

Trabajo Fin de Máster (Curso 2021/2022)

La **adjudicación** del Trabajo Fin de Máster (TFM) se realizará según dos opciones:

Opción A. Elección de un TFM de la lista propuesta.

La lista de TFM propuestos la podéis encontrar al final de este documento. El alumnado podrá solicitar tres TFMs ofertados en orden de prioridad escribiendo un email a la dirección omicas@us.es antes del **19 de Noviembre de 2021**. Durante la semana siguiente tendrá que concertar una cita para una entrevista con las personas que propongan los tres TFMs solicitados. Basado en esta entrevista el investigador(a)/profesor(a) correspondiente dará una puntuación de 0 a 2. El resto de la puntuación se obtendrá multiplicando por 0.8 el expediente académico del alumno/a aportado durante el proceso de admisión al máster. La asignación final de TFM se hará basada en esta puntuación y se comunicará al alumnado antes del **3 Diciembre de 2021**.

Opción B. Propuesta de un TFM por parte del alumnado.

Si un alumno/a acuerda con un/a investigador/a un TFM no ofertado en la lista anterior no entrará en competición con el resto del alumnado para realizarlo. Será suficiente escribir un email a omicas@us.es con el título y descripción del TFM a realizar además de los datos del supervisor/a correspondiente antes del 19 de Noviembre. Será necesario que o bien el supervisor/a sea un/a profesor/a del máster o el TFM sea cotutorizado por un/a profesor/a del máster.

Las fechas para la **entrega** del TFM son:

- **Primera Convocatoria: 20 de Junio de 2022**
- **Segunda Convocatoria: 5 de Septiembre de 2022**
- **Convocatoria extendida de Septiembre: Fecha a concretar.**

La entrega del TFM se realizará a través de la aplicación **Terminus** de la Secretaría Virtual de la Universidad de Sevilla <https://sevius4.us.es/> (es necesario pinchar en la opción **Estudiantes** que se encuentra en la barra superior de la página).

Las fechas para la **defensa** del TFM son:

- **Primera Convocatoria: 27 de Junio – 1 de Julio de 2022**
- **Segunda Convocatoria: 14 – 16 de Septiembre de 2022.**
- **Convocatoria extendida de Septiembre: Fecha a concretar.**

Título: Identificación de elementos reguladores que actúan en cis y sus correspondientes factores de transcripción en el genoma completo de *Ostreococcus tauri* asociados a la regulación circadiana de su transcriptoma

Tutor/es:

Francisco J. Romero-Campero

Afiliación:

Departamento de Ciencias de la Computación e Inteligencia Artificial

Descripción:

Las interacciones entre factores de transcripción (TF) y elementos reguladores que actúan en cis (CRE) proporcionan información crucial sobre la regulación de la expresión génica. Sin embargo, la determinación experimental de CREs, además de presentar limitaciones específicas, es costosa y requiere mucho tiempo. Esto hace necesario el uso de técnicas computacionales complementarias que asistan en la identificación, predicción y anotación in silico de CREs y sus TFs asociados a partir de datos genómicos, transcriptómicos y proteómicos.

En este proyecto se usará la microalga marina modelo *Ostreococcus tauri* para desarrollar un flujo de trabajo automático que reciba como entrada la secuencia del genoma, la anotación de los TFs, bases de datos de sitios de unión conocidos de TFs, datos de expresión génica y abundancia de proteínas. La salida de este flujo de trabajo automático consistirá en la identificación, anotación y visualización de CREs junto con sus TFs asociados y los perfiles de expresión de los posibles genes dianas.

Este flujo de trabajo se basará en dos algoritmos desarrollados para otros sistemas modelo para la identificación de CRE biológicamente significativos en genes coexpresados. Los CREs se identificarán a partir de las secuencias de ADN ubicadas aguas arriba desde los sitios de inicio de la transcripción, las regiones no traducidas (UTR) y los intrones. La importancia biológica de cada CRE se determinará realizando análisis asociativos entre su ausencia/presencia en cada gen, el perfil de expresión de dicho gen y el perfil de abundancia de proteína del TF asociado.

Se considerará la posibilidad de realizar trabajo en el laboratorio asociado al desarrollo de un transformante de *Ostreococcus tauri* para generar datos de ChIP-seq de un TF en concreto que serviría de validación a la predicción realizada por el flujo de trabajo desarrollado en este proyecto.

Datos y software:

Además del genoma de *Ostreococcus tauri* se proporcionará un conjunto de datos transcriptómicos basados en RNA-seq y proteómicos basados en SWATH-MS consistente en 84 muestras tomadas bajo ciclos diurnos completos en condiciones estacionales diferentes. También se proporcionarán datos de ChIP-seq para varios TFs que se usarán como determinación experimental de CREs. Esto permitirá estimar la sensibilidad, precisión y poder predictivo de la metodología.

El desarrollo de este flujo de trabajo automático estará basado en paquetes de R de integración de múltiples ómicas como mixomics (<http://mixomics.org/>) y software para la identificación de motivos de ADN como HOMER (<http://homer.ucsd.edu/homer/motif/>) y MEME (<https://meme-suite.org/meme/>).

Referencias:

González, I. & Lê Cao, K.-A. & Davis, M.J. and Déjean, S. (2013), Visualising associations between paired 'omics' data sets, *J. BioData Mining* 5:19

Ho CL, Geisler M. (2019) Genome-Wide Computational Identification of Biologically Significant Cis-Regulatory Elements and Associated Transcription Factors from Rice, *Plants (Basel)* 23;8(11):441.

Título: Desarrollo de una Herramienta para la Visualización Interactiva de los Resultados de Análisis de RNA-seq y ChIP-seq para Microalgas.

Tutor/es (contacto):

Francisco J. Romero-Campero (fran@us.es)

Afiliación:

Departamento de Ciencias de la Computación e Inteligencia Artificial

Descripción:

MARACAS (MicroAlgae RnA-seq and Chip-seq Analysis) es un pipeline automático para el análisis de datos de RNA-seq y ChIP-seq generados en estudios ómicos aplicados a microalgas. Este pipeline toma como entrada los datos brutos de secuenciación, RNA-seq o ChIP-seq, en formato fastq y genera como salida informes estáticos en formato html y pdf. Estos informes contienen diferentes tablas con los genes, procesos biológicos y rutas significativamente alterados en los correspondientes estudios. Además, se pueden encontrar gráficos de dispersión, boxplots, dot plots, emaplots y cnetplots. Aunque muy informativa esta visualización es estática y no es posible interactuar con ella para cambiar parámetros que mejorarían la interpretabilidad de los resultados. En este proyecto se propone extender las funcionalidades de MARACAS para incluir informes finales interactivos en formato html desarrollados usando, entre otros, los paquetes de R Glimma (<https://www.bioconductor.org/packages/release/bioc/html/Glimma.html>) y shiny (<https://shiny.rstudio.com/>). En estos informes el usuario tendrá la opción de cambiar parámetros como el fold-change y el nivel de significancia para determinar los conjuntos de genes expresados de forma diferencial, en el caso de análisis de datos de RNA-seq, o para determinar las regiones genómicas ocupadas según análisis de datos de ChIP-seq. También será posible interactuar con los distintos gráficos además de para cambiar aspectos estéticos como color para acceder a información asociada a cada punto, línea o nodo.

Datos y software:

MARACAS se encuentra disponible libremente en este repositorio de Github:

<https://github.com/fran-romero-campero/MARACAS>

Para testear el correcto funcionamiento de las nuevas funcionalidades implementadas en MARACAS se usarán datos de RNA-seq y ChIP-seq libremente disponibles en la base de datos GEO (<https://www.ncbi.nlm.nih.gov/geo/>).

Referencias:

AlgaeFUN with MARACAS (<https://greennetwork.us.es/AlgaeFUN/>).

Glimma (<https://www.bioconductor.org/packages/release/bioc/html/Glimma.html>)

Shiny (<https://shiny.rstudio.com/>)

Título: Construcción de redes de co-expresión génica y análisis de DAP-seq para identificar reguladores de la ruta cannabinoide en *Cannabis sativa*

Tutor/es (contacto):

José Tomás Matus. e-mail: tomas.matus@uv.es. <http://tomsbiolab.com/>

Afiliación:

Grupo de Regulación Transcripcional del Metabolismo. Institute for Integrative Systems Biology (I²SysBio, Universitat de València-CSIC)

Descripción:

Las plantas fabrican metabolitos especializados que ayudan en la interacción y la supervivencia con su entorno, los que además constituyen una fuente rica de beneficios para la salud y nutrición humana y representan los componentes básicos en el desarrollo de nuevos productos farmacéuticos. En nuestro grupo de investigación estamos interesados en la regulación transcripcional del metabolismo de cannabinoides en *Cannabis sativa*, estudiando cómo los factores de transcripción (FTs) regulan la síntesis de estos compuestos. El Trabajo de Fin de Máster ofertado se centra en la construcción de redes de co-expresión (gene co-expression networks, o ‘GCN’) mediante métodos de agregación de redes a partir de datos transcriptómicos públicos. Se prevé el uso de diversos lenguajes de programación (Phyton, R y Bash) para la construcción de las redes y para la implementación de algoritmos de machine learning que permitan evaluar la robustez de estas redes. Para aquellos FTs cuyas redes se vean relacionadas con genes de la ruta cannabinoide se estudiarán mediante la técnica *in vitro* DAP-seq (DNA-Affinity Purification Sequencing) con el fin de determinar sus cistromas y demostrar una regulación directa de la vía.

Datos y software:

Se dispondrá de varios genomas de *C. sativa* productoras de CBDA (tipo II y II) y THCA (tipo I) con anotaciones de los genes de la ruta manualmente curados por nuestro grupo. Se proporcionará un conjunto de datos transcriptómicos basados en experimentos de RNA-seq públicos provenientes de distintos tejidos productores y no de cannabinoides. También se proporcionarán datos de DAP-seq para diversos TFs de la familia MYB, WRKY y AP2. Se utilizarán una serie de softwares y paquetes de R para todas las actividades propuestas:

-EGAD (<https://www.bioconductor.org/packages/release/bioc/html/EGAD.html>)

-GEM (<https://groups.csail.mit.edu/cgs/gem/>)

-ChIPpeakAnno (<https://bioconductor.org/packages/release/bioc/html/ChIPpeakAnno.html>)

Referencias:

O’Malley et al., 2016, Cistrome and Epicistrome Features Shape the Regulatory DNA Landscape Cell 165, 1280–1292 May 19, 2016 a 2016 Elsevier Inc.

<http://dx.doi.org/10.1016/j.cell.2016.04.038>

Liesecke, F., De Craene, JO., Besseau, S. et al. Improved gene co-expression network quality through expression dataset down-sampling and network aggregation. Sci Rep 9, 14431 (2019). <https://doi.org/10.1038/s41598-019-50885-8>

Título: Identificación de la transición floral en la planta modelo *Arabidopsis thaliana* usando técnicas de *deep learning* implementadas en R/Python.

Tutor/es (contacto):

Francisco J. Romero-Campero (fran@us.es)

Afiliación:

Departamento de Ciencias de la Computación e Inteligencia Artificial

Descripción:

La transición floral consiste en el momento del desarrollo vegetal en el cual las plantas transitan de la fase vegetativa, donde se generan hojas, a la fase reproductiva, donde se generan los órganos florales que eventualmente darán lugar a frutos. Es de especial interés determinar el momento exacto de la transición floral ya que permite a los agricultores hortícolas establecer protocolos de cuidado especiales para las plantas en floración y estimar el momento exacto cuando se desarrollará el fruto y la recolección. Sin embargo, la determinación de este momento requiere una inspección humana ocular experta propensa a cometer errores. Esta problemática se podría aliviar con el desarrollo de aplicaciones computacionales basadas en técnicas de inteligencia artificial como el *deep learning* que permitan detectar la transición floral a partir de imágenes con una tasa de error controlada.

En este proyecto se busca aplicar las técnicas de *deep learning* implementadas en el paquete de R *keras* (<https://cran.r-project.org/web/packages/keras/index.html>) para desarrollar una aplicación que permita determinar si plantas de *Arabidopsis thaliana* (planta modelo) se encuentran realizando la transición floral a partir de imágenes.

Datos y software:

Se proporcionará un conjunto de datos formados por más de 100 fotos de plantas de *Arabidopsis thaliana* donde se visualiza el botón floral marcando el momento de la floración y más de 100 fotos correspondientes a plantas donde se observa el meristemo vegetativo antes del momento de la floración.



Figura 1.

Ejemplo formado por dos plantas de *Arabidopsis thaliana* donde se observa el meristemo vegetativo responsable del desarrollo de nuevas hojas (A y B) y dos plantas donde se observa el botón floral.

En este proyecto se aplicarán las técnicas de *deep learning* implementadas en el paquete de R y Python *keras* (<https://cran.r-project.org/web/packages/keras/index.html> y <https://keras.io/>)

Referencias:

Chollet, F., Allaire, J (2018) Deep Learning with R, ISBN 9781617295546

(<https://livebook.manning.com/book/deep-learning-with-r/about-this-book/>)

Scripts de R en Keras (<https://github.com/jjallaire/deep-learning-with-r-notebooks>)

Beysolow, T (2017) Introduction to Deep Learning using R, 978-1-4842-2733-6

(<https://link.springer.com/book/10.1007/978-1-4842-2734-3>)

Título: Segmentación y caracterización de tejidos epiteliales en tres dimensiones aplicando técnicas de Deep Learning y otras herramientas computacionales.

