

TESIS DOCTORAL 2022

Pedro de los Reyes Rodríguez



**CONTROL DEL RELOJ CIRCADIANO
POR LA SEÑAL FOTOPERIÓDICA
EN PLANTAS**

Ilustraciones por Israel Barranco



Tesis doctoral

Control del reloj circadiano por la señal fotoperiódica en plantas

Memoria presentada por

Pedro de los Reyes Rodríguez
para optar al título de Doctor por
la Universidad de Sevilla

Bajo la supervisión de

Dr. Federico Valverde Albacete

Dr. Francisco José Romero Campero

Prof. José María Romero Rodríguez

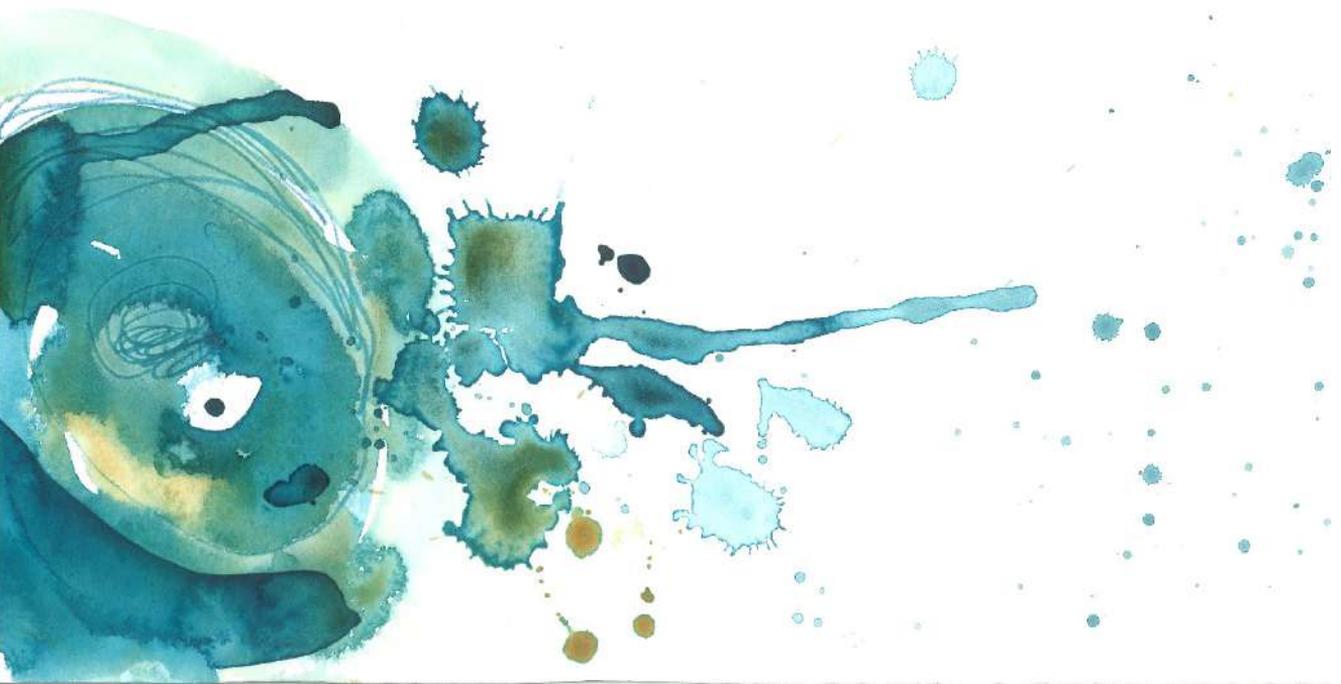


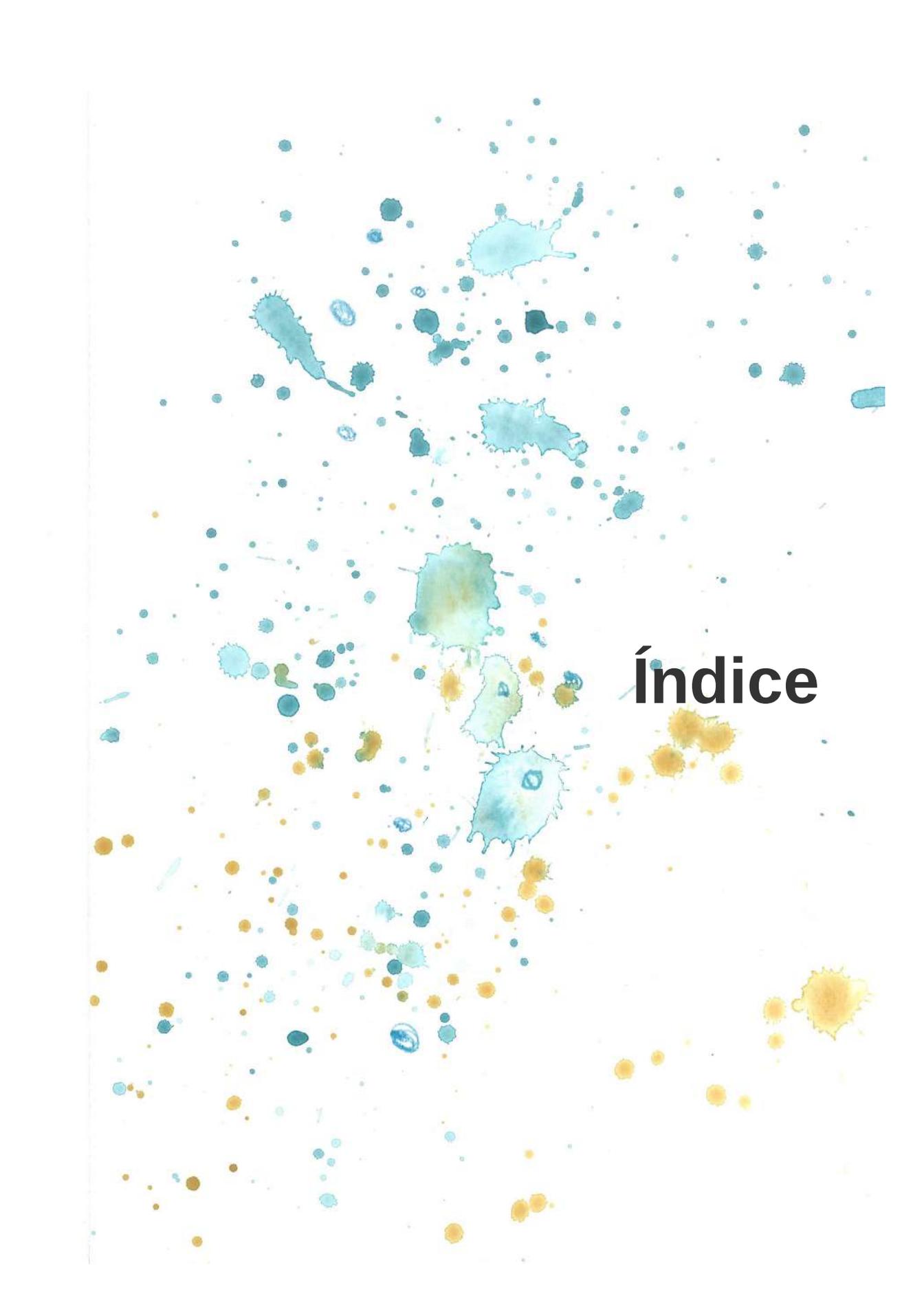
Los trabajos desarrollados en esta investigación han sido financiados por los proyectos BIO2017-83629-R B025BB y PID2020-117018RB-100 a través del Ministerio de Ciencia e Innovación, anteriormente Ministerio de Economía, Industria y Competitividad.

El autor recibió financiación de la Federación de Sociedades Europeas de Bioquímica (FEBS) para la realización de una estancia predoctoral en el Instituto Max Planck para la Mejora Vegetal (MPIPZ, Colonia).

A mis padres

plantas
gen día nodos ostreococcus
arabidopsis
amanecer proteínas
red luz genes redes
análisis condiciones _
procesos datos control
noche circadiano familia
factores reloj
génica
chlamydomonas
expresión
transcripcion

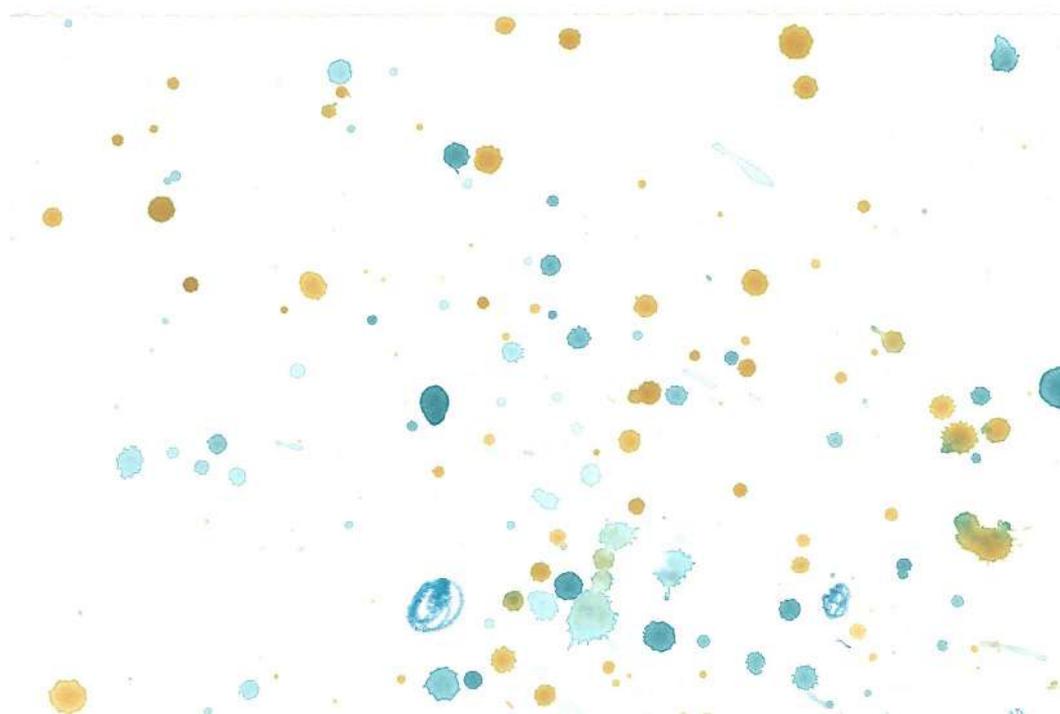




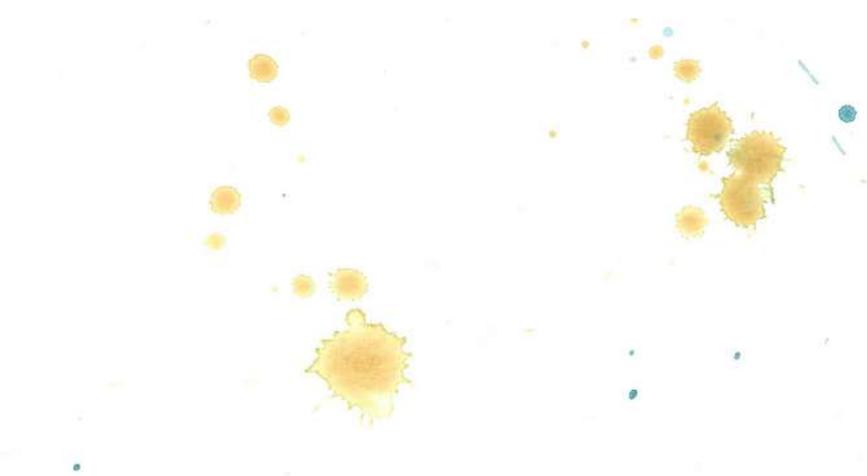
Índice

Resumen	15
Introducción	21
1 La señal fotoperiódica en plantas	23
2 Floración	34
3 Técnicas ómicas	39
4 Biología de sistemas de plantas	43
Hipótesis y objetivos	51
Materiales y métodos	55
1 Recolección y análisis de datos ómicos	57
2 Construcción y análisis de redes génicas	61
3 Análisis de patrones circadianos	66
4 Significancia de solapamiento de conjuntos génicos	67
5 Análisis de conservación de módulos	67
6 Generación de aplicaciones web usando shiny R	68
7 Material vegetal	69
8 Bacterias	70
9 Clonación mediante el sistema Gateway	72
10 Extracción de RNA y síntesis de cDNA	74
11 Análisis de expresión génica mediante QPCR en tiempo real.	75
12 Extracción de proteínas	76
13 Inmunoprecipitación de la cromatina	77
14 Expresión transitoria en <i>Nicotiana benthamiana</i>	80
15 Tiempo de floración	82
16 Movimiento de hojas	82
Conservación de patrones diurnos en el linaje verde	85
1 La mayoría de proteínas de <i>Ostreococcus</i> presentan un ortólogo en <i>Arabidopsis</i> o <i>Chlamydomonas</i> , pero no ocurre lo mismo entre las otras dos especies.	87
2 Grandes familias de factores de transcripción de <i>Arabidopsis</i> tienen un único ortólogo en microalgas	93
3 La mayoría de genes en <i>Ostreococcus</i> y en <i>Chlamydomonas</i> exhiben un patrón rítmico, al contrario que en <i>Arabidopsis</i>	97
4 Los grupos de genes rítmicos confieren separación temporal de procesos biológicos	102

5 Existe un nivel alto de conservación entre las dos microalgas y un nivel moderado entre éstas y <i>Arabidopsis</i>	111
ATTRACTOR, una herramienta para el estudio del reloj circadiano en <i>Arabidopsis</i>.	117
1 Flujo de trabajo computacional	119
2 ATTRACTOR integra reguladores circadianos centrales y señalizadores lumínicos	121
3 Herramienta web ATTRACTOR	122
4 El reloj circadiano en <i>Arabidopsis</i> es robusto a ataques aleatorios y sensible a ataques dirigidos	127
5 El reloj es especialmente activo a ZT0	128
6 Diferentes motivos de red regulan procesos fisiológicos esenciales	131
Control del reloj circadiano por la vía fotoperiódica en <i>Arabidopsis</i>	135
1 Existe una alta interconexión entre el reloj circadiano y la vía fotoperiódica en <i>Arabidopsis</i>	137
2 CONSTANS altera el reloj circadiano bajo condiciones de Día Largo	140
3 CONSTANS se une a promotores de genes centrales del reloj circadiano.	144
4 CO, PRRs y HY5 comparten sitios de unión al genoma en los promotores de genes regulados por el reloj	151
5 CO, PRR5 y HY5 interaccionan <i>in vivo</i>	153
Discusión	157
1 El reloj circadiano regula más procesos transcripcionales en algas que en plantas	160
2 Algas y plantas presentan una alta conservación en la expresión génica en las transiciones luz/oscuridad, a pesar de la gran distancia evolutiva que las separa	161
3 Amplificación y diversificación génica de microalgas hacia plantas superiores	162
4 Características del reloj circadiano en <i>Arabidopsis</i>	163
5 Fotoperiodo y reloj circadiano en <i>Arabidopsis</i>	165
Conclusiones	177
Bibliografía	181
Anexos	203



Resumen



Los organismos fotosintéticos son muy eficientes en la adaptación de su desarrollo a las condiciones ambientales. Para conseguir esta adaptación monitorizan continuamente los estímulos externos que les llegan del ambiente y promueven cambios fisiológicos, como ocurre con las modificaciones del transcriptoma. La luz solar es esencial para la supervivencia de los organismos fotosintéticos, ya que constituye su principal fuente de energía y controla múltiples aspectos de su fisiología. En la percepción de la luz en plantas y algas participan un conjunto de receptores y factores de transcripción que conectan las señales procedentes de la luz con cambios específicos en la expresión génica. La existencia de eventos externos cíclicos muy predecibles como la sucesión de días/noches o de las estaciones, permite coordinar y anticipar de manera muy robusta los procesos biológicos. Esta medida del tiempo es llevada a cabo por un oscilador interno denominado reloj circadiano. Por otra parte, la señalización de la vía fotoperiódica permite a los organismos vegetales medir la longitud del día y así obtener información estacional para controlar complejos procesos que son clave para su supervivencia como la floración. Uno de los genes centrales en esta vía es *CONSTANS*, que está regulado a nivel transcripcional y postraduccional por el reloj circadiano y por señales luminosas. En esta tesis doctoral se ha seguido un enfoque de Biología Molecular de Sistemas para entender cómo han evolucionado estos procesos a lo largo del linaje verde, se han generado herramientas para su estudio y finalmente se ha descrito una nueva conexión entre el reloj circadiano y la vía fotoperiódica en la planta modelo *Arabidopsis thaliana*.

En una primera aproximación investigamos la evolución de la expresión génica diaria en el linaje vegetal, empleando las microalgas *Ostreococcus tauri* y *Chlamydomonas reinhardtii* y la planta superior *Arabidopsis thaliana*. Inicialmente, realizamos un estudio de ortología que ha revelado por una parte la aparición de genes específicos en *Chlamydomonas* y *Arabidopsis* y por otra parte la amplificación y diversificación de familias génicas. Además, hemos investigado cómo ha cambiado la influencia de los ciclos día/noche sobre el transcriptoma a lo largo de la evolución vegetal mediante la construcción de redes de co-expresión en ciclos de luz/oscuridad. Nuestros resultados indican una mayor dependencia de los ciclos diurnos en microalgas que en plantas. En cuanto a la evolución de los patrones de expresión,

Resumen

hemos observado un alto nivel de conservación en genes que se expresan en las transiciones luz/oscuridad a pesar de la gran distancia evolutiva entre algas y plantas. Estos resultados pueden ser explorados en la herramienta CircadiaNet.

En segundo lugar, construimos la red transcripcional ATTRACTOR que integro datos transcriptómicos y datos de ChIP-seq de factores de transcripción involucrados en el reloj circadiano y en la señalización por luz. De esta forma, el estudio de la red permite investigar la regulación coordinada del reloj circadiano y la señalización lumínica sobre la expresión génica en *Arabidopsis*. Mediante el análisis topológico de ATTRACTOR hemos descrito que estos programas transcripcionales poseen la característica conocida como “robusto pero frágil”, es decir, son robustos frente a ataques aleatorios y frágiles frente a ataques dirigidos a genes altamente conectados. Por otra parte, se ha demostrado que los genes cuya regulación es más dependiente del reloj circadiano se expresan durante las primeras horas de la mañana, modulando procesos concretos como respuesta inmune, respuesta a hormonas o crecimiento. Además, se ha descrito que algunos motivos de red compuestos por varios factores de transcripción pueden explicar la regulación de determinados aspectos del desarrollo, dando lugar a propiedades emergentes. Por ejemplo, el motivo formado por CCA1, PIF y PRR5 participa en la respuesta a sequía y frío. Para este fin, se ha desarrollado una aplicación web para su exploración por parte de la comunidad científica.

Por último, utilizando estas herramientas, hemos demostrado un nuevo papel para CO en la regulación del reloj circadiano en *Arabidopsis*. CO formaría parte de una señalización retrógrada desde la vía fotoperiódica hacia el oscilador central, proporcionando información estacional al reloj. En condiciones de día largo, CO se une a genes centrales del reloj circadiano para alterar su perfil de expresión. Uno de estos genes es el *PSEUDO RESPONSE REGULATOR 5 (PRR5)*, con el que además comparte sitios de unión en el genoma. Por ello, CO y PRR5 establecen un motivo de doble retroalimentación con salida múltiple regulando genes diana en común. Además, hemos encontrado que CO se une a motivos de DNA G-box en estos promotores, probablemente a través del factor de transcripción bZIP LONG HYPOCOTIL 5 (HY5).

Photosynthetic organisms are highly efficient in the adaptation of their development to environmental conditions. To achieve this, they continuously track the external stimuli and promote physiological changes, which can be readily checked by transcriptome modifications. Sunlight is essential for plant survival as it is the main energy source and controls multiple aspects of their physiology. In plants and algae, light perception is carried out by several receptors and transcription factors that connect light signals with specific changes in gene expression. The existence of highly predictable cyclical external events, such as the succession of days/nights or seasons, allows them to coordinate and anticipate biological processes in a robust way. This daily time measurement is performed by an internal oscillator called circadian clock. Alternatively, the photoperiodic regulatory pathway allows plants and algae to measure the day length and obtain seasonal information to control complex processes that are key to their survival, such as flowering. One of the key genes in this pathway is *CONSTANS*, which is regulated at transcriptional and post-translational levels by the circadian clock and by light signals. In this doctoral thesis, a Molecular Systems Biology approach was chosen to try to understand how these processes have evolved along the green lineage. Several tools have been generated for their study and a new connection between the circadian clock and the photoperiodic pathway has been described in the model plant *Arabidopsis thaliana*.

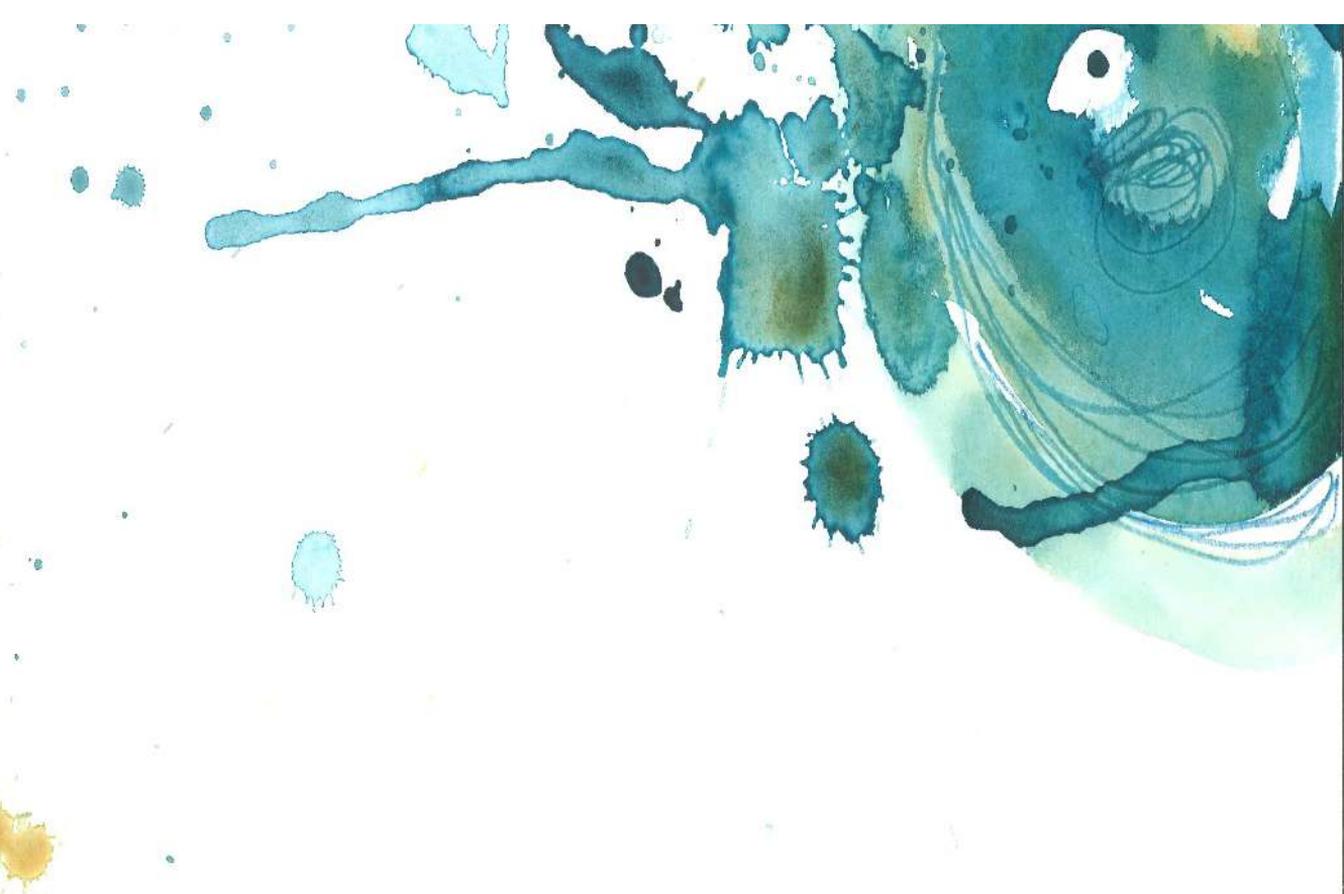
Firstly, we studied the diurnal gene expression evolution along the plant lineage, using the microalgae *Ostreococcus tauri* and *Chlamydomonas reinhardtii* and the spermatophyte *Arabidopsis thaliana*. Initially, we have carried out an orthology analysis that has revealed, on the one hand, the emergence of specific genes in *Chlamydomonas* and *Arabidopsis* and, on the other hand, the amplification and diversification of gene families between plants and algae. In addition, we have investigated the influence raised by day/night cycles on the transcriptome throughout evolution by constructing gene networks of light/dark samples. Our results indicate a greater dependence on diurnal cycles in microalgae than in plants. Regarding the evolution of expression patterns, we have observed a high level of conservation in genes that are expressed in light/dark transitions despite the great evolutionary

Resumen

distance between algae and plants. These results can be explored in the CircadiaNet web.

Second, the ATTRACTOR transcriptional network allows to elucidate the coordinated regulation of circadian clock and light signalling on *Arabidopsis* gene expression. Through the topological analysis of the network, we have shown that these transcriptional programs have the property known as “robust but fragile”, meaning that they are robust against random failures and fragile against attacks directed at highly connected genes. Furthermore, we have shown that those genes whose regulation is more dependent on the circadian clock are expressed during the first hours of the morning, modulating specific processes such as the plant pathogen interaction, response to hormones or growth. In addition, it has been described that network motifs constituted by several transcription factors may help to explain the regulation of certain aspects of development, giving rise to emergent properties. For example, the CCA1/PIF5/PRR5 network motif participates in the response to drought and cold. For this reason, a web application has been developed and is available for the scientific community.

Finally, using these tools, we have demonstrated a new role for CO in the regulation of the circadian clock in *Arabidopsis*. CO would participate in a retrograde signaling from the photoperiodic pathway to the central oscillator, providing seasonal information to the system. Under long-day conditions, CO could bind the promoter of core genes of the circadian clock to alter its expression profile. One of these genes is the *PSEUDO RESPONSE REGULATOR 5 (PRR5)*, which also shares CO binding sites in the genome. Therefore, CO and PRR5 establish a double feedback motif with multiple output regulating common target genes. Furthermore, we have found that CO binds to G-box DNA motifs in these promoters, probably through a bZIP transcription factor, such as LONG HYPOCOTIL 5 (HY5).



Introducción

1 La señal fotoperiódica en plantas

1.1 Señalización lumínica o percepción lumínica

En organismos fotosintéticos es especialmente necesario detectar la calidad, cantidad y dirección de la luz para modular el desarrollo y otras funciones fisiológicas (Li y Mathews, 2016). Por este motivo poseen los llamados fotoreceptores, que son proteínas que median la percepción de las señales lumínicas. A rasgos generales pueden clasificarse en tres categorías; los fitocromos (PHYs), que detectan la luz azul, roja y roja lejana, y los receptores de luz azul/UV-A, que es absorbida por los criptocromos (CRYs) y las fototropinas (PHOTs). Además, existen fotoreceptores específicos de plantas, como las proteínas con dominio LOV ZTL (ZEITLUPE), FKF1 (FLAVIN BINDING, KELCH REPEAT, F-BOX1) y LKP2 (LOV KELCH PROTEIN2), que perciben la luz azul. Recientemente se ha descubierto que la luz ultravioleta B (UV-B) puede ser percibida por la proteína UVR8 (UV RESISTANCE LOCUS8), que está conservada en algas verdes, briofitas, licofitas y angiospermas (Rizzini et al., 2011; Wu et al., 2012; Fernández et al., 2016).

Los ecosistemas marinos tienen propiedades únicas en cuanto a la transmisión de la luz, ya que se atenúa la longitud de onda larga en los primeros metros de la columna de agua. Por tanto, los organismos acuáticos han evolucionado para percibir luz de onda corta, como la luz azul, que es recibida por fotoreceptores específicos. Esto les permite regular importantes funciones celulares como expresión génica, ciclo sexual, fototaxis o fotosíntesis.

En la picroalga marina *Ostreococcus tauri* se han identificado proteínas que participan en el reloj y que perciben señales lumínicas. Varias proteínas de la familia criptocromo/fotoliasa (CPF; cry/photolyase family) (HEIJDE et al., 2010) y una histidina quinasa LOV-domain (LOV-HK) (Djouani-Tahri et al., 2011) han sido descritas como receptores de luz azul. Estos fotoreceptores están bajo control circadiano y además transmiten al reloj información lumínica (Figura 1).

