, 169–200. <https://doi.org/10.1002/jez.1401820203>
- Wourms, J. (1972c). The developmental biology of annual fishes. III. Pre-embryonic diapause of variable duration in the eggs of annual fishes. *Journal of Experimental Zoology*, 182, 389–414. <https://doi.org/10.1002/jez.1401820310>
- Zhou, J., Pham, P., Han, H., Wang, W., Cho, K. (2022). Engagement of Foxh1 in chromatin regulation revealed by protein interactome analyses. *Development, growth & differentiation*, 64(6), 297-305. <https://doi.org/10.1111/dgd.12799>

## ANEXOS:

### Kraken

Krona : gráfico de anotaciones por el programa Kraken:



**Fig A1.** En el programa Kraken el 86% de los genes no fueron encontrados en la base de datos del programa ya que la base de datos del programa está enfocado en análisis de metagenómica y encontrar bacterias y arqueas en análisis genómicos. Aunque también tiene bases de datos eucariotas en organismos tradicionales (homosapiens, rattus, etc.)

**Tabla A1.** Anotaciones de Kraken. Valores de anotaciones de bacteria y eucariotas en cada estadio por triplicado y las proporciones porcentuales de los mismos.

<u>Muestra</u>		<u># Reads</u>	<u># Reads sin Bact</u>	<u>% seq class</u>	<u>% bacteria</u>	<u>#Rd bact</u>	<u>% no bact</u>	<u>% Bac/Anotado</u>
1_2.fastq.gz	Materno_rep1	31.216.564	27.364.225	18,70	12,34	3.852.339	6,36	66%
<b>2_2.fastq.gz</b>	Materno_rep2	32.202.767	28.161.320	19,55	12,55	4.041.447	7,00	64%
3_2.fastq.gz	Materno_rep3	25.985.177	23.729.664	15,40	8,68	2.255.513	6,72	56%
4_2.fastq.gz	Sin_Ind_rep1	27.368.160	18.760.874	37,14	31,45	8.607.286	5,69	85%
5_2.fastq.gz	Sin_Ind_rep2	27.707.234	21.700.306	27,25	21,68	6.006.928	5,57	80%
<b>6_2.fastq.gz</b>	Sin_Ind_rep3	28.803.804	21.723.829	30,94	24,58	7.079.975	6,36	79%
7_2.fastq.gz	IndDI_rep1	25.316.548	24.650.723	13,54	2,63	665.825	10,91	19%
8_2.fastq.gz	IndDI_rep2	23.079.692	21.302.556	19,39	7,70	1.777.136	11,69	40%
<b>9_2.fastq.gz</b>	IndDI_rep3	30.028.030	28.673.766	14,36	4,51	1.354.264	9,85	31%
10_2.fastq.gz	Epibolia_rep1	29.160.839	22.214.727	30,64	23,82	6.946.112	6,82	78%
<b>11_2.fastq.gz</b>	Epibolia_rep2	30.263.788	22.882.450	41,13	24,39	7.381.338	16,74	59%
12_2.fastq.gz	Epibolia_rep3	30.662.305	19.522.690	42,12	36,33	11.139.615	5,79	86%
<b>13_2.fastq.gz</b>	DIA_Lab_rep1	34.316.360	28.074.214	29,69	18,19	6.242.146	11,50	61%
14_2.fastq.gz	<b>DIA_Lab_rep2</b>	31.221.966	28.134.114	16,85	9,89	3.087.852	6,96	59%
15_2.fastq.gz	DIA_Lab_rep3	27.142.736	20.275.624	38,21	25,30	6.867.112	12,91	66%
<b>16_2.fastq.gz</b>	<b>DIB_Lab_rep1</b>	28.772.670	28.243.253	63,67	1,84	529.417	61,83	3%
<b>17_2.fastq.gz</b>	DIB_Lab_rep2	28.004.477	21.101.373	31,36	24,65	6.903.104	6,71	79%
18_2.fastq.gz	DIB_Lab_rep3	27.172.897	22.776.322	22,36	16,18	4.396.575	6,18	72%
19_2.fastq.gz	Reagregado_rep1	25.652.693	20.770.986	23,83	19,03	4.881.707	4,80	80%
<b>20_2.fastq.gz</b>	Reagregado_rep2	32.206.691	25.469.051	25,37	20,92	6.737.640	4,45	82%
21_2.fastq.gz	Reagregado_rep3	33.438.804	23.316.878	33,41	30,27	10.121.926	3,14	91%
22_2.fastq.gz	DIA_Campo_rep1	28.887.704	28.546.829	6,29	1,18	340.875	5,11	19%
<b>23_2.fastq.gz</b>	DIA_Campo_rep2	27.270.416	26.869.541	7,89	1,47	400.875	6,42	19%
24_2.fastq.gz	DIA_Campo_rep3	30.988.988	30.700.790	8,24	0,93	288.198	7,31	11%
25_2.fastq.gz	DIB_Campo_rep1	25.594.172	25.271.685	6,96	1,26	322.487	5,70	18%
<b>26_2.fastq.gz</b>	DIB_Campo_rep2	33.553.122	33.046.470	9,19	1,51	506.652	7,68	16%
27_2.fastq.gz	DIB_Campo_rep3	26.759.840	26.564.493	11,81	0,73	195.347	11,08	6%