Tutor/es (contacto):

Luis María Escudero Cuadrado (lmesudero-ibis@us.es)
Pedro Gómez Gálvez

Afiliación:

Departamento de Biología Celular

Descripción:

Al comienzo del desarrollo embrionario los tejidos se pliegan y adopta formas tridimensionales (3D) muy complejas que dan lugar a la formación de los órganos adultos. Las células epiteliales son los bloques de construcción a lo largo de este proceso, y cuyas interacciones son clave para controlar la organización celular y la arquitectura del órgano. Recientemente, se ha descubierto que las células epiteliales adoptan una forma geométrica especial (“escutoide” [1]) que facilita que los tejidos se plieguen durante el desarrollo. Por lo tanto, queda demostrado que no es suficiente el estudio bidimensional de la organización epitelial para extraer información relevante de los procesos biológicos durante la morfogénesis. De aquí surge la necesidad de estudiar el empaquetamiento 3D de los epitelios de una forma realista en distintas etapas del desarrollo embrionario. Para ello utilizamos nuevos métodos de análisis de imagen, con segmentación basada en Deep Learning (usando Convolutional Neuronal Network, por ejemplo, U-Net 3D or ResNet) y extracción de características geométricas y topológicas, que caractericen de una manera realista la formación de los tejidos embrionarios. De esta manera, podremos comparar como se empaquetan los tejidos epiteliales en distintas condiciones, por ejemplo, con inhibición/sobreexpresión de alguna molécula que afecte el empaquetamiento o modificando las condiciones de su entorno durante el desarrollo (temperatura, niveles de oxígeno, etc.).

Datos y software:

Se proporcionarán “stacks” de imágenes tomadas en el microscopio confocal, donde se encontrarán teñidas ciertas moléculas presentes en la membrana celular (permitirá visualizar los contornos celulares), y los núcleos celulares. Estas imágenes se procesarán con el uso de Fiji, Matlab y Python. Dependiendo de la necesidad computacional de las tareas, se hará uso de Google Colab donde aprovecharemos el uso de GPU, y de los distintos notebooks de ZeroCostMic4 [2] en los que se llevan a cabo tareas de procesamiento de imagen con el uso de técnicas de Deep Learning.

Referencias:

[1] Gómez-Gálvez, P., Vicente-Munuera, P., Tagua, A. et al. Scutoids are a geometrical solution to three-dimensional packing of epithelia. Nat Commun 9, 2960 (2018). <https://doi.org/10.1038/s41467-018-05376-1>

[2] von Chamier, L., Laine, R.F., Jukkala, J. et al. Democratising deep learning for microscopy with ZeroCostDL4Mic. Nat Commun 12, 2276 (2021). <https://doi.org/10.1038/s41467-021-22518-0>

Título: Clasificación de regiones de cromatina abierta en base a la topografía de la señal de ATAC-seq.

Tutor/es (contacto):

Jose Carlos Reyes (jose.reyes@cabimer.es) y Jose Antonio Guerrero-Martínez (jose.guerrero@cabimer.es)

Afiliación:

Centro Andaluz de Biología Molecular y Medicina Regenerativa (CABIMER).
Consejo Superior de Investigaciones Científicas

Descripción:

La técnica de ATAC-seq (*Assay for Transposase-Accessible Chromatin* seguida de secuenciación) se utiliza para la identificación de regiones de cromatina accesible, o cromatina abierta (1, 2). La mayoría de las regiones reguladoras activas del genoma, como los promotores o los elementos reguladores distales en cis (*enhancers*) son relativamente accesibles y, por lo tanto, pueden ser identificadas mediante esta técnica genómica (3). Sin embargo todas las regiones regulatorias no tienen las mismas características en cuanto a patrón de accesibilidad, extensión, intensidad de señal de ATAC, distribución de picos...etc. La figura 1 muestra varios ejemplos de enhancers con diferentes topografía. En este Trabajo de Fin de Master proponemos investigar y clasificar mediante métodos de clustering las distintas topografías de accesibilidad encontradas en enhancers caracterizados en nuestro laboratorio (4) y correlacionarlo con la actividad del enhancer, el patrón de modificaciones de sus histonas, la unión de factores transcriptionales...etc. Investigaremos las características funcionales y moleculares asociadas a las distintas categorías de enhancers identificadas.

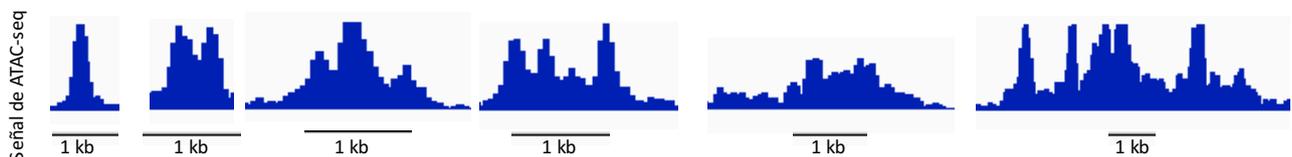


Figura 1

Datos y software:

Utilizaremos datos obtenidos en nuestro laboratorio de ATAC-seq, DNase-seq, ChIP-seq y RNA-seq (4). Utilizaremos software habituales para el preprocesamiento y procesamiento de los datos genómicos indicados en R/Bioconductor: Rsubread, csaw, diffbind, edgeR, ...etc.

También se analizará la adecuación de software utilizados para la caracterización de picos de ChIP-seq según su forma a nuestros datos (FunChIP; <http://bioconductor.org/packages/FunChIP>) y software para análisis digital de footprinting genómico como pyDNase (<https://github.com/jpiper/pyDNase>). Finalmente, también usaremos software para la identificación de motivos de ADN como HOMER (<http://homer.ucsd.edu/homer/motif/>) y MEME (<https://memesuite.org/meme/>).

Referencias:

1. J. D. Buenrostro *et al.* *Nat Methods* **10**, 1213-1218 (2013).
2. F. Yan *et al.* *Genome Biol* **21**, 22 (2020).
3. D. Shlyueva, G. Stampfel, A. Stark. *Nat Rev Genet* **15**, 272-286 (2014).
4. J. A. Guerrero-Martínez *et al.* *Nature communications* **11**, 6196 (2020).

Title:

The role of network structure in validating cophylogenetic congruence as evidence for cospeciation.

Supervisor:

Miguel A. Fortuna (@labfortuna)

Institutional affiliation:

Computational Biology Lab (<https://fortunalab.org>)
Estación Biológica de Doñana (EBD-CSIC)

Description:

In evolutionary biology we try to find out what happened in the past by looking at the footprint that evolutionary history has left in the present. Here, we will test to what extent the evolutionary history of interacting lineages inferred from the current cophylogenetic congruence depends on the structure of the species interaction networks. We will first unravel the macroevolutionary events leading to the architecture currently found in species interaction networks. Second, we will quantify the congruence between the phylogenies of the interacting lineages resulting from that evolutionary history (i.e., to what extent species from one lineage tend to interact with species occupying a similar position on the phylogeny of the other lineage). Finally, we will explore whether a high cophylogenetic congruence can be interpreted as evidence for cospeciation in compartmentalized networks (i.e., parallel speciation in interacting lineages), and as evidence for sorting in nested networks (i.e., a lineage tracks one of its newly emerged partners without speciating itself).

Computational approach:

We will use a method for the generation of cophylogenies based on stochastic growth models for binary trees to trace the evolutionary history of two lineages of interacting species. We will consider the most commonly used events to describe host-parasite evolution: cospeciation, sorting, duplication, and host switch. Then, we will characterize the resulting network structure and quantify the cophylogenetic congruence of the phylogenies associated to the species involved in the network.

References:

- Hommola, K., Smith, J. E., Qiu, Y., and Gilks, W. R. (2009). A permutation test for host-parasite cospeciation. *Mol. Biol. Evol.*
- Keller-Schmidt, S. (2012). *Stochastic tree models of macroevolution*. PhD Thesis. University of Leipzig, Germany.
- Keller-Schmidt, S., Tugrul, M., Eguiluz, V. M., Hernandez-Garcia, E., and Klemm, K. (2010). An age dependent branching model for macroevolution. *arXiv*.
- Legendre, P., Desdevises, Y., and Bazin, E. (2002). A statistical test for host-parasite coevolution. *Syst. Biol.*
- Ronquist, F. (2002). Parsimony analysis of coevolving species associations. In: Page, R. D. M. (ed.), *Tangled trees: phylogeny, cospeciation, and coevolution*. University of Chicago Press.

Title:

The ecological mechanisms driving genetic assimilation.

Supervisor:

Miguel A. Fortuna (@labfortuna)

Institutional affiliation:

Computational Biology Lab (<https://fortunallab.org>)

Estación Biológica de Doñana (EBD-CSIC)

Description:

Phenotypic plasticity is the tendency for organisms with the same genotype to produce different, but repeatable, phenotypes under different environmental conditions (e. g., predator avoidance, insect polyphenism, osmoregulation in fish, and the timing of metamorphosis in amphibians). In spite of the prevalence of phenotypic plasticity, adaptation by natural selection will erode plasticity in evolving populations. One of the mechanisms responsible for the loss of phenotypic plasticity is known as *genetic assimilation*. Genetic assimilation occurs when a trait that was originally triggered by the environment loses this environmental sensitivity and ultimately becomes fixed or expressed in a population. This process might promote the origins of new traits and possibly fuel speciation and adaptive radiation. We will explore computationally the ecological mechanisms leading to genetic assimilation and its role in fostering future adaptation.

Computational approach:

Phenotypic plasticity is also a pervasive feature of digital organisms (self-replicating computer programs that mutate and evolved within a user-defined computational environment). The environment can change the phenotype of a digital organism either directly—by providing different inputs to the same sequence of executed instructions—or indirectly—by modifying the sequence of instructions that is actually executed (i.e. transcriptome). We will use the Avida software platform for the study of evolution to: 1) show experimentally that genetic assimilation takes place in evolving populations; 2) shed light on the ecological mechanisms by which some phenotypes become more sensitive to be genetically assimilated.

References:

- Ehrenreich, I. M. and Pfennig, D. W. (2016). Genetic assimilation: a review of its potential proximate causes and evolutionary consequences. *Annals of Botany*.
- Ofria, C. and Wilke, C. O. (2004). Avida: A software platform for research in computational evolutionary biology. *Artificial Life*.
- Pigliucci, M. (2001). *Phenotypic plasticity: beyond nature and nurture*. Johns Hopkins University Press.
- Pigliucci, M., Murren, C. J., and Schlichting C. D. (2006). Phenotypic plasticity and evolution by genetic assimilation. *J. Exp. Biol.*
- Waddington, C. H. (1953). Genetic assimilation of an acquired character. *Evolution*.

Título:

Análisis proteómico de la señalización del Sulfuro mediante persulfuración en *Arabidopsis*.

Tutor/es (contacto):

Angeles Aroca Aguilar

Afiliación:

Instituto de Bioquímica Vegetal y Fotosíntesis. Universidad de Sevilla

Descripción:

En los sistemas animales, el sulfuro ya está reconocido como una molécula de señalización importante, comparable a CO y NO (Wang, 2012); y se ha demostrado que el mecanismo de acción del sulfuro se explica mediante la persulfuración de proteínas. En plantas, se ha cambiado recientemente el concepto de sulfuro de una molécula tóxica a una molécula de señalización de la misma importancia que NO y H₂O₂ (García-Mata y Lamattina, 2013). La persulfuración es la modificación postraduccional de residuos de cisteína en proteínas mediada por sulfuro, alterando el grupo tiol (-SH) para formar un grupo persulfuro (-SSH) (Mustafa et al., 2009). Las proteínas modificadas por persulfuración muestran cambios funcionales en actividades enzimáticas, estructuras y localizaciones subcelulares (Mustafa et al., 2009, Aroca et al., 2015; Kimura, 2015; Paul y Snyder, 2015; Aroca et al., 2017). Se sabe que la persulfuración juega un papel clave no sólo en la actividad funcional de las proteínas modificadas; también puede regular la localización celular de proteínas con consecuencias significativas en los sistemas de las plantas. Recientemente se ha demostrado la implicación de la persulfuración en la regulación de la autofagia, un proceso conservado en células eucariotas (Laureano et. al, 2002; Aroca et al. 2021). Además, se han identificado un gran número de proteínas implicadas en la fotorrespiración de las plantas como dianas de persulfuración. El rol de esta modificación y la medida en que una proteína determinada se somete a esta modificación oxidativa es actualmente objeto de intensas investigaciones. Además, se cree que la señalización mediante esta PTM representa un mecanismo de protección frente a los sistemas de reducción celular. De ahí la importancia del estudio proteómico de la implicación de esta modificación en la señalización mediada por sulfuro de procesos tan importantes como la autofagia y la fotorrespiración de las plantas.

Datos y software:

Se disponen de datos proteómicos cuantitativos e identificativos de varias control y muestras con mitofagia inducida, así como un proteoma de persulfuración de *Arabidopsis* en condiciones de alto CO₂ y aire normal.

Referencias:

- Wang, R. (2012). Physiological Implications of Hydrogen Sulfide: A Whiff Exploration That Blossomed. *Physiol. Rev.* 92, 791.
- Garcia-Mata, C., and Lamattina, L. (2013). Gasotransmitters are emerging as new guard cell signaling molecules and regulators of leaf gas exchange. *Plant Sci* 201-202, 66-73.
- Mustafa, A.K., Gadalla, M.M., Sen, N., Kim, S., Mu, W., Gazi, S.K., Barrow, R.K., Yang, G., Wang, R., and Snyder, S.H. (2009). H₂S Signals Through Protein S-Sulphydration. *Sci. Signal.* 2, ra72-.

- Aroca, A., Benito, J.M., Gotor, C., and Romero, L.C. (2017a). Persulfidation proteome reveals the regulation of protein function by hydrogen sulfide in diverse biological processes in *Arabidopsis*. *Journal of Experimental Botany* 68, 4915-4927.
- Aroca, A., Schneider, M., Scheibe, R., Gotor, C., and Romero, L.C. (2017b). Hydrogen Sulfide Regulates the Cytosolic/Nuclear Partitioning of Glyceraldehyde-3-Phosphate Dehydrogenase by Enhancing its Nuclear Localization. *Plant and Cell Physiology* 58, 983-992.
- Aroca, A., Serna, A., Gotor, C., and Romero, L., C. (2015). S-sulfhydration: a cysteine post-translational modification in plant systems. *Plant Physiology* 168, 334-342
- Kimura, H. (2015). Physiological Roles of Hydrogen Sulfide and Polysulfides. *Handb. Exp. Pharmacol.* 230, 61.
- Paul, B.D., and Snyder, S.H. (2015). H₂S: A Novel Gasotransmitter that Signals by Sulfhydration. *Trends in Biochemical Sciences* 40, 687-700.
- Aroca A, Yruela I, Gotor C, Bassham DC. 2021. Persulfidation of ATG18a regulates autophagy under ER stress in *Arabidopsis*. *Proc Natl Acad Sci U S A* 118, e2023604118.
- Laureano-Marín AM, Aroca A, Perez-Perez ME, Yruela I, Jurado-Flores A, Moreno I, Crespo JL, Romero LC, Gotor C. 2020. Abscisic Acid-Triggered Persulfidation of the Cys Protease ATG4 Mediates Regulation of Autophagy by Sulfide. *Plant Cell* 32, 3902-3920.