En el modelo de alga clorofita *Chlamydomonas reinhardtii* se observa una mayor complejidad en la percepción de luz. Poseen un criptocromo de plantas (plant-

Introducción

like CRY) pCRY (Reisdorph y Small, 2004), uno similar al de los animales (animal-like CRY) aCRY (Beel et al., 2012b), y dos DASH-CRYs (*Drosophila*, *Arabidopsis*, *Synechocystis*, *Homo*) (Beel et al., 2012a, 2012b). A diferencia del resto de criptocromos, aCRY puede responder tanto a luz azul como a luz roja. *Chlamydomonas* contiene una sólo copia de fototropina (PHOT) que responde a luz azul y ultravioleta A (UV-A) (Huang et al., 2002), así como un fotoreceptor UVR8 que media la percepción de la luz ultravioleta B (UV-B) (Figura 1) (Tilbrook et al., 2016).

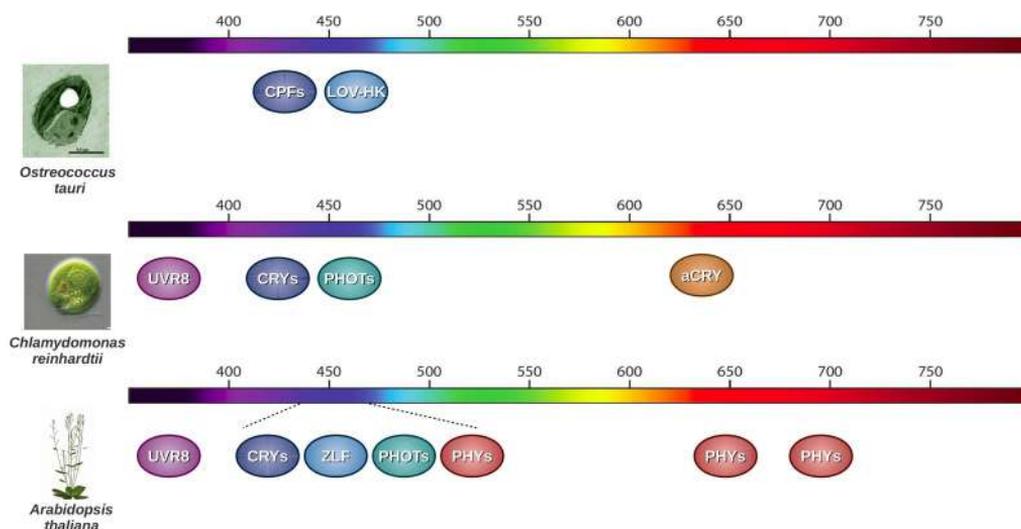


Figura 1. Fotoreceptores que llevan a cabo la percepción lumínica en *Ostreococcus*, *Chlamydomonas* y *Arabidopsis*. Se observa una mayor complejidad y diversidad en fotoreceptores desde microalgas hacia plantas superiores. *Ostreococcus* sólo posee proteínas de la familia CPFs y una LOV-HK que percibe luz azul. *Chlamydomonas* además puede percibir la luz UV-B con una proteína UVR8 y luz roja con un CRY similar al de animales. *Arabidopsis*, por su parte, ha desarrollado fitocromos que perciben luz azul, roja y roja lejana, además de una diversificación en proteínas con dominio LOV como ZTL, LKP2 y FKF1 (ZLF).

Las plantas superiores, además de los anteriores, presentan los fitocromos (PHYs), que son fotoreceptores capaces de recibir luz roja o luz roja lejana, lo que ha permitido entre otras respuestas la competición con plantas vecinas por la luz roja fotosintéticamente activa, proceso esencial en organismos sésiles (Rockwell et al., 2014). Además, se ha sugerido que un mayor número de copias de fitocromos con distintas funciones se relaciona con una mejora de la adaptación. *Arabidopsis* posee 5 fitocromos (PHYs A-E) con funciones distintas y redundantes (Sharrock y Quail, 1989),

la luz azul y UV-B es percibida por dos PHOTs, tres CRYs y las proteínas ZTL/FKF1/LKP2. Por último, una proteína UVR8 media la señalización de la luz UV-B (Figura 1). Excelentes revisiones han sido publicadas sobre la percepción lumínica y la variedad de fotoreceptores existentes en *Arabidopsis* (Fankhauser y Staiger, 2002; Li y Mathews, 2016; Serrano-Bueno et al., 2017; Sanchez et al., 2020).

La estructura y función de estos fotoreceptores ha sufrido una importante evolución a lo largo del linaje vegetal haciendo frente a grandes cambios ambientales tales como la transición de las plantas al medio terrestre.

1.2 El reloj circadiano en plantas y algas

Los organismos vivos han adaptado sus procesos fisiológicos esenciales a los ciclos de luz y oscuridad provocados por la rotación de la tierra. Así han desarrollado un reloj interno que les permite anticipar el ciclo solar y sincronizar diversos procesos fisiológicos. Estos ritmos circadianos (del latín, *circa dies*) se mantienen bajo condiciones constantes con un periodo de aproximadamente 24 h y están ampliamente extendidos en la naturaleza, presentes en casi todos los eucariotas. A pesar de la gran distancia evolutiva entre los diferentes fitoeucariotas, el mecanismo básico del reloj está compuesto por un sistema de bucles de retroalimentación conectados.

El reloj circadiano puede considerarse como un sistema de tres componentes: Un *input* o entrada que sincroniza a un oscilador central y *outputs* que conectan el oscilador con diversos procesos fisiológicos, metabólicos y de desarrollo (Figura 2). Sin embargo, en los últimos años se ha ido abandonando este concepto lineal del reloj por un sistema mucho más complejo en el que procesos regulados por el reloj tienen influencia sobre éste (McClung et al., 2002), con bucles de retroalimentación.

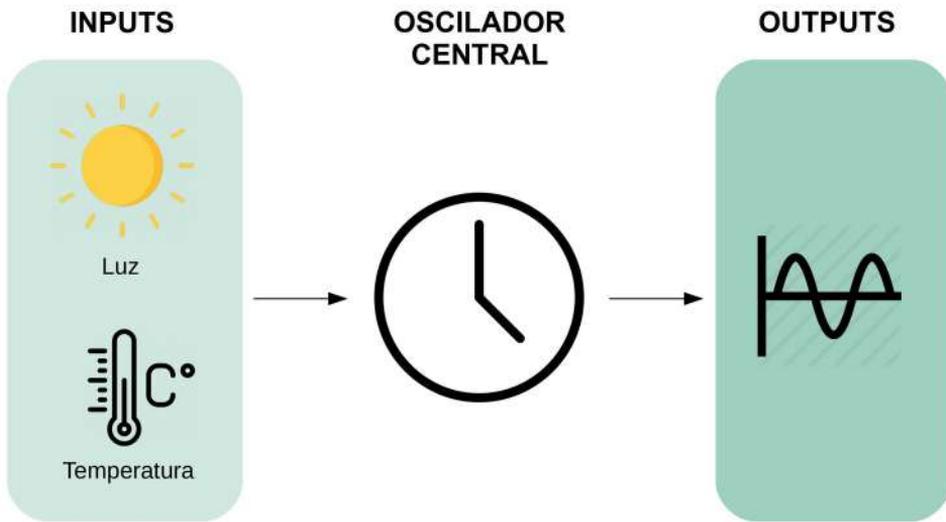


Figura 2. Esquema del concepto clásico del oscilador o reloj circadiano. Existe un oscilador central que se sincroniza con estímulos externos cíclicos o *inputs* como la luz o la temperatura. Este oscilador genera pulsos que permiten sincronizar y anticipar procesos fisiológicos.

En eucariotas, el reloj circadiano celular se basa principalmente en bucles de retroalimentación transcripcionales y postraduccionales donde las proteínas regulan su propia expresión y la de sus genes diana (Dunlap, 1999). En plantas y algas, el reloj circadiano regula gran parte del transcriptoma (Michael y McClung, 2003; de Los Reyes et al., 2017).

En los últimos años se han usado aproximaciones genómicas, bioquímicas y computacionales para avanzar en la comprensión de la estructura molecular del reloj circadiano en plantas. La mayoría de los componentes del reloj son represores transcripcionales que están interconectados a lo largo del día constituyendo bucles de retroalimentación. En *Arabidopsis* se han identificado tres bucles transcripcionales, el bucle central, el de la mañana y el de la tarde (McClung, 2014) (Figura 3). El bucle de la mañana se compone principalmente por los factores de transcripción tipo MYB CCA1/LHY (CIRCADIAN CLOCK ASSOCIATED 1 y LATE ELONGATED HYPOCOTYL) y los PSEUDO RESPONSE REGULATORS 5/7/9 (PRR5/7/9). Durante el amanecer, CCA1 y LHY reprimen la expresión génica de genes que se expresan por la tarde a través de la unión al motivo de DNA *Evening Element* (EE). Así CCA1 y

LHY actúan como represores transcripcionales de los PRRs5/7/9 (Kamioka et al., 2016). Además, se ha descrito que una proteína kinasa CK2 tiene un papel en el reloj interfiriendo en la unión de CCA1 a los promotores de los genes regulados (Portolés y Más, 2010). Posteriormente, desde por la mañana hasta la primera parte de la noche PRR9, PRR7 y PRR5 secuencialmente reprimen la transcripción de *CCA1* y *LHY* a través del motivo de DNA caja G (*G-box*; CACGTG) (Nakamichi et al., 2010, 2012; Liu et al., 2016) completándose el bucle de retroalimentación negativa.

Por la noche, TOC1 (TIMING OF CAB1), también conocido como PRR1, se acumula y participa en la represión de *CCA1* y *LHY* uniéndose a G-boxes en sus promotores (Gendron et al., 2012; Huang et al., 2012). También reprime otros genes del reloj que se expresan desde por la mañana hasta por la tarde (*CCA1*, *LHY*, *PRR9*, *PRR7*, *TOC1*, *ELF4*, y *LUX*). La familia de los PRRs se une a las G-boxes a través de los factores de transcripción PIFs (PHYTOCHROME-INTERACTING FACTORS) especialmente en condiciones de día corto (SD, ciclo de 16 horas de oscuridad y 8 de luz) (Soy et al., 2016; Martín et al., 2018). Recientemente se ha descubierto que el factor de transcripción TCP denominado CHE (*CCA1* HIKING EXPEDITION) actúa como represor de *CCA1*, que a su vez reprime la expresión de *CHE* (Pruneda-Paz et al., 2009). Estas proteínas constituyen el bucle central.

En el bucle de la tarde, GIGANTEA (GI) regula positivamente a TOC1. Inicialmente se pensó que esto lo hacía junto a un factor desconocido Y (Locke et al., 2006; Pokhilko et al., 2010; Locke et al., 2005) pero más tarde se describió al denominado *Evening Complex*, revelándose la conexión desconocida entre GI y TOC1 (Pokhilko et al., 2012; Nusinow et al., 2011). Este complejo está formado por las proteínas LUX ARRHYTHMO (LUX) y EARLY FLOWERING 3 y 4 (ELF3 y ELF4). La activación de su expresión por la tarde se debe al descenso de los niveles de *CCA1* y *LHY*. Estas proteínas forman un complejo que reprimen la expresión de *PRR9* y *LUX* uniéndose a sitios de unión de LUX en sus promotores (Dixon et al., 2011; Helfer et al., 2011; Nusinow et al., 2011; Herrero et al., 2012).

Por otra parte, al final de la noche TOC1 y PRR5 son gradualmente degradados por la E3 ubiquitín ligasa ZTL, lo que permite la acumulación de nuevo de

Introducción

CCA1/LHY. GI estabiliza a ZTL en presencia de luz azul, y este complejo marca a TOC1 para su degradación (Kim et al., 2007). Asimismo, PRR3 y PRR5 interaccionan con ZTL, lo que impide su unión a TOC1, previniendo a esta última de la degradación y resultando en una mayor amplitud en la expresión cíclica (Para et al., 2007; Fujiwara et al., 2008; Kiba et al., 2007).

Alternativamente a todos estos bucles de retroalimentación negativa, existen también activadores que intervienen en los principales bucles del reloj. LIGHT REGULATED 1 y 2 (LWD1/LWD2) son dos proteínas que participan en la señalización lumínica y que regulan el oscilador central activando a *PRR9*, *PRR5* y *TOC1*. Además, *PRR9* y *PRR7* regulan positivamente los niveles de *LWD1* y *LWD2*, por lo que estas proteínas constituyen un bucle de retroalimentación positiva (Wu et al., 2008; Wang et al., 2011c).

Se ha descrito que el factor de transcripción MYB REVEILLE 8 (RVE8) tiene un papel activador en el reloj circadiano uniéndose al motivo EE. Esta proteína se acumula varias horas después del amanecer e induce la expresión de genes de la mañana y de la tarde como *PRR5*, *TOC1*, *GI*, *ELF4*, y *LUX*. A su vez, la transcripción de *RVE8* está inhibida por *PRR5* (Rawat et al., 2011).

Además, dos miembros de la familia LNK (NIGHT LIGHT-INDUCIBLE AND CLOCK-REGULATED), *LNK1* y *LNK2*, integran señales lumínicas al núcleo del oscilador (Rugnone et al., 2013). De forma similar a *RVE8*, también participan en la activación de genes de la mañana y la tarde en el reloj, concretamente *PRR5*, *ELF4*, y *FKF1* (Xie et al., 2014). *RVE8* interacciona con estas dos proteínas por lo que actuarían como co-activadores. Sorprendentemente, los LNKs actúan como represores en los promotores de *CCA1*, *LHY* y *LNKs* para inhibir la vía de señalización de las antocianinas (Pérez-García et al., 2015). Todas estas complejas relaciones que incluyen varios *loops* de retroalimentación pueden observarse en la Figura 3

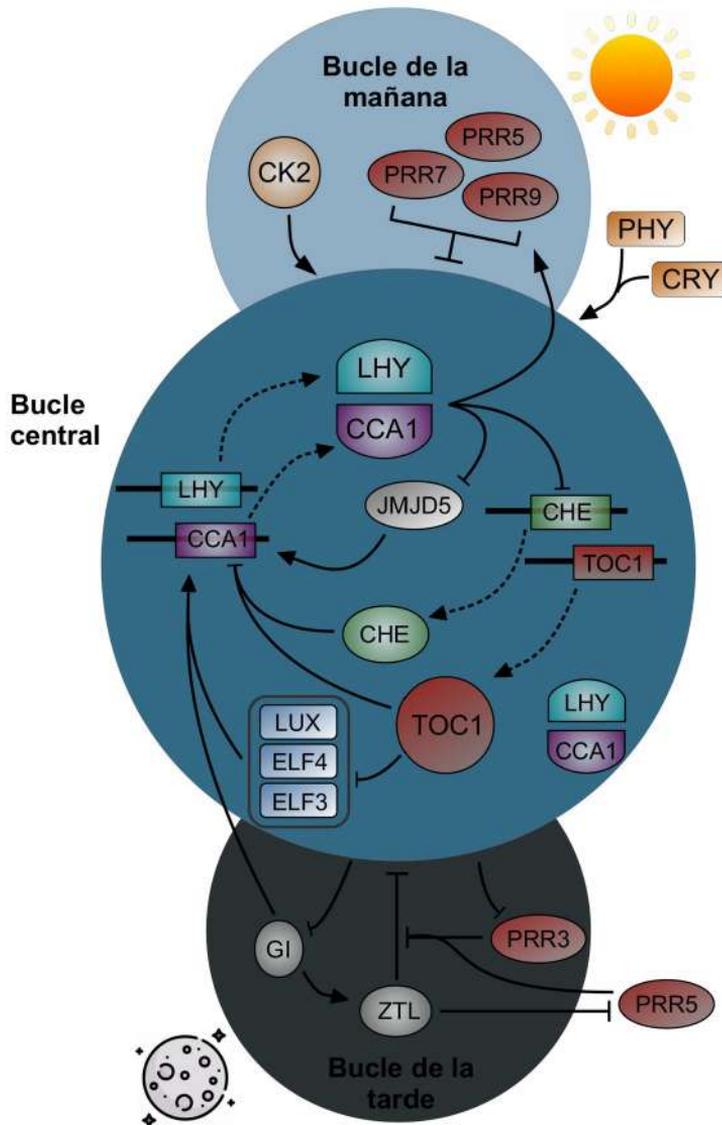


Figura 3. El reloj circadiano en *Arabidopsis thaliana* está compuesto principalmente por tres bucles transcripcionales. El bucle central consiste en los factores de transcripción tipo MYB CCA1/LHY que reprimen la expresión de *TOC1* (*PRR1*) y éste a su vez por la noche inhibe la expresión de *CCA1/LHY*. El bucle de la mañana lo componen *PRR5/7/9* y *CCA1/LHY*. Por la mañana, *CCA1* y *LHY* inhiben la expresión de los genes *PRR5/7/9* y por la tarde éstos se activan y reducen los niveles de *CCA1/LHY*. En el bucle de la tarde, *GI* junto al *Evening Complex* (*ELF3*, *ELF4* y *LUX*) regulan positivamente a *TOC1*. Al final de la noche *TOC1* y *PRR5* son degradados por *ZTL*, lo que permite de nuevo la acumulación de *CCA1/LHY*. (Adaptado de Chen, 2013).

Introducción

La importancia del reloj circadiano en plantas reside en que controla multitud de facetas de la fisiología y el desarrollo, permitiendo anticipar eventos periódicos. Los factores de transcripción que componen el reloj, además de reprimir transcripcionalmente a otros componentes del oscilador, controlan la expresión de multitud de genes salida regulando así el ritmo circadiano de procesos esenciales. Algunos procesos regulados en *Arabidopsis* por el reloj son la fotosíntesis, la apertura y cierre de estomas, la elongación del hipocotilo, el movimiento de hojas, la fosforilación de proteínas o la floración (Barak et al., 2000).

La multicelularidad de las plantas superiores, unido a la redundancia génica, complica el estudio del reloj circadiano, pero modelos unicelulares más simples han surgido como herramientas útiles para el estudio de estos sistemas. *Chlamydomonas* se ha establecido en los últimos años como el organismo modelo fotosintético unicelular, sin embargo el tamaño de su genoma es similar al de *Arabidopsis* lo que complica aproximaciones de genética reversa (Cerutti et al., 1997; León y Fernández, 2007). Por otra parte, varios genomas de prasinofíceas han sido secuenciados, entre ellos el de la microalga *Ostreococcus*, considerado el eucariota de vida libre más pequeño del mundo (Courties et al., 1994). *Ostreococcus* es una microalga no flagelada que posee un genoma compacto con pocas regiones repetitivas y no codificantes.

En *Ostreococcus* se han identificado dos homólogos de genes centrales del reloj de plantas superiores, *TOC1-like* y *CCA1-like*, en adelante *osttaTOC1* y *osttaCCA1* (Figura 4). Se han observado características típicas de componente del reloj en *osttaTOC1*, ya que participa en su propia regulación y en la de *osttaCCA1*, y controla un gen salida que se expresa rítmicamente (*CAB/LHCB*). Por su parte, *osttaCCA1* se une a un elemento EE esencial para la regulación del promotor de *osttaTOC1* (Corellou et al., 2009).

En *Chlamydomonas*, el reloj circadiano controla multitud de procesos clave para el desarrollo y la fisiología del alga, como la fototaxis, la quimiotaxis, la división celular, adhesión a superficies, contenido en almidón, sensibilidad a luz UV y

metabolismo del nitrógeno (BRUCE, 1970; Byrne et al., 1992; Goto y Johnson, 1995; Straley y Bruce, 1979; Ral et al., 2006; Nikaido y Johnson, 2007; Iliev et al., 2006).

En un análisis genético se identificaron en torno a 30 *loci* como posibles reguladores del reloj circadiano en *Chlamydomonas* denominados “*RHYTHM OF CHLOROPLAST*” (ROC) debido a que los mutantes sufrían una alteración del ritmo circadiano en la bioluminiscencia del cloroplasto. Finalmente la mutación en seis genes, *ROC15*, *ROC40*, *ROC55*, *ROC66*, *ROC75*, y *ROC114* causó defectos severos en el ritmo circadiano de varios procesos y sus mutaciones fueron complementadas con el fragmento de DNA genómico del propio gen (Matsuo et al., 2008).

La expresión de los genes *ROC15*, *ROC40*, *ROC66* y *ROC75* oscila cíclicamente con diferentes fases, por lo que aparentemente estas proteínas forman bucles transcripcionales de retroalimentación. Por otra parte, *CHLAMY1* es una proteína de unión a RNA que se une a las regiones 3' UTR de varios RNA mensajeros, especialmente en los que codifican proteínas implicadas en metabolismo del carbono y del nitrógeno (Waltenberger et al., 2014). Está formada por dos subunidades, C1 y C3 y forma parte del mecanismo central del reloj en *Chlamydomonas* actuando como represor traduccional (Iliev et al., 2006).

Aunque las secuencias de aminoácidos de las proteínas de *Chlamydomonas* suelen ser más largas que las de *Arabidopsis*, cuatro de estos factores de transcripción (*ROC15*, *ROC40*, *ROC66*, y *ROC75*) muestran homología con proteínas de *Arabidopsis* implicadas en el reloj circadiano, específicamente en los dominios de unión al DNA. *ROC15* y *ROC75* presentan un dominio GARP también presente en *LUX* y el dominio único MYB que aparece en los factores de transcripción *CCA1* y *LHY* es similar al que posee *ROC40* (Figura 4) (Matsuo et al., 2008; Onai y Ishiura, 2005; Hazen et al., 2005; Wang et al., 1997).

Por otra parte, *ROC66* es similar a *CONSTANS* (CO) y a algunos CO-LIKEs (*COL1* y *COL9*) ya que posee un dominio CCT similar. CO es el gen central de la vía fotoperiódica y se ha demostrado que algunos COLs participan en el reloj (Figura 4) (Ledger et al., 2001).

Introducción

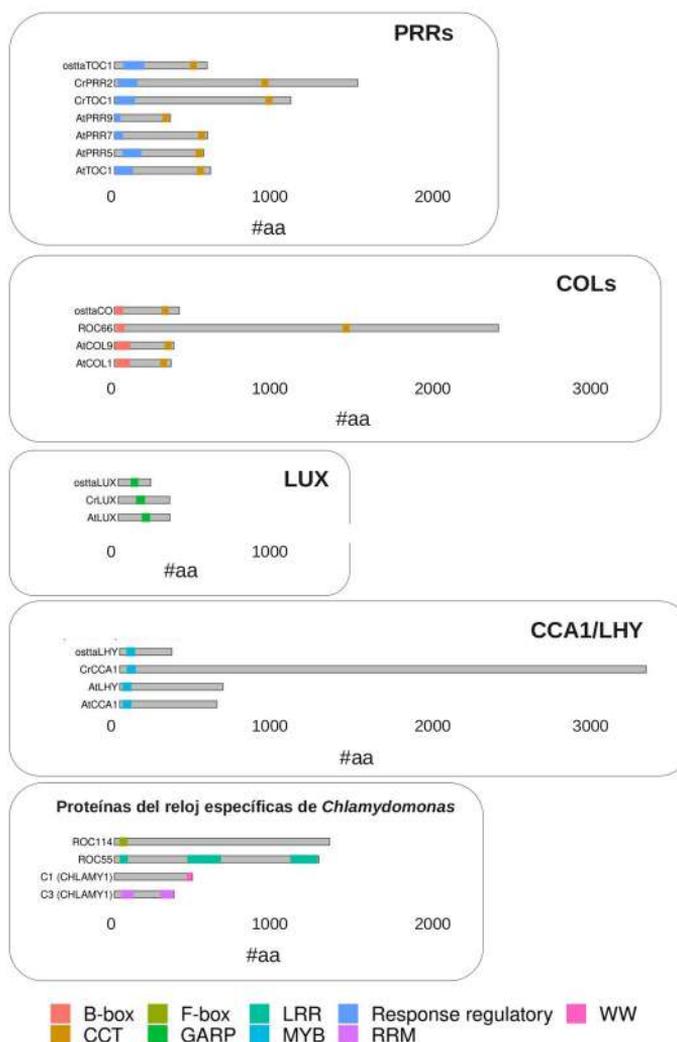


Figura 4. Componentes del reloj circadiano en *Arabidopsis*, *Chlamydomonas* y *Ostreococcus*. Se muestran las proteínas del reloj de cada organismo agrupadas por familias. En *Ostreococcus* se han descrito dos homólogos de TOC1 (ostta13g01890, osttaTOC1) y de CCA1/LHY (ostta06g01320, osttaLHY) con alta conservación entre los dominios PRR, CCT y MYB. Se ha descrito también una proteína con dominio GARP similar al que posee LUX. En *Chlamydomonas* se observa una mayor complejidad en el oscilador central. El dominio único MYB de los factores de transcripción CCA1/LHY es similar al que aparece en CrCCA1 (Cre12.g514400). Encontramos dos miembros de la familia PRR, CrTOC1 y CrPRR2 (Cre02.g094150 y Cre16.g676421). Además, ROC66 (Cre06.g278200) codifica un factor de transcripción de la familia CO-like implicado en el reloj. También se ha descrito una proteína con un dominio GARP como el que aparece en LUX. Por otra parte, hay proteínas que no existen en otros organismos modelos vegetales, como ROC55, ROC114 y CHLAMY1.

En otros casos la homología puede ser menor, por ejemplo se encontraron en *Chlamydomonas* dos homólogos de TOC1, CrPRR1 y CrPRR2, pero son distintos al TOC1 de angiospermas y aún no han sido estudiados en profundidad (de Los Reyes et al., 2017; Satbhai et al., 2011).

Este microalga además posee elementos que no aparecen en el linaje verde; ROC55 Y ROC114 no presentan homólogos en *Arabidopsis* ni en otros organismos fotosintéticos modelo. Tampoco las subunidades C1 y C3 de CHLAMY1 se parecen a ninguna otra proteína del reloj conocida de plantas, pero comparten homología con una proteína de rata presente en la zona del cerebro donde reside el reloj de mamíferos (Zhao et al., 2004).

El análisis comparativo del reloj en el linaje verde parece mostrar que la similitud entre los relojes de *Arabidopsis*, *Chlamydomonas* y *Ostreococcus* responde a que poseen un ancestro común, pero cada organismo ha evolucionado hasta alcanzar características específicas, como se muestra en la Figura 4.

1.3 La vía fotoperiódica

La vía del fotoperiodo regula procesos fisiológicos y de desarrollo en respuesta a la longitud del día. Este sistema involucra una detección de la luz a través de fotorreceptores y una medida del tiempo llevada a cabo por el reloj circadiano. En *Arabidopsis* la vía fotoperiódica es uno de los principales factores que regulan la floración, donde CO juega un papel central activando la expresión del florigeno *FT* (*FLOWERING LOCUS T*). La familia de factores de transcripción DOFs (DNA-BINDING WITH ONE FINGER) actúa como *output* directo del reloj, restringiendo la expresión de *FT* al momento correcto del día.

El módulo CO-FT se encuentra ampliamente distribuido en plantas y constituye uno de los elementos activadores de la floración más conservados (Böhlenius et al., 2006). La familia de homólogos de CO o *CONSTANS-like* evolucionó rápidamente y se encuentra en plantas y algas (Lagercrantz y Axelsson, 2000).

En *Chlamydomonas*, procesos fundamentales como el crecimiento, la división celular o la biosíntesis de lípidos está regulada por el fotoperiodo, reforzando al reloj

Introducción

circadiano, especialmente en condiciones de baja intensidad de luz (Serrano et al., 2009; Deng et al., 2015). La expresión de *CrCO* (*Cre06.g278159*) es fotoperiódica y regulada por el reloj (Matsuo et al., 2008). Además, se ha comprobado que es capaz de complementar la mutación de *co* en *Arabidopsis* (Serrano et al., 2009).

También se ha identificado una conservación en el módulo DOF-CO en *Chlamydomonas*; en día corto, CrDOF activa la expresión de *CrCO* uniéndose a su promotor, mientras que en día largo (LD, ciclo de 16 horas de luz y 8 de oscuridad) inhibe la expresión de genes implicados en la división celular (Lucas-Reina et al., 2015).

Por tanto, la respuesta a la longitud del día o respuesta fotoperiódica surgió hace millones de años permitiendo a los primeros organismos fotosintéticos predecir la variabilidad estacional y ha evolucionado hasta la compleja vía molecular que gobierna la transición floral en angiospermas.

2 Floración

Una de las decisiones clave durante el desarrollo vegetal es la transición floral, en la que se producen cambios moleculares y metabólicos que inducen el paso del estado vegetativo al reproductivo (Piñeiro y Jarillo, 2013). Esta transición está extremadamente regulada por muchos factores, ya que florecer en el momento correcto garantiza el éxito reproductivo.