## Estudio de Ontología Génica y enriquecimiento.

**Tabla A2.** Listado de genes grupos azul y magenta

Grupo azul

número de genes	Categoría GO	Genes
47	Unión de compuestos cíclicos orgánicos	myh9 eif1b gata4 abca1 rad51c aacs cpsf6 hnf4a clcn5 papolg matr3 tanc2 hip1r gfi1b incenp alx4 gnl3l ddx21 acaca pif1 limk1 lrp1 slc6a6 tyk2 rab38 cers5 rab1a angel2 fmo5.2 nphs1 myh11 nfyb foxh1 csde1 blk hic2 pin4 rap2a cchcr1 atf4 ube2b eif3d snai1 nr6a1 sf3b3 gapdh rac1
47	Unión de compuestos heterocíclicos	myh9 eif1b gata4 abca1 rad51c aacs cpsf6 hnf4a clcn5 papolg matr3 tanc2 hip1r gfi1b incenp alx4 gnl3l ddx21 acaca pif1 limk1 lrp1 slc6a6 tyk2 rab38 cers5 rab1a angel2 fmo5.2 nphs1 myh11 nfyb foxh1 csde1 blk hic2 pin4 rap2a cchcr1 atf4 ube2b eif3d snai1 nr6a1 sf3b3 gapdh rac1
33	Unión de iones	myh9 gata4 abca1 rad51c aacs hnf4a pcdh1 clcn5 papolg fh12 matr3 hip1r incenp gnl3l lsr ddx21 acaca pif1 limk1 lrp1 tyk2 rab38 rab1a calm1 csrp1 fmo5.2 myh11 blk rap2a ube2b iscu nr6a1 rac1
29	Unión de proteínas	myh9 itprid2 slc26a7 tanc2 hip1r pkp3 cd2ap ano5 ankrd28 wdr75 limk1 lrp1 slc6a6 cav1 tfg nphs1 birc6

		fhod3 myh11 nfyb blk fmn12 hic2 crk sf3b3 rac1 cpsf7 naca hbegf
24	Unión de moléculas pequeñas	myh9 abca1 rad51c aacs clcn5 papolg hip1r incenp gnl3l ddx21 acaca pif1 limk1 tyk2 rab38 rab1a fmo5.2 myh11 blk rap2a ube2b nr6a1 gapdh rac1
22	Unión de derivados de carbohidratos	myh9 abca1 rad51c aacs clcn5 papolg hip1r incenp gnl3l ddx21 acaca pif1 limk1 tyk2 rab38 rab1a myh11 blk rap2a ube2b nr6a1 rac1
12	Actividad del factor de transcripción que se une al ADN	Gata4 HNF4A GF11B ALX4 LRP1 SLC6A6 NPHS1 NFYB FOXH1 HIC2 ATF4 SNAIL1
11	Actividad de hidrolasa	abca1 rad51c hip1r incenp pif1 rab38 rab1a angel2 rap2a nr6a1 rac1
10	Actividad de la transferasa	dgat1 papolg limk1 tyk2 cers5 peli2 birc6 blk ube2b gcn1
7	Actividad del transportador	abca1 clcn5 ano5 slc6a6 XB5730115 uaca clic5
6	Actividad del transportador transmembrana	abca1 clcn5 ano5 slc6a6 XB5730115 clic5
5	Actividad de la oxidoreductasa	limk1 rdh7 fmo5.2 XB5730115 gapdh
5	Unión de complejos que contienen proteínas	myh9 eif1b fhod3 myh11 fmn12