Título:

Caracterización de un mutante de autofagia en la microalga modelo *Chlamydomonas reinhardtii*.

Tutor/es (contacto):

María Esther Pérez Pérez (eperez@ibvf.csic.es).
Manuel Jesús Mallén Ponce.

Afiliación:

Instituto de Bioquímica Vegetal y Fotosíntesis (IBVF) (CSIC-Universidad de Sevilla).

Descripción:

La autofagia es un proceso degradativo mediante el cual las células eucariotas eliminan compuestos celulares dañados o innecesarios para mantener la homeostasis celular y afrontar condiciones de estrés [1, 2, 3]. La autofagia se caracteriza por la formación de vesículas de doble membrana denominadas autofagosomas en las que se engloba el material que será finalmente degradado en la vacuola o el lisosoma [1]. La proteína ATG8 desempeña una función esencial tanto en la formación del autofagosoma como en la selección del cargo o material a degradar [4]. Por ello, ATG8 ha sido utilizada en distintos organismos, incluida el alga modelo *Chlamydomonas reinhardtii*, como un marcador celular y molecular de autofagia [2]. Recientemente, hemos descrito que existe una fuerte conexión entre la activación de la autofagia, la funcionalidad del cloroplasto y el metabolismo de lípidos [5]. Con objeto de investigar la función de ATG8 en la respuesta a estrés en el cloroplasto, hemos generado un mutante carente de ATG8 mediante la tecnología CRISPR-Cas9 en *Chlamydomonas*. Nuestros datos indican que el mutante *atg8* es hipersensible a la cerulenina, un inhibidor específico de la síntesis de ácidos grasos que genera daños en el cloroplasto y activa fuertemente la autofagia [5]. Mediante análisis de RNA-Seq, se estudiará la expresión diferencial en células WT y *atg8* de *Chlamydomonas* tratadas con cerulenina a distintos tiempos.

Datos y software:

Datos de RNA-Seq de nuestro análisis comparativo.
PuTTY para cliente SSH y Winscp para transferencia de archivos a servidor.

Referencias:

- [1] Ichimura Y *et al.* (2000) *Nature* 408: 488-492.
- [2] Pérez-Pérez ME, Florencio FJ, Crespo JL (2010) *Plant Physiol.* 152(4): 1874-1888.
- [3] Pérez-Pérez ME, Lemaire SD, Crespo JL (2021) *J. Exp. Bot.* 72(9): 3340-3351.
- [4] Marshall RS, Vierstra RD (2018) *Annu. Rev. Plant Biol.* 69: 173-208.
- [5] Heredia-Martínez LG, Andrés-Garrido A, Martínez-Force E, Pérez-Pérez ME, Crespo JL (2018) *Plant Physiol.* 178(3): 1112-1129.

Título:

Impacto de la homeostasis redox del cloroplasto en el transcriptoma nuclear de plantas

Tutor/es (contacto):

Juan Manuel Pérez Ruiz (jperez4@us.es)

Afiliación:

Departamento de Bioquímica Vegetal y Biología Molecular (Universidad de Sevilla)

Descripción:

La fotosíntesis oxigénica, junto a otros procesos metabólicos esenciales, tiene lugar en cloroplastos. En estos orgánulos, la actividad de numerosas enzimas está regulada mediante control redox, modificaciones postraduccionales basadas en intercambios tiol-disulfuro entre residuos de cisteína de estas proteínas. Resultados previos de nuestro grupo han puesto de manifiesto que la homeostasis redox del cloroplasto es esencial en el crecimiento y desarrollo, temprano y adulto, de las plantas (Cejudo et al., 2021). En este mecanismo de regulación juega un papel central la proteína NADPH- Tiorredoxina Reductasa C (NTRC). Mutantes de la planta modelo *Arabidopsis thaliana* deficientes en NTRC muestran desregulación en el control redox de las enzimas cloroplastídicas, resultando en ineficiencia de la fotosíntesis e inhibición del crecimiento (Ojeda et al., 2017; Pérez-Ruiz et al., 2017). La función de los cloroplastos coordina el crecimiento y desarrollo de las plantas, así como su respuesta a situaciones de estrés. Numerosas evidencias establecen que el estado funcional de estos orgánulos es transmitido al núcleo, resultando en cambios en la expresión de genes nucleares. Este mecanismo de comunicación, conocido como señalización retrograda, está mediado por moléculas señalizadoras producidas en el cloroplasto, incluyendo metabolitos de diferente naturaleza y el estado redox del orgánulo. El objetivo central de este Trabajo Fin de Máster es identificar genes nucleares que responden de manera específica a cambios en la homeostasis redox del cloroplasto, así como elementos reguladores responsables de esta reprogramación transcripcional. Los resultados obtenidos permitirán establecer procesos, intra y extra-cloroplastídicos, controlados a nivel transcriptómico por el balance redox del orgánulo.

Datos y software:

El genoma de la planta modelo *Arabidopsis thaliana* se encuentra disponible, así como numerosos recursos bioinformáticos. Se proporcionará un conjunto de datos transcriptómicos basados en RNA-seq obtenidos de plantas silvestres y mutantes deficientes en NTRC.

Referencias:

Cejudo, F.J., González, M. and Pérez-Ruiz J.M. (2021) Redox regulation of chloroplast metabolism. *Plant Physiology* 186: 9-21

Ojeda, V., Pérez-Ruiz J.M., González, M., Nájera, V.A., Sahrawy, M., Serrato, A.J., Geigenberger, P., and Cejudo, F.J.* (2017) NADPH thioredoxin reductase C and thioredoxins act concertedly in seedling development. *Plant Physiology* 174: 1436-1448

Pérez-Ruiz J.M., Naranjo, B., Ojeda, V., Guinea, M. and Cejudo, F.J. (2017) NTRC-dependent redox balance of 2-Cys peroxiredoxins is needed for optimal function of the photosynthetic apparatus. *Proceedings of the National Academy of Sciences* 114: 12069-12074

Título: Identificación de posibles secuencias reguladoras en promotores controlados por estrés traduccional en la cianobacteria *Anabaena* sp. PCC 7120

Tutor/es (contacto): Ignacio Luque Romero (ignacio.luque@ibvf.csic.es)

Afiliación: Instituto de Bioquímica Vegetal y Fotosíntesis

Descripción:

Las bacterias de vida libre se encuentran frecuentemente sometidas a condiciones de estrés a las que deben aclimatarse para sobrevivir. El estrés traduccional es aquél estado celular en el que el funcionamiento de la maquinaria de traducción se encuentra parcialmente inhibido pero no totalmente inactivo. Este estado es provocado por diversas condiciones medioambientales, y en el laboratorio puede inducirse por exposición a numerosos factores de estrés, entre los que se incluyen los antibióticos que inhiben al ribosoma. Un trabajo reciente de nuestro laboratorio ha puesto en evidencia que la cianobacteria *Anabaena* reacciona a condiciones de estrés traduccional mediante una respuesta global, según la cual la mitad de los genes del genoma, unos 3000 genes, aumentan o disminuyen su expresión al menos dos veces. Estos genes se dividieron en categorías según su patrón de expresión, de modo que se distinguió una categoría de genes de represión rápida, otro de genes de inducción rápida y otro de inducción progresiva. El análisis de la identidad y la posible función de estos genes llevó a proponer que la respuesta de *Anabaena* a condiciones de estrés nutricional es un fenómeno en dos fases. La primera de ellas va asociada a una ralentización de las funciones fundamentales de la célula como la replicación, la división celular, la traducción o la generación de energía por fotosíntesis. Esto provoca la entrada de la célula en un estado de dormancia en el que el metabolismo es mínimo. En una segunda etapa, se inducen genes dirigidos a reconstituir algunas funciones metabólicas necesarias para la aclimatación a las condiciones de estrés. En particular se observa una recuperación de la traducción que probablemente tiene como fin la síntesis de proteínas específicas de estrés.

La respuesta al estrés traduccional es una respuesta global que probablemente tiene un modo de regulación complejo orquestado por numerosos factores de transcripción que pueden actuar simultánea o secuencialmente. Estos factores, que aún no han sido identificados, probablemente se unen a secuencias específicas de los promotores que regulan. La caracterización de dichas secuencias puede ayudar a definir el conjunto de genes que las portan, que son por tanto potenciales dianas de un factor de transcripción común. ello contribuiría a entender mejor la sucesión de acontecimientos y la regulación de la respuesta al estrés traduccional. Además puede proporcionar información valiosa para la identificación mediante medios experimentales de los factores de transcripción que se unen a las secuencias resultantes del análisis.

Para la identificación de las posibles secuencias reguladoras se utilizará la información de experimentos de RNA-Seq previamente generada por nuestro grupo y por otros, y se utilizarán programas para agrupar los genes en redes de co-expresión que definirán grupos de genes que se inducen o reprimen en determinados momentos de la respuesta y que podrían estar regulados por factores de transcripción comunes. Las secuencias promotoras de estos genes se someterán a análisis con programas para el reconocimiento de patrones para identificar posibles secuencias reguladoras en estos promotores.

Datos y software:

Se utilizarán datos de RNA-Seq generados en nuestro laboratorio que se analizarán con programas para la elaboración de redes de co-expresión (se requiere el manejo del paquete R para la ejecución de programas como Bowtie, EdgeR, limma, Cytoscape o Mfuzz) y programas para la detección o enriquecimiento de motivos en las regiones promotoras (MEME, GLAM2, STREME, YAMDA)

Referencias:

1. Ritchie, M.E. et al. (2015) *limma* powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* 43: e47.
2. Kumar, L., and Futschik, M. (2007) *Mfuzz: a software package for soft clustering of microarray data*. *Bioinformatics* 2: 5–7.
3. Bailey, T.L. (2011) *DREME: motif discovery in transcription factor ChIP-seq data*. *Bioinformatics* 27: 1653–1659.
4. Bailey, T.L. (2011) *DREME: motif discovery in transcription factor ChIP-seq data*. *Bioinformatics* 27: 1653–1659.
5. Brenes-Álvarez et al. (2019) *Elements of the heterocyst-specific transcriptome unravelled by co-expression analysis in Nostoc sp. PCC 7120*. *Environ Microbiol* 21:2544-2558.

Título: Análisis funcional e integración multi-ómica del efecto causado por el coronavirus SARS-CoV-2 a nivel olfatorio

Tutor/es (contacto) y afiliación:

Director: Enrique Santamaría Martínez (enrique.santamaria.martinez@navarra.es), Unidad de Neuroproteómica Clínica, NAVARRABIOMED.

Tutora: Ángela Fontán Lozano (afontan@us.es), Departamento de Fisiología, Universidad de Sevilla.

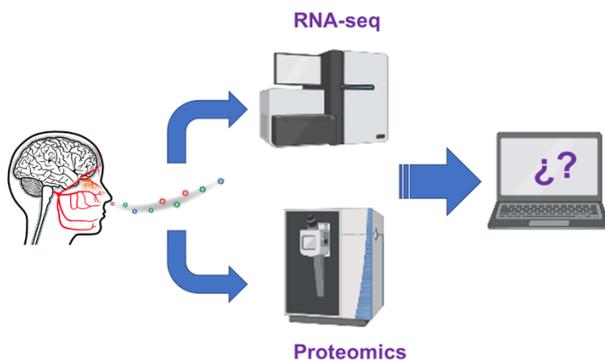
Descripción:

Durante el último año, se han utilizado estrategias -ómicas para elucidar los efectos inducidos por la infección por coronavirus SARS-CoV-2 en diferentes sistemas biológicos. Sin embargo, a pesar de que uno de los síntomas más precoces de infección por SARS-CoV-2 es la pérdida repentina de olfato, no existen estudios que profundicen en el entendimiento de la alteración metabólica global que conlleva una disfunción olfatoria en pacientes COVID-19. Por otra parte, la mayoría de los estudios han utilizado la partícula viral completa, lo que dificulta el entendimiento de cuales son los efectos moleculares inducidos por cada una de las proteínas que componen el virus.

En este proyecto, se persigue caracterizar el efecto de las proteínas más inmunogénicas del SARS-CoV-2 (proteínas estructurales M, N, S y E) en células olfatorias. Para ello, se han realizado transfecciones, experimentos de transcriptómica junto con análisis proteómicos con el objetivo de: i) identificar interactomas olfatorios modulados por cada proteína, ii) analizar similitudes y diferencias respecto a las alteraciones a nivel pulmonar ampliamente estudiadas, iii) contextualizar respecto al interactoma conocido de cada una de las proteínas estructurales del SARS-CoV-2 y iv) integración multi-ómica específica de cada proteína coronaviral.

Datos y software:

Se proporcionarán datos de RNA-seq (generados con plataforma Hiseq 1500 de Illumina) así como datos proteómicos (generados con un espectrómetro de masas 5600 Triple-TOF) obtenidos en células olfatorias que expresan cada una de las proteínas estructurales del SARS-CoV-2. Se utilizarán un abanico de herramientas de caracterización funcional, pathways, generación de interactomas y outputs generados mediante herramientas de Biología de Sistemas.



Referencias:

Gordon DE et al. A SARS-CoV-2 protein interaction map reveals targets for drug repurposing. **Nature**. 2020 Jul;583(7816):459-468.

Bouhaddou M et al. The Global Phosphorylation Landscape of SARS-CoV-2 Infection. **Cell**. 2020 Aug 6;182(3):685-712.e19.

Lachén-Montes M et al. Proteomics Insights Into the Molecular Basis of SARS-CoV-2 Infection: What We Can Learn From the Human Olfactory Axis. **Front Microbiol**. 2020 Sep 22;11:2101.

Stukalov A et al. Multilevel proteomics reveals host perturbations by SARS-CoV-2 and SARS-CoV. **Nature**. 2021 Jun;594(7862):246-252.