Se han identificado en *Arabidopsis* cuatro vías principales que promueven la transición floral y por tanto el tiempo de floración, la vía fotoperiódica, la vía autónoma, la vernalización y la vía de las giberelinas. Todas ellas convergen en los integradores florales SUPPRESSOR OF OVEREXPRESSION OF CO 1 (SOC1) y FT (Figura 5).

La vernalización es la promoción de la floración debido a una exposición prolongada a bajas temperaturas, lo que reduce los niveles de expresión de *FLC* (*FLOWERING LOCUS C*), que es un represor floral que inhibe la transcripción de *FT* (Michaels y Amasino, 1999). Por otra parte, los componentes de la vía autónoma regulan el tiempo de floración independientemente de la longitud del día u otros estímulos externos (Simpson, 2004). Por último, varias vías hormonales contribuyen a

la plasticidad del proceso de floración, entre las que destaca la vía de las giberelinas. Esta hormona está conectada con otras vías que influyen en la floración a través de las proteínas DELLA (Blazquez et al., 1998; Wang et al., 2016; Conti, 2017).

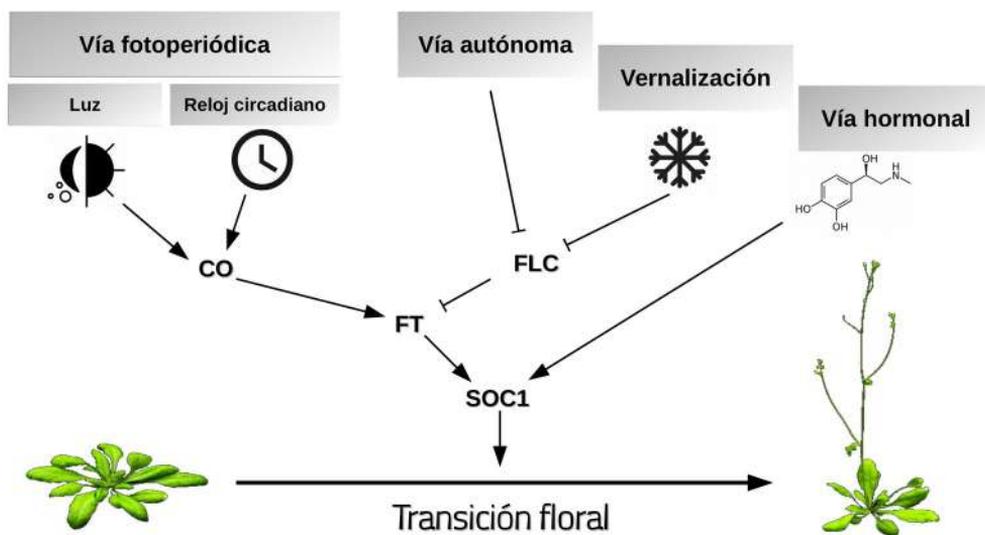


Figura 5. Modelo simplificado de las vías principales que regulan el tiempo de floración en *Arabidopsis thaliana*. La vía fotoperiódica induce la expresión de CO en LD, cuyos niveles están regulados por la luz y por el reloj circadiano. La vía autónoma regula negativamente la expresión del represor floral FLC, que también se inhibe independientemente por la vía de la vernalización. Finalmente, la vía hormonal contribuye a la floración, destacando la vía de las giberelinas.

La vía fotoperiódica es uno de los principales agentes que controlan la floración y está regulada por sensores lumínicos, así como por el reloj circadiano. La percepción del fotoperiodo permite a las plantas, entre otras cosas, predecir cambios estacionales y adaptarse a diferentes latitudes.

2.1 Regulación de la función de CO

El gen central de esta vía en *Arabidopsis* es *CONSTANS* que actúa como integrador de ambas señales. CO codifica una proteína que presenta dos dominios altamente conservados y un dominio intermedio más variable. CO contiene dos cajas B (B-boxes) localizadas en el extremo amino que participan en interacciones proteína-proteína y un dominio CCT (CONSTANS, CO LIKE, TIMING OF CAB1) situado en el

Introducción

extremo carboxilo. Se ha demostrado que el dominio CCT está implicado en la interacción con otras proteínas y con el DNA, además de actuar como señal transportadora al núcleo (Laubinger et al., 2006). CO es el regulador central de la vía fotoperiódica y activa la expresión del florigeno *FT*, que dispara la diferenciación floral (Valverde, 2011).

Para explicar la floración por fotoperiodo, se ha postulado el modelo de la coincidencia externa en el que se induce la floración fotoperiódica cuando coinciden altos niveles del transcrito de *CO* (mediado por el reloj circadiano) con la estabilización de la proteína por la luz, únicamente en condiciones de LD (Suárez-López et al., 2001; Yanovsky y Kay, 2002; Valverde et al., 2004). Por tanto, la expresión génica de *CO* y la estabilidad de la proteína están finamente reguladas por el reloj circadiano y señales lumínicas (Putterill et al., 1995; Samach et al., 2000; Suárez-López et al., 2001; Mizoguchi et al., 2005).

El complejo formado por GI y FKF1 participa en este control regulando la transcripción de *CO* a través de la unión a su promotor al final de la tarde e induciendo la degradación de los CDFs, que actúan como represores de la expresión de *CO* (Imaizumi et al., 2005; Sawa et al., 2007). Las proteínas FBH1, FBH2, FBH3 y FBH4 (FLOWERING BHLHs) son también activadores de *CO* uniéndose a elementos E-box en su promotor, además, los factores de transcripción TCPs (microRNA319-regulated TEOSINTE BRANCHED/CYCLOIDEA/PCF) interaccionan con los FBHs y también se unen al promotor de *CO* para activar su expresión (Liu et al., 2017). Por otra parte, se ha demostrado que los factores de transcripción PRR9, PRR7 y PRR5 inducen indirectamente la floración fotoperiódica, ya que reprimen la expresión de *CDF1* (Nakamichi et al., 2007).

Aunque el transcrito de *CO* se expresa a altos niveles desde la tarde hasta el amanecer, se ha descrito que en términos de estabilidad de la proteína, *CO* se degrada en la oscuridad y se estabiliza durante el día, mediado por diferentes fotorreceptores y E3 ubiquitín ligasas (Valverde et al., 2004). Durante las primeras horas del día, la ubiquitín ligasa HOS1 (HIGH EXPRESSION OF OSMOTICALLY RESPONSIVE GENES1) interacciona con *CO* y lo marca para ser degradado,

inhibiendo la activación de *FT* en el momento incorrecto del día y previniendo una floración precoz en *Arabidopsis* (Lazaro et al., 2012). Por la noche, COP1 y SPA promueven la degradación de CO (Laubinger et al., 2006; Jang et al., 2008).

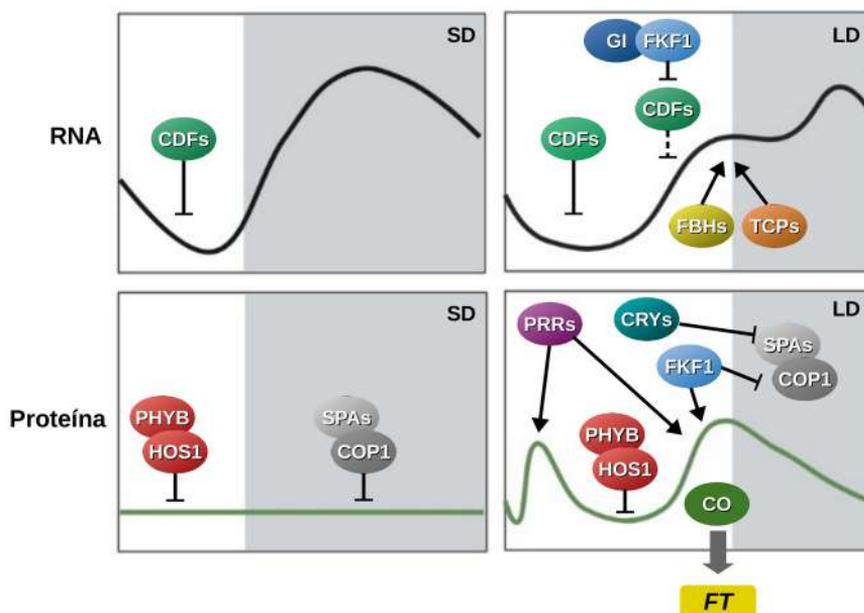


Figura 6. Modelo de coincidencia externa por el cual CO induce a *FT* únicamente en condiciones de día largo. La floración fotoperiódica en *Arabidopsis* está mediada por el control transcripcional y postraduccional de CO. En día corto, su expresión está reprimida por los CDFs durante la mañana. Al final del día los niveles del mRNA de CO empiezan a subir para alcanzar el máximo durante la noche, sin embargo, no se traduce en niveles de proteína puesto que durante el día PHYB y HOS1 promueven la degradación de CO y por la noche lo hacen SPA y COP1. De esta forma se previene la floración en día corto. En cambio, en LD los CDFs son degradados durante la tarde por GI/FKF1 en respuesta a luz azul, y en ese momento FBHs y TCPs inducen la expresión de CO. Este aumento de expresión durante la tarde en LD permite la acumulación de la proteína puesto que CO se estabiliza por FKF1 y los PRRs. Además, los CRYs y FKF1 promueven la degradación de COP1 y SPA. Por tanto, sólo cuando hay altos niveles de expresión génica (regulada principalmente por el reloj circadiano) en una ventana de luz que permite la estabilidad de la proteína se promueve la transcripción de *FT* (Adaptado de Shim y Jang, 2020).

El patrón diurno de CO (Figura 6) se genera por las acciones redundantes de reguladores postraduccionales. Así, desde el amanecer, varios fotorreceptores comienzan a estabilizar a CO. En respuesta a la luz azul, CRY1 y CRY2 interaccionan

Introducción

con COP1/SPA1 para evitar la degradación de CO por este complejo (Yanovsky y Kay, 2002; Liu et al., 2008; Zuo et al., 2011). Asimismo, durante el día, dos fitocromos participan en la estabilidad de CO de diferentes formas: En respuesta a la luz roja, PHYB interacciona con HOS1 promoviendo la degradación de CO, pero por la tarde, PHYA consigue disgregar el complejo CO1-SPA1 estabilizando así a CO en presencia de luz roja lejana (Lazaro et al., 2012; Endo et al., 2013; Sheerin et al., 2015; Hajdu et al., 2015).

Además de las vías de regulación de fotorreceptores descritas, existen proteínas del reloj que regulan la acumulación de CO en respuesta a la luz azul. Durante la mañana, ZTL se asocia con GI para inhibir la estabilidad de CO (Song et al., 2014), pero por la tarde, la luz azul aumenta los niveles de CO a través de su interacción con FKF1 (Song et al., 2012). De hecho, se ha descrito que FKF1, GI y CO podrían formar un complejo que estabiliza a CO (Song et al., 2014). Además, recientemente se ha demostrado que los PRRs estabilizan a CO y promueven la floración en respuesta al fotoperiodo (Hayama et al., 2017).

Al igual que la estabilidad de la proteína, su papel como regulador transcripcional está afectado por diferentes interactores (Figura 6). Las proteínas DELLA, la inmunofilina FKBP12, la proteína RING-DOMAIN BOI, las B-boxes BBX19 y miP1a y el factor remodelador de la cromatina PICKLE pueden regular la función transcripcional de CO por diferentes mecanismos (Xu et al., 2016; Wang et al., 2016; Serrano-Bueno et al., 2019; Nguyen et al., 2015; Wang et al., 2014; Graeff et al., 2016; Jing et al., 2019).

En conjunto, la función de CO está altamente regulada por multitud de factores, lo que contribuye a la precisión en la transcripción de *FT* y por tanto de la transición floral. En cuanto a la participación de CO en otros procesos biológicos, se ha descrito que puede modificar el metabolismo del almidón a través de la activación del gen *GBSS1* (*GRANULE BOUND STARCH SYNTHASE*) (Isabel Ortiz-Marchena et al., 2014) y que también podría estar implicado en la apertura de estomas (Ando et al., 2013). Sin embargo, poco se conoce sobre otros procesos regulados por CO.

Actualmente no existen pruebas de que CO ejerza un control sobre el reloj circadiano. Sin embargo, ya que presenta un dominio CCT en su extremo carboxilo terminal al igual que otras proteínas centrales del reloj como TOC1, PRR5, PRR7 o PRR9 (Nakamichi et al., 2010), se ha sugerido que podría tener una función reguladora sobre el reloj (Strayer et al., 2000). Siguiendo esta dirección, en este estudio se muestran evidencias de que CO, además de estar controlado por el reloj circadiano, también regula genes centrales de éste. Por tanto, los programas transcripcionales que controlan la transición floral y el reloj están interconectados constituyendo un sistema complejo de bucles de retroalimentación. En los últimos años, con el objetivo de descifrar este sistema, se han aplicado técnicas ómicas basadas en secuenciación de altas prestaciones. Sin embargo, estos datos están normalmente fragmentados por lo que se hace necesaria una aproximación basada en biología molecular de sistemas para integrarlos y obtener una comprensión global de las interacciones entre la vía fotoperiódica y el reloj circadiano en *Arabidopsis*.

3 Técnicas ómicas

3.1 Transcriptómica

Tras el desarrollo de la genómica, surgieron otras técnicas ómicas han surgido, como la metabolómica, la proteómica o la transcriptómica, siendo esta última ampliamente utilizada para medir la expresión génica. La transcriptómica estudia el transcriptoma, que se define como el conjunto de todas las moléculas de RNA o transcritos presentes en la célula en un momento concreto. La molécula de RNA transmite la información necesaria desde el DNA para sintetizar las proteínas, por tanto, el estudio del transcriptoma nos permite conocer la actividad de células, tejidos y organismos. Sin embargo, debido a la complejidad de la regulación postraduccional de las proteínas y sus diferentes tasas de degradación, la correlación entre la expresión de un transcrito y la cantidad de proteína no es siempre directa.

Para capturar el estado del transcriptoma se han desarrollado varias tecnologías que emplean principalmente dos aproximaciones: basada en hibridación (*microarrays*) y en secuenciación (secuenciación de RNA o RNAseq).

Introducción

La aparición de los *microarrays* transformó la biología molecular al permitir realizar investigaciones globales de la actividad celular y aunque tienen multitud de aplicaciones, se han usado predominantemente para medir niveles de expresión génica (Hoheisel, 2006). De forma general, un *microarray* de DNA consiste en un portaobjetos de vidrio o chip en el que se unen secuencias de DNA específicas en lugares determinados. Después, estas secuencias se usan como “sondas” que hibridarán con ácidos nucleicos marcados en la muestra. El apareamiento de las bases complementarias entre la muestra y las secuencias del *microarray* produce una cantidad de luz que se puede medir. Esta unión es usada para medir la concentración de dichas especies de ácidos nucleicos en la muestra (Schulze y Downward, 2001).

Al contrario que los *microarrays*, los métodos basados en secuenciación como el RNAseq determinan directamente la secuencia de cDNA. Además, también tiene ventajas frente a los métodos basados en etiquetas (SAGE, CAGE, MPSS), que usan la tecnología de secuenciación de Sanger y presentan problemas en el mapeo de las secuencias al genoma. La tecnología RNAseq usa métodos de secuenciación de altas prestaciones para secuenciar una parte o el RNA total presente en una muestra (Wang et al., 2009).

La metodología consiste en una primera extracción de RNA de la muestra y su conversión a una genoteca de cDNA. Después, cada molécula de cDNA se secuencia por uno de los extremos (*single-end*) o por ambos (*paired-end*). Más tarde, durante el análisis de los datos, estas secuencias cortas se mapean contra el genoma de referencia. Por tanto, además de conocer la abundancia de cada transcrito, es posible realizar análisis de maduración del RNA (*splicing*) alternativo.

El RNAseq ofrece algunas ventajas frente a tecnologías previas. En primer lugar, esta metodología no está limitada a secuencias conocidas del genoma, lo que permite secuenciar transcriptomas completos *de novo* (Barbazuk et al., 2007), conocer cómo están conectados los exones o descubrir variaciones de nucleótido simple (SNPs) en las regiones transcritas. Por otra parte, la secuenciación de RNA presenta un ruido (*background*) bajo, debido a que las secuencias de DNA difícilmente pueden mapear en varios sitios del genoma. Además, no existen problemas de saturación,

como ocurre en los *microarrays*, ofreciendo un intervalo de detección mucho más amplio que las tecnologías previas (Wang et al., 2009).

3.2 Cistrómica

La regulación de la expresión génica en eucariotas es compleja. Sobre el genoma actúan factores de transcripción que se unen en regiones cis (cistroma) y RNA polimerasas para generar RNAs mensajeros. Además, el transcriptoma está afectado en cada célula por el epigenoma, que consiste en el conjunto de moléculas y proteínas que marcan el ADN sin cambiar su secuencia pero que controlan la activación o desactivación de genes.

Las proteínas de unión al DNA pueden ser factores de transcripción u otras asociadas a la cromatina, como las histonas. Los factores de transcripción son proteínas que se unen a elementos reguladores del genoma (elementos cis), normalmente localizados cadena arriba de los genes que regulan, para modular su tasa de transcripción. Un factor de transcripción típico de plantas contiene una región de unión al DNA, un sitio de oligomerización, un dominio de activación de la transcripción y módulos señalizadores al núcleo. Las regiones de unión al DNA y de interacción con otras proteínas a veces están cercanas en la proteína y sus funciones pueden solaparse. En las células vegetales, algunos genes se expresan constitutivamente mientras que la transcripción de otros responde a estímulos específicos, todo ello orquestado por factores de transcripción, la interacción entre ellos con otras proteínas y su unión a secuencias reguladoras específicas (Liu et al., 1999)

El cistroma, por tanto, se define como el conjunto completo de sitios de unión de factores de transcripción. Se han desarrollado métodos para construir mapas del cistroma permitiendo generar redes transcripcionales que controlan procesos claves de desarrollo o de respuesta a diferentes estímulos. La exploración los sitios de unión de estas proteínas al genoma comenzó hace ya cuatro décadas con el desarrollo de los ensayos de cambio de movilidad electroforética o *EMSA*, por sus siglas en inglés (Garner y Revzin, 1981; Fried y Crothers, 1981). Posteriormente, la aparición de la

Introducción

técnica de inmunoprecipitación de la cromatina (ChIP; *Chromatin Immunoprecipitation*) permitió estudiar estas interacciones *in vivo*.

Los pasos básicos del ChIP son fijación, sonicación, inmunoprecipitación y análisis del DNA inmunoprecipitado (Figura 7). La fijación o *crosslinking* se realiza exponiendo a las células a formaldehído u otros fijadores, que unen la proteína covalentemente a la cromatina. Después, el DNA se fragmenta por sonicación o tratamientos enzimáticos para obtener fragmentos pequeños. Más tarde, el factor de transcripción unido a regiones del genoma específicas se inmunoprecipita usando anticuerpos nativos o contra una etiqueta fusionada a la proteína. Finalmente, el DNA se aísla y se analiza principalmente por diferentes métodos: amplificación de secuencias de interés mediante PCR o qPCR (ChIP-QPCR), hibridación en *microarray* (ChIP-on-chip) o secuenciación de altas prestaciones (ChIP-seq) (Das et al., 2018). Estas secuencias cortas se mapean contra el genoma de referencia y usando el control adecuado, es posible determinar sitios de unión enriquecidos o “picos” de unión.

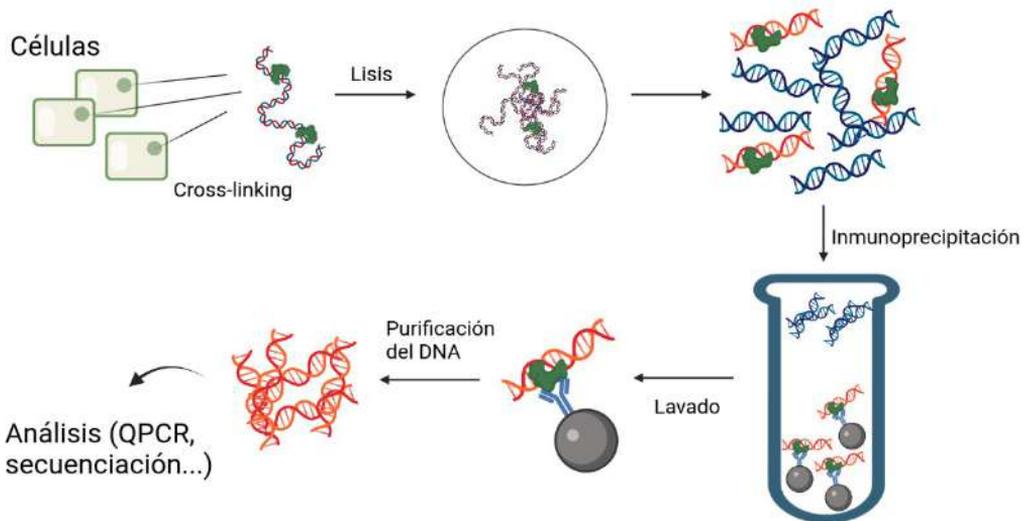


Figura 7. Protocolo esquematizado de la inmunoprecipitación de la cromatina (ChIP). Los principales son la fijación o *cross-linking*, la lisis de las células y sonicación del DNA, inmunoprecipitación, el lavado de los complejos anticuerpo-proteína y purificación del DNA.

La técnica de CHIP-seq ha sido ampliamente utilizada en biología vegetal y ha permitido estudiar a fondo la dinámica de la cromatina, el papel de numerosos factores de transcripción en la regulación génica, la metilación del DNA o modificaciones de histonas (Kaufmann et al., 2010).

4 Biología de sistemas de plantas

Los genes contienen la información necesaria para codificar proteínas, que en última instancia son las que llevan a cabo todos los procesos necesarios en los organismos vivos. Históricamente, se han estudiado estos genes y proteínas como elementos aislados, pero desde hace años se ha ido abandonando esa aproximación reduccionista para dar paso a una visión global de los sistemas vivos (Matus et al., 2019). Este cambio de paradigma ha venido apoyado por la aparición de las tecnologías de secuenciación masiva de nueva generación (NGS) que permiten disponer de grandes cantidades de datos de secuenciación de ácidos nucleicos (Lister et al., 2009). Estos datos masivos se almacenan en repositorios de libre acceso, y su análisis, integración e interpretación hace necesario el uso de recursos computacionales. Esto no es exclusivo de la biología, sino que ocurre en multitud de áreas científicas, promoviendo el desarrollo de la Ciencia de Datos o *Big Data*. Específicamente, en el campo de la biología molecular, esto ha llevado a la aparición de la Biología Molecular de Sistemas (Matus et al., 2019).

La biología de sistemas se define como una rama interdisciplinar de la ciencia centrada en el estudio de seres vivos desde una perspectiva global u holística que combina disciplinas como matemáticas, computación, estadística o biología (Figura 8). Busca describir o explicar fenómenos biológicos usando una visión global a través de las interacciones de todas las entidades moleculares, en lugar de tratarlas individualmente (Mesarović, 1968).

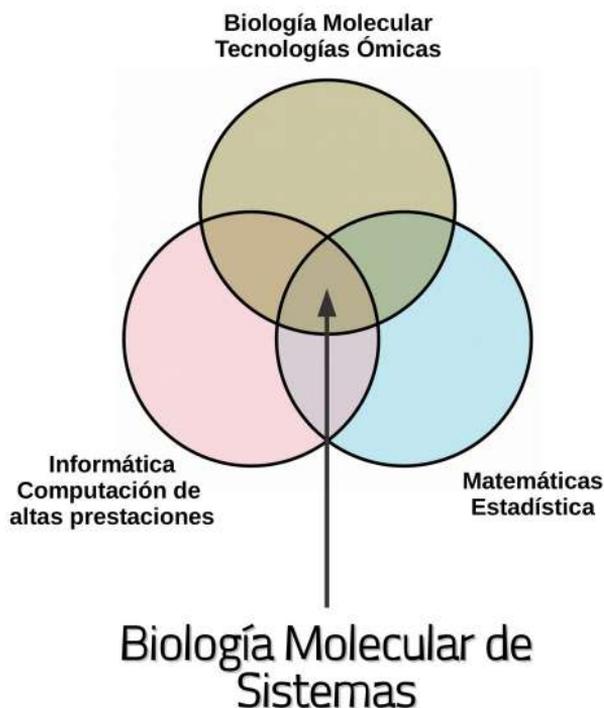


Figura 8. Representación esquemática de la Biología Molecular de Sistemas como una disciplina en la intersección entre computación, matemáticas y biología. Adaptado de (Matus et al., 2019).

La metodología típica en biología molecular de sistemas comienza con la generación de datos ómicos y su integración con datos existentes. Estos datos son analizados con métodos estadísticos y recursos computacionales, con el objetivo de generar modelos que permitan predecir fenotipos o respuestas en sistemas vivos (Sheth y Thaker, 2014). Este flujo de trabajo incluye el trabajo experimental usando técnicas de biología molecular con el fin de confirmar nuevas hipótesis (Figura 9).

Por tanto, esta aproximación busca la integración y no la disociación de elementos. De esta forma, es posible construir redes compuestas por genes, sus productos, las interacciones entre ellos y estudiar así las estructuras de diseño básicas de los seres vivos. A nivel subcelular es posible estudiar diferentes redes moleculares, como redes de coexpresión génica, redes transcripcionales, redes de interacción proteína-proteína o redes metabólicas. La ciencia de redes se ha

establecido como el paradigma central de esta disciplina, ya que ofrece un entorno de modelado eficaz (Li et al., 2015).

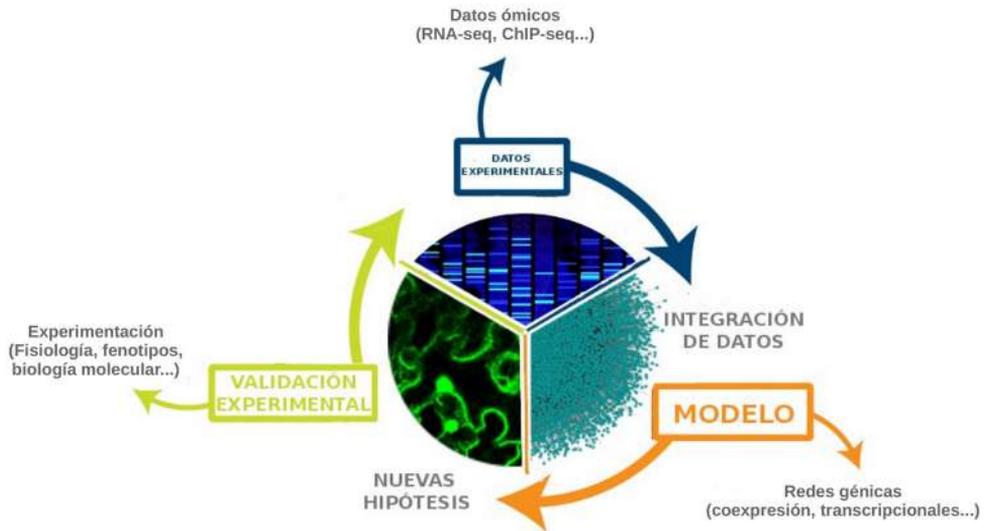


Figura 9. Flujo de trabajo típico en Biología Molecular de Sistemas. Los datos ómicos experimentales deben ser analizados e integrados para generar un modelo, que pueden estar compuestos por redes génicas. La integración de estos datos genera propiedades emergentes que dan lugar a nuevas hipótesis. Estas hipótesis son analizadas mediante trabajo experimental y eventualmente conducen la generación de nuevos datos.

4.1 Ciencia de redes

Estamos rodeados de sistemas complejos en nuestro día a día. Podemos tomar como ejemplos la red telefónica mundial, las redes sociales, o nuestro cerebro, formado por millones de neuronas. Todos ellos requieren de la cooperación de sus elementos para su correcto funcionamiento. Se denominan complejos debido a que no es posible entender su comportamiento a partir de la exploración individual de sus componentes, sino que existen propiedades que emergen de las interacciones entre ellos. Estos sistemas pueden ser capturados y estudiados como redes o grafos (Barabási, 2015).