4	Actividad del transductor molecular	hnf4a il1r2 tshr tyk2
4	Actividad del receptor de señalización	hnf4a il1r2 tshr tyk2
2	Actividad reguladora de la traducción	eif1b eif3d
2	Regulador de la función molecular	birc6 rap2a
2	Actividad de la ligasa	aacs acaca
2	Actividad reguladora de la traducción, unión de ácidos nucleicos	eif1b eif3d
1	Actividad de la proteína G	rap2a
1	Actividad de transferencia de electrones	XB5730115
1	Actividad del reservorio de nutrientes	uaca
1	Actividad de la citocromo-c oxidasa	XB5730115
1	Actividad del receptor nuclear	hnf4a
1	Unión de lípidos	cadera1r

1	Actividad de la isomerasa	pin4
1	Actividad reguladora de enzimas	birch6
1	Actividad del transportador transmembrana xenobiótico	slc6a6
1	Encuadernación de grupos metálicos	escu
1	Actividad adaptadora molecular	Cav1

Grupo magenta:

Número de genes	Categoría GO de alto nivel	Genes
4	unión de proteínas	timp2 timp3 ppib xbp1
4	Unión de enzimas	timp2 timp3 ppib xbp1
3	Unión de iones	timp2 as timp3
3	Unión de cationes	timp2 as timp3
2	actividad reguladora enzimática	timp2 timp3

2	Actividad catalítica, que actúa sobre una proteína	as ppib
2	actividad inhibidora de enzimas	timp2 timp3
2	actividad reguladora de la peptidasa	timp2 timp3
1	unión de la cromatina	xbp1
1	actividad del factor de transcripción de unión al ADN	xbp1
1	actividad hidrolasa	as
1	actividad de isomerasa	ppib
1	unión amida	ppib
1	unión de compuestos cíclicos orgánicos	xbp1
1	unión de compuestos heterocíclicos	xbp1
1	Actividad de factor de transcripción con unión al ADN, específica de la ARN polimerasa II.	xbp1
1	Unión de ácidos nucleicos	xbp1
1	actividad peptidasa	as
1	actividad de la isomerasa cis-trans	ppib
1	Unión de péptidos	ppib
1	Unión de proteínas idéntica	xbp1

1	Unión de aniones	as
1	Actividad de dimerización de proteínas	xbp1
1	Respuesta al daño del ADN	line1

**Tabla A3.** Datos de los análisis GO y enriquecimiento de genes de grupo azul y magenta respectivamente.

Enriquecimiento o FDR	nGenes	Genes de la vía	Enriquecimiento en categorías	Funciones biológicas
1.80E-04	14	1508	4.6	Actividad del factor de transcripción de unión al ADN
5.00E-02	5	306	4.6	Unión de filamentos de actina
1.50E-03	12	1423	4.2	Actividad de factor de transcripción de unión al ADN, específica de la ARN polimerasa II
2.50E-03	11	1373	4.2	Unión de ADN específica de secuencia de la región reguladora de la transcripción de la
1.30E-02	9	1056	3.9	Unión de ADN específica de secuencia de la región cis-reguladora de la ARN polimerasa II
1.50E-02	9	1094	3.7	Unión de ADN específica de secuencia de la región reguladora cis
9.80E-03	11	1480	3.4	Unión de la región cis-reguladora de la transcripción
9.80E-03	11	1480	3.4	Unión de ácidos nucleicos a la región reguladora de la transcripción
5.00E-02	8	495	3	unión de GTP
5.00E-02	8	525	3	Unión de nucleótidos de guanilo
5.00E-02	8	525	3	Unión de ribonucleótidos de guanilo
2.00E-02	11	1529	3	Unión de ADN bicatenario específica de secuencia
3.10E-02	12	1683	2.6	Unión de ADN específica de secuencia
4.60E-02	11	1608	2.6	Unión de ADN de doble cadena
2.00E-02	15	1972	2.4	Actividad del regulador de la transcripción
4.60E-02	19	2678	1.9	unión al ADN
3.70E-02	31	4006	1.6	Unión de ácidos nucleicos