Título:

ANÁLISIS PROTEÓMICO CUANTITATIVO DEL EFECTO DEL SULFURO DE HIDRÓGENO EN LA PERSULFURACIÓN DE PROTEÍNAS EN CONDICIONES DE SEQUÍA.

Tutor/es (contacto):

LUIS C. ROMERO GONZALEZ (lromero@ibvf.csic.es)

CECILIA GOTOR MARTÍNEZ (gotor@ibvf.csic.es)

Afiliación:

INSTITUTO DE BIOQUÍMICA VEGETAL Y FOTOSÍNTESIS

A.E. CONSEJO SUPERIOR DE INVESTIGACIONES CIENTÍFICAS Y UNIVERSIDAD DE SEVILLA

Descripción:

El análisis masivo de proteínas basado en espectrometría de masas es esencial para comprender de forma global los cambios de abundancia de proteínas y como la diversidad de modificaciones post-traduccionales (PTM) contribuye a la respuesta de las plantas a cambios en el desarrollo o a estímulos medioambientales. El análisis proteómico ha evolucionado de ser descriptivo y solo generar una lista de proteínas o PTMs a aproximaciones más funcionales y cuantitativas para comprender los mecanismos moleculares que regulan la actividad celular.

El H₂S es una molécula señalizadora que regula numerosos procesos fisiológicos y patológicos tanto en mamíferos como en sistemas vegetales. El mecanismo molecular que subyace a la función señalizadora del H₂S en la modificación de residuos de cisteína (R-SH) para formar un persulfuro (R-SSH). Esta PTM es denominada persulfuración (o S-sulfidratación)

El análisis de proteínas persulfuradas es abordado mediante un marcaje quimioselectivo denominado tag-switch que permite el etiquetado de las proteínas modificadas con biotina para su posterior purificación con estreptavidina inmovilizada. Este método ha sido modificado para mejorar su selectividad y el enriquecimiento de las proteínas. Se desea analizar el proteoma de persulfuración mediante dos métodos de marcaje diferente y comparar la calidad de los resultados de dichos métodos.

El H₂S desempeña una función protectora frente a procesos de estrés abióticos como la sequía, salinidad, altas temperaturas o contaminación por metales pesados, entre otros. Con objeto de identificar la función reguladora de H₂S se analizará el perfil de proteínas persulfuradas en condiciones de sequía con objeto de detectar proteínas dianas de esta modificación. Este análisis se abordará mediante una aproximación cuantitativa libre de marcaje (label free quantification, LFQ) basada en comparar la cantidad de péptidos en diferentes separaciones LC-MS.

Datos y software:

El sistema biológico para estudiar son plantas de Arabidopsis crecidas en condiciones normales de crecimiento comparadas con plantas sometidas a condiciones de sequía. Las proteínas persulfuradas presentes en extractos proteicos de tres réplicas biológicas de cada condición se purificarán mediante dos métodos diferentes de marcaje, se digieren con tripsina y se analizarán mediante LC-MS/MS con una aproximación LFQ.

Los datos que se prevén obtener son valores de intensidad/área de las proteínas identificadas en muestras de plantas crecidas cada condición.

Las proteínas estadísticamente diferentes en cada condición serán estudiadas mediante su clasificación ontológica o su función en rutas metabólicas para detectar la presencia de rutas metabólicas o procesos biológicos sobrerrepresentado en el estudio de cuantificación.

Referencias:

Jurado-Flores, A., Romero, L.C., Gotor, C. (2021). Label-Free Quantitative Proteomic Analysis of Nitrogen Starvation in Arabidopsis Root Reveals New Aspects of H₂S Signaling by Protein Persulfidation. *Antioxidants (Basel)* **10**: 508.

Laureano-Marin, A.M., Aroca, A., Perez-Perez, M.E., Yruela, I., Jurado-Flores, A., Moreno, I., Crespo, J.L., Romero, L.C., Gotor, C. (2020). Abscisic Acid-Triggered Persulfidation of the Cys Protease ATG4 Mediates Regulation of Autophagy by Sulfide. *Plant Cell* **32**: 3902-3920.

Zhang, D., Macinkovic, I., Devarie-Baez, N.O., Pan, J., Park, C.M., Carroll, K.S., Filipovic, M.R., Xian, M. (2014). Detection of protein S-sulfhydration by a tag-switch technique. *Angew Chem Int Ed Engl* **53**: 575-581.

Zivanovic, J., Kouroussis, E., Kohl, J.B., Adhikari, B., Bursac, B., Schott-Roux, S., Petrovic, D., Miljkovic, J.L., Thomas-Lopez, D., Jung, Y., Miler, M., Mitchell, S., Milosevic, V., Gomes, J.E., Benhar, M., Gonzalez-Zorn, B., Ivanovic-Burmazovic, I., Torregrossa, R., Mitchell, J.R., Whiteman, M., Schwarz, G., Snyder, S.H., Paul, B.D., Carroll, K.S., Filipovic, M.R. (2019). Selective Persulfide Detection Reveals Evolutionarily Conserved Antiaging Effects of S-Sulfhydration. *Cell Metab* **30**: 1152-1170.e1113.

Título: Papel de la microbiota como herramienta en la prevención de enfermedades neuropediátricas a través de la suplementación de la dieta materna.

Tutor/es (contacto): Dra Patricia Ybot Gonzalez (pybot-ibis@us.es) y Dra Marta Reyes (mreyesc-ibis@us.es).

Afiliación: Instituto de Biomedicina de Sevilla. Sevilla

Descripción: La neurulación es un proceso morfogénico muy complejo dentro del desarrollo embrionario de los vertebrados que culmina en la formación del tubo neural, precursor del sistema nervioso central (médula espinal y encéfalo). La complejidad molecular y celular del proceso de neurulación está evidenciada por el gran número de mutantes en ratón que desarrollan defectos del tubo neural, defectos congénitos causados por fallo en la formación del tubo neural. En humanos, los “defectos del tubo neural” (DTN) son la segunda causa más frecuente de anomalías congénitas, afectando a uno por cada mil embarazos.

En la actualidad, no hay cura para estas enfermedades neuropediatricas causando una alta morbilidad. El adecuado suministro de nutrientes aportado por la dieta materna es crucial para el correcto desarrollo del embrión. Una estrategia que ha probado su eficiencia en la prevención de enfermedades neuropeditatricas es la suplementación a través de la dieta materna, como es el caso del ácido fólico para prevenir DTN. En nuestro grupo investigamos el papel de diferentes suplementos de la dieta materna dirigida a la prevención de los defectos del tubo neural.

Para apoyar los altos niveles de proliferación celular y el correcto funcionamiento de los procesos del desarrollo embrionario se requiere, vía materna, un aporte equilibrado de nutrientes. Aunque el aporte de nutrientes está determinado por el acceso a comida y la elección de dieta, los niveles de nutrientes circulantes no siempre vienen determinados por la calidad de la comida. La microbiota intestinal influye en la síntesis de vitaminas, por lo tanto, alterando la disponibilidad de los nutrientes. La microbiota materna no sólo tiene la capacidad de cambiar el sistema metabólico de la madre, sino también el del feto.

Mediante modelos de ratón, estudiamos la prevención de enfermedades neuropediátricas a través de la suplementación de la dieta materna, como influyen estas dietas en la composición de la microbiota intestinal y el papel de la microbiota en la biodisponibilidad de los nutrientes y en los sistemas metabólicos de la madre y del feto.

Datos y software:

El proyecto que se oferta para este master es el estudio bioinformático de los resultados tras la secuenciación de las librerías genéticas de las regiones de ADN V3-V4 del gen 16S ARN de la microbiota de las madres gestantes con dieta con y sin suplemento.

Este análisis nos proporcionará un perfil del contenido metagenómico de las muestras fecales donde se analizará la taxonomía bacteriana y la abundancia relativa de cada grupo taxonómico en cada uno de los grupos de ratones. De esta forma será posible comparar la microbiota fecal antes y después del tratamiento expuesto previamente.

Las secuencias obtenidas de cada pase de MiSeq™ serán categorizadas a OTUs (usando QIIME™ v.2). Como técnicas de control se usarán: secuenciación de paralelo de comunidades Mock (mezcla conocida de ADN bacteriano), técnicas de secuenciación negativa y muestras de controles. Muestras con menos de 1000 lecturas se eliminarán. La alpha diversidad (número de OTUs únicas) y la beta diversidad (distancias UniFrac intra-individuales) serán calculadas e implementadas mediante QIIME™. La diversidad de la microbiota vendrá dada por el valor de la mediana de las distancias UniFrac individuales y se resumirá de la siguiente manera: valores altos representarán comunidades de composición distinta de otros individuos y valores bajos comunidades estables.

Referencias:

- 1.- Bernabeu A, et al. Effect of the vaginal microbiome on the pregnancy rate in women receiving assisted reproductive treatment. J Assist Reprod Genet. 2019 Oct;36(10):2111-2119.
- 2.- Wang X, et al Longitudinal investigation of the swine gut microbiome from birth to market reveals stage and growth performance associated bacteria. Microbiome. 2019 Jul 30;7(1):109.

Título: Estudio de la redundancia metabólica y sus implicaciones en el funcionamiento de comunidades microbianas de ambientes extremos

Tutor/es (contacto):

Blanca Vera Gargallo
vera@us.es

Afiliación:

Departamento de Microbiología y Parasitología. Facultad de Farmacia. Universidad de Sevilla.

Descripción:

En las comunidades microbianas naturales, los datos genómicos y metagenómicos sugieren que muchas especies microbianas tienen metabolismos altamente redundantes y aparentemente son capaces de utilizar los mismos sustratos o sintetizar metabolitos energéticamente costosos¹. Se ha propuesto que esta redundancia funcional fomenta la resistencia y la estabilidad ecológicas, ya que una función de un ecosistema con varias especies redundantes (las que realizan funciones similares) debería soportar mejor la pérdida de especies individuales². La coexistencia de organismos metabólicamente redundantes es, sin embargo, paradójica, ya que la teoría indica que las especies que requieren un recurso común deberían competir entre sí y la auxotrofia (la incapacidad de un organismo para sintetizar un compuesto orgánico concreto necesario para su crecimiento) es ventajosa cuando compuestos metabólicamente costosos están disponibles en el medio (por ejemplo, porque otro miembro de la comunidad lo sintetiza). Las diferencias entre la redundancia funcional evaluada a nivel genómico (nichos fundamentales) y transcriptómico (nicho efectivo) pueden contribuir a explicar la coexistencia de especies redundantes a nivel metabólico³ y, por tanto, proporcionar información sobre las interacciones y la estabilidad de la comunidad. Este proyecto pretende avanzar en la comprensión de la coexistencia e interacciones microbianas en ambientes extremos. Para ello, se recuperarán metagenomas de ambientes extremos disponibles en las bases de datos públicas (NCBI, EBI, IMG/M ER, MG-RAST) para posteriormente evaluar su redundancia metabólica. Se explorarán vías metabólicas generales, pero también específicas (como la síntesis de la vitamina B12 y las vías implicadas en auxotrofías en estos hábitats) para determinar su grado de redundancia. Se evaluará el número y la diversidad filogenética de los taxones capaces de realizar una determinada función y se comparará entre los diferentes ambientes extremos estudiados. A continuación, utilizando los ambientes hipersalinos como modelo, evaluaremos hasta qué punto la diferencia en los nichos fundamentales (evaluados a partir de los metagenomas) y efectivos (evaluados mediante metatranscriptómica) contribuye a explicar esta duplicidad metabólica.

Datos y software:

Metagenomas de diversos ambientes extremos disponibles en las bases de datos públicas
Metagenomas y metatranscriptoma de ambientes hipersalinos acuáticos (salinas solares)
Los programas GraftM⁴, Metage2Metabo⁵, HUMAnN2⁶ y VB₁₂Path⁷ se utilizarán para anotar las funciones metabólicas funcional y taxonómicamente e identificar auxotrofías en datos metagenómicos y metatranscriptómicos.

Referencias:

1. Hester ER, Jetten MSM, Welte CU, Lückner S. 2019. Metabolic Overlap in Environmentally Diverse Microbial Communities. *Front Genet.* 10:989. doi: 10.3389/fgene.2019.00989.
- Biggs, C. R., L. A. Yeager, D. G. Bolser, C. Bonsell, A. M. Dichiera, Z. Hou, S. R. Keyser, A. J. Khursigara, K. Lu,

2. Muth AF, Negrete B, Erisman BE. 2020. Does functional redundancy affect ecological stability and resilience? A review and meta-analysis. *Ecosphere* 11(7):e03184. [10.1002/ecs2.3184](https://doi.org/10.1002/ecs2.3184).
3. Nuccio EE, Starr E, Karaoz U, Brodie EL, Zhou J, Tringe SG, Malmstrom RR, Woyke T, Banfield JF, Firestone MK, Pett-Ridge J. 2020. Niche differentiation is spatially and temporally regulated in the rhizosphere. *ISME J.* 14(4):999-1014. doi: [10.1038/s41396-019-0582-x](https://doi.org/10.1038/s41396-019-0582-x).
4. Boyd JA, Woodcroft BJ, Tyson GW. 2018. GraftM: a tool for scalable, phylogenetically informed classification of genes within metagenomes. *Nucleic Acids Res.* 46(10):e59. doi: [10.1093/nar/gky174](https://doi.org/10.1093/nar/gky174).
5. Belcour A, Frioux C, Aite M, Bretaudeau A, Hildebrand F, Siegel A. 2020. Metage2Metabo, microbiota-scale metabolic complementarity for the identification of key species. *Elife.* 9:e61968. doi: [10.7554/eLife.61968](https://doi.org/10.7554/eLife.61968).
6. Franzosa EA, McIver LJ, Rahnavard G, Thompson LR, Schirmer M, Weingart G, Lipson KS, Knight R, Caporaso JG, Segata N, Huttenhower C. 2018. Species-level functional profiling of metagenomes and metatranscriptomes. *Nat Methods.* 15(11):962-968. doi: [10.1038/s41592-018-0176-y](https://doi.org/10.1038/s41592-018-0176-y).
7. Zhou J, Yu X, Liu J, Qin W, He Z, Stahl D, Jiao N, Zhou J, Tu Q. 2021. VB₁₂Path for Accurate Metagenomic Profiling of Microbially driven cobalamin synthesis pathways. *mSystems* 6:e00497-21. <https://doi.org/10.1128/mSystems.00497-21>.