Por tanto, para entender estos sistemas complejos es esencial explorar las redes que los gobiernan. Así nace a principios del siglo XXI la ciencia de redes, cuya

Introducción

base matemática es la teoría de grafos, que tiene su origen en 1735 cuando el matemático alemán Leonhard Euler resolvió el problema de los siete puentes de Königsberg (Euler, 1741). La ciencia de redes ha sufrido un rápido desarrollo debido a que poseemos las herramientas para generar, coleccionar y analizar grandes cantidades de datos correspondientes a cualquier sistema. Es posible generar redes que capturen cualquier sistema biológico o no, como la red de co-citas de *Nature* (Figura 10). Esta aproximación se define como interdisciplinaria, ya que ofrece un lenguaje universal sin importar qué tipo de red sea objeto de estudio, utiliza datos empíricos, es cuantitativa y tiene carácter computacional (Barabási, 2015).

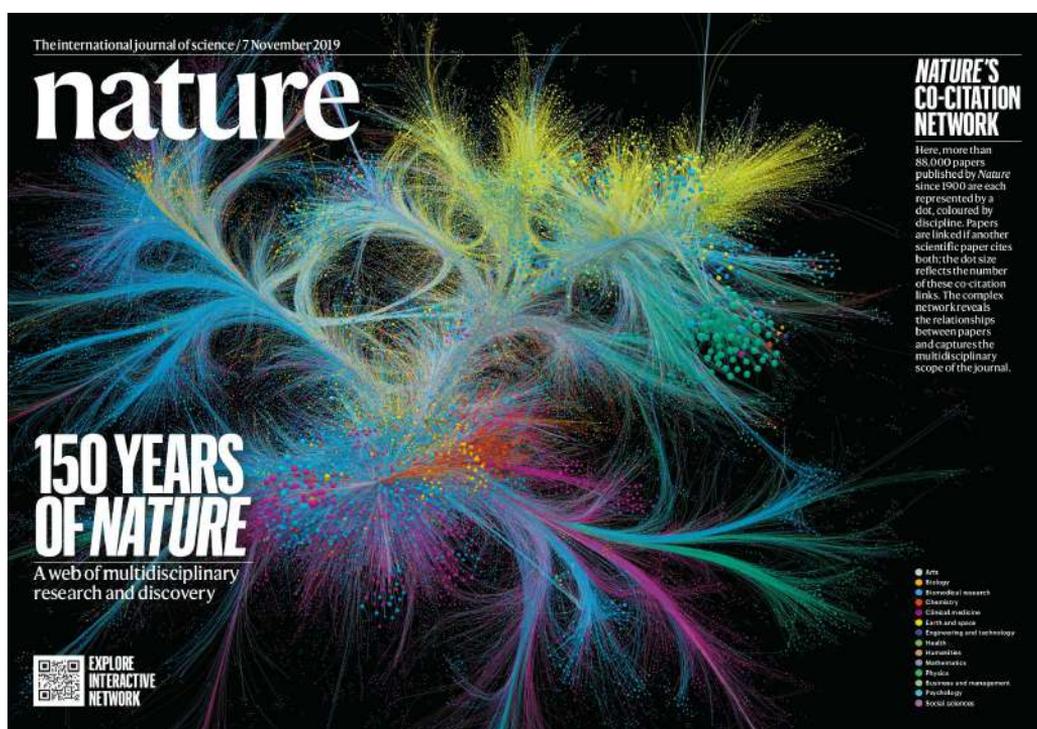


Figura 10. Red de co-citas de la revista científica *Nature*. Cada nodo de la red representa un artículo publicado en *Nature*. Dos artículos se enlazan con una arista si otro cita a ambos. El tamaño del nodo representa el número de esas co-citas (Gates et al., 2019).

Una red se define como un conjunto de nodos que representan los elementos del sistema, unidos por aristas que constituyen las interacciones entre ellos. Dependiendo de si existe direccionalidad en las aristas, podemos encontrar redes

dirigidas y redes no dirigidas. Un parámetro importante a tener en cuenta es el grado de un nodo, que se define como el número de vecinos de dicho nodo.

Las redes biológicas son dinámicas por lo que se hace necesario estudiar su evolución en el tiempo, por cambios producidos debido a factores externos o internos, como ataques dirigidos o fallos aleatorios. En oposición a redes aleatorias, en las que la mayoría de nodos tienen un grado similar (Figura 11A), se ha descrito que la mayoría de las redes biológicas exhiben una estructura determinada no aleatoria, denominada propiedad libre de escala, en las cuales la distribución del grado de los nodos sigue una ley potencial. Esto significa que unos pocos nodos (denominados *hubs* o nodos concentradores) están altamente conectados mientras que el resto presentan pocas conexiones (Barabási y Oltvai, 2004). Esto las hace robustas frente a fallos aleatorios y frágiles a ataques dirigidos. Además, se puede observar una rápida propagación de la información dentro de la red debido a la existencia de *hubs* (Figura 11B).

Una característica común de las redes complejas es que dos nodos escogidos aleatoriamente pueden ser conectados por pocos pasos, son las denominadas redes de mundo pequeño. Esta propiedad ha sido observada en multitud de sistemas, como la red informática mundial (*World Wide Web*). Sin embargo, en las redes libre de escala, el camino que conecta dos nodos aleatorios es menor de lo que se predice para las redes de mundo pequeño, mostrando un efecto "*ultra-small world*" (Cohen y Havlin, 2003).

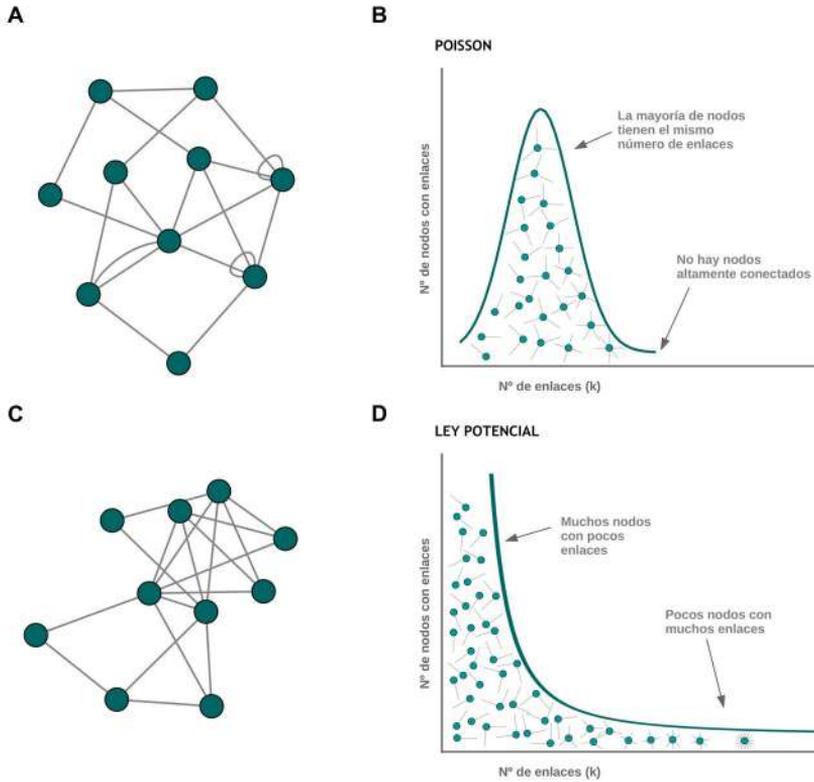


Figura 11. Comparación de redes aleatorias y redes libres de escala. (A) Representación gráfica de una red aleatoria. **(B)** La distribución del grado de los nodos de una red aleatoria sigue una distribución de Poisson, similar a una curva de Bell. Todos los nodos tienen un grado similar y no existen nodos altamente conectados. **(C)** Representación gráfica de una red libre de escala. **(D)** La distribución del grado en redes libres de escala se ajusta a una potencial negativa en la que la mayoría de los nodos tienen pocos enlaces. Sin embargo, todos estos nodos están estrechamente relacionados a través de *hubs* o nodos altamente conectados.

Las redes biológicas tienden a presentar un alto agrupamiento, lo que encaja con el hecho de que las funciones celulares se lleven a cabo de una forma modular (Hartwell et al., 1999). Un módulo o *cluster* es un grupo de moléculas (nodos) que están conectadas y trabajan juntos para realizar una función específica (Wall et al., 2004; Alon, 2003; Ravasz y Barabási, 2003). Esta propiedad se puede observar en redes de todo tipo, desde grupos de amigos en redes sociales o grupos de páginas web relacionadas dentro de la *World Wide Web*. En biología encontramos muchos

ejemplos de modularidad, determinados grupos de moléculas gobiernan procesos específicos como síntesis de ácidos nucleicos o degradación de proteínas. Además, también existe modularidad temporal, de manera que distintas fases del desarrollo están reguladas por diferentes conjuntos de moléculas.

Se ha observado que redes que realizan las mismas tareas o procesos muestran patrones recurrentes a pesar de ser evolutivamente distantes, estas interacciones entre elementos se denominan motivos de red. Técnicamente, un motivo de red es un subgrafo que aparece significativamente con más frecuencia en redes biológicas que en redes aleatorias. Motivos específicos aparecen en redes de desarrollo diferentes a los que se han descrito en redes de procesamiento de información. Por tanto, se ha sugerido que a pesar de su gran complejidad, las redes están gobernadas por elementos simples (Milo et al., 2002).

En redes transcripcionales, la interacción más básica es la regulación simple, que puede ser negativa o positiva. Si ocurre esto se traza una arista del gen Y al gen X, lo que significa que el factor de transcripción codificado por el gen Y se une al promotor del gen X. Un motivo de red que es típico de redes de desarrollo es la autorregulación, en la que un factor de transcripción regula la expresión de su propio gen. Una autorregulación negativa produce una aceleración de la respuesta en redes transcripcionales (Rosenfeld et al., 2002) y mantiene estables los niveles de proteínas entre diferentes células. En cambio, la autorregulación positiva disminuye la velocidad de respuesta, puede incrementar la variabilidad de proteína entre células y sirve de memoria para mantener la expresión génica después de una señal activadora. Si avanzamos en complejidad podemos encontrar los bucles de retroalimentación o *feedback loops*, y los bucles de anteroalimentación o *feedforward loops*, que debido a su dinámica permiten a los sistemas filtrar señales espurias o generar pulsos de expresión (Figura 12) (Alon, 2007a).

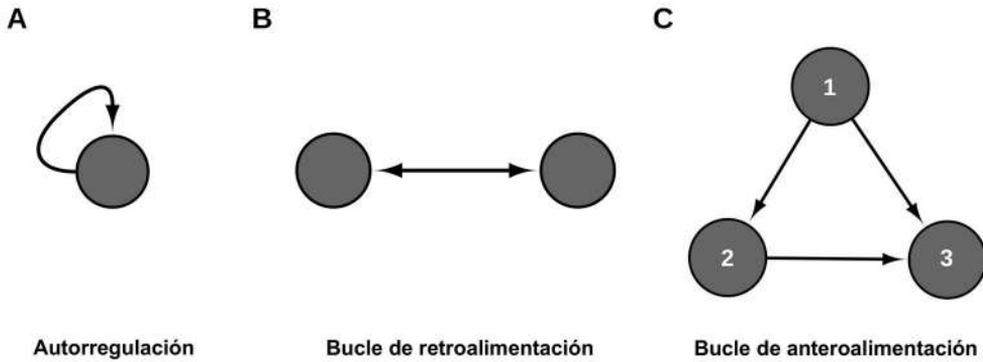
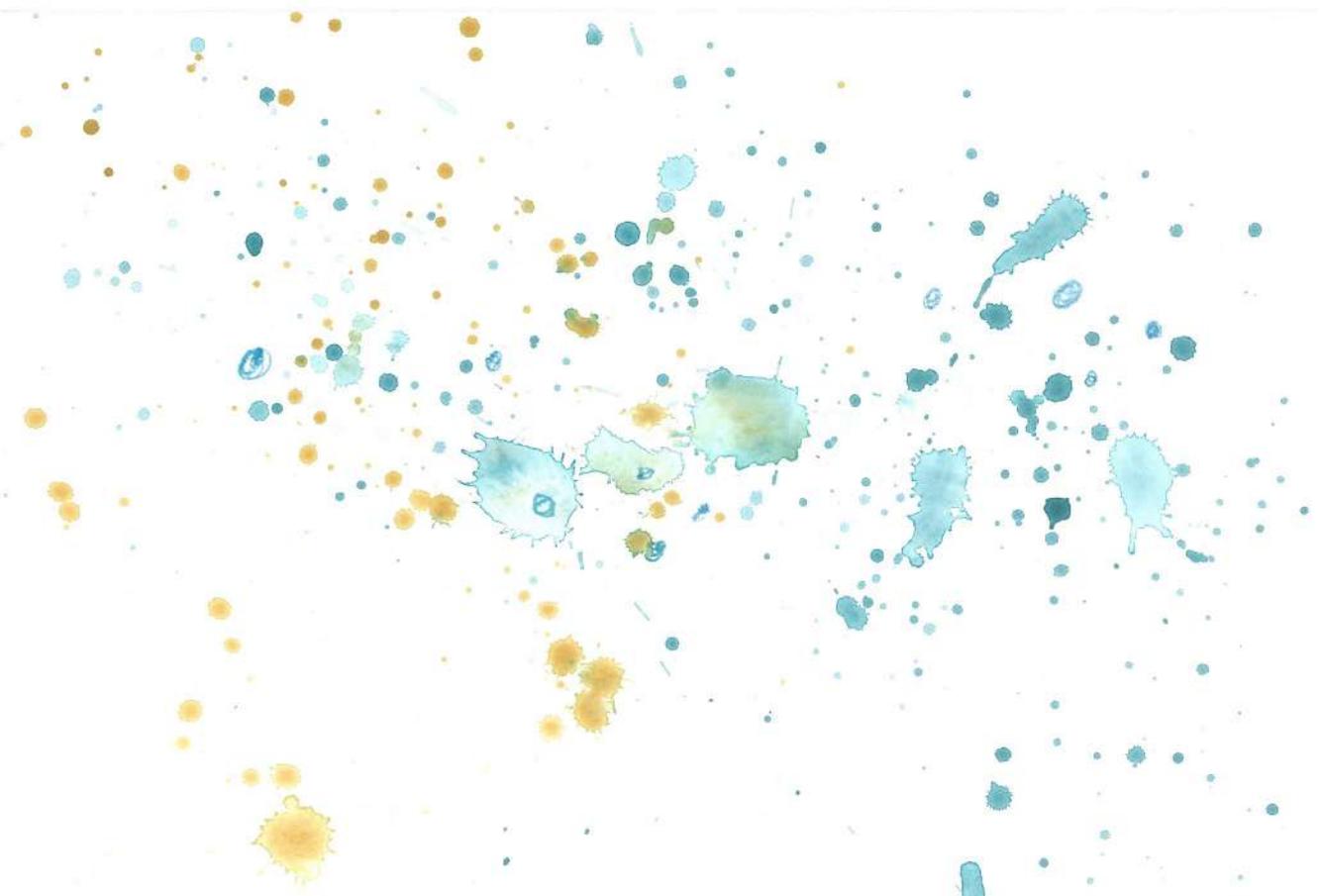


Figura 12. Varios motivos de red encontrados en redes biológicas. (A) El motivo de autorregulación es típico de redes que controlan programas de desarrollo. En el caso de redes transcripcionales consiste en un gen que se autorregula. (B) El bucle de retroalimentación consiste en dos genes que se regulan uno a otro. (C) El bucle de retroalimentación está formado por un gen principal [1] que regula a dos genes [2,3], a su vez el gen 2 regula al gen de salida [3]. Este motivo permite a los sistemas filtrar señales espurias o generar pulsos de expresión.

Durante esta tesis doctoral se han empleado varias técnicas ómicas para entender cómo funciona la regulación fotoperiódica de la expresión génica en algas y plantas. Se ha generado información novedosa de carácter evolutivo, desarrollando nuevas herramientas para el estudio de la fisiología vegetal y encontrándose una conexión hasta la fecha desconocida entre el fotoperiodo y el reloj circadiano.

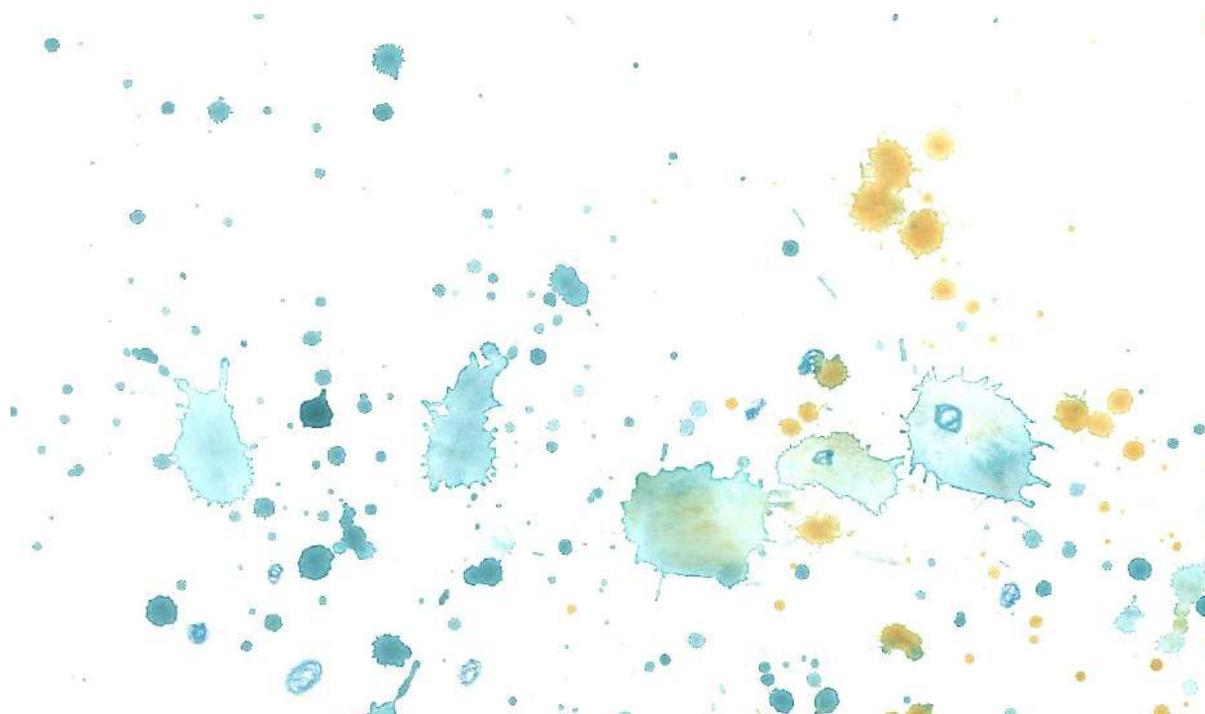


Hipótesis y objetivos



El reloj circadiano en organismos vegetales controla multitud de procesos esenciales para la fisiología y el desarrollo. Se ha descrito una alta interconexión entre el reloj circadiano y la vía fotoperiódica en organismos fotosintéticos. No obstante, no se conoce una señalización retrógrada de la respuesta fotoperiódica que regule ritmos circadianos en *Arabidopsis*. En este trabajo se aborda esta hipótesis desde una perspectiva evolutiva. Para ello, se ha empleado una aproximación de Biología Molecular de Sistemas con el fin de estudiar cómo han evolucionado estos programas transcripcionales en el linaje fotosintético, analizar las propiedades del reloj en plantas superiores y profundizar en el control de los ritmos circadianos por la vía fotoperiódica. Por lo tanto, los objetivos de esta tesis doctoral se resumen en los siguientes apartados:

1. Estudio de la conservación y la evolución de los patrones de expresión génica entre microalgas y plantas superiores.
2. Análisis de la regulación coordinada ejercida por la señalización lumínica y el reloj circadiano sobre el transcriptoma de *Arabidopsis*.
3. Estudio del efecto de la vía fotoperiódica sobre el reloj circadiano a través de CO.



Materiales y métodos



1 Recolección y análisis de datos ómicos

1.1 Análisis de datos de microarray

Los datos de microarray se procesaron usando el paquete R *affy* (Gautier et al., 2004), que realiza un control de calidad, corrección del ruido de fondo, normalización y estimación de los niveles de expresión mediante el algoritmo RMA (*Robust Multiarray Average*). Para identificar los genes diferencialmente expresados se empleó el paquete *limma* (Ritchie et al., 2015), se consideraron aquellos que presentaban el doble de expresión cuando se comparaban con su correspondiente control (*fold-change* > 2).

1.2 Análisis de datos de RNAseq

En este estudio se han usado datos de RNAseq disponibles en la base de datos pública *Sequence Read Archive* (SRA) y datos generados en este laboratorio. Para *Arabidopsis*, se analizaron datos de RNAseq provenientes del estudio GSE43865 (Rugnone et al., 2013), que comprende 22 muestras recogidas en régimen LD durante un día completo. Por otra parte, se analizaron datos de microarrays del estudio GSE3416 (Bläsing et al., 2005), recogidos bajo condiciones de día neutro (ND, 12 horas de luz y 12 horas de oscuridad). Por otra parte, también se emplearon datos de RNAseq generados en el laboratorio de plantas Col-0 y 35S:CO recogidas a ZT16 bajo condiciones de LD. (Anexo 1). Los datos de *Chlamydomonas* analizados consisten en un estudio de RNAseq (GSE71469; Zones et al., 2015) en el que se recogen muestras en condiciones ND en intervalos de 1 h y 0,5 h durante 24 h (Anexo 2). Por último, para *Ostreococcus* se analizó el estudio de microarray GSE16422 (Monnier et al., 2010). Este estudio recoge muestras en ciclos ND cada 3 h (Anexo 3).

Los genomas de referencia utilizados fueron *Arabidopsis* TAIR10 (Lamesch et al., 2012), *Chlamydomonas reinhardtii* v5.6 (Merchant et al., 2007) (descargados ambos del portal Phytozome) y el genoma ensamblado de *Ostreococcus tauri* v2 (Blanc-Mathieu et al., 2014) descargado de JGI.

Materiales y métodos

Para el análisis de datos de RNAseq se siguió el protocolo *Tuxedo* (Trapnell et al., 2012) (Figura 13A), que emplea los paquetes de código abierto *Tophat* (Trapnell et al., 2009) y *Cufflinks* (Trapnell et al., 2010). En primer lugar se realizó un análisis de calidad y filtrado de lecturas de baja calidad con la herramienta *FastQC*. A continuación las lecturas fueron alineadas al genoma de referencia correspondiente usando *Tophat*, que a su vez utiliza el mapeador *bowtie* (Langmead et al., 2009). El alineamiento de las lecturas fue almacenado en archivos BAM (*Binary Alignment Maps*). Después, se usaron las lecturas mapeadas y los transcritos conocidos del genoma anotado para ensamblar los transcriptomas para cada muestra empleando el paquete *Cufflinks*. Como unidad de medida se utilizó el FPKM (*Fragments Per Kilobase of transcript per million Mapped fragments*) (Mortazavi et al., 2008; Trapnell et al., 2010). Además, ajustando los parámetros del paquete *Cuffdiff*, se tuvieron en cuenta métodos de normalización que reducen el sesgo debido a la distribución no uniforme de las lecturas mapeadas en los transcritos (Roberts et al., 2011). Estas normalizaciones eliminan los sesgos en los datos a la vez que conservan la variación de la expresión génica que ocurre de forma natural en la transcripción, permitiendo la comparación de los niveles de expresión entre varios experimentos. Finalmente, el resto del análisis se realizó usando el paquete de R *CummeRbund* (Goff et al., 2014).

Posteriormente, se utilizó el protocolo “nuevo *Tuxedo*” o *Ballgown* (Pertea et al., 2016) para reanalizar los datos de RNAseq de 35S:CO y Col-0, ya que incorpora programas más rápidos y eficientes (Figura 13B). Este protocolo usa el alineador *HISAT2* (Kim et al., 2019) para mapear las lecturas al genoma de referencia. El ensamblado de los transcritos y la estimación de la expresión génica se realiza en este protocolo con los paquetes *Stringtie* (Kovaka et al., 2019) y *Ballgown* (Frazee et al., 2015) respectivamente. *Ballgown* no realiza ninguna normalización de los datos por lo que se optó por el método de normalización del cuartil superior. Por último, se usó el paquete *limma* para determinar los genes diferencialmente expresados (Ritchie et al., 2015).

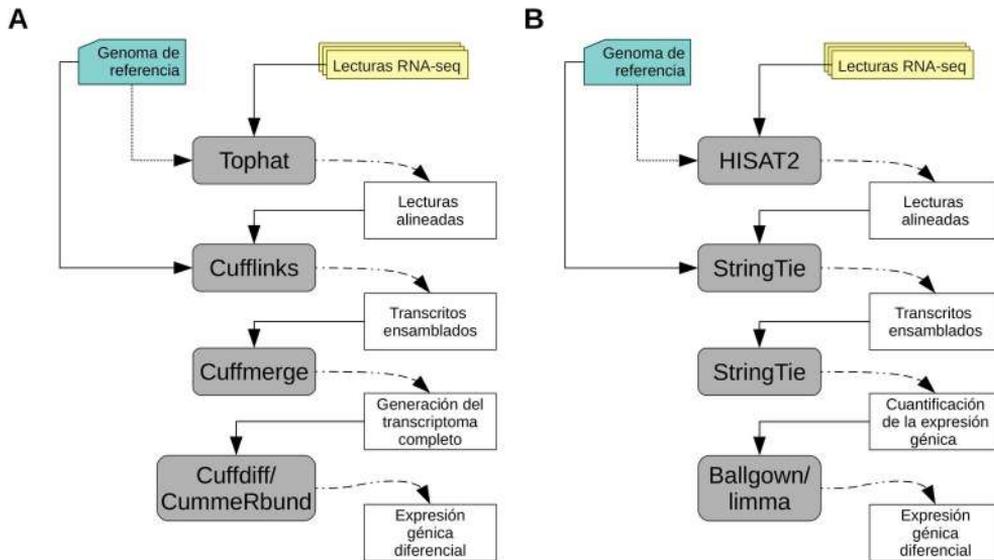


Figura 13. Flujo de trabajo empleado para analizar datos de RNAseq. (A) Siguiendo el protocolo *Tuxedo*, las lecturas se alinean al genoma de referencia correspondiente utilizando el alineador *Tophat* y se ensamblan los transcritos con *Cufflinks*. El transcriptoma completo se ensambla mediante la herramienta *Cuffmerge*. Por último, *Cuffdiff* (incluido en *Cufflinks*) y *CummeRbund* permite realizar el análisis de expresión génica diferencial. (B) El protocolo “nuevo *Tuxedo*” alinea las lecturas mediante la herramienta *HISAT*. Posteriormente *Stringtie* ensambla los transcritos y cuantifica la expresión génica. Finalmente, con los paquete de R *Ballgown* y *limma* se determinan los genes expresados diferencialmente.

1.3 Análisis de datos de ChIP-seq

El flujo de trabajo seguido para el análisis de ChIP-seq se ilustra en la Figura 14. En primer lugar, se llevó a cabo el análisis de calidad usando el software *FASTQC*. Posteriormente se usó el mapeador *bowtie* para alinear las lecturas cortas al genoma de referencia de *Arabidopsis* TAIR10. Después se llevó a cabo la determinación de áreas enriquecidas con respecto a una muestra control o picos usando la herramienta *MACS2* (*Model-based Analysis for ChIP-seq*) (Zhang et al., 2008), que mejora la resolución espacial de los sitios de unión. Esta herramienta emplea una distribución de Poisson para tener en cuenta sesgos por los que determinadas regiones del genoma

pueden presentar inherentemente más lecturas mapeadas, eliminando así falsos positivos.

Posteriormente se empleó la herramienta *PeakAnnotator* (Salmon-Divon et al., 2010) para identificar elementos funcionales próximos a los picos, y el resultado de este análisis se usó para determinar los genes diana mediante un script de R, siguiendo el criterio del gen más cercano cadena abajo, con un límite de 2000 pb.

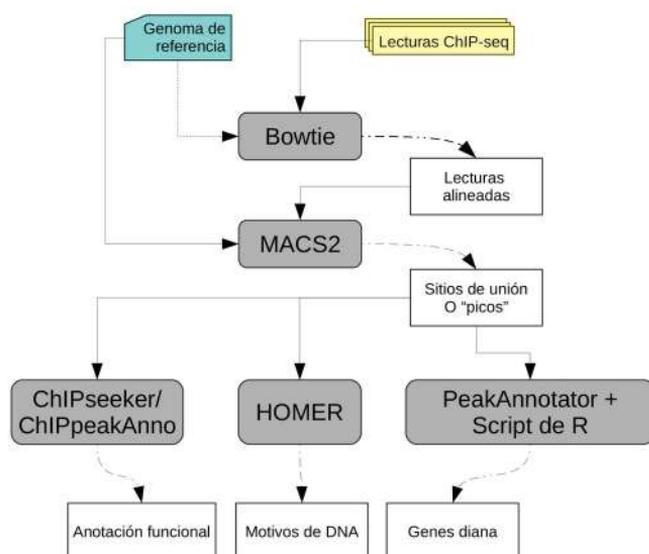


Figura 14. Flujo de trabajo para el procesamiento de datos de ChIP-seq. Las lecturas cortas provenientes de la secuenciación se mapean al genoma de referencia TAIR10 de *Arabidopsis* mediante la herramienta *bowtie*. A partir de este alineamiento se identifican los picos o sitios de unión empleando la herramienta *MACS*. Una vez obtenidos los picos, se anotan (*ChIPseeker/ChIPpeakAnno*), se identifican los motivos de DNA enriquecidos (*HOMER*) y se determinan los genes diana para el factor de transcripción analizado (*PeakAnnotator/script de R*).