Enriquecimiento FDR	nGenes	Genes de la vía	Enriquecimiento en categorías	Funciones biológicas
1.70E-02	1	5	346.4	Unión de iones cloruro
1.70E-02	1	12	346.4	actividad metalodipeptidasa
2.10E-03	2	23	173.2	actividad inhibidora de la metaloendopeptidasa
2.50E-02	1	50	173.2	actividad carboxipeptidasa
2.50E-02	1	15	173.2	Actividad de dipeptidasa
5.70E-03	2	69	63	unión de proteasa
5.70E-03	2	144	53.3	actividad inhibidora de la endopeptidasa
5.70E-03	2	155	53.3	actividad inhibidora de peptidasas
5.70E-03	2	154	49.5	Actividad reguladora de la endopeptidasa
6.80E-03	2	191	40.8	Actividad reguladora de la peptidasa
1.70E-02	2	308	21.7	Actividad inhibidora de enzimas
3.70E-02	2	835	12.8	Unión de iones de zinc
5.70E-03	4	1605	7.3	Unión enzimática

**Figura A2:** Alineamientos en Mega X, de el resto de los codones que poseen selección diversificadora episódica

### codón 148

✓1. TRINITY DN12722 c0 g1 i7 FOXH1	P P A D ? - - - ? G T S R S P V R S G D P S P
✓2. PREDICTED: Austrofundulus limnaeus forkhead box H1 (foxh1)	P P A S ? - - - ? A T S L S P A Q P A D L P L
✓3. PREDICTED: Nematolebias whitei forkhead box H1 (foxh1) mRNA	P L A D ? - - - ? V T S R S P V R S G D P S P
✓4. PREDICTED: Nothobranchius furzeri forkhead box H1 (foxh1) DE	T P P D ? - - - ? V T S V P P M R S E D P S P