Título: Desarrollo de herramientas bioinformáticas para el análisis de datos multi-ómicos

Tutor/es (contacto):

Sonia Tarazona (sotacam@eio.upv.es)

Teresa Rubio

Ángeles Arzalluz

Pedro Salguero (pedsalga@upv.edu.es)

Afiliación:

Universitat Politècnica de València

Descripción:

Una de las principales líneas de investigación del Grupo de Genómica de la Expresión Génica (<http://conesalab.org/>) es el diseño o adaptación de metodologías estadísticas para el análisis integrativo de datos ómicos y su implementación en software bioinformático: paquetes de R, herramientas web, etc.

Algunos ejemplos de herramientas en las que estamos trabajando actualmente son:

- MORE (<https://github.com/ConesaLab/MORE>). Librería de R para la construcción de redes de regulación multi-ómicas mediante la utilización de modelos lineales generalizados.
- MOSim (<http://bioconductor.org/packages/release/bioc/html/MOSim.html>). Librería de R para generar datos simulados de experimentos multi-ómicos.
- PaintOmics (<http://www.paintomics.org/>). Herramienta web para la visualización y análisis de datos multi-ómicos sobre rutas biológicas de KEGG, Reactome, etc.
- MultiPower (<https://github.com/ConesaLab/MultiPower>). Librería de R para la estimación del tamaño muestral óptimo y el estudio de la potencia estadística de datos multi-ómicos.

Proponemos como posibles temas de Trabajos Final de Máster:

- Incorporación de nuevas funcionalidades de análisis, interpretación o visualización a las distintas herramientas del grupo.
- Adaptación de dichas herramientas a otros tipos de datos: single-cell, datos clínicos, etc.
- Aplicación y comparación de las herramientas con otros métodos publicados en diferentes conjuntos de datos multi-ómicos de distinta naturaleza.

Datos y software:

Aunque los temas propuestos son metodológicos, se utilizarán tanto datos simulados como reales para testar las herramientas/funcionalidades desarrolladas. El estudiante propondrá uno o más conjuntos de datos multi-ómicos públicos que sean susceptibles de ser analizados con las herramientas implementadas. Por tanto, una parte importante del trabajo será también la búsqueda y pre-procesamiento de dicho set de datos.

Referencias:

Sonia Tarazona; Héctor Carmona; Ana Conesa; Marta Llanola; Vicente Felipo. A multi-omic study for uncovering molecular mechanisms associated with hyperammonemia-induced cerebellar function impairment in rats. *Cell Biology and Toxicology*. 2021. ISSN 0742-2091. DOI: <https://doi.org/10.1007/s10565-020-09572-y>

Teresa Rubio Martínez-Abarca; Vicente Felipo; Sonia Tarazona; Roberta Pastorelli; Desamparados Escudero-García; Joan Tosca; Amparo Urios; Ana Conesa; Carmina Montoliu. Multi-omic analysis unveils biological pathways in peripheral immune system associated to minimal hepatic encephalopathy appearance in cirrhotic patients. *Scientific Reports*. 11, pp. 1 - 14. 2021. ISSN 2045-2322. DOI: 10.1038/s41598-020-80941-7

Nuria Planell; Vincenzo Lagani; Patricia Sebastián León; Frans van der Kloet; Ewoud Ewing; Nestoras Karathanasis; Arantxa Urdangarin; Imanol Arozarena; Maja Jagodic; Ioannis Tsamardinos; Sonia Tarazona; Ana Conesa; Jesper Tegner; David Gómez Cabrero. STATegra: Multi-Omics Data Integration - A Conceptual Scheme With a Bioinformatics Pipeline. *Frontiers in Genetics*. 12, 2021. ISSN 1664-8021. DOI:

<https://doi.org/10.3389/fgene.2021.620453>

Carmen Nuño; Manuel Ugidos; Sonia Tarazona; Manuel Martín Expósito; Alberto Ferrer; Susana Rodriguez Navarro; Ana Conesa. A multi-omics dataset of heat-shock response in the yeast RNA binding protein Mip6. *Scientific data*. 2020. ISSN 2052-4463. DOI:

<https://doi.org/10.1038/s41597-020-0412-z>

Sonia Tarazona; Leandro Balzano; David Gómez Cabrero; Andreas Schmidt; Axel Imhof; Thomas Hankemeier; Jesper Tegner; Johan Westerhuis; Ana Conesa. Harmonization of quality metrics and power calculation in multi-omic studies. *Nature Communications*. 11, pp. 1 - 13. 2020. ISSN 2041-1723. DOI: 10.1038/s41467-020-16937-8

Manuel Ugidos; Sonia Tarazona; José Manuel Prats Montalbán; Alberto Ferrer; Ana Conesa. MultiBaC: A strategy to remove batch effects between different omic data types. *Statistical Methods in Medical Research*. 29, pp. 2851 - 2864. 2020. ISSN 0962-2802. DOI: 10.1177/0962280220907365

Sonia Tarazona; Elena Bernabeu; Héctor Carmona; Belén Gómez Jiménez; Javier García Planells; Pim E.G. Leonards; Stephan Jung; Ana Conesa; Vicente Felipo; Marta Llansola. A Multiomics Study To Unravel the Effects of Developmental Exposure to Endosulfan in Rats: Molecular Explanation for Sex-Dependent Effects. *ACS Chemical Neuroscience*. 2019. ISSN 1948-7193

David Gómez Cabrero; Sonia Tarazona; Isabel Ferreiros Vidal; Ricardo N. Ramirez; Carlos Company; Andreas Schmidt; Theo Reijmers; Veronica von Saint Paul; Francesco Marabita; Javier Rodríguez Ubreva; Antonio Garcia Gomez; Thomas Carroll; Lee Cooper; Ziwei Liang; Gopuraja Dharmalingam. STATegra, a comprehensive multi-omics dataset of B-cell differentiation in mouse. *Scientific data*. 6, 2019. ISSN 2052-4463. DOI: 10.1038/s41597-019-0202-7

Víctor Sánchez-Gaya; Salvador Casaní-Galdón; Manuel Ugidos; Zheng Kuang; Jane Mellor; Ana Conesa; Sonia Tarazona. Elucidating the Role of Chromatin State and Transcription Factors on the Regulation of the Yeast Metabolic Cycle: A Multi-Omic Integrative Approach. *Frontiers in Genetics*. 2018. ISSN 1664-8021.

Rafael Hernández de Diego; Sonia Tarazona; Carlos Martínez Mira; Leandro Balzano; Pedro Furió; Georgios Pappas; Ana Conesa. PaintOmics 3: a web resource for the pathway analysis and visualization of multi-omics data. *Nucleic Acids Research*. 46, pp. 503 - 509. 2018. ISSN 0305-1048. DOI: 10.1093/nar/gky466

Pedro Furió; Ana Conesa; Sonia Tarazona. RGmatch: matching genomic regions to proximal genes in omics data integration. *BMC Bioinformatics*. 17, pp. 1293. 2016. ISSN 1471-2105. DOI: 10.1186/s12859-016-1293-1

Sonia Tarazona; Leandro Balzano; Ana Conesa. Multiomics Data Integration in Time Series Experiments. *COMPREHENSIVE ANALYTICAL CHEMISTRY. Data Analysis for Omic Sciences: Methods and Applications*. 18, pp. 505 - 532. Elsevier, 2018. ISBN 978-0-444-64044-4

Carlos Martínez-Mira, Ana Conesa, Sonia Tarazona. MOSim: Multi-Omics Simulation in R. <https://www.biorxiv.org/content/10.1101/421834v1>

Tarazona, S., Arzalluz-Luque, A. & Conesa, A. Undisclosed, unmet and neglected challenges in multi-omics studies. *Nat Comput Sci* 1, 395–402 (2021). <https://doi.org/10.1038/s43588-021-00086-z>

Título: Mejora del modelo metabólico a escala genómica *iFP764* con capacidad para simular la adaptación a la salinidad en la bacteria extremófila *Chromohalobacter salexigens*

Tutor/es (contacto):

Montserrat Argandoña Bertrán (montseab@us.es)¹

Manuel Salvador de Lara (manuel.salvador@idener.es)²

Afiliación:

¹ Departamento de Microbiología y Parasitología. Facultad de Farmacia. Univ. Sevilla

² Optimización orientada a la sostenibilidad S.L (IDENER).

Descripción:

Chromohalobacter salexigens es una γ -proteobacteria halófila considerada como un modelo biológico para estudiar los procesos de osmoadaptación en bacterias, pero también es considerada productor natural de ectoínas, para lo cual la bacteria adapta su metabolismo para sintetizar estos compuestos en función de la salinidad del medio (1). Las ectoínas se sintetizan y acumulan en el citoplasma de algunas bacterias halófilas en respuesta al estrés hiperosmótico y/o por elevadas temperaturas, y poseen un gran potencial en Biomedicina por sus propiedades bioestabilizadoras, antiinflamatorias, antiagregantes o antioxidantes, entre otras. Como productor natural de ectoínas, *C. salexigens* se ha propuesto como una buena alternativa a *Halomonas elongata* para la producción industrial de estos solutos. Para el diseño racional y obtención de nuevas cepas de *C. salexigens* optimizadas en la hiperproducción de ectoínas así como para estudiar la adaptación metabólica en respuesta a la salinidad, se ha desarrollado y validado un modelo metabólico a escala genómica, *iFP764* (2). Este modelo es capaz de predecir los flujos metabólicos a baja y elevada salinidad, mostrando las rutas que se activan/inhíben en cada condición y estableciendo, además, relaciones entre distintos subsistemas metabólicos del metabolismo nitrogenado y carbonado, así como con la síntesis y catabolismo de las ectoínas. Además, diferentes análisis ómicos han puesto de manifiesto que, entre otros aspectos de la fisiología, se producen cambios en la bioenergética, relacionados con los generadores de gradientes de protones, y una adaptación de la cadena respiratoria, y además han permitido identificar nuevas rutas metabólicas y sistemas clave (3). Es por ello que se hace necesario ampliar dicho modelo metabólico para así mejorar su capacidad predictiva, objetivo del Trabajo de fin de Master propuesto.

Datos y software:

Se dispone del modelo metabólico *iFP764* (2) así como datos de proteómica y RNAseq (3) en respuesta a salinidad y en diferentes condiciones de producción de ectoínas para localizar los genes/proteínas y rutas metabólicas no incluidos en *iFP764*.

El flujo de trabajo para la ampliación del modelo metabólico y simulaciones de flujos estará basado en el uso diferentes herramientas dentro del paquete CobraToolbox (<http://opencobra.sourceforge.net>) (4, 5) e implementado a través de MATLAB

Referencias:

(1) Pastor *et al.* (2013). J Biol Chem. 288:17769-81

(2) Piubeli *et al.* (2018). Microb Cell Fact. 17:2.

(3) Salvador *et al.*, (2018). Front Microbiol. 9:1845.

(4) Heirendt *et al.*, (2019). Nat. Prot. 14: 639-702

(5) Price *et al.*, (2003). Trends in Biotechnol. 21:162-9

Título: Análisis de las redes de regulación implicadas en la adaptación metabólica a la salinidad y su conexión con la síntesis de ectoínas en la bacteria extremófila *Chromohalobacter salexigens*

Tutor/es (contacto):

Montserrat Argandoña Bertrán (montseab@us.es)¹

Manuel Salvador de Lara (manuel.salvador@idener.es)²

Afiliación:

¹ Departamento de Microbiología y Parasitología. Facultad de Farmacia. Univ. Sevilla

² Optimización orientada a la sostenibilidad S.L (IDENER).

Descripción:

Chromohalobacter salexigens es una bacteria extremófila, halófila moderada, considerada como un modelo biológico para estudiar los procesos de osmoadaptación en bacterias y un productor natural de ectoínas, solutos compatibles de gran interés biotecnológico. Para ello posee mecanismos que controlan las rutas de síntesis, degradación y transporte de ectoínas, lo que le permite acumular en su citoplasma la cantidad óptima de estos solutos en función de las condiciones externas (1). Además, la bacteria adapta su metabolismo en función de la salinidad para poder sintetizar la cantidad necesaria de estos compuestos (2). Para desarrollar nuevas estrategias basadas en Biología de Sistemas que permitieran tener un conocimiento más global sobre los mecanismos de regulación que controlan la adaptación metabólica en función de las condiciones ambientales y su conexión con la síntesis de ectoínas en *C. salexigens*, se ha iniciado la reconstrucción de la red de regulación a escala genómica que controla la producción de estos compuestos y su conexión con el metabolismo, proceso complejo, ya que implica a su vez la integración de diversas redes, la obtención de datos experimentales y análisis *in silico*. Para ello, se necesita obtener información de aquellos sistemas que detectan las señales ambientales (sistemas de dos componentes) y cómo se relacionan con determinados reguladores transcripcionales, así como los genes metabólicos controlados por ambos sistemas de regulación. Recientemente se ha construido mediante una estrategia combinada de análisis *in silico* y datos experimentales (3) una red de regulación transcripcional (TRNcore) a escala genómica, y se ha obtenido mediante RNAseq el transcriptoma diferencial en respuesta a salinidad de mutantes en diversos reguladores globales clave, implicados en señalización/detección de la salinidad externa y en la regulación transcripcional. El objetivo del trabajo fin de Master propuesto es refinar, validar y ampliar la red TRNcore, así como analizar las redes implicadas en el control de la adaptación metabólica a salinidad y su conexión con la síntesis de ectoínas así como la identificación de reguladores transcripcionales clave.

Datos y software:

Se dispone de la red TRNcore, construida previamente, así como datos de RNAseq del transcriptoma diferencial en respuesta a salinidad de diferentes reguladores globales clave de *C. salexigens*.