Finalmente, para identificar los motivos de DNA de unión de los factores de transcripción se usó la herramienta *HOMER* (*Hypergeometric Optimization of Motif Enrichment*) (Heinz et al., 2010), que analiza las regiones genómicas buscando secuencias de DNA enriquecidas en torno a los picos. Para ello se usó el comando *findMotifsGenome.pl*. Por último, se emplearon los paquetes de R *ChIPseeker* (Yu et al., 2015) y *ChIPpeakAnno* (Zhu et al., 2010) para extraer información útil del conjunto de picos identificados.

1.4 Análisis de enriquecimiento en términos de ontología génica y rutas metabólicas

Los términos de ontología génica o GO (*GO terms*) para *Arabidopsis* y *Chlamydomonas* se obtuvieron del portal web Phytozome, mientras que para *Ostreococcus* fueron descargados de ORCAE. El análisis de enriquecimiento en términos de ontología génica sobre cualquier grupo de genes se llevó a cabo usando el paquete de R *topGO* (Alexa y Rahnenfuhrer, 2021). Como *universo* (total de genes considerados en el test) se empleó el conjunto de genes para cada especie. Para evaluar la significancia del enriquecimiento se utilizó el test exacto de Fisher con un umbral de $\alpha = 5\%$. La herramienta online REVIGO (Supek et al., 2011) se usó para eliminar la redundancia de los términos significativamente enriquecidos. Por otra parte, para identificar las vías enriquecidas en *Arabidopsis* se empleó la función *enrichKEGG* del paquete de R *clusterProfiler* (Wu et al., 2021). Se realizó una corrección del testeo múltiple con el método *Benjamin-Hochberg* y se estableció un umbral para los q-valores de 0,05.

2 Construcción y análisis de redes génicas

2.1 Conceptos básicos en teoría de redes

Una red N se define como un par ordenado $N=(V,E)$ formado por dos conjuntos: V es el conjunto de nodos o vértices que constituyen los elementos del sistema y E es el número de aristas que representan las interacciones entre ellos. El orden de una red es número (n) de nodos y el tamaño se define como el número de aristas (m).

$$V=\{v_1,\dots,v_n\}$$

$$E=\{e_k=\{v_i,v_j\} \text{ donde } v_i,v_j \in V; k=1,\dots,m\}$$

Se pueden distinguir dos tipos de redes. Las redes no dirigidas no presentan direccionalidad en las aristas mientras que las redes dirigidas sí.

Para realizar el análisis computacional de las redes se emplean matrices de adyacencia, que se define como una matriz cuadrada $n \times n$ donde la componente i,j es 1 cuando existe interacción entre los nodos $i - j$, y 0 en caso de no haber relación.

Adicionalmente, existen redes ponderadas en las que las aristas tienen un peso asociado que no se reducen a un valor booleano (1 ó 0).

Dos nodos son vecinos si los une una arista. Así, el camino mínimo entre dos nodos cualesquiera es aquel camino entre dichos elementos con la mínima longitud posible. La distancia entre ellos $d(i,j)$ se define como la longitud del camino mínimo.

Uno de los conceptos más importantes en teoría de redes es el grado de un nodo, que es el número de nodos unidos a él. Por otra parte, la transitividad o el coeficiente de agrupamiento de un nodo (*clustering coefficient*) varía entre 1 y 0, dependiendo de si todos sus vecinos están conectados entre sí o no lo están, respectivamente. Por tanto, la transitividad representa el grado de agrupamiento o conexión alrededor de un nodo.

Formalmente, la distribución del grado de los nodos se define como la probabilidad de encontrar un nodo de grado k aleatoriamente en la red, $P(dv = k)$. Para representar la frecuencia relativa de los grados de una red se emplean histogramas. Cuando la distribución del grado de los nodos se ajusta a una potencial negativa, hablamos de redes libres de escala. En este tipo de redes la probabilidad de encontrar un nodo con bajo grado es alta debido a que la mayoría de ellos presentan pocas conexiones, mientras que existen pocos nodos concentradores o *hubs* que tienen una alta conexión.

Por otra parte, la distribución de la longitud del camino mínimo entre nodos de una red hace referencia a la probabilidad de encontrar un camino de longitud k de forma aleatoria en la red. De forma similar, se usan histogramas para mostrar la frecuencia relativa de la longitud de caminos mínimos de la red. Es una medida de la eficiencia de transferencia de información de la red. Cuando la longitud media del camino mínimo de una red es significativamente menor de lo esperado, se denomina red de mundo pequeño, en las que dos nodos aleatorios se pueden conectar por pocos pasos.

2.2 Construcción y análisis de redes de co-expresión génica

El flujo de trabajo seguido para la construcción y el análisis de las redes de coexpresión génica se ilustra en la Figura 15. Para medir la coexpresión génica se usó la correlación entre los perfiles de expresión en 24 h, considerando que dos genes estaban coexpresados cuando el coeficiente de correlación de Pearson entre los perfiles era mayor que 0,95.

Se generó una red de coexpresión génica para cada una de las tres especies en estudio, en las que los nodos representan genes rítmicos y se traza una arista entre ellos si están coexpresados de acuerdo al criterio establecido. Para visualizarlas se empleó Cytoscape, software empleado para representar y analizar redes complejas (Shannon et al., 2003). Concretamente se aplicó el algoritmo de visualización *Prefuse Force-Directed Layout*, que distribuye los elementos en la red basándose en la simulación de dos fuerzas que interaccionan. Los nodos se repelen mientras que las aristas se atraen, por lo que los diferentes elementos se localizan en la red de modo que la energía sea mínima (Kobourov, 2012).

El análisis de las redes génicas se llevó a cabo usando el paquete de R *igraph* (Csardi, 2006). La propiedad libre de escala se evaluó mediante regresión lineal sobre la transformada logarítmica de la distribución del grado de los nodos. Las transformadas logarítmicas convierten las potenciales negativas en rectas con pendiente negativa. Así, después de hacer la transformación, se puede comprobar que la correspondiente R^2 es suficientemente alta y que el correspondiente p-valor es suficientemente bajo.

Por otra parte, la propiedad de mundo pequeño se examinó generando 10000 redes libres de escala con las mismas propiedades topológicas, mediante la función *barabasi.game* del paquete *igraph*, y posteriormente comparando el camino mínimo entre nodos con las redes simuladas, con el propósito de estimar un p-valor.

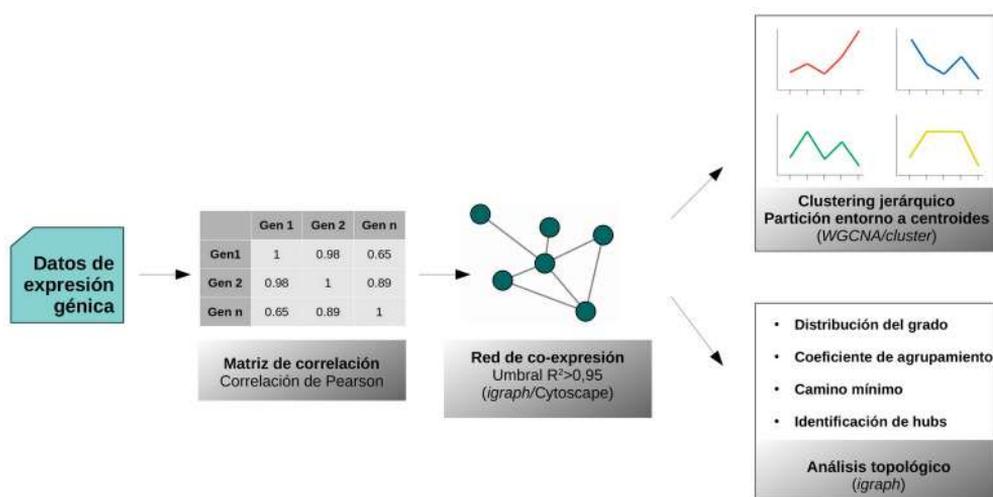


Figura 15. Flujo de trabajo computacional seguido para la construcción y análisis de redes de co-expresión génica. Estas redes se basan en datos transcriptómicos masivos obtenidos en este caso de microarrays y RNAseq en los que podemos encontrar la expresión de genes a lo largo de 24 horas. Con estos datos se genera una matriz de correlación que contiene los coeficientes de correlación de Pearson entre cada elemento. A partir de esta matriz se construye y visualiza la red de co-expresión con las herramientas *igraph* y *Cytoscape*. Finalmente, se realiza un agrupamiento mediante dos métodos, *clustering* jerárquico y PAM (partición entorno a centroides) utilizando los paquetes *WGCNA* y *cluster*, así como un análisis topológico empleando el paquete *igraph*.

Los análisis de agrupamiento o *clustering* se llevaron a cabo mediante dos métodos, el método jerárquico y el de partición entorno a centroides, con las funciones de `hclust` y `pam`, del paquete de R *cluster*. Posteriormente se eligió el método y el número de grupos con el que se obtuvo la mejor silueta, que es un valor que representa la bondad de la clasificación.

2.3 Construcción y análisis de redes transcripcionales

La construcción y el análisis de las redes transcripcionales se llevó a cabo según el flujo de trabajo mostrado en la Figura 16. Una vez obtenidas las dianas de cada factor de transcripción se procedió a construir la red transcripcional. Para ello se utilizó un *script* propio de R, *network_construction.R*, el cual utiliza el paquete de R *igraph*. Este *script* toma como entrada todos los archivos con las dianas de cada factor de transcripción. En primer lugar se crea una matriz de adyacencia en la que las

columnas y las filas representan los genes. A continuación, mediante una función de R se asigna uno al elemento i,j si el gen j es diana del factor de transcripción codificado por el gen i ; se asigna cero en caso contrario. Para construir la red transcripcional se utilizó la función *graph.adjacency*, que toma como dato de entrada la matriz de adyacencia y el tipo de red, en este caso dirigida. De esta forma se construyeron dos grafos, uno con los 33 factores de transcripción y sus dianas y el subgrafo inducido compuesto solamente por los factores de transcripción.

Las redes generadas se representaron gráficamente en Cytoscape (Smoot et al., 2011), y de nuevo se aplicó el método *Prefuse Force Layout*.

Adicionalmente se realizó un análisis de la topología de la red. En este caso, para determinar si la red era libre de escala se empleó el Test de Kolmogorov-Smirnov. Usando la función *degree* se determinó la distribución del grado de los nodos de la red. Mediante la función *power.law.fit*, también del paquete *igraph*, se realizó el análisis estadístico basado en el mencionado test. Esta función recibe como entrada la distribución del grado de los nodos y devuelve el objeto *Ks.p*, que corresponde al p-valor de rechazar que dicha distribución sigue una potencial negativa.

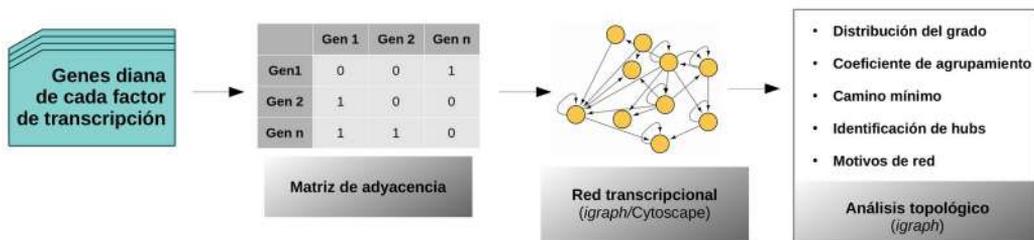


Figura 16. Flujo de trabajo computacional seguido para la construcción y análisis de redes transcripcionales. Se parte de datos de unión de factores de transcripción al genoma. Así, ficheros con genes diana de cada factor de transcripción se integran generando una matriz de adyacencia en la que se escribe el valor 1 para el elemento $A_{i,j}$ si hay una arista desde el nodo i hasta el nodo j . A partir de esta matriz se construye y visualiza la red usando las herramientas *igraph* y Cytoscape. Por último se realiza el análisis topológico de la red.

A continuación, se determinaron los nodos concentradores o *hubs* de acuerdo a dos metodologías distintas. Por una parte, se consideraron como *hubs* el 5% de los nodos con mayor grado o número de vecinos. Por otra parte, se utilizó la función

hub.score de *igraph*, que toma como entrada la red y asigna una puntuación entre 0 y 1 a cada nodo, según el algoritmo HITS (*Hyperlynked-Induced Topic Search*) (Kleinberg, 1997). De igual forma se seleccionaron como *hubs* el 5% de nodos con mayor puntuación.

Asimismo, se calculó el coeficiente de agrupamiento de la red usando la función *transitivity* y el camino medio entre nodos con la función *average.path.length*.

2.4 Motivos de red

Para identificar motivos de red presentes en nuestras redes transcripcionales, se evaluó la significancia estadística de todos los subgrafos posibles formados por uno, dos o tres nodos. En primer lugar, se generaron grafos aleatorios usando el modelo *Erdős-Rényi* implementado en la función de R *erdos.renyi.game* del paquete *igraph*. Se generaron 10000 grafos aleatorios con las mismas propiedades topológicas.

Se comparó la frecuencia con la que aparecía cada motivo en la red de estudio con las veces que aparecía en cada una de las redes aleatorias para estimar el p-valor correspondiente para cada motivo.

3 Análisis de patrones circadianos

La detección de patrones periódicos significativos en los datos de expresión génica analizados se realizó usando *RAIN* (*Rhythmicity Analysis Incorporating Nonparametric methods*). Este paquete de *Bioconductor* contiene funciones que implementan métodos no paramétricos robustos para la detección de ritmos con formas arbitrarias y periodos previamente especificados (Thaben y Westermarck, 2014). La principal función, *rain*, se usó con los siguientes parámetros: una matriz numérica con los niveles de expresión génica de las diferentes réplicas biológicas en un periodo de 24 h; un intervalo de muestreo de 4 h para *Arabidopsis* y de 3 h para *Chlamydomonas* y *Ostreococcus*; un periodo de 24 h; y el número de réplicas, para *Arabidopsis* y *Ostreococcus* tres réplicas y para *Chlamydomonas* dos. Se estableció un nivel de significancia de $\alpha = 1\%$ para asumir que un determinado gen presenta un patrón rítmico a lo largo del día. La forma de la curva de expresión de cada gen periódico se

caracterizó usando su pico (punto temporal donde la expresión es máxima) y su mínimo. Se dividió el día en cuatro periodos, denominados en este trabajo Amanecer, Día, Atardecer y Noche. Así, los genes se clasificaron en 16 *clusters* diferentes de acuerdo al momento donde presentaban el máximo y el mínimo en su curva de expresión.

4 Significancia de solapamiento de conjuntos génicos

Para evaluar la significancia estadística del solapamiento de dos conjuntos génicos se recurrió a la prueba de Fisher, que es un método exacto para estudiar si existe asociación entre dos variables cualitativas. Se utilizó la función de R `fisher.test`. En el caso de más de dos conjuntos de genes, se empleó el paquete de R *SuperExactTest* (Wang et al., 2015) que calcula la probabilidad exacta de múltiples intersecciones basándose en una distribución hipergeométrica para determinar enriquecimiento. Los resultados se visualizaron mediante diagramas de Venn.

Para evaluar la asociación entre dos conjuntos de regiones génicas, se optó por emplear tests de de permutaciones. La intersección entre dos grupos de regiones se calculó con el paquete de R *ChIPpeakAnno* (Zhu et al., 2010) y después se randomizó uno de los dos conjuntos creando un grupo de sitios de unión totalmente aleatorio. A continuación, se calculó de nuevo el solapamiento entre el conjunto aleatorio y el real. Se repitió el proceso 10000 veces para determinar si el solapamiento original se debía al azar o no, estimándose un p-valor.

5 Análisis de conservación de módulos

Se consideró que dos genes potenciales ortólogos según el algoritmo *MBBH* exhibían un patrón circadiano conservado cuando ambos pertenecían al mismo *cluster* (máximo y mínimo de expresión en la misma franja horaria) o cuando el coeficiente de correlación de Pearson entre sus perfiles de expresión era mayor que 0,98. Sin embargo, la conservación entre patrones de co-expresión de dos grupos diferentes de genes se calculó de acuerdo al estadístico de preservación compuesta *Zsummary*

(Langfelder et al., 2011). Un valor de *Zsummary* menor que 2 indica que no hay conservación, mientras que un *Zsummary* mayor que 10 constituye evidencia de un gran nivel de conservación. Para calcular este parámetro se usó el paquete de R *WGCNA* (Langfelder y Horvath, 2012) y la función implementada en el mismo.

6 Generación de aplicaciones web usando shiny R

Shiny es un paquete de R que permite construir aplicaciones web directamente desde código R. Por ello no son necesarios conocimientos de desarrollo web.

La estructura de una aplicación de *Shiny R* es la siguiente:

- *ui (user interface)*: especifica la interfaz gráfica de usuario de de la aplicación. En esta sección de código se definen y establecen las entradas (*inputs*) y salidas (*outputs*). Existen multitud de funciones para diseñar la interfaz e incluir multitud de elementos como paneles, etiquetas o pestañas.
- *Server (server function)*: contiene las instrucciones que se necesitan para construir la aplicación. Aquí se calculan las salidas (*outputs*) a partir de las entradas (*inputs*) especificadas por el usuario en la interfaz gráfica realizándose los cálculos y computaciones necesarias para generarlos.
- *shinyApp()*: función que crea el objeto aplicación *Shiny*.

Los bloques *ui* y *server* se comunican, el primero comunica al *server* los parámetros introducidos por el usuario y entonces éste genera los resultados que se muestran en la interfaz gráfica (*ui*). Uno de las características más importantes de *Shiny* es que usa *programación reactiva*. Esto significa que las funciones del *server* reaccionan automáticamente a cambios que se hacen en el *ui*, permitiendo generar aplicaciones interactivas para el usuario.

7 Material vegetal

7.1 Organismos empleados

En este trabajo se han empleado como organismos modelo la planta superior *Arabidopsis thaliana* y las microalgas *Ostreococcus tauri* y *Chlamydomonas reinhardtii*.

Arabidopsis thaliana es una planta herbácea de pequeño tamaño perteneciente a la familia *Brassicaceae*. Su genoma tiene 120 Mb distribuidas en 5 cromosomas. Las estirpes utilizadas se detallan en la Tabla 4.

Tabla 4. Estirpes de *Arabidopsis* utilizadas en este trabajo.

Mutante/Sobre-expresor	Ecotipo	Gen	Locus
<i>co-10</i>	Col-0	AT5G15840	<i>CO</i>
35S: <i>CO</i>	Col-0	AT5G15840	<i>CO</i>
<i>pr5</i>	Col-0	AT5G24470	<i>PRR5</i>
35S: <i>GFP:HY5 (hy5-211)</i>	Col-0	AT5G11260	<i>HY5</i>
<i>hy5-2</i>	Col-0	AT5G11260	<i>HY5</i>
<i>hyh</i>	Col-0	AT5G11260	<i>HYH</i>
<i>hy5-2/hyh</i>	Col-0	AT5G11260	<i>HY5/HYH</i>

Ostreococcus tauri es una microalga de la familia *Prasinophyceae* y es el eucariota de vida libre más pequeño conocido. Se caracteriza por tener un genoma pequeño y compacto. Se ha utilizado la estirpe RCC745.

Chlamydomonas reinhardtii pertenece a la familia *Chlorophyceae* y es un modelo clásico para el estudio de organismos fotosintéticos. Su genoma está compuesto de 17 cromosomas con un total de 110 Mb. La estirpe empleada ha sido la mutante de pared celular *cw15* (considerada silvestre en este estudio).

7.2 Cultivo de microalgas en matraz

La cepa de *Chlamydomonas cw15* se cultivó en matraces Erlenmeyer con agitación bajo una intensidad lumínica de $75 \mu\text{Em}^{-2}\text{s}^{-1}$ y un intervalo de temperaturas de 18°C

Materiales y métodos

(noche) a 22°C (día). Se usó medio mínimo Sueoka. Por otra parte, *Ostreococcus* se cultivó en agua marina esterilizada por filtración y autoclavada complementada con vitaminas B2, bajo las condiciones previamente descritas para *Chlamydomonas*.

La recogida de microalgas se realizó por centrifugación a 2000 g durante 4 min. Posteriormente se lavó el precipitado con 1 ml de PBS y se volvió a centrifugar. El precipitado resultante se guardó a -80°C hasta su procesamiento.

7.3 Cultivo de plantas en placas de Petri

Las semillas de *Arabidopsis* fueron estratificadas durante 4 días a 4°C en oscuridad antes de sembrarlas en placas de medio MS (Murashige & Skoog) (Murashige y Skoog, 1962) en condiciones LD. Los puntos temporales se indican como *Zeitgeber Time N* (ZTN), indicando el número de horas N tras encenderse las luces en el fitotrón o cámara de cultivo (ZT0), simulando el amanecer.

Para el estudio expresión génica realizado en el segundo capítulo de Resultados, plántulas de 12 días de edad se recogieron cada 4 h durante un periodo de 24 h. Para los análisis de expresión génica y los experimentos de inmunoprecipitación de la cromatina del tercer capítulo se recogieron plántulas a los 10 días tras la germinación, en la ventana comprendida entre los puntos ZT12 y ZT16.

8 Bacterias

8.1 Estirpes utilizadas

Las estirpes utilizadas en este trabajo se detallan en la Tabla 5. La cepa de *Escherichia coli* DH5 α se empleó para las clonaciones. La estirpe DB3.1 se utilizó para la propagación de plásmidos *Gateway* originales.

La cepa de *Agrobacterium tumefaciens* GV1301 pmp90 se usó para la transformación de *Arabidopsis* y para estudios de expresión transitoria de proteínas en hojas de *Nicotiana*.

Tabla 5. Estirpes de bacterias utilizadas en este trabajo.

Bacteria	Estirpe	Genotipo	Referencia
<i>E. coli</i>	DH5 α	supE44 Δ lac U169 (Φ 80 lacZ Δ M15) hsdR17 recA1 endA1 gyrA96 thi-1 relA1	(Hanahan, 1983)
<i>E. coli</i>	DB3.1	gyrA462 endA1 Δ (sr1-recA) mcrB mrr hsdS20 glnV44 (=supE44) ara14 galK2 lacY1 proA2 rpsL20 xyl5 leuB6 mtl1	(Bernard y Couturier, 1992)
<i>A. tumefaciens</i>	GV301 (pmp90)	R ^r Gn ^r	(Watson et al., 1975)

8.2 Cultivo en medio líquido y sólido

Todas las bacterias se cultivaron en medio Luria-Bertani (LB) (Sambrook et al., 1989) en agitación, a 37 °C y a 28 °C, para *E. coli* y *A. tumefaciens*, respectivamente. Para el cultivo en medio líquido se empleaba medio LB solidificado con Bacto-agar 1,5% (p/v) (SIGMA).

Para la selección de bacterias portadoras de plásmidos se suplementaba el medio de cultivo con el antibiótico correspondiente (Tabla 6).

Tabla 6. Plásmidos utilizados en este trabajo.

Plásmido	Características	Referencia o fuente
pDONR207	Vector inicial del sistema Gateway® (Gm ^R)	Invitrogen
pYFN43	Vector de expresión génica en plantas (Km ^R en bacteria e Hyg ^R en plantas) utilizado para la clonación mediante sistema Gateway®. Fusiona la mitad N-terminal de la proteína fluorescente amarilla (YFP) para ensayos de BiFC.	(Belda-Palazón et al., 2012)
pYFC43	Vector de expresión génica en plantas (Km ^R en bacteria e Hyg ^R en plantas) utilizado para la clonación mediante sistema Gateway®. Fusiona la mitad C-terminal de la YFP.	(Belda-Palazón et al., 2012)
pGWB644	Vector de expresión génica en plantas (Spect ^R en bacteria y BASTA ^R en plantas) utilizado para la clonación mediante sistema Gateway®. Produce una fusión C-terminal de la proteína fluorescente azul (CFP).	(Nakamura et al., 2010)
pGWB642	Vector de expresión génica en plantas (Spect ^R en bacteria y BASTA ^R en plantas) utilizado para la clonación mediante sistema Gateway®. Produce una fusión C-terminal de la YFP	(Nakamura et al., 2010)
pGWB645	Vector de expresión génica en plantas (Spect ^R en bacteria y BASTA ^R en plantas) utilizado para la clonación mediante sistema Gateway®. Produce una fusión N-terminal de la CFP.	(Nakamura et al., 2010)
pGWB515	Vector de expresión génica en plantas (Spect ^R en bacteria e Hyg ^R en plantas) utilizado para la clonación mediante sistema Gateway®. Produce una fusión N-terminal del epítipo HA (Human influenza hemagglutinin).	(Nakagawa et al., 2007)

9 Clonación mediante el sistema Gateway

Los plásmidos utilizados en este trabajo se generaron mediante el sistema Gateway® *Technology* (Invitrogen). En primer lugar se amplificaron mediante PCR los fragmentos de DNA a clonar. Para ello se usaron cebadores que incluyesen en su secuencia los sitios de reconocimiento attB1 y attB2. Debido a la longitud de estos cebadores, se realizaron las PCRs en dos pasos. En la primera reacción se incluyeron parte de las secuencias attB utilizándose cebadores con una secuencia específica del fragmento de DNA y una parte de la secuencia de recombinación. En la segunda reacción se

amplificaban los fragmentos resultantes de la primera para añadir la secuencia restante. Esto permitió estandarizar la segunda reacción con unos cebadores genéricos y utilizar unos más simples en la primera reacción.

Tras purificar el producto final de la PCR, se realizaron las reacciones de recombinación BP, según las instrucciones del fabricante, con el objetivo de introducir el fragmento de DNA en el vector de entrada pDONR207. Una vez terminada la reacción, se transformaron células de *E.coli* DH5 α . Las bacterias transformadas se sembraron en placas de LB suplementadas con gentamicina 20 μ g/ml y se incubaron durante una noche a 37°C. De forma general se seleccionaron 4 colonias de cada reacción para la purificación del plásmido usando el kit *ISOLATE II Plasmid Mini Kit* (Bioline). Los plásmidos se verificaron por secuenciación y se usó uno de ellos en la segunda reacción de recombinación (LR) para transferir el inserto clonado al vector de destino. Con este vector se transformaron células de *Agrobacterium* para realizar ensayos de expresión transitoria en hojas de *Nicotiana* (Apartado 14).

Los cebadores empleados para la clonación de fragmentos de DNA se detallan en la Tabla 7. El cDNA de *CO* se encontraba ya en un pDONR221 existente en el laboratorio y el cDNA de *PRR5* se obtuvo de Addgene contenido en el plásmido pENTR/D-TOPO.

Tabla 7. Cebadores empleados en la clonación de los genes *HY5* y *FLM*. También se incluyen los cebadores utilizados para añadir la parte externa de las secuencias attB en la segunda reacción de PCR.

	Secuencia directa	Secuencia reversa
Cebadores externos sitios attB	GGGGACAAGTTTGTACAAAAAAGCAG GCT	GGGGACCACTTTGTACAAGAAAG CTGGGT
<i>HY5</i>	AAAAGCAGGCTTCATGCAGGAACAAG CGACTAGC	AGAAAGCTGGGTCTCAAAGGCTT GCATCAGC
<i>HY5</i> sin codón de STOP	AAAAGCAGGCTTCATGCAGGAACAAG CGACTAGC	AGAAAGCTGGGTCAAGGCTTGCA TCAGCATTAGA
<i>FLM</i>	AAAAGCAGGCTGCATGGGAAGAAGA AAAATCGAG	AGAAAGCTGGGTCTTAATTGAGCA GCGGGAGAGT

10 Extracción de RNA y síntesis de cDNA

Durante la manipulación del RNA es necesario adoptar una serie de medidas de precaución para evitar su degradación por las RNAsas. Así, todo el material y soluciones fueron esterilizados en un horno a 200°C y en un autoclave, respectivamente. Por otra parte, todas las soluciones se prepararon con agua Milli-Q suplementada con dietilpicrocarbonato (DEPC) al 0,1% (v/v). Una vez recogido el material vegetal, se congeló inmediatamente en nitrógeno líquido para evitar su degradación.