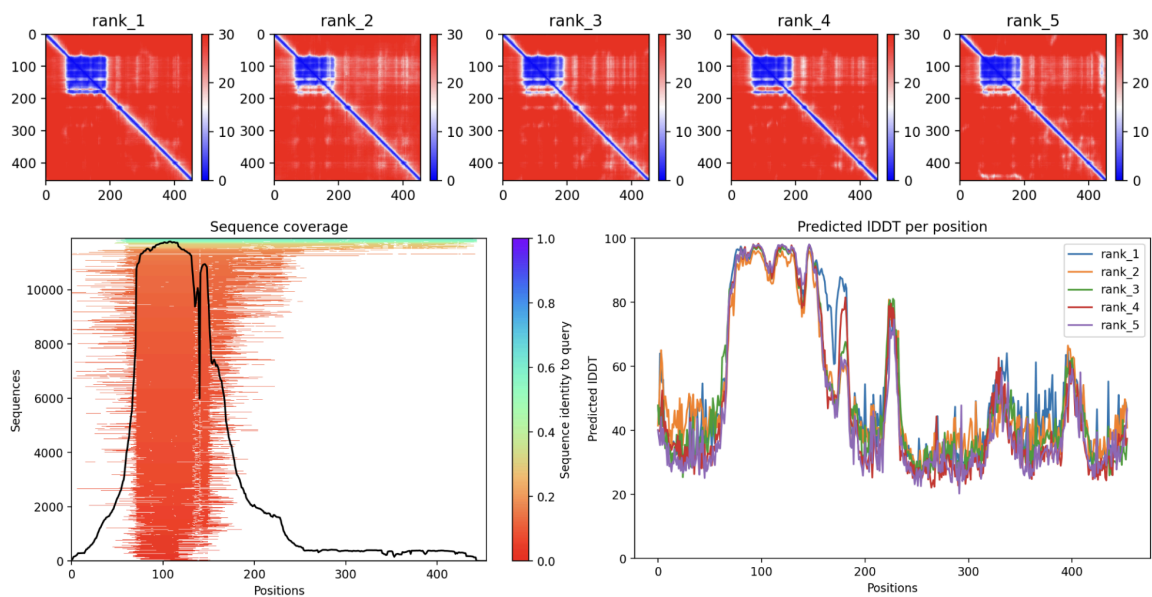
### codón 189

✓1. TRINITY DN12722 c0 g1 i7 FOXH1	H S L R P S P A S G D ? ? A A A G D C W G E A
✓2. PREDICTED: Austrofundulus limnaeus forkhead box H1 (foxh1)	H S L R P A N A S G D G D V A V G D C W G E E
✓3. PREDICTED: Nematolebias whitei forkhead box H1 (foxh1) mRNA	H S L R P S P A S G D G E V A A G D C W G E T
✓4. PREDICTED: Nothobranchius furzeri forkhead box H1 (foxh1) DE	H S L R P S T A S A E V D K ? - - ? C L G E M

### codón 199

✓1. TRINITY DN12722 c0 g1 i7 FOXH1	G D ? ? A A A G D C W G E A E C P R ? A P P R
✓2. PREDICTED: Austrofundulus limnaeus forkhead box H1 (foxh1)	G D G D V A V G D C W G E E E R P R ? T P P R
✓3. PREDICTED: Nematolebias whitei forkhead box H1 (foxh1) mRNA	G D G E V A A G D C W G E T E R P C ? T P P R
✓4. PREDICTED: Nothobranchius furzeri forkhead box H1 (foxh1) DE	A E V D K ? - - ? C L G E M E R P R P S P P P

### Plots for FoxH1\_4f808\_0



**Figura A3:** Resultados de la predicción estructural de la proteína FoxH1-i7, obtenidos mediante AlphaFold2–ColabFold, incluyendo las matrices de contactos, la cobertura del alineamiento múltiple de secuencias (MSA) y los puntajes de confianza por posición (pLDDT). **a)** Mapas de distancia / contacto (rank\_1 a rank\_5). La fila superior presenta los cinco modelos estructurales generados por AlphaFold, ordenados según su puntaje de confianza (“rank”). En cada mapa, la intensidad del color representa la distancia promedio entre residuos: Azul oscuro / celeste, contactos o distancias cortas, indicando regiones plegadas y estructuralmente bien definidas. Rojo, distancias mayores, generalmente regiones flexibles o desordenadas. Las primeras ~150–180 posiciones muestran un patrón consistente de contactos (zona azul intensa), lo que indica un núcleo estructural estable, correspondiente al dominio Forkhead de FoxH1. El resto de la proteína se muestra con menor densidad de contactos, coherente con regiones intrínsecamente desordenadas típicas de factores de

transcripción. **b)** Cobertura del alineamiento de secuencias (MSA). El panel inferior izquierdo muestra la cobertura del MSA utilizado como entrada para AlphaFold: El eje vertical representa el número de secuencias obtenidas por MMseqs2 (más de 10.000 secuencias). Las líneas horizontales muestran la presencia o ausencia de cada residuo en esas secuencias. El color indica la identidad con respecto a la proteína consultada (rojo = baja identidad; verde = alta). La curva negra indica cuántas secuencias contribuyen para cada posición. Se observa una cobertura muy alta entre los residuos ~50–180, coincidiendo con el dominio Foldhead, lo cual refuerza la robustez de la predicción estructural en esa región. **c)** pLDDT por posición (Predicted IDDT). El panel inferior derecho muestra el puntaje pLDDT, una medida de confianza del modelo (0–100): > 90, región muy confiable, estructura rígida, 70–90, buena confianza, 50–70, probable región flexible, < 50, desorden intrínseco o estructura no definida. Todos los modelos (rank\_1 a rank\_5) son consistentes y muestran: Altos valores de pLDDT (80–95) en el dominio Forkhead (~70–160 aa), indicando alta confianza en su estructura. Baja confianza (20–50) en los extremos N-terminal y C-terminal, lo que sugiere regiones desordenadas, algo típico en factores de transcripción con múltiples regiones reguladoras flexibles.