Para el refinado y ampliación se seguirá la estrategia descrita por Imam y col. 2014 (3) y se podrán utilizar diversas herramientas de integración de datos que permiten la reconstrucción de redes de regulación e inferencia de interacciones a escala genómica a partir de datos experimentales de RNAseq, así como la validación de las mismas, como CoRegNet (4), WGCNA (<https://horvath.genetics.ucla.edu/html/CoexpressionNetwork/Rpackages/WGCNA/>), entre otras.

Referencias:

(1) Pastor *et al.*, (2010) *Biotechnol Adv.* 2010. 28(6):782-801

(2) Pastor *et al.* (2013). *J Biol Chem.* 288:17769-81

(3) Imam *et al.* (2015). *Plos Computational Biology.* 11(2):e1004103

(4) Nicolle *et al.* (2015). *Bioinformatics.* 3:3066-8.

Título: Caracterización de progenitores endoteliales en circulación en un modelo de daño neurovascular neonatal en ratón empleando estudio transcripcional de célula única, *Single-Cell RNA-seq*.

Tutor/es (contacto):

María José Sánchez/Patricia Ybot (mjsansan@upo.es pybot-ibis@us.es)

Afiliación:

Grupo Desarrollo Sistema Hemato-Vascular y Células Madre del CABD/ Grupo Neuro-desarrollo del IBIS

Descripción:

Durante un daño vascular se activan mecanismos de reparación incluyendo la activación de células endoteliales progenitoras (EPC). Las EPC junto con otras poblaciones celulares del sistema inmunológico salen al torrente sanguíneo desde donde migran a localizaciones tisulares y comienzan un proceso de revascularización y recuperación funcional de los tejidos. La participación de estos progenitores está documentada en las enfermedades neurovasculares en adultos, pero menos se conoce en procesos de daño neurovascular en neonatos. Las enfermedades neurovasculares en la infancia están reconocidas como causa más común de muerte y discapacidad y se asocian con deterioros motores, sensoriales y cognitivos. Existe por tanto gran interés en determinar los componentes celulares involucrados en respuesta regenerativa que puedan contribuir al establecimiento de un tratamiento neuroprotector adecuado.

En este proyecto usaremos un modelo de enfermedad neurovascular en ratón neonatal donde estudiaremos con tecnología single cell transcriptomics (scRNA-seq) los cambios en poblaciones celulares circulantes relacionados con la inducción del daño vascular intentando establecer una correlación de la extensión y progresión del daño cerebral con la función de los diferentes tipos celulares circulantes. La información producida proporcionará las bases guía para los estudios de funcionalidad inmunológica y regenerativa facilitando la obtención de biomarcadores tempranos y de evolución de daño cerebral esenciales para el tratamiento en la población infantil

Datos y software:

El proyecto que se oferta para este master es el estudio bioinformático de los resultados tras la secuenciación de librerías de cDNA de muestras ya disponibles de sangre control y tratados (inducción hipoxia/isquemia, HI) y la participación en la generación de librerías “single cell gene expression” y estudio bioinformático de dos nuevas muestras enriquecidas por citometría para EPCs.

Para el análisis bioinformático se empleará la plataforma proporcionada por el sistema 10x Genomics, que incluye los programas: CELLRANGER para verificación, cuantificación de transcritos y demultiplexing; LOUPE para visualización interactiva (agrupar células en cuanto a expresión de genes, encontrar genes que difieran significativamente entre clusters y niveles de expresión); el paquete de R SEURAT para identificación de clusters y el análisis de expresión diferencial.

Este análisis nos proporcionará información sobre los cambios en frecuencia de tipos celulares, la aparición de nuevos subtipos de células progenitoras circulantes y la abundancia de cada transcrito antes y después del daño. De esta forma será posible comparar el cambio en poblaciones celulares progenitoras, pro-inflamatorias y antiinflamatorias asociadas a un daño vascular con implicaciones en ictus neonatal con especial foco en determinar cambios en expresión génica de EPC implicada en recuperación de daño vascular.

Además de los datos del transcriptoma SC se proporcionarán un conjunto de datos basados en citometría y proteómica en muestras de sangre seca con paneles de múltiples marcadores que complementan y validan el estudio de sc RNAseq.

Referencias:

1-Single Cell Analysis in Vascular Biology. Chavkin NW and Hirschi KK. *Frontiers in*

Cardiovascular Medicine, 2020.

<https://www.frontiersin.org/articles/10.3389/fcvm.2020.00042/full>

2- Endothelial Progenitor Cells for Ischemic Stroke: Update on Basic Research and Application.

Liao S, Stem Cell Int. 2017. [https://pubmed.ncbi.nlm.nih.gov/28900446/-](https://pubmed.ncbi.nlm.nih.gov/28900446/)

3- Characterization of a Fetal Liver Cell Population Endowed with Long-Term Multiorgan Endothelial Reconstitution Potential. Cañete A, Stem Cells, 2017.

<https://pubmed.ncbi.nlm.nih.gov/27615355/>

4-Hicks SC, Biostatistics, 2018. Missing data and technical variability in single-cell RNA-sequencing experiments, <https://doi.org/10.1093/biostatistics/kxx053>

5- LOUPE browser.

<https://support.10xgenomics.com/single-cell-gene-expression/software/downloads/latest#loupe>

Título:

Genome-wide analysis of MYB gene family in *Dianthus broteri* (Caryophyllaceae) and identification of members regulating scent biosynthesis

Tutor/es (contacto):

Francisco Balao (fbalao@us.es)

Afiliación:

Departamento de Biología Vegetal y Ecología

Descripción:

Los factores de transcripción (TFs) son proteínas que poseen dos dominios funcionales relacionados con la unión al DNA y la activación/represión de la expresión génica en respuesta a estímulos endógenos o exógenos. La identificación de los TFs nos permite comprender mejor las vías de traducción y la respuesta funcional de las plantas. Los TFs se pueden dividir en diferentes familias basándose en los dominios de unión al DNA. MYB es una de las familias más diversas de TFs en eucariotas y están relacionada con procesos muy importantes como el control del crecimiento, actividades fisiológicas y respuesta a estímulos ambientales. Entre los MYB, ODORANT1 (ODO1), EMISSION OF BENZENOID II (EOBII), EOBI, y LATE ELONGATED HYPOCOTYL (LHY) regulan la producción de aromas florales mediante la activación de los genes de las rutas de biosíntesis en diversas especies (Abbas et al 2021). La identificación de los genes MYB controlando el aroma floral en otras especies es muy importante para conocer la evolución de las plantas con flor, así como tiene ya que tiene importantes aplicaciones en horticultura.

Dianthus broteri es un complejo poliploide de claveles silvestres que emiten una amplia variedad de aromas florales incluyendo cetonas, α -farneseno, β -cariofileno (sesquiterpenos), y bencenoides (Figura 1, Balao et al 2010, Balao et al 2011). Si en embargo no existe información sobre el rol de los factores de transcripción MYB en esta especie. En este trabajo se identificarán los factores de transcripción MYB en el genoma del clavel silvestre *Dianthus broteri* 2x. Se realizarán estudios filogenéticos para clasificarlos, así como análisis de sintenia con genomas relacionados de otras especies. Por último, se estudiará la relación entre la expresión de los distintos MYB y la emisión de volátiles florales mediante datos de un experimento de RNAseq y los perfiles cromatográficos de gases (GC-MS).

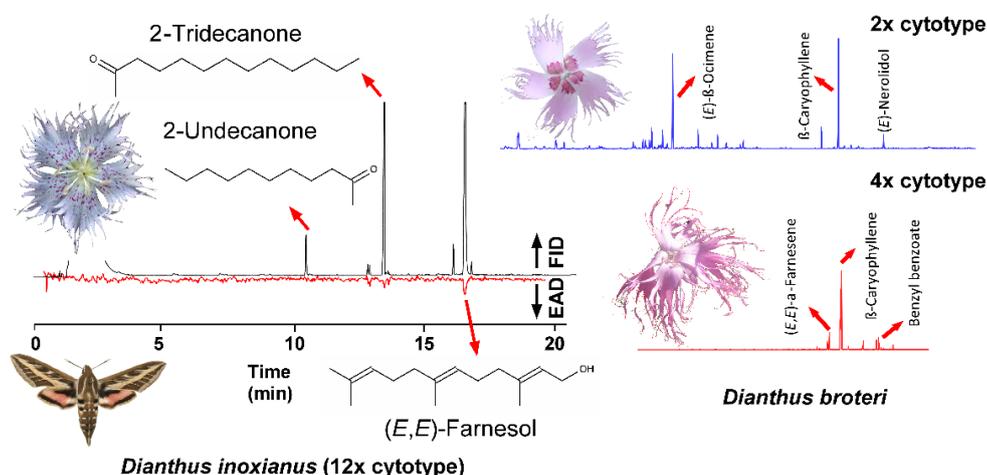


Figura 1. Perfiles cromatográficos del aroma floral de citotipos de *Dianthus broteri*

Datos y software:

Se explorará el genoma recientemente ensamblado de *Dianthus broteri* y un dataset de 12 muestras de RNAseq. Se llevará a cabo la identificación y anotación de TF con HMMER y PlantTFDB. La alienación de secuencias y los análisis filogenéticos se realizarán con Muscle y RAxML. Se utilizará diversos paquetes de R/Bioconductor para analizar las correlaciones entre los aromas florales y la expresión de los TFs. Para los análisis de sintenia con otros genomas se utilizará MCScanX, mientras que se estudiarán las secuencias promotoras con PlantCARE.

Referencias:

- Balao, F., Valente, L. M., Vargas, P., Herrera, J., & Talavera, S. (2010). Radiative evolution of polyploid races of the Iberian carnation *Dianthus broteri* (Caryophyllaceae). *New Phytologist*, 187(2), 542–551. <https://doi.org/10.1111/j.1469-8137.2010.03280.x>
- Balao, F., Herrera, J., Talavera, S., & Dötterl, S. (2011). Spatial and temporal patterns of floral scent emission in *Dianthus inoxianus* and electroantennographic responses of its hawkmoth pollinator. *Phytochemistry*, 72(7), 601–609. <https://doi.org/10.1016/j.phytochem.2011.02.001>
- Abbas, F., Ke, Y., Zhou, Y., Yu, Y., Waseem, M., Ashraf, U., Wang, C., Wang, X., Li, X., Yue, Y., Yu, R., & Fan, Y. (2021). Genome-Wide Analysis Reveals the Potential Role of MYB Transcription Factors in Floral Scent Formation in *Hedychium coronarium*. *Frontiers in Plant Science*, 12(February), 1–19. <https://doi.org/10.3389/fpls.2021.623742>
- Yarahmadov, T., Robinson, S., Hanemian, M., Pulver, V., & Kuhlemeier, C. (2020). Identification of transcription factors controlling floral morphology in wild *Petunia* species with contrasting pollination syndromes. *Plant Journal*, 104(2), 289–301. <https://doi.org/10.1111/tpj.14962>

Título:

The mobilome of *Dianthus broteri* (Caryophyllaceae): comparison of transposable element landscapes among a wild and a cultivated carnation

Tutor/es (contacto):

Francisco Balao (fbalao@us.es) y Ovidiu Paun (ovidiu.paun@univie.ac.at)

Afiliación:

Departamento de Biología Vegetal y Ecología (US) / Department of Botany and Biodiversity Research (Universidad de Viena)

Descripción:

El mobiloma, el conjunto de elementos genéticos móviles o elementos transponibles (TEs), es un importante motor de la evolución de los genomas de plantas (Baidouri & Panaud 2013). Los TEs se clasifican en diferentes familias y su prevalencia difiere entre organismos dependiendo de múltiples factores como una actividad de trasposición variable o distintas presiones selectivas (Chalopin et al 2015). Los TEs son una fuente de mutaciones, pueden interrumpir genes cuando se insertan o escinden y sus efectos pueden ser negativos o beneficiosos (Quadrana et al 2016). A lo largo de la evolución, los TEs pueden ser domesticados para actuar como factores de transcripción o estar implicados en regulación epigenética causando cambios fenotípicos. En este trabajo se identificarán los (TE) en el genoma del clavel silvestre *Dianthus broteri* y el clavel cultivado *Dianthus caryophyllus* para investigar el rol del mobiloma en el proceso de domesticación del clavel. Se realizarán estudios filogenéticos para clasificarlos en diferentes familias. Se estudiará la evolución de las diferentes familias de TE (inserciones y deleciones) y su posible efecto sobre los genes en el proceso de domesticación.

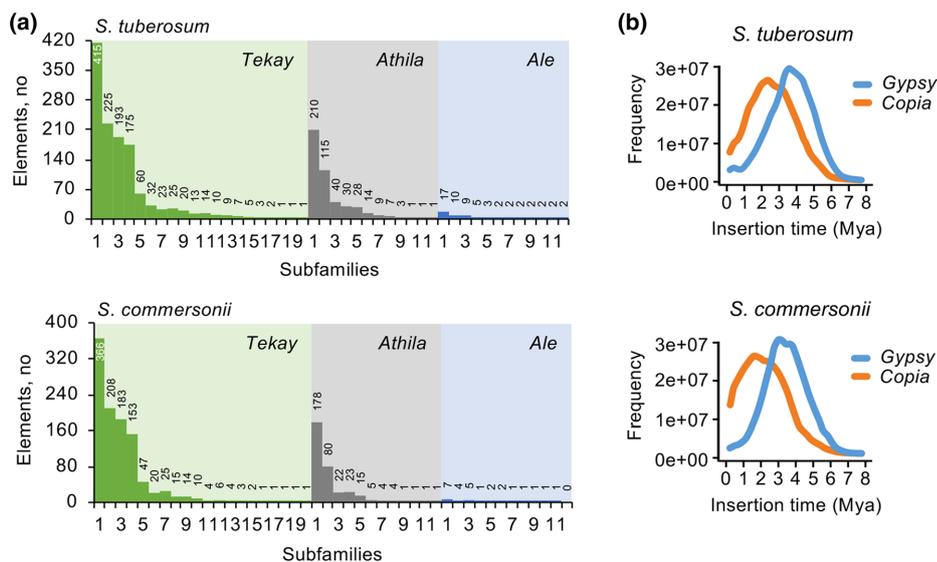


Figura 1. Abundancia de TEs y frecuencia de inserción en *Solanum* (Espósito et al, 2019). Fotos del clavel cultivado (*Dianthus caryophyllus*) y una especie silvestre (*D. broteri*)

Datos y software:

Se utilizarán los genomas de *Dianthus broteri* y *Dianthus caryophyllus* (Yagi et al. 2014) para la búsqueda y clasificación de TE con distintos softwares (TE-greedy-nester y RepeatMasker). Se llevarán a cabo análisis filogenéticos mediante Muscle (alineamiento de secuencias) y RAxML. Los análisis comparativos evolutivos se realizarán usando diferentes paquetes de Bioconductor en Rstudio.