La extracción de RNA se llevó a cabo con el método basado en Tioisocianato de guanidina [*Trizol*® (Invitrogen)], según el protocolo descrito por Chomczynski y Sacchi (Chomczynski y Sacchi, 1987). Tras triturar el material vegetal (en el caso de *Arabidopsis*), se añadió 1 ml de Trizol, se agitó en vórtex para homogeneizarlo y se incubó 5 min en hielo. Posteriormente se añadió 200 µl de cloroformo y se volvió a mezclar la muestra antes de centrifugarla en una microcentrífuga a 15000g durante 15 min a 4°C. Después de la centrifugación, se transfirió el sobrenadante a un tubo nuevo y se añadieron 500 µl de isopropanol. Se incubó la muestra 15 min a temperatura ambiente y se centrifugó de nuevo a 15000 g durante 10 min a 4°C. A continuación, para precipitar selectivamente el RNA, se añadieron 750 µl de cloruro de litio 3M y se incubó 10 min en hielo para volver a centrifugar las muestras a 15000 g durante 10 min a 4°C. Finalmente, se lavó el precipitado con 500 µl de etanol al 85% (v/v) y se centrifugaron a máxima velocidad durante 10 min a 4°C. El precipitado final se secó al aire para eliminar los restos de etanol y se resuspendió en 22 µl de agua Milli-Q suplementada con DEPC 0,1% (v/v). Se usaron 2 µl para medir la concentración de RNA en un espectrofotómetro nanodrop® (ND-1.000 Spectrophotometer, Nanodrop Technologies).

La reacción de retrotranscripción se realizó con el kit de QIAGEN *Quantitect Reverse Transcription* siguiendo las instrucciones del fabricante. En las reacciones se partió de 500 ng de RNA para una mezcla con 0,5 µl de retrotranscriptasa, 1 µl de tampón de reacción y agua hasta alcanzar un volumen final de 10 µl. Después de

inactivar la retrotranscriptasa, se añadió agua Milli-Q hasta alcanzar 40 µl de volumen final. Las posteriores reacciones de QPCR se llevaron a cabo usando 1 µl de cDNA.

11 Análisis de expresión génica mediante QPCR en tiempo real.

Para la cuantificación relativa de la cantidad de las moléculas de RNA mensajero se llevaron a cabo PCRs cuantitativas en tiempo real (QPCR). Las reacciones de QPCR se llevaron a cabo utilizando el fluoróforo SYBR® Green (SensiFAST™ SYBR® & Fluorescein Kit) y se realizaron en un volumen final de 10 µl al que se le añadió: 1 µl de CDNA, 0,2 µM de cada cebador, 5 µl de la solución SensiFAST y agua Milli-Q hasta alcanzar el volumen total. Las reacciones se realizaron por triplicado en un termociclador iQ5 de BioRad, siguiendo el programa indicado en la Tabla 8.

Tabla 8. Programa utilizado para las reacciones de QPCR

CICLOS	TEMPERATURA	TIEMPO
CICLO Paso 1	95 °C	2 minutos
CICLO 2 (40x) Paso 1 Paso 2 Paso 3	95 °C 60 °C 72 °C	5 segundos 10 segundos 6 segundos
CURVA DE FUSIÓN Paso 1 Paso 2 Paso 3 (81x)	95 °C 95 °C 55 °C	30 segundos 1 minuto 30 segundos

La cantidad de transcrito se estimó a partir del ciclo umbral (*Threshold Cycle*, Ct) (Ginzinger, 2002) usando el algoritmo ddCt (Zhang et al., 2016) y se relativizó frente a los niveles del gen de expresión constitutiva *UBIQUITINA 10 (UBQ10)*. Los cebadores empleados para estimar la expresión génica se muestran en la Tabla 9.

Tabla 9. Cebadores utilizados para medir la expresión génica en las reacciones de QPCR.

	Secuencia directa	Secuencia reversa	Tamaño del fragmento amplificado (pb)
Genes de <i>Ostreococcus</i>			
<i>OtEF1α</i>	GACGCGACGGTGGATCAA	CGACTGCCATCGTTTTACC	202
<i>OtHY5</i>	GAGGGAAGATTGGGAAGAAGAA	CCTTCTCCAACGCCTGAAT	81
Genes de <i>Chlamydomonas</i>			
<i>CrTUB</i>	GTTGCATCGTTAGCGTGGACG	GCAGCAGCCAATGTTTCAGACT	170
<i>CrHY5</i>	GGCATGAACCCTAGCTTCC	CATCATGTCCTCGCTGATGT	144
Genes de <i>Arabidopsis</i>			
<i>UBQ10</i>	GAAGTTCAATGTTTCGTTTCATGT	GGATTATACAAGGCCCAAAA	145
<i>HY5</i>	GCTGAAGAGGTTGTTGAGGA	TCTCCAAGTCTTTCACTCTGTTT	101
<i>PRR5</i>	GCAATCTCTTCAACGAGAAGCCGC	CGAACGAATTGGCCTTTGATTCG	138
<i>GI</i>	CCAGCAACAATACGGTGCC	CGATAGGACGGACTATTCATCCG	158
<i>PRR7</i>	GGAAGTGGTAGCGGAACTTGG	CGTACCTTCTTCGGAAGCACC	110
<i>BAM9</i>	CTTGATGGGAAGACTCCTATGG	GTGATTCCCCTGATTGTGTTG	98

12 Extracción de proteínas

12.1 Aislamiento de proteínas en *Nicotiana benthamiana*

Las hojas de *Nicotiana* se trituraron en presencia de nitrógeno líquido hasta obtener un polvo fino blanquecino. En ese momento se añadieron 50 ml de tampón de extracción de proteínas [50 mM Na-phosphate pH 7.4, 150 mM NaCl, 10% glycerol, 5 mM EDTA, 1 mM DTT, 0.1% Triton X-100, 50 μ M MG-132, 2 mM Na₃VO₄, 2 mM NaF, PIC (*Protease Inhibitor Cocktail*) 1:1000] (Song et al., 2012) junto con el nitrógeno líquido, y se volvió a triturar. El material homogeneizado se dejó descongelar a temperatura ambiente y después se centrifugó durante 10 minutos a 4°C a una velocidad de 5000 rpm en microfuga. El sobrenadante se filtró a través de una membrana de nylon (SIGMA) de 100 μ m de diámetro de poro. La solución resultante se usó inmediatamente para los ensayos de co-inmunoprecipitación de proteínas o se congeló en nitrógeno líquido para su conservación a -80°C.

12.2 Aislamiento de cromatina en *Arabidopsis*

Para la extracción de núcleos de plántulas de *Arabidopsis* se trituró en un mortero el tejido en presencia de nitrógeno líquido hasta conseguir un polvo ligero blanquecino. Después, este tejido se mezcló con el tampón de extracción 1 [sacarosa 400 mM, Tris-HCl 10mM pH 8,0, MgCl₂ 10mM, β-mercaptoetanol 5mM] y se incubó a 4°C hasta su completa homogeneización. Esta solución se filtró a través de dos capas de Miracloth (Merck) y se centrifugó a 3200g durante 20 min a 4°C. El precipitado se resuspendió en tampón de extracción 2 [sacarosa 250 mM, Tris-HCl pH 8,0 10 mM, MgCl₂ 10 mM, 1% (v/v) Tritón X-100, β-mercaptoetanol 5 mM] y la solución resultante se transfirió a un nuevo microtubo que se centrifugó a 5000g durante 10 min a 4°C. Se descartó el sobrenadante y el precipitado se resuspendió en 300 µl tampón de lisis nuclear [NaCl 150 mM, SDS 0,1% (v/v), Tritón X-100 1% (v/v), EDTA 2 mM, Tris-HCl pH 8,0 20 mM] con la ayuda de un *green pestle*.

Los tampones se suplementaron antes de ser utilizados con los siguientes reactivos para preservar la integridad de las proteínas: (PIC; SIGMA) 1:10000, PMSF 0,1 mM, MG132 (SIGMA) 1:10000, NaVO₃ 1:10000, y NaF 1:1000.

Una vez resuspendidos los núcleos, se sonicaron en un Bioruptor® Pico durante 15 ciclos de 30 s de encendido y 1 min de apagado, para conseguir fragmentos de entre 200 y 500 pb. Tras la sonicación se incubó en rotación a 4°C durante 30 min. La cromatina sonicada se centrifugó durante 10 min a 16000g para precipitar restos de la fracción nuclear, y el sobrenadante (cromatina) se utilizó para pasos posteriores o se congeló en nitrógeno líquido y se guardó a -80°C.

13 Inmunoprecipitación de la cromatina

La inmunoprecipitación de la cromatina (ChIP) se llevó a cabo siguiendo el protocolo de Bowler (Bowler et al., 2004) con algunas modificaciones para mejorar el enriquecimiento, recomendadas para la inmunoprecipitación de factores de transcripción que no son muy abundantes (Lau y Bergmann, 2015). Las plántulas fueron cultivadas durante 10 días bajo condiciones LD y se recogieron 9 g de tejido en la ventana de tiempo ZT12-ZT14. Después, se realizó una doble fijación o *crosslinking*

Materiales y métodos

para capturar complejos de proteínas que no están directamente unidas al DNA. El tejido vegetal se fijó en un concentrador de vacío en presencia de DSG 1 mM durante 10 min, en 1% (v/v) de formaldehído 20 min y finalmente se añadió glicina a una concentración de 0,125 M para detener la fijación. Seguidamente se lavó el tejido con agua Milli-Q, se eliminó el exceso de líquido y se congeló en nitrógeno líquido.

El tejido se trituró en presencia de nitrógeno líquido y se obtuvo la cromatina como se describe en el apartado 12.2. Se trabajó con alícuotas de 3 g para no perder eficiencia hasta tener aislada la cromatina, momento en el que se combinaron. Tras obtener la cromatina, se llevó a cabo su cuantificación para partir de la misma cantidad en la muestra y en el control. Para ello se transfirieron 50 µl de cromatina a un nuevo microtubo al que se añadieron 450 µl de tampón de cuantificación de cromatina [Tris-HCl pH 7,4 10 mM, EDTA 1 mM, SDS 0,25% (v/v)]. Esta muestra se incubó con 1 µl de RNAsa A (10 mg/mL) durante 15 min a 37°C. A continuación, se añadieron 3 µl de proteinasa K (20 mg/mL) y se incubó 4 h a 65°C en un termobloque con agitación (300 rpm) para revertir el *crosslinking*. Seguidamente se añadieron 500 µl de fenol-cloroformo-isoamílico (25:24:1), se agitó con ayuda de un vórtex y se centrifugó a 17000g durante 5 min. Al sobrenadante se le añadieron 10 µg de glicógeno y 2,5 volúmenes de etanol para precipitar el DNA durante 2 h a -80°C. A continuación, se centrifugó a máxima velocidad durante 30 min en una microfuga a 4°C y el precipitado restante se lavó dos veces con etanol al 70% (v/v). Finalmente se resuspendió en 50 µl de tampón TE [Tris-HCl 10 mM pH 8,0, EDTA 1 mM]. El DNA purificado se cuantificó en un nanodrop (DO 260 nm) o en un fluorómetro QUBIT para mayor sensibilidad. De esta forma, las distintas muestras se equilibraron de acuerdo a su concentración antes de realizar la inmunoprecipitación.

Un 10% de la cromatina se reservó como *input* y el resto se utilizó para la inmunoprecipitación. Los complejos proteína-DNA se inmunoprecipitaron durante una noche usando anticuerpo anti-CO (Valverde et al., 2004) y agarosa magnética GFP-Trap[®] (anti-GFP acoplada a bolas magnéticas) (Chromotek) para capturar las proteínas CO y GFP:HY5, respectivamente. En el caso de la inmunoprecipitación de CO, la solución se incubó además durante 3 h con gránulos magnéticos Dynabeads (Invitrogen) para capturar los inmunocomplejos.

En ambos casos, se realizaron tres lavados de las perlas con soluciones que contenían concentraciones decrecientes de sal: el tampón de lisis nuclear, el tampón de lavado de alta sal [NaCl 500 mM, SDS 0,1% (v/v), Tritón X-100 1% (v/v), EDTA 2 mM, Tris-HCl 20 mM pH 8], el tampón de lavado de LiCl [LiCl 250 mM, NP-40 1% (v/v), deoxicolato de sodio 1% (p/v), EDTA 1mM, Tris-HCl 10 mM pH 8] y por último dos lavados con tampón TE.

Tras el último lavado, los inmunocomplejos se eluyeron de las partículas con 300 µl de una solución que contenía un 1% (v/v) de SDS y NaHCO₃ al 0,1% (v/v), incubándolos 30 min a 65°C en agitación (1200 rpm). Una vez recogido el eluato, se procedió a la purificación del DNA. El *crosslinking* entre las proteínas y el DNA se revirtió mediante la incubación durante una noche con 5 µl de NaCl 5M a 65°C. Al día siguiente, se eliminaron los restos de RNA añadiendo 10 ug de RNAsa A incubando la muestra 30 min a 37°C. Después, se añadieron 50 ug proteinasa K para eliminar las proteínas a 45°C durante 1 h. El DNA se aisló con el método de fenol-cloroformo-isoamílico descrito anteriormente y se resuspendió en 50 µl de tampón TE.

En el experimento de CHIP para CO se usaron las líneas 35S:CO y *co-10* como control. Por otra parte, para identificar los sitios de unión de HY5 se utilizaron las líneas 35S:*GFP:HY5 (hy5-211)* y Col-0 como control.

Posteriormente, para determinar el enriquecimiento en los sitios de unión del genoma, se usó 1 µl del DNA aislado para las reacciones de QPCR. El programa utilizado fue el mismo que se describe en la Tabla 8, pero además se incluyeron reacciones utilizando como molde 3 diluciones seriadas de DNA genómico, con el objetivo de obtener una curva de calibración. Así, se tuvieron en cuenta las diferencias en la eficiencia de los cebadores y se pudo comparar el enriquecimiento obtenido en distintos amplicones. Los cebadores usados se muestran en la Tabla 10.

Tabla 10. Cebadores empleados para medir el enriquecimiento de unión de los factores de transcripción en las diferentes regiones del DNA.

	Secuencia directa	Secuencia reversa	Tamaño del fragmento amplificado (pb)
TSS <i>ACT2</i>	GCGATGTTTGAGTTTCAATAAAC GCTGC	CTCACCTTCACCATTCCAGTTCCA	286
G-box <i>GI</i>	ATCACGAATCGTATGGAGATCAT TA	CAATCAGAGTGGAGCAAGAGAT	110
G-box <i>PRR5</i>	AGGTGAAAGACTGTGTCAGATAA	TGTTGGGTGGGATTAGATTAGAG	106
CORE1/2 <i>FT</i>	GTGGCTACCAAGTGGGAGAT	TAACTCGGGTCGGTGAAATC	201
CORE3 <i>FT</i>	GCTCAAACATGTTGCTCGAA	TGCGATCAGTAAAATACACAGACA	287

La secuenciación se realizó en la unidad de genómica del Cabimer (Centro Andaluz de Biología Molecular y Medicina Regenerativa). Se usaron dos réplicas biológicas para la preparación de las librerías, usando el kit *ThruPLEX DNA-seq* de *Takara Bio Europe*. Las librerías se secuenciaron en el secuenciador *NextSeq500* de *Illumina*. El análisis de los datos de ChIP-seq de CO se realizó como se describe en el apartado 1.3 de Materiales y Métodos.

14 Expresión transitoria en *Nicotiana benthamiana*

14.1 Agroinfiltración

Las cepas de *Agrobacterium* se incubaron en medio LB suplementado con los antibióticos correspondientes para la selección durante una noche a 30°C en agitación. Después, se centrifugaron los cultivos durante 5 min a 3200g y se resuspendió el precipitado en 3 ml de medio MEX [MES 0,97% (p/v) pH 5,6; Glucosa 0,5% (p/v); NaH₂PO₄ 0,024% (p/v); acetosiringona 0,2 mM; AB SALT 5X (v/v)] siendo AB Salts 20X: 20 g NH₄Cl, 6 g MgSO₄ 7H₂O, 3 g KCl, 200 mg CaCl₂, 50 mg FeSO₄ 7H₂O por litro. Este lavado se realizó un total de 3 veces y a continuación se resuspendió el precipitado de bacterias en 3 ml de medio MES 10 mM suplementado con acetosiringona 0,2 mM. Para equilibrar los distintos cultivos que se iban a

agroinfiltrar, se diluyeron con solución MES 10 mM hasta una DO (600 nm) final de 0,5. Se realizaron las combinaciones necesarias para los distintos experimentos añadiendo además el cultivo de *Agrobacterium* que expresa la proteína p19 (Voinnet et al., 2003). Una vez preparados los cultivos, se realizó una pequeña herida en la parte abaxial de cada hoja de tabaco a infiltrar con ayuda de una aguja estéril y seguidamente, empleando una jeringuilla sin aguja se forzó la entrada de la solución en el tejido manteniendo una presión uniforme.

14.2 Ensayos de BiFC

Durante los ensayos de BiFC, las secuencias de cDNA de *CO*, *HY5*, *PRR5* y *FLM* se clonaron en los vectores Gateway pYFN43 y pYFC43 (Belda-Palazón et al., 2012), que fusionan las partes N y C-terminal de la YFP (*Yellow Fluorescent Protein*) al extremo carboxilo de las proteínas. La estirpe GVG3101 pmp90 de *Agrobacterium* se transformó con estas construcciones y se infiltraron en hojas de *Nicotiana* de 4 semanas según lo descrito en el apartado 14.1 de Materiales y Métodos. A los 2-3 días se visualizaron en el microscopio confocal Leica TCS SP2/DMRE. La YFP reconstituida se excitó con un láser de argón ajustado a 514 nm con una potencia del 20%, y su fluorescencia se detectó entre 515 y 565 nm.

14.3 Ensayos de colocalización y FRET

Para visualizar la colocalización de 3 proteínas y para los ensayos de Transferencia de energía de resonancia de Förster (FRET, por sus siglas en inglés), las secuencias de cDNA de *HY5* y *PRR5* se fusionaron a los fragmentos C- y N- terminales de la YFP, respectivamente. Por otra parte, los cDNAs de *CO* y *FLM* se clonaron en los vectores pGWB644 y pGWB642, que producen fusiones C-terminales de las proteínas fluorescentes CFP e YFP (Shyu et al., 2008). Se agroinfiltraron hojas de *Nicotiana* siguiendo el protocolo descrito en el apartado 14.1 y 2-3 días después se visualizaron en el microscopio confocal. La proteína CFP se excitó con una línea de láser de 458 nm y se recogió su emisión en una ventana de 465-479 nm. La YFP reconstituida se excitó con un láser de 514 nm y su emisión se detectó entre 520 y 545 nm.

La colocalización se midió usando el coeficiente de correlación de Pearson, implementado en el *plugin coloc2* de Fiji (distribución de ImageJ) (Schindelin et al., 2012).

Los valores de FRET se midieron mediante la técnica *sensitized emission* usando un software implementado en el microscopio Olympus y los análisis posteriores se realizaron utilizando el *plugin PixFret* para Fiji (Feige et al., 2005). Se determinó el *bleed-through* para el aceptor y el donador, el fondo o ruido se sustrajo y se calcularon los valores N_{FRET} para cada núcleo, ya que este método permite normalizar las diferencias en los niveles de proteína fluorescente (Xia y Liu, 2001).

15 Tiempo de floración

Para medir el tiempo de floración, se sembraron las plantas en semillas en tierra después de estratificarlas durante 4 días a 4°C. Cada línea analizada se sembraba junto con la cepa silvestre y se anotaba en cada caso el número de hojas de roseta en el momento de la aparición del botón floral

16 Movimiento de hojas

Para monitorizar el movimiento de hojas de *Arabidopsis*, se estratificaron las semillas durante 3-5 días a 4°C en oscuridad. Después se germinaron en microtubos de 5 ml rellenos de sustrato (perforados por el fondo), y colocados sobre un pequeño depósito de agua con el objetivo del riego continuado y así minimizar los movimientos producidos por la entrada y salida de agua. Las plántulas se crecieron en LD con una intensidad lumínica de $120 \mu\text{Em}^{-2}\text{s}^{-1}$ durante 10 días, para transferirse posteriormente a una cámara de cultivo en luz continua (LL).

Las dos primeras hojas fueron fotografiadas cada 6 min durante 6 días usando una cámara acoplada a un ordenador de placa reducida Raspberry Pi. El movimiento de las hojas se obtuvo usando el software TRIP (Greenham et al., 2015) como el movimiento vertical en función del tiempo. El periodo del movimiento y otros parámetros se analizaron con el software BioDare (que hace uso de la Transformada

Rápida de Fourier seguida del método de mínimos cuadrados no lineales; FFT-NLLS) (Zielinski et al., 2014) y con el paquete de R *CircaCompare* (Parsons et al., 2020).

Los scripts utilizados en este trabajo de investigación se encuentran disponibles en GitHub a través del siguiente link: <https://github.com/pedrodelosreyes>.



Capítulo 1

Conservación de patrones diurnos en el linaje verde



1 La mayoría de proteínas de *Ostreococcus* presentan un ortólogo en *Arabidopsis* o *Chlamydomonas*, pero no ocurre lo mismo entre las otras dos especies.

La identificación de potenciales ortólogos entre dos especies es uno de los principales cuellos de botella en la genómica y transcriptómica comparativa (Dessimoz et al., 2012). El algoritmo *Bidirectional Best Hit* (BBH) es uno de los métodos más usados para la identificación automática de potenciales ortólogos. A pesar de su simplicidad, el algoritmo BBH es muy preciso manejando genomas bacterianos y de arqueas (Wolf y Koonin, 2012). Sin embargo, este algoritmo no funciona de forma óptima en genomas con alta tasa de duplicación génica como en plantas y animales, perdiendo hasta el 60% de ortólogos (Dalquen y Dessimoz, 2013). En este trabajo, para identificar ortólogos entre especies evolutivamente distantes, hemos desarrollado una variante del algoritmo BBH llamada *Multiple Best Bidirectional Hit* (MBBH). Este método asigna a cada gen g_i^1 de la primera especie $k \leq N$ ortólogos potenciales de la segunda especie, g_1^2, \dots, g_k^2 , basado en similitud de secuencia de las proteínas codificadas si, y sólo si, el gen de interés g_i^1 está entre los N genes más similares de la primera especie (Figura 17).

Usando la herramienta MBBH, más del 97% de las proteínas de *Ostreococcus* pudieron asociarse a un ortólogo potencial de *Arabidopsis* o *Chlamydomonas* (Tabla Online 1). Este resultado es consistente con el reducido y compacto genoma de *Ostreococcus* (Palenik et al., 2007) y sugiere, como había sido propuesto anteriormente (Raven et al., 2013), que este picoeucariota es el mínimo organismo fotosintéticamente activo y que su expresión génica está casi completamente regulada por el reloj circadiano. Confirmando esta idea, el análisis de enriquecimiento en términos GO realizado sobre el grupo de genes sin potencial ortólogo en *Arabidopsis* o *Chlamydomonas* no produjo ningún resultado significativo. De hecho, la mayor parte de genes sin ortólogo se localizan en los cromosomas 2 y 19. Algunos de estos genes codifican para varias proteínas no anotadas (ostta02g03690, ostta02g05500, ostta02g00800) y un grupo de transposasas mariner/Tc1-like (ostta02g01245,

osta02g01247, osta02g02355). Esto apoya resultados previos sobre la heterogeneidad del genoma de *Ostreococcus*, en los que se describen los cromosomas 2 y 19 como diferentes en términos de contenido en GC, uso de codones y número de elementos transponibles (Derelle et al., 2006).

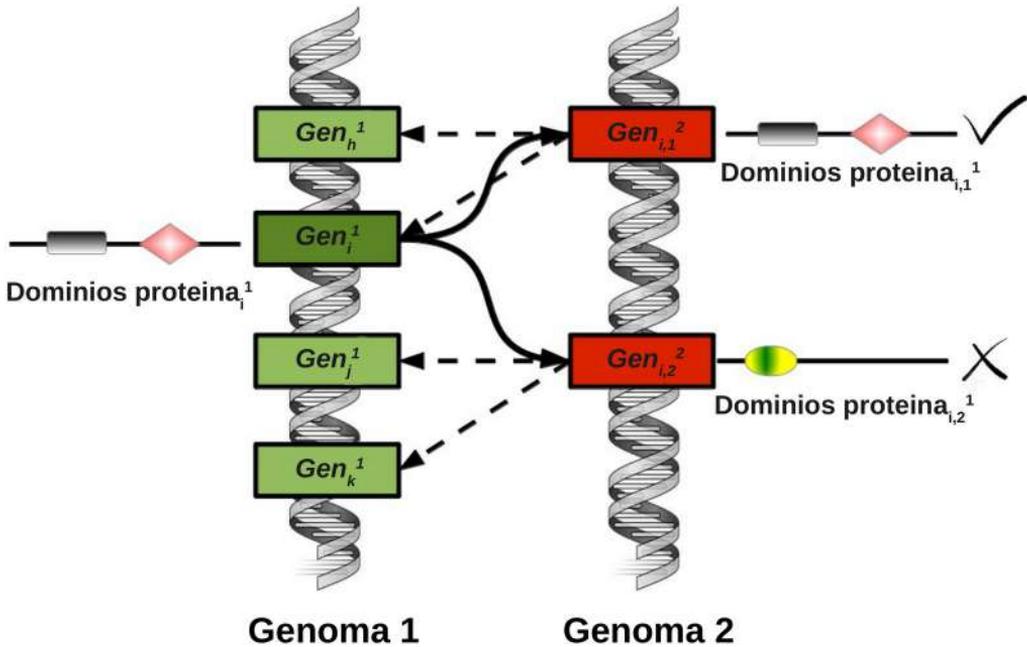


Figura 17. Diagrama explicativo de la herramienta *Multiple Bidirectional Best Hits* (MBBH). Para la $Proteína_i^1$ codificada por el Gen_i^1 de la primera especie (verde oscuro), el algoritmo selecciona los genes (rojo) que codifican las N proteínas que muestran la mayor similitud de secuencia ($proteína_{i,1}^2$, $proteína_{i,2}^2$, ... $proteína_{i,n}^2$). En este estudio se ajustó N a 20. A estas 20 proteínas se les denomina *initial best hits*. Se realiza el mismo proceso para la segunda especie. Por cada *initial hit* encontrado en de esta especie (rojo), se seleccionan los N genes (verde claro) que codifican las proteínas con mayor similitud de secuencia. Entonces los *initial best hits* que se encuentran para el el gen que codifica la $proteína_i^1$ de la primera especie se filtran. Sólo se mantienen aquellos genes que tuvieron como *initial best hit* a la $proteína_i^1$ (marcado con el símbolo ✓). Por tanto, se requiere bidireccionalidad. Además, los *best bidirectional hits* que no comparten al menos un dominio proteico se eliminan (marcado con el símbolo X).

Siguiendo el mismo análisis comparativo, aproximadamente el 85% de las proteínas de *Chlamydomonas* presentaron potenciales ortólogos en *Ostreococcus* y *Arabidopsis* (Tabla Online 2). Por tanto, un grupo de en torno a 2600 genes de *Chlamydomonas* que codifican para proteínas específicas no pudieron ser asociadas a

ningún ortólogo en las otras dos especies. El análisis de enriquecimiento de ontología génica llevado a cabo sobre este grupo de genes usando el paquete de R *topGO* y resumido con *REVIGO* reveló que los 3 términos más enriquecidos y no redundantes fueron “*Movement of cell or subcellular component*”, “*Organic cyclic compound metabolism*” y “*DNA integration*” (Figura 18A). El mismo grupo de genes fueron asociados con los términos GO “*Movement of cell or subcellular component*”, “*Anatomical structure homeostasis*” y “*single-organism process*”. Por tanto, los últimos dos términos GO se consideraron redundantes y sólo “*Movement of cell or subcellular component*” fue considerado. De forma similar, los términos “*Organic cyclic compound metabolism*” y “*cellular aromatic compound metabolism*” incluyeron la misma lista de genes que “*Organic cyclic compound metabolism*” por lo que sólo el último se tomó en consideración.

Diferentes genes que codifican proteínas flagelares como *flagellar inner arm dynein 1* (Cre14.g624950) y *axonemal dynein heavy chain 6* (Cre14.g627576) (Figura 18B) son representativos del grupo “*Movement of cell or subcellular component*”. Las células de *Chlamydomonas* exhiben dos flagelos polares mientras que en las células de *Ostreococcus* o *Arabidopsis* estas estructuras no están presentes, lo que explica la ausencia de ortólogos en estos organismos. Además, los genes específicos de *Chlamydomonas* anotados con el término GO “*Organic cyclic compound metabolism*” incluyen las familias ciclasas de adenilil y guanilil clase III, incluyendo genes como la *adenilato/guanilato ciclasa CYG11* (Cre07.g320700) (Figura 18C). Estas enzimas que catalizan en animales la síntesis de GMP y AMP cíclico no existen en plantas, explicando la ausencia de potenciales ortólogos en *Ostreococcus* y *Arabidopsis*. De hecho, esto coincide con la idea de que algunos genes de *Chlamydomonas* están más cercanos a animales que a plantas (Merchant et al., 2007). Por otra parte, el único retrotransposón *gag-pol-related* (Cre01.g045850) (Figura 18D) es un ejemplo de los genes específicos de *Chlamydomonas* asociados al término de GO “*DNA integration*”, lo que sugiere que este grupo de virus patógenos que infectan *Chlamydomonas* es específico para esta microalga.

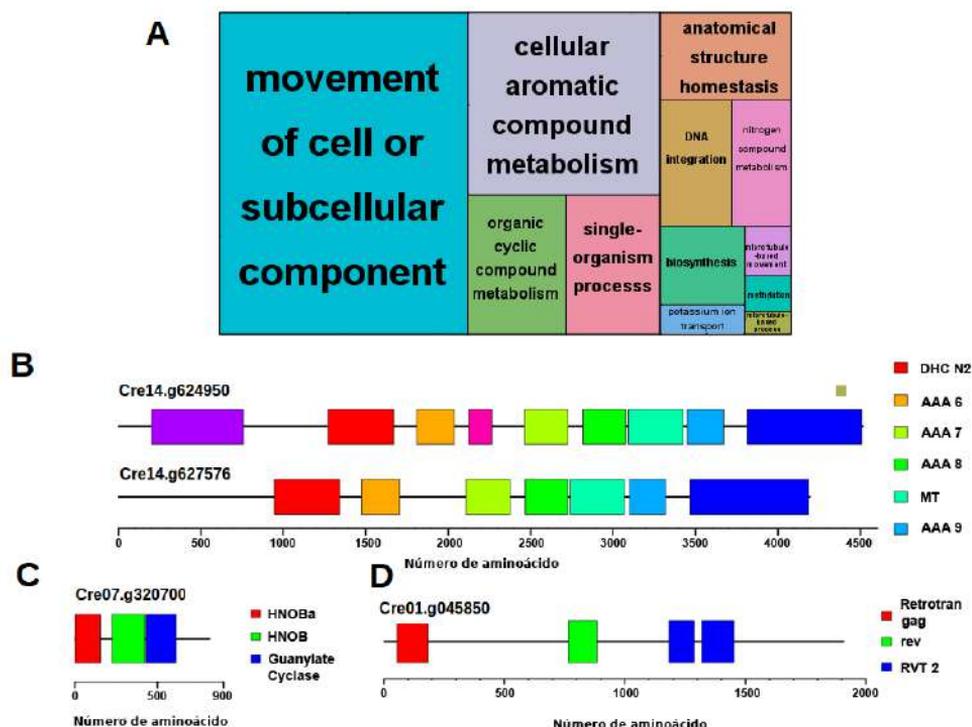


Figura 18. Anotación funcional de genes identificados exclusivamente en *Chlamydomonas*. No se encontraron potenciales ortólogos en *Ostreococcus* o *Arabidopsis* para el 15% de las proteínas de *Chlamydomonas*. (A) Enriquecimiento en términos GO sobre el grupo de genes de *Chlamydomonas* sin potencial ortólogo en las otras dos especies, encontrando los términos “movement of cell or subcellular component”, “organic cyclic compound metabolism” y “DNA integration”. Cada área en el treemap representa el $-\log_{10}(\text{p-valor})$ para el correspondiente término GO. (B) Estructura de las proteínas Flagellar inner arm dynein 1 (Cre14.g624950) y axonemal dynein heavy chain 6 (Cre14.g627576), dos proteínas específicas de *Chlamydomonas* involucradas en movimiento celular. (C) Estructura proteica de la ciclase de guanilil y adenilil clase III codificada por el gen *CYG11* (Cre07.g320700), exclusivo de *Chlamydomonas* e involucrado en el metabolismo de componentes cíclicos orgánicos. (D) Estructura de proteínas del retrotransposón *gag-pol-related* (Cre01.g045850), un ejemplo de proteínas específicas de *Chlamydomonas* asociadas al término “DNA integration”. Las cajas de color en las secuencias de aminoácidos representan dominios identificados en la base de datos *Pfam* y cuyo código aparece a la izquierda de cada representación proteica.

Nuestro análisis también identificó potenciales ortólogos en *Ostreococcus* y *Chlamydomonas* para aproximadamente el 75% de las proteínas de *Arabidopsis* (Tabla Online 3), sugiriendo que un cuarto de estas proteínas no están presentes en

microalgas y han podido ser adquiridas durante la evolución hacia plantas superiores. De acuerdo con esta idea, un análisis de enriquecimiento en términos GO mostró que estas proteínas específicas de *Arabidopsis* están involucradas en los siguientes tres términos significativos no redundantes: “*Cell communication*”, “*Cell wall organization*”, y “*Multiorganism reproductive process*” (Figura 19A). De forma similar a los resultados previos, el término de GO “*Signaling*” compartía la misma lista de genes que “*Cell communication*”; “*Cellular glucan metabolism*” presentó los mismos genes asociados que “*Cell wall organization*” y el grupo de genes asignados a “*Biological regulation*” y “*Multiorganism reproductive process*” también fueron idénticos. Por tanto, sólo se discute el segundo término GO mencionado. Estos términos GO corresponden a complejas características vegetales propias de la multicelularidad que están ausentes en microalgas. Así, proteínas clave involucradas en biogénesis de la pared celular y metabolismo celular del glucano en *Arabidopsis* como Pectin Methylesterase 1 (PME1, At1g53840) no presentan potenciales ortólogos en *Chlamydomonas* u *Ostreococcus* (Figura 19B). En efecto, la historia evolutiva de los genes *PMR* ha sido estudiada previamente, estableciéndose su aparición en las algas Charofitas multicelulares, lo que explica su ausencia en algas más primitivas (Wang et al., 2013). Un caso concreto de proteína asociada a “*Multi-cellular processes*” sin potenciales ortólogos en *Chlamydomonas* y *Ostreococcus* fue la NAC domain containing protein 98 de *Arabidopsis* (ANAC098, At5g53950) (Figura 19C). La familia de factores de transcripción NAC es una de las más grandes en plantas y está involucrada en múltiples procesos clave como el desarrollo floral. De nuevo, esta familia de factores de transcripción se originó en Charofitas (Zhu et al., 2012) y no aparece en *Chlamydomonas* y *Ostreococcus*. Finalmente, la Recognition of Peronospora Parasitica 1 (RPP1, At3g44480) es un ejemplo específico de una proteína involucrada en “*Cell communication/signaling*” y es parte del grupo específico de proteínas de *Arabidopsis* relacionadas con resistencia a patógenos y muerte celular programada (Figura 19D) (Schreiber et al., 2016).

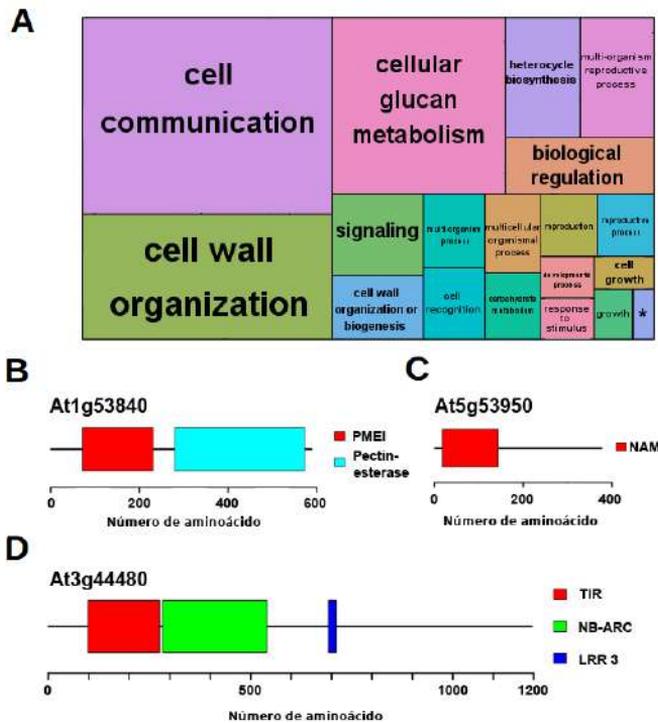


Figura 19. Anotación funcional de los genes identificados exclusivamente en *Arabidopsis*. No se encontraron ortólogos en *Chlamydomonas* o en *Ostreococcus* para un cuarto de las proteínas de *Arabidopsis*. (A) El enriquecimiento en términos GO para el grupo de genes específicos de *Arabidopsis* indica que están involucrados en “cell communication/signaling”, “cell wall organization”, “cellular glucan metabolism” y “multiorganism reproductive process.” (*) “single-organism developmental process.” Cada rectángulo del *treemap* representa el $-\log_{10}(\text{p-valor})$ para el correspondiente término GO. (B) Estructura proteica codificada por el gen *Pectin Methyl-esterase 1*, (*PME1*, At1g53840), uno de los genes específicos involucrados en organización de la pared celular y metabolismo del glucano. (C) Un caso de proteína específica de *Arabidopsis* anotada con el término GO “multi-organism reproductive process”, AC domain containing protein 98 (ANAC098, At5g53950), que participa en el desarrollo floral. (D) Estructura de la proteína codificada por el gen *Recognition of Peronospora Parasitica 1* (*RPP1*, At3g44480), uno de los genes exclusivos de *Arabidopsis* involucrado en señalización celular. Las cajas de color en las secuencias de aminoácidos representan dominios identificados en la base de datos *Pfam*.

2 Grandes familias de factores de transcripción de *Arabidopsis* tienen un único ortólogo en microalgas

El método MBBH para definir potenciales ortólogos puede detectar varios candidatos para cualquier gen de interés, identificando múltiples genes homólogos que pueden haber aparecido por procesos de duplicación génica. En *Arabidopsis*, de media, 5,19 y 7,83 genes pueden ser asociados a un único homólogo en *Chlamydomonas* y *Ostreococcus*, respectivamente, indicando que las familias génicas de *Arabidopsis* son de media 5 a 8 veces más grandes que en las de las microalgas estudiadas. Dicho resultado concuerda con los eventos de duplicación de genoma completo y duplicación génica que fueron cruciales en la evolución de familias de genes de plantas (Romero-Campero et al., 2013; Rensing, 2014). Curiosamente, este proceso es particularmente frecuente en familias de factores de transcripción y la herramienta MBBH ha detectado eficientemente dominios funcionales en las secuencias proteicas de las tres especies estudiadas e identificado la familia de factores de transcripción a la que pertenecen. En general, sus tamaños coinciden con los datos disponibles en la *Plant Transcription Factor Database*, PlantTFDB (Jin et al., 2017), confirmando la fiabilidad de esta aproximación. En promedio, menos de 4 y 7 proteínas forman cada familia de factores de transcripción en *Ostreococcus* y *Chlamydomonas*, respectivamente, mientras que una media de 30 miembros constituyen las familias de *Arabidopsis*, lo que apoya la idea de múltiples eventos de duplicación.

Dos claros ejemplos de amplificación en factores de transcripción son las familias de proteínas DOF y COL (Figura 20A,B) que presentan un único miembro en *Chlamydomonas* y dos miembros en *Ostreococcus*, respectivamente, mientras en *Arabidopsis* existen 47 proteínas DOF y 22 COL. Un subgrupo de genes *DOF* están regulados por el reloj (*CYCLING DOF FACTORS*, *CDFs*), y a su vez controlan la expresión de *CO* en *Arabidopsis* y patata durante la floración fotoperiódica (Imaizumi et al., 2005; Fornara et al., 2009; Kloosterman et al., 2013). Esta ruta constituye un módulo DOF-CO conservado en espermatofitas que también está presente en *Chlamydomonas* (Lucas-Reina et al., 2015). En *Ostreococcus* podría ser diferente. Mientras que los ortólogos de *Chlamydomonas* y *Arabidopsis* presentaron un único

dominio proteico DOF, en una de las dos proteínas DOF de *Ostreococcus* (OtDOF2, ostta04g02850) se identificó un dominio *Response Regulator* N-terminal adicional con un potencial motivo *phosphoacceptor aspartic acid-aspartic acid-lysine* (DDK) (Figura 20A). Esto sugiere que OtDOF2 puede ser parte de un sistema fosforelay (*phosphorelay*) (Djouani-Tahri et al., 2011). La familia génica *COL* (Figura 20B) parece haber surgido en microalgas, con un sólo miembro representativo en *Chlamydomonas* (*CrCO*, Cre06.g278159) (Valverde, 2011) y dos potenciales ortólogos en *Ostreococcus* (*OtCOL1*, ostta04g03620 y *OtCOL2*, ostta09g01510). En *Chlamydomonas*, *CrCO* está involucrado en ritmo circadiano, control del ciclo celular, síntesis de almidón, contenido lipídico y mecanismos de fotoprotección del cloroplasto (Serrano et al., 2009; Deng et al., 2015; Gabilly et al., 2019; Tokutsu et al., 2019; Matsuo et al., 2008) y algunas de estas funciones han sido conservadas en algunos ortólogos de *Arabidopsis* (Romero-Campero et al., 2013; Isabel Ortiz-Marchena et al., 2014).

Las familias de los factores de transcripción MADS-box y PRR (Figura 20C,D) constituyen otros casos interesantes. La familia MADS-box es un ejemplo de amplificación frecuente a lo largo de la evolución vegetal, ya que una única copia está presente en *Chlamydomonas* y *Ostreococcus* mientras que en *Arabidopsis* esta familia tiene 146 miembros. Las proteínas MADS-box han sido clasificadas en diferentes subfamilias dependiendo del reclutamiento de nuevos dominios proteicos además del dominio MADS (MCM1, AGAMOUS, DEFICIENS, SRF1) de unión al DNA, como las K-box (contienen una región conservada llamada caja K; Figura 20C), explicándose así la gran diversidad funcional de esta familia de factores de transcripción en plantas superiores. Por otra parte, los más recientes PRRs (Figura 20D) han sufrido una amplificación génica. Estas proteínas presentan un dominio N-terminal Pseudo Response Regulator y un dominio CCT C-terminal, pero no ocurre lo mismo en microalgas. Únicamente encontramos una copia en *Ostreococcus* (*OtPPR*, ostta13g01820) y dos en *Chlamydomonas* (*CrPPR1*, Cre02g.094150; *CrPPR2*, Cre16.g676421). Además, estas proteínas presentan un potencial motivo fosfoceptor DDK.

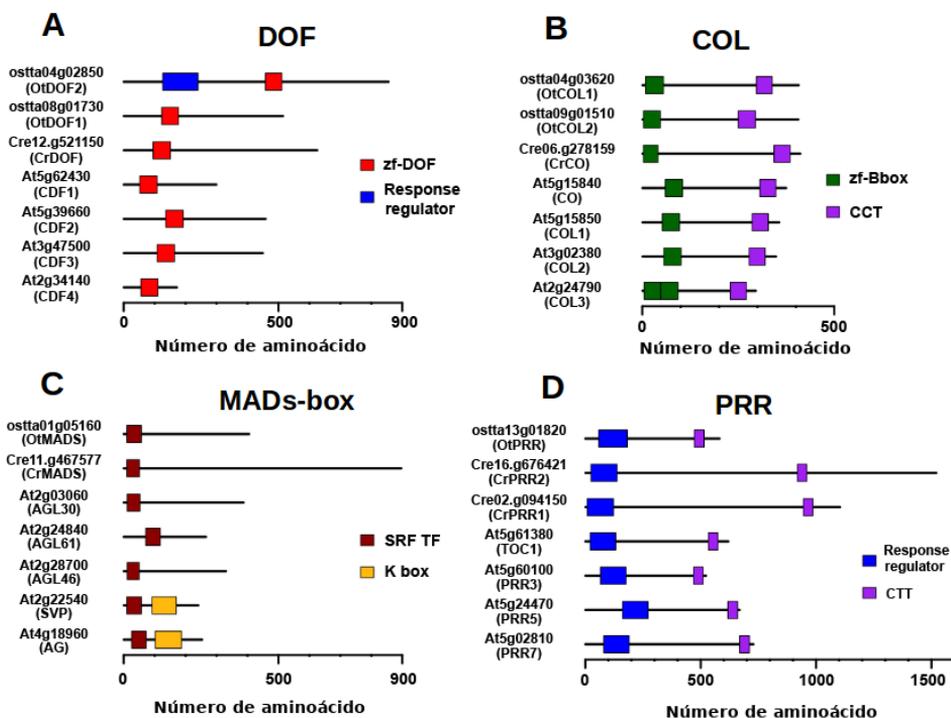


Figura 20. Amplificación de familias de factores de transcripción durante la evolución. Las familias de *Arabidopsis* son en promedio 5 y 8 veces más grandes que las de *Chlamydomonas* y *Ostreococcus*, respectivamente, lo que indica la existencia de eventos de duplicación génica. (A) Estructura de las proteínas DOF. *Ostreococcus* tiene dos miembros, uno de ellos posee un dominio *Response Regulator* en extremo N-terminal (*OtDOF2*), mientras que en *Chlamydomonas* sólo encontramos uno. De los 47 miembros que aparecen en *Arabidopsis*, los más similares a los DOFs de algas son los CDFs, involucrados en la floración (B) Estructura proteica de la familia de factores de transcripción COL. Tiene dos miembros en *Ostreococcus*, uno sólo en *Chlamydomonas* y 22 en *Arabidopsis*. Se muestra CO y sus homólogos más similares. (C) Estructura proteica de los factores de transcripción de la familia MADs-box. Esta familia ha sufrido un intenso proceso de amplificación desde una sólo copia en *Ostreococcus* y *Chlamydomonas* hasta una familia multigénica de 146 miembros en *Arabidopsis*. (D) Estructura de proteínas de la familia PRR. A pesar de su relativo pequeño tamaño en *Arabidopsis* (10 miembros), también ha experimentado un proceso de amplificación génica, desde una sólo copia en *Ostreococcus* y dos en *Chlamydomonas*. Las cajas de color en las secuencias de aminoácidos representan dominios identificados en la base de datos *Pfam*.

Con el objetivo de analizar la relación entre la conservación en la secuencia de aminoácidos - estructura de proteínas y la conservación de los perfiles de expresión, se analizaron datos transcriptómicos de las tres especies a lo largo de 24 h. Como era de esperar, se observaron diferencias en los perfiles de expresión de genes de *Arabidopsis*; algunos genes mantienen el mismo patrón que los de *Chlamydomonas* y *Ostreococcus* mientras que otros mostraban otro completamente nuevo (Figura 21). En este sentido, en la familia *DOF* (Figura 21A), los genes de *Arabidopsis* *CDF1* y *CDF2* han conservado un perfil de expresión similar al de *CrDOF* y *OtDOF1*, con un máximo en torno a ZT0 y un mínimo en torno a ZT12. Sin embargo, *CDF4* y *OtDOF2* alcanzan su punto máximo a ZT21, mientras que *CDF3* lo hace a ZT3, con un mínimo a ZT18, lo que no se observa en genes de *Chlamydomonas* u *Ostreococcus*. Una situación parecida se observa para las familias de factores de transcripción *COL* (Figura 21B), *MADS-box* (Figura 21C) y *PRR* (Figura 21D). Los genes de *Arabidopsis* *COL1*, *COL2* y *COL3* presentan un perfil de expresión muy similar a *OtCOL1* y *CrCO*, con un mínimo a ZT12 (Figura 21B). Por otra parte, en los perfiles de expresión de *OtCOL2* y *CO* se observan diferencias en cuanto al momento donde alcanzan su nivel mínimo de expresión, a ZT0 y ZT6, respectivamente. La familia *MADS-box* presenta una gran diversificación en el perfil de expresión de sus miembros, con grandes diferencias entre los patrones génicos de *Arabidopsis*, *Chlamydomonas* y *Ostreococcus* (Figura 21C). En la familia *PRR*, *OtPRR* presenta el mismo perfil de expresión que *PRR1* (*TOC1*) y *PRR3* en *Arabidopsis*, con el punto máximo a ZT12, mientras que *PRR5* y *PRR7* lo presentan antes, en torno a ZT9. Sin embargo, *CrPRR1* y *CrPRR2* muestran patrones completamente diferentes, con el máximo a ZT21 y ZT0, respectivamente (Figura 21D). Esto implicaría que, en general, la amplificación génica viene acompañada de una diversificación en el perfil de expresión y por tanto, de una diversificación funcional (Romero-Campero et al., 2013).

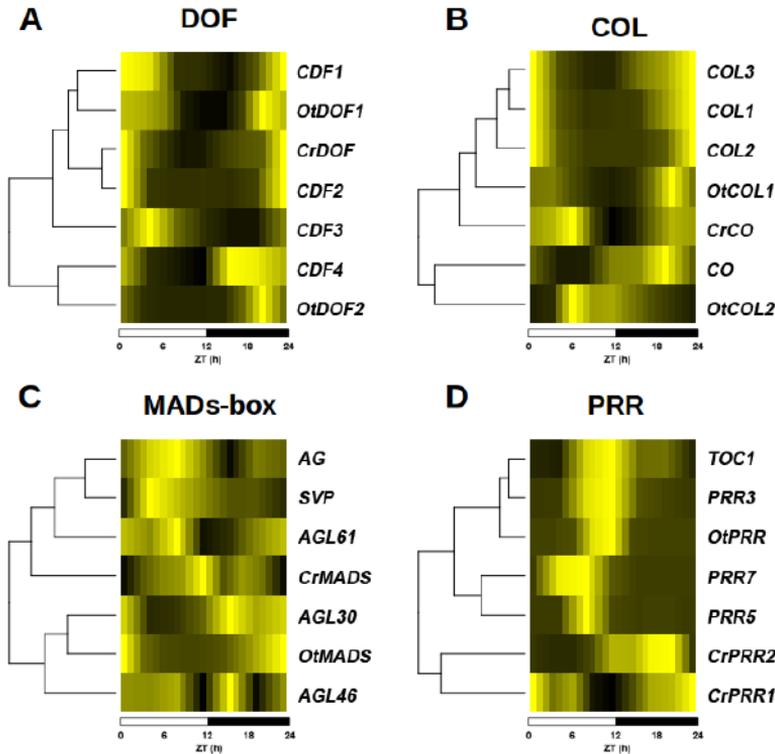


Figura 21. Diversificación de los perfiles de expresión tras la amplificación génica. Se muestran mapas de calor representando la expresión génica de las familias de factores de transcripción de la Figura 20 en condiciones ND. Se usa una escala de color desde amarillo (máxima expresión) hasta negro (mínima expresión). Se empleó un *clustering* jerárquico para ordenar los genes de acuerdo a su expresión génica. **(A)** Perfil de expresión de la familia génica *DOF*. **(B)** Perfil de expresión de la familia génica *COL*. **(C)** Perfil de expresión de la familia génica *MADS-box*. **(D)** Perfil de expresión de la familia génica *PRR*. Debajo de cada gráfico, se indica la escala de tiempo *Zeitgeber Time* (ZT) en horas (h).

3 La mayoría de genes en *Ostreococcus* y en *Chlamydomonas* exhiben un patrón rítmico, al contrario que en *Arabidopsis*

Recientemente, se han generado datos masivos transcriptómicos de organismos fotosintéticos crecidos en diversas condiciones ambientales y fisiológicas. Uno de las señales ambientales más estudiadas es el cambio en los ciclos de día y noche. No obstante, el análisis de cambios diurnos en el transcriptoma de diferentes

especies se llevó a cabo independientemente en diferentes estudios, lo que hace necesario la aplicación de técnicas de Biología Molecular de Sistemas para integrarlos y compararlos. En este estudio se han analizado datos de *microarray* generados a lo largo de un periodo de 24 h en condiciones ND para *Ostreococcus* (Monnier et al., 2010) y *Arabidopsis* (Bläsing et al., 2005), datos de RNAseq de *Chlamydomonas* (Zones et al., 2015) en condiciones ND y *Arabidopsis* en LD (Rugnone et al., 2013). En la sección de Materiales y Métodos se describen los detalles del procesamiento de los datos. Para determinar el efecto del ciclo de periodos de luz y oscuridad sobre el transcriptoma, identificamos genes que mostraban un patrón de expresión periódico o rítmico con un periodo aproximado de 24 h en las tres especies. Se emplearon métodos no paramétricos implementados en el paquete de R de *Bioconductor RAIN* para detectar patrones significativos de expresión génica con formas de onda arbitrarias y un periodo de 24 h (Thaben y Westermarck, 2014).

El análisis reveló que más del 90% de los genes de *Ostreococcus* detectables en el *microarray* presentaron patrones rítmicos diarios (Tabla Online 1). Por otro lado, no se identificó ningún término GO significativamente enriquecido en los genes no periódicos. Esto sugiere que prácticamente el transcriptoma completo de *Ostreococcus* se regula de manera periódica y que la mayoría de procesos biológicos en esta microalga está altamente afectada por los ciclos día/noche. Sin embargo, en *Chlamydomonas*, aproximadamente el 70% de los genes mostraron perfiles significativos de expresión génica rítmica diaria (Tabla Online 2). En el análisis de enriquecimiento en términos de GO llevado a cabo sobre los genes de regulación no periódica sólo se identificaron unos pocos procesos biológicos significativos. Los dos más significativos fueron “DNA integration” y “Defense response to virus” incluyendo genes como la transcriptasa reversa Cre05.g235102, la ligasa de DNA Cre06.g277801 y la 2'-5' oligoadenilato-sintetasa Cre15.g641050. Se trata probablemente de procesos biológicos inducidos por estímulos bióticos y por tanto, independientes de los ciclos de luz/oscuridad. Finalmente, sólo el 43,18% de los genes de *Arabidopsis* mostraron patrones circadianos significativos de acuerdo a nuestro análisis (Tabla Online 3), sugiriendo que, durante la evolución de las plantas superiores, muchos procesos biológicos fueron desacoplados de la influencia externa de los ciclos de luz y

oscuridad. Estos porcentajes de genes rítmicos van en consonancia con resultados previos (Bläsing et al., 2005; Monnier et al., 2010; Zones et al., 2015). El análisis de enriquecimiento en rutas metabólicas llevado a cabo sobre los genes rítmicos de *Arabidopsis* reveló varias rutas clave significativas influenciadas por cambios rítmicos (Tabla 11). Como se esperaba, la ruta más significativa fue la denominada “*Circadian rhythms in plants*”. También fue esperable encontrar enriquecidas las rutas “*Porphyrin and chlorophyll metabolism*” y “*Pentose phosphate pathway*”, ya que están involucradas en la fotosíntesis y por tanto es probable que estén altamente reguladas por ciclos luz/oscuridad presentando patrones rítmicos. Para algunas rutas metabólicas esenciales como “*Fatty acid degradation*” y “*Starch and sucrose metabolism*” también se detectó enriquecimiento significativo en los genes rítmicos, lo que indica que el metabolismo de *Arabidopsis* aún se regula estrechamente por periodos alternativos de luz y oscuridad. Más sorprendente fue la identificación de “*Alpha-Linolenic acid metabolism*” y “*Plant hormone signal transduction*” entre los términos GO enriquecidos, sugiriendo, bien que estas vías son antiguas o que durante la especiación vegetal estas rutas nuevas involucradas en síntesis de hormonas o señalización fueron reclutadas a un patrón rítmico luz/oscuridad.

Tabla 11. Vías significativamente enriquecidas en los genes de expresión circadiana de Arabidopsis en este trabajo.