Referencias:

- Yagi, M., Kosugi, S., Hirakawa, H., Ohmiya, A., Tanase, K., Harada, T., Kishimoto, K., Nakayama, M., Ichimura, K., Onozaki, T., Yamaguchi, H., Sasaki, N., Miyahara, T., Nishizaki, Y., Ozeki, Y., Nakamura, N., Suzuki, T., Tanaka, Y., Sato, S., ... Tabata, S. (2014). Sequence analysis of the genome of carnation (*Dianthus caryophyllus* L.). DNA Research, 21(3), 231–241. <https://doi.org/10.1093/dnares/dst053>
- Lexa, M., Jedlicka, P., Vanat, I., Cervenansky, M., & Kejnovsky, E. (2021). TE-greedy-nester: Structure-based detection of LTR retrotransposons and their nesting. Bioinformatics, 36(20), 4991–4999. <https://doi.org/10.1093/bioinformatics/btaa632>
- Baidouri, M. El, & Panaud, O. (2013). Comparative genomic paleontology across plant kingdom reveals the dynamics of TE-driven genome evolution. Genome Biology and Evolution, 5(5), 954–965. <https://doi.org/10.1093/gbe/evt025>
- Chalopin, D., Naville, M., Plard, F., Galiana, D., & Volff, J. N. (2015). Comparative analysis of transposable elements highlights mobilome diversity and evolution in vertebrates. Genome Biology and Evolution, 7(2), 567–580. <https://doi.org/10.1093/gbe/evv005>
- Quadrana, L., Silveira, A. B., Mayhew, G. F., Leblanc, C., Martienssen, R. A., Jeddelloh, J. A., & Colot, V. (2016). The Arabidopsis thaliana mobilome and its impact at the species level. 1–25. <https://doi.org/10.7554/eLife.15716>
- Esposito, S., Barteri, F., Casacuberta, J., Mirouze, M., Carputo, D., & Aversano, R. (2019). LTR-TEs abundance, timing and mobility in *Solanum commersonii* and *S. tuberosum* genomes following cold-stress conditions. Planta, 250(5), 1781–1787. <https://doi.org/10.1007/s00425-019-03283-3>

Título:

Análisis transcriptómico durante la biogénesis del cloroplasto en la microalga *Chlamydomonas reinhardtii*.

Tutor/es (contacto):

M. Águila Ruiz Sola (mrsola@us.es)

Afiliación:

Instituto de Bioquímica Vegetal y Fotosíntesis. Departamento de Bioquímica Vegetal y Biología Molecular. Universidad de Sevilla.

Descripción:

Además de ser el orgánulo que alberga la fotosíntesis, el cloroplasto es un sensor medioambiental fundamental capaz de regular la expresión de genes nucleares en respuesta a diferentes estreses. Así por ejemplo, además de absorber la luz y usarla para producir energía vía la fotosíntesis, es capaz de informar al núcleo de los posibles efectos negativos de la exposición exagerada a una situación de alta luz. Esta comunicación cloroplasto-núcleo se denomina señal retrógrada (SR), y es esencial para el correcto crecimiento y desarrollo de todos los organismos fotosintéticos (1). La señal retrógrada es clave durante la biogénesis de cloroplasto, sin embargo, aún se desconoce cuáles son los genes targets de este control y en qué momento ocurre. En el laboratorio hemos caracterizado un mutante que nos permite estudiar la biogénesis del cloroplasto en el alga unicelular *Chlamydomonas reinhardtii*, uno de los organismos modelo vegetales más importante. Las células del mutante *yellow-in-the-dark* (*y4*) (2) crecidas en oscuridad presentan un plasto no fotosintético, que no sintetiza clorofilas y muestran un color amarillo en cultivo líquido, característico de la acumulación de carotenoides (figura 1A). Sin embargo, cuando exponemos el mutante a la luz, las células rápidamente comienzan a acumular clorofilas y en una ventana de 48h horas son capaces de desarrollar un cloroplasto completo y fotosintético (figura 1A, B y C).

El objetivo de este proyecto es describir los cambios transcripcionales que ocurren durante la biogénesis del cloroplasto y analizar el efecto de la señal retrógrada en este proceso. Para ello, se analizarán datos de RNAseq que acabamos de generar de células *y4* crecidas en oscuridad y expuestas a la luz durante un *time-course*, en presencia y ausencia de una droga lincomicina, que activa la señal retrógrada y bloquea la biogénesis del cloroplasto (figura 1C, D).

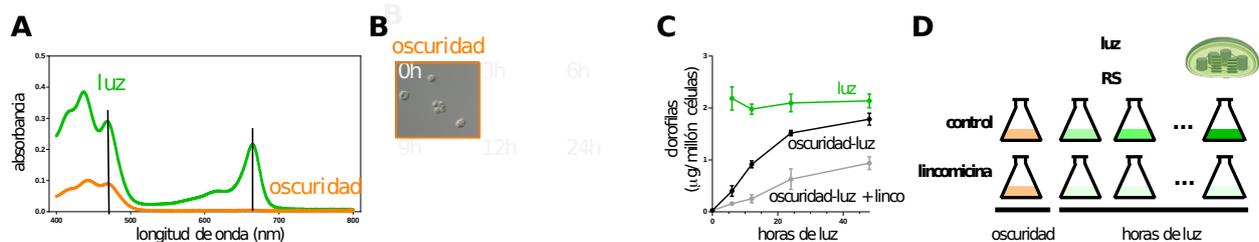


Figura 1: el mutante *y4* es un modelo para estudiar la biogénesis del cloroplasto y la señal retrógrada en el alga unicelular *Chlamydomonas reinhardtii*.

Datos y software:

Genoma de *Chlamydomonas reinhardtii* y datos transcriptómicos basados en RNA-seq.

Referencias:

- Hernández-Verdeja T, and Strand, A (2018). Retrograde signals navigate the path to chloroplast development. *Plant Physiol.* 176: 967–976.
- Suzuki JY, and Bauer CE (1992). Light-independent chlorophyll biosynthesis: involvement of the chloroplast gene *chlL* (*frxC*). *The Plant Cell* 4:929-940.

Título: Modelado matemático y simulación numérica de modelos de expresión génica con ecuaciones diferenciales

Tutor/es (contacto):

Daniel Franco Coronil (franco@us.es) y Faustino Maestre Caballero (fmaestre@us.es).

Afiliación:

Departamento de Ecuaciones Diferenciales y Análisis Numérico, Universidad de Sevilla.

Descripción:

La expresión génica es el proceso mediante el cual la información almacenada en el ADN se convierte en un producto génico funcional, como una proteína. Las dos funciones principales que componen el proceso de expresión génica son la transcripción y la traducción. La transcripción y la traducción están controladas por el número de ARNm (ARN mensajero) y proteínas en la célula. La expresión génica se puede representar como un sistema de ecuaciones diferenciales de primer orden para la tasa de cambio de ARNm y proteínas. Estas ecuaciones involucran ciclos de transcripción, traducción, degradación y retroalimentación.

El objetivo de este trabajo es estudiar un sistema de ecuaciones diferenciales de primer orden propuesto por Hunt et al. en [1], que modela la expresión génica, en concreto, la síntesis de proteínas en bacterias procariotas. Se tratará de analizar el comportamiento del modelo a través de soluciones de equilibrio y bifurcaciones, para obtener, entre otras propiedades de interés, información sobre la interrelación de la transcripción y la traducción en células procariotas.

La simulación numérica de este modelo nos ayudará a comprender su comportamiento y corroborar las propiedades que hemos analizado teóricamente.

Datos y software:

Realizaremos test numéricos de validación del modelo estudiado, comparando nuestros resultados con los existentes en la bibliografía sobre el tema, para ello emplearemos el software matemático Matlab u otro software similar.

Referencias:

[1] J. Hunt, E. Miller, J. Pham, L. Laplace. A continuous model of gene expression. California State Polytechnic University, Pomona (2005).

[2] Chen T, He HL, Church GM. Modeling gene expression with differential equations. Pacific Symposium of Biocomputing. pp 29-40 (1999).

[3] Echevarría, Rosa. “Apuntes de MATLAB orientados a métodos numéricos elementales”. <https://personal.us.es/echevarria/documentos/ApuntesMATLABMOM.pdf>

Título: Sistemas dinámicos sobre redes complejas.

Tutor/es (contacto): José Antonio Langa Rosado (langa@us.es) y José Ramón Portillo Fernández (josera@us.es).

Afiliación: Departamento de Ecuaciones Diferenciales y Análisis Numérico y Departamento de Matemática Aplicada I.

Descripción: El estudio de fenómenos reales a partir de su dinámica sobre redes complejas es una de las aproximaciones científicas más exitosas en la actualidad en multitud de disciplinas. En efecto, abstraer un problema real de las Ciencias de la Vida a una red de nodos y conexiones, sobre las que podemos definir una dinámica en forma de ecuaciones diferenciales, permite modelar multitud de problemas muy complejos desde un marco abstracto, abierto así al análisis riguroso desde las matemáticas. Esta interacción entre la Teoría de Sistemas Dinámicos y la Teoría de Grafos está siendo actualmente explotada en numerosas áreas de la Física, la Biología y la Neurociencia, por citar algunas. En este TFM, con un foco esencial en la formación matemática, que permita abordar con mejor fundamento algunos fenómenos aplicados, pretendemos abordar los temas siguientes:

1. Teoría de estabilidad para sistemas diferenciales.
2. Introducción a la moderna Teoría de Sistemas Dinámicos.
3. Introducción a la Teoría de Grafos.
4. Modelización y dinámica sobre redes complejas.
5. Aplicaciones e implementación. Experimentos computacionales.

Datos y software: podremos usar Matlab, Mathematica, Python, C, herramientas de supercomputación (SLURM) para los experimentos, tanto para la resolución de sistemas de ecuaciones diferenciales como la generación de grafos de dimensión alta. Estas son las herramientas básicas que nos harán falta en las modelizaciones propuestas.

Referencias:

[1] A. N. Carvalho, J. A. Langa, J. C. Robinson, *Attractors for Infinite-Dimensional Non-autonomous Dynamical Systems*, Applied Mathematical Series 182, Springer, New York, (2013).

[2] S. Boccaletti and V. Latora and Y. Moreno and M. Chavez and D.-U. Hwang, *Complex networks: Structure and dynamics*, Physics Reports 424 (4), 175-308, (2006)

[3] G. Guerrero, J. A. Langa, A. Suárez, *Attracting Complex Networks*, in *Complex Networks and Dynamics-Social and Economic Interactions*, Lecture Notes in Economics and Mathematical Systems, Springer, (2016).

[4] J. Grilli, M. Adoriso, S. Suweis, G. Barabás, J. R. Banavar, S. Allesina, and A. Maritan, "Feasibility and coexistence of large ecological communities," *Nature Communications*, vol. 8, Feb. 2017.

Título: Estudio de modelos diferenciales aplicados a procesos biológicos.

Tutor/es (contacto): Francisco Guillén González (guillen@us.es), María Ángeles Rodríguez Bellido (angeles@us.es)

Afiliación: Departamento de Ecuaciones Diferenciales y Análisis Numérico, Universidad de Sevilla.

Descripción:

El objetivo de este trabajo es introducir al alumno en el estudio de algunos procesos biológicos a través del modelado matemático mediante sistemas diferenciales. La finalidad última es conseguir entender y reproducir la realidad, intentando, entre otras cosas, reproducir las principales características de las situaciones analizadas y predecir su comportamiento. Para ello, varias son las herramientas que se tendrán en cuenta:

- Diseño de sistemas diferenciales que modelen, por ejemplo, modelos de poblaciones, modelos genéticos, ...
- Resolución de dichos modelos usando software matemático como Matlab o software libre de similares características
- Uso, análisis y comparación de datos: análisis a través de la interpolación y ajuste
- Estabilidad y comportamiento asintótico desde un punto de vista numérico

Datos y software:

Se usará el software matemático Matlab y/o paquetes similares de software libre. Eventualmente se efectuarán comprobaciones con bases de datos propias o bibliográficas relativas a los modelos analizados.

Referencias:

[1] Bacaër, Nicolas. “A Short History of Mathematical Population Dynamics”, Springer-Verlag London Limited 2011. [versión en español http://www3.uah.es/rafael_bravo/BreveHistoria.pdf]

[2] Echevarría, Rosa. “Apuntes de MATLAB orientados a métodos numéricos elementales”. <https://personal.us.es/echevarria/documentos/ApuntesMATLABMOM.pdf>

[3] Neuhauser, Claudia. “Matemáticas para Ciencias.” PEARSON - PRENTICE HALL, Madrid, 2004.

Título:

Comparing the transcriptional activity of transposable elements in frogs with hugely different genome size.

Tutor/es (contacto):

H. Christoph Liedtke
Iván Gómez Mestre

**Afiliación:**

Estación Biológica de Doñana - CSIC

Descripción:

When in the 1950's it was discovered that genomes varied in size, but that this size does not correlate with organismal complexity, it was seen as a paradox. Now, we know that it is the non-coding regions of genomes that vary the most in size. In particular, we are discovering that some of the largest genomes in the animal kingdom have accumulated massive amounts of repetitive genomic DNA, most prominently, transposable elements (TEs). However, why this accumulation is occurring at different rates across different vertebrate species, and how TEs can be purged from the genome to reduce genome size is still not well understood.

Among vertebrates, amphibians have some of the largest variation in genome size, ranging from less than 1 Gbp to more than 130 Gbp, making them an excellent model for studying how genomes grow larger or smaller over evolutionary time. Their extraordinary size means that there are only very few assembled amphibian genomes available to date. However, *de novo* transcriptome sequencing is a powerful tool for studying molecular mechanisms in non-model systems. For this project, we wish to compare the activity of transposable elements through comparative transcriptomics of two closely related species of frogs with widely varying genome sizes: the Western- (*Pelobates cultripes*) and Mexican- (*Scaphiopus couchii*) Spadefoot toads. As seen among amphibians more widely, this difference in genome size is also associated with extremely different developmental rates, with *S. couchii* (small genome) developing to adulthood in a fraction of the time it takes *P. cultripes* (large genome). There are therefore also interesting ecological and life history implications to genome size.

With this project, we aim to quantify the transcriptional activity of TEs (SINE and LINE retroelements) as well as genes involved in TE silencing, such as *Ago* and *Piwi* subfamily genes, during the larval stages of these species. We expect to find that TE activity correlates with genome size and that *S. couchii*, with its severely reduced genome, is experiencing a much higher silencing activity of transposons. This study will greatly improve our understanding of how genome architecture can differ wildly even between closely related species, and how TE activity interplays with organismal traits, such as developmental rate.

The student will also be fully integrated in our research group, taking part in our weekly lab meetings and scientific discussions to experience what a career in research is like.

Datos y software:

The student working on this project will have at their disposal already sequenced RNAseq data for both species. The project will have 3 major components:

- 1) Assemble *de novo* transcriptomes of both species using composite assembly pipelines such as TransPi [<https://github.com/palmuc/TransPi>]

- 2) Identify and quantify the relative expressed repetitive elements (DNA transposons, LTR retroelements, LINEs, SINEs etc.) using software such as RepeatMasker [www.repeatmasker.org] and RepeatScout [<https://github.com/mmcco/RepeatScout>], and compare this across species.
- 3) Quantify the expression of TE silencing genes, such as those belonging to the *Ago* and *Piwi* subfamilies, using software such as DESeq2 [<https://bioconductor.org/packages/release/bioc/html/DESeq2.html>].

Beyond this, the student will be using R as their primary programming language and will use command line (bash) to launch tasks on our in-house linux computing cluster.

Referencias sugeridas:

Carducci, F., Carotti, E., Gerdol, M., Greco, S., Canapa, A., Barucca, M. & Biscotti, M.A. (2021). Investigation of the activity of transposable elements and genes involved in their silencing in the newt *Cynops orientalis*, a species with a giant genome. *Scientific Reports*, 1–11.

Lamichhaney, S., Catullo, R., Keogh, J.S., Clulow, S., Edwards, S.V. & Ezaz, T. (2021). A bird-like genome from a frog: Mechanisms of genome size reduction in the ornate burrowing frog, *Platyplectrum ornatum*. *Proceedings of the National Academy of Sciences, USA*, 1–12.

Liedtke, H.C., Gower, D.J., Wilkinson, M. & Gomez-Mestre, I. (2018). Macroevolutionary shift in the size of amphibian genomes and the role of life history and climate. *Nature Ecology & Evolution*, **2**, 1792–1799.

Liedtke, H.C., Harney, E. & Gomez-Mestre, I. (2021). Cross-species transcriptomics uncovers genes underlying genetic accommodation of developmental plasticity in spadefoot toads. *Molecular Ecology*, **30**, 2220-2234

Título:

Impacto del transportador de Na SOS1 sobre la respuesta de adaptación al estrés salino de plantas de arroz

Tutor/es (contacto):

Fco. Javier Quintero Toscano y José M. Pardo Prieto

Afiliación:

Instituto de Bioquímica Vegetal y Fotosíntesis (CSIC).

Descripción:

La proteína SOS1 es un transportador de Na que juega un papel esencial en la respuesta adaptativa de las plantas a condiciones de salinidad mediante el control de los flujos de sodio, lo que permite reestablecer una correcta homeostasis iónica en la planta (1). Además, es posible, que SOS1 actúe como sensor de Na, desarrollando funciones más allá de la de transportador iónico. Un análisis transcriptómico realizado en plantas de arroz mutantes que no expresan el transportador SOS1 reveló que numerosos genes que se inducen por estrés salino se encontraban reprimidos en este mutante, lo que sugiere que la ruta de señalización está bloqueada en el mutante *sos1* (2). Para investigar este fenómeno con mayor profundidad. Para ello, se van a realizar estudios masivos de expresión génica (RNAseq) en plantas de arroz que contienen una mutación de SOS1 con pérdida de función (2), así como en plantas que contienen alelos de SOS1 con ganancia de función (3). La expresión génica en estas plantas se comparará con la de plantas silvestres.

Datos y software:

Además del genoma de arroz (*Oryza sativa*, v. *japonica*) se proporcionará un conjunto de datos transcriptómicos basados en RNA-seq consistente en muestras de raíz y parte aérea de cuatro genotipos diferentes (silvestre, mutante *ko* y dos mutantes hiperactivos) crecidas en medio hidropónico control o suplementado con 75 mM de NaCl.

Referencias:

(1) Shi, S., Quintero, F.J., Pardo, J.M. and Zhu, J.K. (2002) "The putative plasma membrane Na⁺/H⁺ antiporter SOS1 controls long distance Na⁺ transport in plants". *The Plant Cell* 14: 465-477.

(2) El Mahi, H., Perez-Hormaeche, J., De Luca, A., Villalta, I., Espartero, J., Gamez-Arjona, F., Fernandez, J.L., Bundó, M., Mendoza, I., Mieulet, D., Lalanne, E., Lee, S.Y., Yun, D.J., Guiderdoni, E., Aguilar, M., Leidi, E.O., Pardo, J.M. and Quintero, F.J. (2019) "A Critical Role of Sodium Flux via the Plasma Membrane Na⁺/H⁺ Exchanger SOS1 in the Salt Tolerance of Rice". *Plant Physiology* 180: 1046-1065.

(3) Quintero, F.J., Martinez-Atienza J., Villalta, I., Jiang, X., Kim, WY., Ali, Z., Fujii, H., Mendoza, I., Yun, D.J., Zhu, J.K. and Pardo, J.M. (2011) "Activation of the plasma membrane Na⁺/H⁺ antiporter Salt-Overly-Sensitive 1 (SOS1) by phosphorylation of an auto-inhibitory C-terminal domain." *Proc Natl Acad Sci U. S. A.* 108: 2611-2616

Título: Análisis proteómico en lipoproteínas de pacientes obesos e hiperlipémicos. Hacia la búsqueda de patrones de enfermedad

Tutor/es (contacto):

¹Sergio Montserrat de la Paz (delapaz@us.es)

²Ignacio Ortea García (iorteag@gmail.com; nacho.orteag@inibica.es)

Afiliación:

1. Departamento de Bioquímica Médica y Biología Molecular e Inmunología, Facultad de Medicina, Universidad de Sevilla.

2. Grupo de Proteómica Clínica y Traslacional, Instituto de Investigación en Innovación Biomédica de Cádiz (INiBICA)

Descripción:

Las HDLs (*Lipoproteínas de alta densidad*) son un grupo heterogéneo de complejos lipídicos y proteicos. Los niveles de HDL han demostrado una fuerte correlación inversa con la enfermedad cardiovascular. Esta acción ateroprotectora de las HDL se debe principalmente a su promoción del transporte reverso del colesterol y a la supresión de la respuesta pro-inflamatoria. Sin embargo, en los últimos años se ha puesto en evidencia que las HDL en condiciones patológicas pueden perder su funcionalidad debido a cambios en su composición. Es por ello que en el marco del proyecto US-1263458 financiado por la Junta de Andalucía, obtuvimos HDL purificadas de voluntarios sanos, de pacientes obesos, de voluntarios sanos sometidos a diferentes condiciones de hiperlipidemia postprandial. Las muestras fueron analizadas por el Servicio de Proteómica de la Universidad de Córdoba mediante espectrometría de masas de alta resolución en Sistema Orbitrap Fusion y la identificación de proteínas fue llevada a cabo por el software MaxQuant. El objetivo principal de este trabajo es analizar los perfiles proteómicos y establecer patrones de cambio entre las distintas poblaciones de pacientes y condiciones. Esperamos que este proyecto sirva a las lipoproteínas como nuevos biomarcadores de enfermedad.

Datos y software:

Se disponen de datos proteómicos cuantitativos e identificativos de muestras basales de voluntarios sanos y obesos y de voluntarios sanos sometidos a diferentes condiciones de hiperlipidemia postprandial. Se dispone de los datos de cuantificación (LFQ) transformados en log₂. Además, se dispone de los raw files y todos los archivos de la identificación y cuantificación.

Referencias:

Plubell D.L. et al. High-Density Lipoprotein Carriers Markers That Track With Recovery From Stroke. *Circulation Research* **2020**, 127, 1274-1287.

Andraski A.B. et al. Effects of replacing dietary monounsaturated fat with carbohydrate on HDL (High-density lipoprotein) protein metabolism and proteome composition in humans. *Arteriosclerosis, Thrombosis, and Vascular Biology* **2019**, 39, 2411-2430.

Zhang Y. et al. HDL subclass proteomic analysis and functional implication of protein dynamic change during HDL maturation. *Redox Biology* **2019**, 101222.

Gordon S.M. et al. High density lipoprotein proteome is associated with cardiovascular riskfactors and atherosclerosis burden as evaluated by coronary CT angiography. *Atherosclerosis* **2018**, 278, 278-285.

Título:

Análisis de la influencia del fotoperiodo en el crecimiento celular en *Arabidopsis thaliana*

Analysis of the influence of photoperiod on cell growth in *Arabidopsis thaliana*

**Tutor/es (contacto): Federico Valverde Albacete (Tutor/director)
Gloria Serrano Bueno (Directora)**

Afiliación: Plant Development Unit, Institute for Plant Biochemistry and Photosynthesis (IBVF), Spanish National Research Council (CSIC) and University of Seville (US)

Descripción:

The length of the day or photoperiod influences many key developmental processes in plants, such as seed dormancy, juvenility or the floral transition. Recently, our group has also unveiled new roles for the main photoperiodic hub gene CONSTANS (CO) in other processes related with senescence, aging and growth in different organs such as flowers. By obtaining and analyzing RNAseq and proteomic data, the candidate will try to find out the mechanism that triggers the photoperiod response in cell growth and division, an important trait for biotechnological applications. The candidate will get experience in plant molecular biology, biochemistry and bioinformatics in an international academic atmosphere that has shown to be a strong asset for future applied and academic opportunities.

Datos y software:

Transcriptomics, proteomics, physiologic, metabolomics data, statistics. Construction of gene expression networks, clustering, GO terms, database use (Phytozome, TAIR, ChalyNet...) Use of R-based bioinformatics packages (Bowtie, Cufflinks, Cytoscape, topGO, HOMER,...).

Referencias:

- 1- Serrano-Bueno, G., et al (2021) **Front. Plant Sci.** doi: 10.3389/fpls.2021.634393.
- 2- Serrano-Bueno Get al, (2020) **The Plant Journal** 101, 1287-1302.
- 3- Hayama, R et al (2019). **Scientific Reports** 9.
- 4- Serrano-Bueno, G., (2017) **Current Opinion in Plant Biology** 37,10-17.
- 5- Romero-Campero1, et al (2016) **BMC Genomics** 17:227. Doi: 10.1186/s12864-016-2564-y

Título: Ecology meets functional genomics: understanding the genetic basis of adaptive plant development in rapidly changing environments

Tutor/es (contacto):

Francisco J. Romero-Campero (fran@us.es)

Xavier Picó Mercader (xpico@ebd.csic.es)

Afiliación:

FJR-C: Dept. de Ciencias de la Computación e Inteligencia Artificial (Univ. Sevilla)

XPM: Estación Biológica de Doñana (EBD-CSIC)

Descripción: The regulation of flowering time has substantial biological significance because it defines the vegetative-to-reproductive transition and determines the length of the post-embryonic life history. The relevance of this trait lies in the fact that variation in flowering time is under selection pressure because it affects different plant fitness components.

The main goal of this TFM is to provide empirical evidence for the **diversity of genetic pathways** used by plants to complete the vegetative-to-reproductive transition in natural environments using wild *Arabidopsis thaliana* genotypes. We still deeply ignore how **genetic regulatory networks** actually function and evolve in the wild. To our knowledge, just a few studies have attempted to assess genome-wide gene expression patterns of *A. thaliana* in natural environments (Richards et al. 2012, Song et al. 2018).

In this TFM, we will address this issue with brand new **transcriptome data** from four locally adapted genotypes (from seaside, forest, and high altitude populations) grown in a common field experiment (Sierra de Grazalema; Nov 2020-Mar 2021). Data encompass **RNA-Seq data** from different developmental stages (vegetative, induced and flowered) and two time points in the day (morning and afternoon). The goal is to detect, through a **differential gene expression analysis** comparing the transcriptome after and before flowering, the sets of expressed genes that are common and unique in each genotype.

Datos y software: The RNA-Seq data includes expression data from ca. 33,000 genes from four genotypes (Bon, Moj, Vel and Col) collected in the morning and in the afternoon over the vegetative, inductive and flowering stages, totaling 84 transcriptomes. This data set has already been checked out for quality and it is ready to be analyzed. The software to analyze RNA-Seq data is extensive, well developed and mostly available as R packages.

Referencias:

Previous studies inspiring this project:

1. Richards CL et al. 2012 Genome-wide patterns of Arabidopsis gene expression in nature. PLoS Genetics 8:e1002662. Doi: <https://doi.org/10.1371/journal.pgen.1002662>.
2. Song YH et al. 2018. Molecular basis of flowering under natural long-day conditions in Arabidopsis. Nature Plants 4:824–835. Doi: <https://doi.org/10.1038/s41477-018-0253-3>.

Own research on adaptive variation and genomic analyses in wild *A. thaliana* populations:

3. Exposito-Alonso M et al. 2018. Spatio-temporal variation in fitness responses to contrasting environments in *Arabidopsis thaliana*. Evolution 72:1570–1586. Doi: <https://doi.org/10.1111/evo.13508>.
4. Tabas-Madrid D et al. 2018. Genome-wide signatures of flowering adaptation to climate temperature: regional analyses in a highly diverse native range of *Arabidopsis thaliana*. Plant, Cell and Environment 41:1806–1820. Doi: <https://doi.org/10.1111/pce.13189>.