Identificador	Descripción	q-valor	Ratio circadiano/total
ath04712	Circadian rhythm - plant	$4,20 \cdot 10^{-6}$	29/36
ath00030	Pentose phosphate pathway	$5,25 \cdot 10^{-3}$	35/58
ath00052	Galactose metabolism	$1,54 \cdot 10^{-2}$	32/55
ath00590	Arachidonic acid metabolism	$1,54 \cdot 10^{-2}$	13/17
ath00966	Glucosinolate biosynthesis	$1,54 \cdot 10^{-2}$	14/19
ath00860	Porphyrin and chlorophyll metabolism	$1,76 \cdot 10^{-2}$	28/48
ath00360	Phenylalanine metabolism	$1,84 \cdot 10^{-2}$	25/42
ath00071	Fatty acid degradation	$2,66 \cdot 10^{-2}$	24/41
ath00500	Starch and sucrose metabolism	$2,66 \cdot 10^{-2}$	66/139
ath00130	Ubiquinone and other terpenoid-quinone biosynthesis	$2,66 \cdot 10^{-2}$	21/35
ath00260	Glycine, serine and threonine metabolism	$3,54 \cdot 10^{-2}$	37/72
ath00592	alpha-Linolenic acid metabolism	$3,54 \cdot 10^{-2}$	21/36
ath01200	Carbon metabolism	$3,62 \cdot 10^{-2}$	114/262
ath01230	Biosynthesis of amino acids	$3,62 \cdot 10^{-2}$	111/255
ath04075	Plant hormone signal transduction	$3,62 \cdot 10^{-2}$	118/273
ath01210	2-Oxocarboxylic acid metabolism	$4,66 \cdot 10^{-2}$	37/74

Con el objetivo de representar los genes rítmicos diarios y sus relaciones complejas de coexpresión, se construyó una red de coexpresión génica para cada especie fotosintética analizada (Figura 22A-C). Los nodos representan los genes rítmicos y las aristas relaciones de coexpresión, de manera que se asume que dos genes circadianos se coexpresan cuando el índice de correlación de Pearson entre los perfiles de 24 h de expresión es mayor que 0,95, trazándose una arista entre ellos. Las tres redes de coexpresión génica se visualizaron usando la disposición *Prefuse Force Directed* implementada en la herramienta software Cytoscape (Shannon et al., 2003).

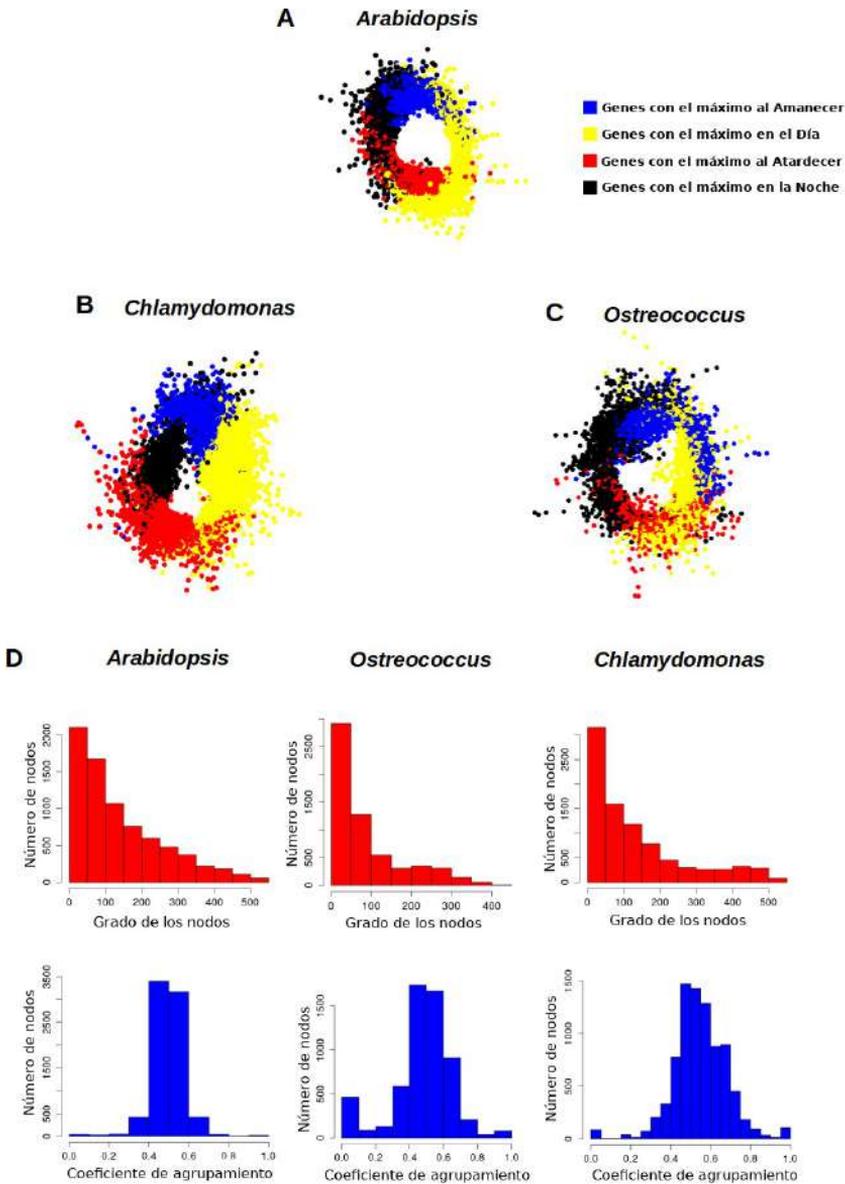


Figura 22. Redes de coexpresión génica circadiana. Los genes se representan por nodos y las aristas indican coexpresión entre ellos. Las tres redes adquirieron forma de anillo reflejando el patrón periódico de los genes rítmicos. (A) Red de coexpresión génica periódica de *Arabidopsis*. (B) Red de coexpresión génica periódica de *Chlamydomonas*. (C) Red de coexpresión génica periódica de *Ostreococcus*. (D) Análisis topológico de las redes de coexpresión indicando que son redes libres de escala y de mundo pequeño. La distribución del grado de sus nodos (en rojo) se ajusta a una potencial negativa. El coeficiente medio de agrupamiento (azul) es significativamente mayor que en redes aleatorias con las mismas propiedades.

Curiosamente, las tres redes de coexpresión adquirieron la misma estructura de anillo, plasmando la relación y el orden cronológico entre los patrones de coexpresión de los genes periódicos. Dos parámetros topológicos básicos, la distribución del grado de los nodos y el coeficiente de agrupamiento fueron calculados para cada red (Figura 22D). La distribución del grado de los nodos de las tres redes sigue una distribución potencial negativa con un p-valor menor que $2,2 \times 10^{-16}$, lo que implica que son redes libres de escala (Barabási, 2015), sugiriendo que los patrones de coexpresión génica para los genes rítmicos diarios en las tres especies son robustos a ataques o mutaciones aleatorios pero frágiles a cambios dirigidos a genes *hub*, aquellos coexpresados con un gran número de genes (Aoki et al., 2007). Además, el coeficiente de agrupamiento de las tres redes fue significativamente alto comparado con el mismo coeficiente de redes libres de escala generadas aleatoriamente con el mismo número de nodos y aristas. Por tanto, las tres redes constituyen redes de mundo pequeño (Barabási, 2015), con caminos relativamente cortos conectando dos nodos cualesquiera. Esta propiedad confiere una rápida propagación de información entre genes con diferentes patrones de expresión génica.

4 Los grupos de genes rítmicos confieren separación temporal de procesos biológicos

El patrón rítmico diario de expresión génica para genes individuales se puede describir por su punto mínimo y por su punto máximo o “pico”, el punto temporal cuando su nivel de expresión es mayor (Figura 23A, Tablas online 1-3). Los datos transcriptómicos usados en las redes de coexpresión fueron recogidos en condiciones ND, pero usando diferentes puntos temporales: mientras que los datos de *Ostreococcus* y *Chlamydomonas* fueron recolectados cada 3 h, los de *Arabidopsis* se recogieron cada 4 h. Para poder comparar los perfiles de expresión de las tres especies, se dividió un día de 24 h en cuatro intervalos temporales. Por tanto, el Amanecer (azul) se definió como el intervalo de tiempo entre ZT21 y ZT3; el Día (amarillo), de ZT3 a ZT9; el Atardecer (rojo) se consideró desde ZT9 a ZT15 y la Noche (negro) se estableció como el periodo desde ZT15 a ZT21 (Figura 23B). Así,

los genes rítmicos diarios se pueden clasificar en 16 grupos diferentes de acuerdo al intervalo donde presentan el máximo y el mínimo.

Debido a la cantidad de datos procesados, la representación gráfica de esos 16 grupos puede resultar confusa, por lo que para facilitar su visualización decidimos unir todos los grupos con el pico en el mismo intervalo en sólo uno. Así, para cada nuevo grupo de genes representamos el perfil de expresión de cada gen y el perfil medio del grupo completo (Figura 23C). Curiosamente, no se observaron diferencias sustanciales entre los perfiles medios de las tres especies analizadas. Por ello, cuando coloreamos los genes de esos *clusters* en las correspondientes redes de coexpresión génica (Figura 22A-C), observamos que éstos presentan la misma localización en las tres redes. Esto constituye una prueba sólida de la veracidad de nuestra aproximación y muestra a simple vista que los genes rítmicos diarios se asocian realmente en grupos temporales de genes desempeñando acciones sincronizadas y reflejando el curso natural de un reloj de 24 h. En este análisis, las fronteras que separan los grupos fueron más marcadas en *Arabidopsis* y *Chlamydomonas* que en *Ostreococcus*, cuya red mostró un mayor grado de mezcla entre los grupos temporales (Figura 22A-C). Esto apoya la idea ya establecida de que el reloj primitivo de *Ostreococcus* no es tan eficiente como el de otros organismos más complejos.

Conservación de patrones diurnos en el linaje verde

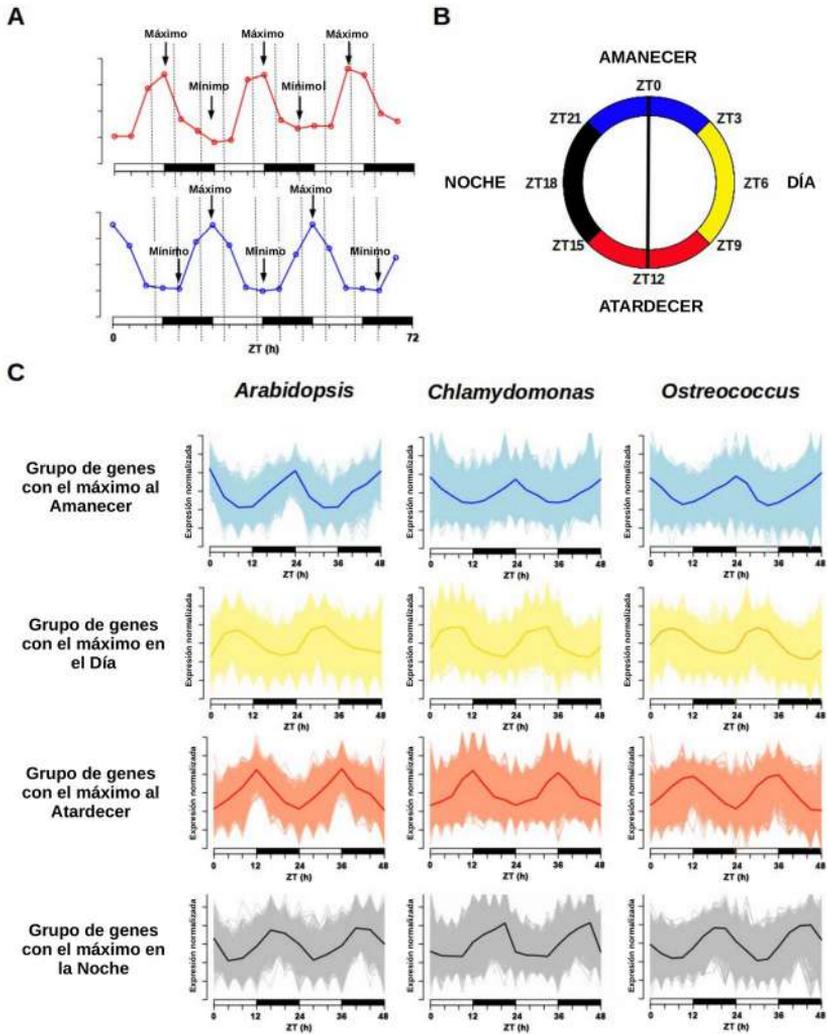


Figura 23. Agrupamiento o clustering rítmico diario. Se describen las ondas de expresión génica por su pico y por su punto mínimo. **(A)** Representación de la expresión génica de dos genes periódicos donde se observan los puntos máximo y mínimo. **(B)** Las 24 h de un día se dividen en cuatro intervalos: “Amanecer” se define como el tiempo entre ZT21 y ZT3 (azul); “Día” se considera de ZT3 a ZT9 (amarillo); “Atardecer” desde ZT9 a ZT15 (rojo) y “Noche” consiste en el intervalo entre ZT15 y ZT21 (negro). **(C)** Cuatro grupos diferentes se definen para la visualización con el mismo código de color: genes con el máximo al Amanecer, en el Día, al Atardecer o en la Noche. Para cada grupo de genes se representa el perfil de expresión normalizada en 48 h, así como la expresión media en un color más intenso.

Los grupos de genes más amplios en *Ostreococcus* y *Chlamydomonas* contienen genes que tienen su máximo de expresión al Atardecer y presentan el

mínimo al Amanecer (1620 y 3751 genes, respectivamente) y aquellos que tienen el máximo al Amanecer y el mínimo al Atardecer (1744 y 2869 genes, respectivamente). Los genes de esos dos grupos constituyen el 56,85 y el 49,98% del total de genes rítmicos diarios en *Ostreococcus* y *Chlamydomonas*, respectivamente, y coinciden con la idea común de que las transiciones luz/oscuridad y oscuridad/luz juegan papeles centrales en la regulación génica de ambas microalgas. Sorprendentemente, en *Arabidopsis* esos dos grupos de genes no son los mayoritarios, siendo sólo el 22,51 % de los genes periódicos. Esto sugiere un grado de desacople entre la regulación de la expresión génica y las transiciones entre luz y oscuridad en plantas superiores. No obstante, el grupo de genes más amplio en *Arabidopsis* contiene 1209 genes que alcanzan su máximo de expresión al Amanecer y tienen su mínimo durante el Día, lo que indica que la transición de oscuridad a luz aún es clave en la regulación de los genes rítmicos en *Arabidopsis*. El segundo grupo más numeroso en *Arabidopsis* contiene 1014 genes cuya expresión presenta el máximo de expresión durante la Noche y el mínimo durante el Día (Anexo 4). Esto sugiere cierta independencia en la regulación de la expresión génica rítmica en *Arabidopsis* puesto que el segundo grupo más importante tiene sus puntos máximo y mínimo en los periodos de luz y oscuridad, y no en las transiciones.

Se realizó un análisis de enriquecimiento en términos de ontología génica sobre los tres grupos génicos específicos con el objetivo de identificar los procesos biológicos llevados a cabo por estos grupos rítmicos. Como era de esperar, se encontraron diferentes términos enriquecidos para cada grupo biológico, indicando una separación temporal entre ellos. Algunos procesos biológicos mostraron enriquecimiento en los mismos grupos para las tres especies diferentes, como “*DNA metabolic process*” (para el grupo de genes con el máximo al Atardecer y el mínimo al Amanecer), “*photosynthesis*” (grupo con el máximo durante el Día y el mínimo por la Noche) y “*carbohydrate catabolic process*” (genes que tienen el máximo al Amanecer y el mínimo al Atardecer), indicando que se conservan patrones en la expresión génica diaria (Tablas S4-S6). Por el contrario, se observa una anticipación, un retraso o directamente un desacople de la expresión periódica en algunos procesos biológicos.

También se encontró enriquecimiento en términos GO relacionados con el ciclo celular, como “DNA metabolic process”, en el grupo de genes con el máximo al Atardecer y mínimo al Amanecer en las tres especies. Los genes *REPLICATION PROTEIN A* (Cre16.g65100, ostta18g01440 y At4g19130), *ORIGIN RECOGNITION COMPLEX 1* (Cre10.g455600, ostta04g05220, y At1g26840) y *CELL DIVISION CYCLE 45* (Cre06.g270250, ostta04g04640, y At3g25100) mostraron perfiles de expresión muy similares en las tres especies (Figura 24A). Curiosamente, los términos GO “DNA replication” y “DNA metabolic process” también están enriquecidos en el grupo de genes con el máximo durante el Día y mínimo al Amanecer, pero únicamente en *Ostreococcus* y *Arabidopsis*, lo que indica una anticipación o unos picos de expresión más amplios en estas dos especies que en *Chlamydomonas*, que muestra picos de expresión más estrechos y localizados en la transición luz/oscuridad (Figura 24A). Por ejemplo, los genes *DNA POLYMERASE A4* (Cre07.g312350, ostta13g02040, y At5g41880), *MINI CHROMOSOME MAINTENANCE 2* (Cre07.g338000, ostta01g02580, y At1g44900), *CYCLIN A1* (Cre03.g207900, ostta02g00150, y At1g44110) muestran un pico estrecho centrado en ZT12 en el caso de *Chlamydomonas*, mientras que *Ostreococcus* y en *Arabidopsis* tienen un pico más extendido centrado en ZT6. Esto podría indicar una mejor sincronización del cultivo en *Chlamydomonas* que en *Ostreococcus*, y una heterogeneidad en los diferentes tejidos de *Arabidopsis*. El término GO “Photosynthesis” se encontró enriquecido en los grupos de genes con el pico de expresión durante el día y el mínimo durante la noche en las tres especies estudiadas (Tablas S4-6). Concretamente *PHOTOSYSTEM I SUBUNIT D* (Cre05.g238332, ostta10g03280, y At1g03130) y *PHOTOSYSTEM II SUBUNIT O* (Cre09.g396213, ostta14g00150, and At3g50820) presentan perfiles de expresión muy similares, más extendidos de nuevo en *Arabidopsis* y *Ostreococcus* que en *Chlamydomonas* (Figura 24B). Solamente se detectaron pequeñas variaciones en la expresión de genes que codifican componentes del complejo ATP sintasa y el complejo del citocromo b6f. Los genes *ATP CHLOROPLAST SYNTHASE 1 (ATPC)*, (Cre06.g259900, ostta09g01080, y At4g04640) y *PHOTOSYNTHETIC ELECTRON TRANSFER C (PETC)*, (Cre11g.467689, ostta07g02450, y At4g03280) muestran un pico al amanecer y un mínimo al atardecer en *Chlamydomonas* y *Ostreococcus*, en

cambio en *Arabidopsis* *ATPC1* alcanza el máximo más tarde y *PETC* se anticipa, presentándolo durante la noche (Figura 24B).

El proceso de generación de ribosomas o ribogénesis es otro caso interesante de conservación y evolución de los patrones de expresión periódicos de un proceso biológico. Efectivamente, el término GO “*Ribosome biogenesis*” parece estar enriquecido principalmente en el grupo de genes con el máximo al Amanecer y el mínimo durante el Día en *Ostreococcus* y *Chlamydomonas*, mientras que en *Arabidopsis* sólo aparece significativamente enriquecido en el grupo que muestra el máximo durante el día y el mínimo durante la noche (Tablas S4-S6). Una investigación detallada en este proceso revela en *Chlamydomonas* un desacople entre la génesis de ribosomas citosólicos, plastídicos y mitocondriales (Zones et al., 2015). Los componentes de los ribosomas citosólicos tienen su máxima expresión al Atardecer (p. ej., *CrRPL5*, Cre14.g621450), los que codifican componentes de los ribosomas plastídicos alcanzan la cota más alta al Amanecer (p. ej., *CrPRPL1*, Cre02.g088900), mientras que los de los ribosomas mitocondriales presentan un pico más extendido al Amanecer (p. ej., *CrMRPL21*, Cre09.g388550) (Figura 23C). Este desacople es menos evidente en *Ostreococcus* ya que los genes que dan lugar a los ribosomas citosólicos (p. ej., *OtRPL5*, ostta15g01160), plastídicos (p. ej., *OtPRPL1*, ostta12g02760) y mitocondriales (p. ej., *OtMRPL21*, ostta01g04030) muestran un pico al Amanecer y el mínimo durante el periodo del Día (Figura 24C). En *Arabidopsis* no se observa una sincronización entre los genes que codifican los ribosomas (Figura 24C). Por tanto, este análisis muestra cómo los ciclos de luz/oscuridad en algas primordiales sincronizaron durante la evolución toda la ribogénesis en un mismo punto. En *Chlamydomonas* este proceso se dividió en tres etapas diferentes dependiendo del genoma nuclear, plastídico o mitocondrial, mientras que en *Arabidopsis* la génesis de ribosomas es prácticamente independiente de los ritmos circadianos.

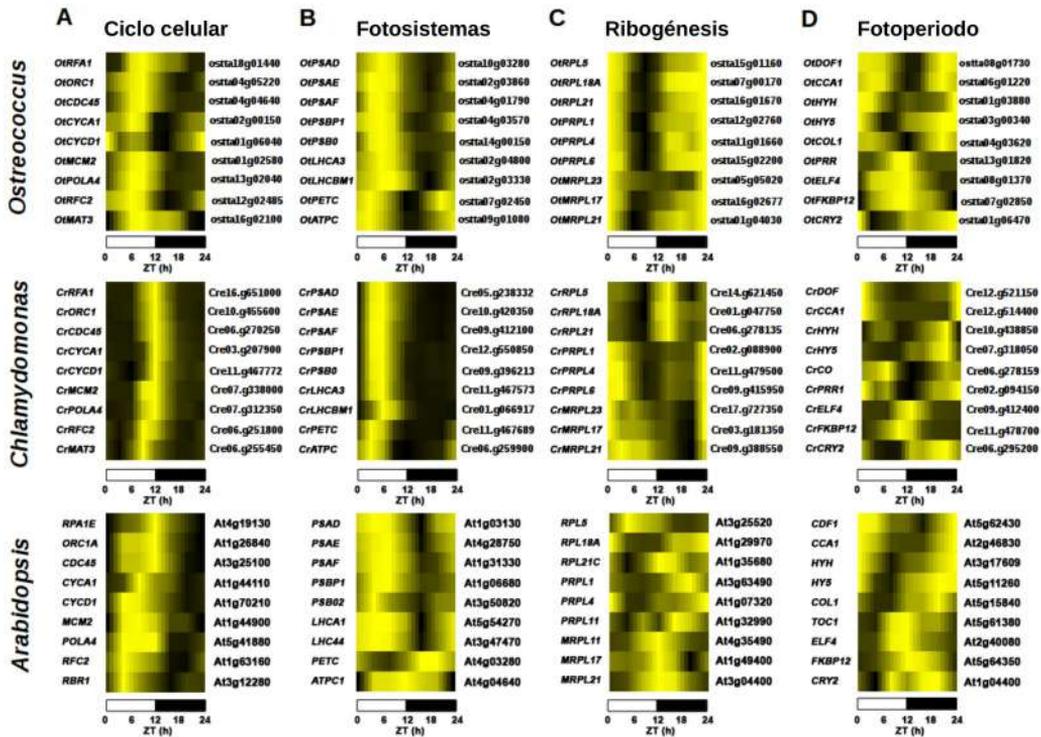


Figura 24. Comparación de los patrones de expresión en procesos biológicos clave en las tres especies analizadas. Se usan mapas de calor para representar la expresión génica y se ordenan los genes empleando un *clustering* jerárquico como en la Figura 21. (A) Varios genes involucrados en el ciclo celular muestran diferencias en su patrón de expresión génica al comparar las tres especies, destacando picos de expresión más localizados en *Chlamydomonas*. (B) Se muestra la expresión génica de genes que codifican componentes del fotosistema. Presentan patrones similares entre las tres especies con algunas excepciones, como los genes *ATPC1* y *PETC* en *Arabidopsis*. (C) Genes que participan en la ribogénesis. Se observa una sincronización entre los genes nucleares, mitocondriales y plastídicos en *Ostreococcus*. En *Chlamydomonas* se observa una diferenciación temporal mientras que en *Arabidopsis* no existe una sincronización aparente. (D) Los reguladores transcripcionales del fotoperiodo y el reloj circadiano muestran perfiles similares en las tres especies. *DOF1*, *CCA1* y *COL* presentan el mismo patrón. En otros casos como *TOC1* y sus ortólogos existe una similitud entre *Arabidopsis* y *Ostreococcus* mientras que *Chlamydomonas* exhibe diferencias en la expresión génica.

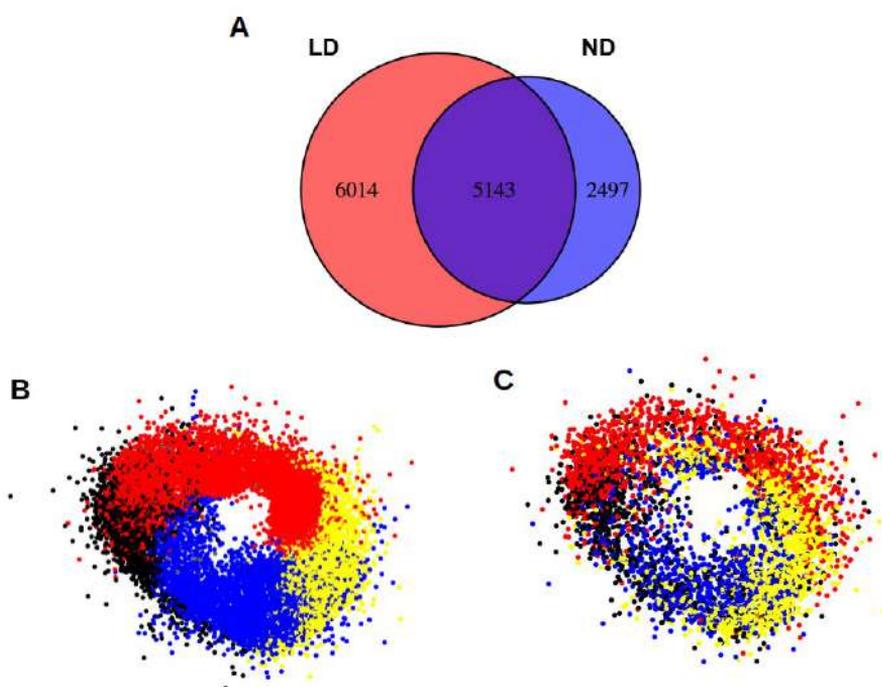
Igualmente se analizó la conservación y evolución de los principales reguladores transcripcionales involucrados en la respuesta al fotoperiodo y reloj circadiano en *Arabidopsis*. Se encontró una fuerte conservación entre los perfiles de expresión de los reguladores clave de estas vías en *Ostreococcus* y *Arabidopsis*, con

algunas variaciones en *Chlamydomonas* (Figura 24D). Por ejemplo, las tres especies mostraron perfiles de expresión similares de la familia de genes centrales en la vía fotoperiódica; *DOF* (Cre12.g51440, ostta04g02850, At5g62430) alcanzan su máximo de expresión al amanecer y tienen su mínimo al atardecer y *COL* (Cre06.g278159, ostta04g03620, At5g15840), que tienen su máximo en la Noche en el caso de *Ostreococcus* y *Arabidopsis*, y durante el Día en *Chlamydomonas* (Figura 24D). Notablemente, dos de los genes claves del reloj circadiano, *CCA1* (At2g46830, ostta06g01220) y *TOC1* (At5g61380, ostta13g01820) exhiben los mismos perfiles de expresión simétricos en *Ostreococcus* y *Arabidopsis*, mientras que en *Chlamydomonas* los ortólogos detectados en nuestro análisis, *CrCCA1* (Cre12.g514400) y *CrTOC1* (Cre02.g094150), muestran el mismo patrón de expresión, con el máximo al amanecer y el mínimo al atardecer (Figura 24D). Estas diferencias apuntan a una evolución divergente del núcleo regulador de los ritmos circadianos en *Chlamydomonas*, como se ha sugerido en estudios previos (Mittag et al., 2005). No obstante, son necesarios análisis más exhaustivos y realizar un mayor trabajo experimental para validar estos resultados.

Algunos genes relevantes involucrados en la fotomorfogénesis como los factores de transcripción bZIP *HY5* y *HY5-HOMOLOG (HYH)* muestran el mismo patrón de expresión en *Chlamydomonas* (Cre07.g318050, Cre10.g438850), *Ostreococcus* (ostta03g00340, ostta01g03880) y *Arabidopsis* (At5g11260, At3g17609), alcanzando el máximo al amanecer y el mínimo durante el día (Figuras 24D, 27C,D).

Al comparar diferentes fotoperiodos (ND y LD) en *Arabidopsis* se observaron aspectos interesantes (Figura 25). Así, cuando comparamos la expresión génica a partir de datos de RNAseq de plantas crecidas en LD (rojo) con la expresión en ND (azul), más del 41% de los genes en LD exhibieron un perfil periódico diario (11157) mientras que el porcentaje se redujo al 37% de los genes en condiciones ND (7640). De los 7640 genes periódicos en ND, 5143 (rojo intenso, 67%) también se identificaron en LD, en cambio, aproximadamente la mitad de los 11157 genes en LD (6014, 54%) no siguieron un patrón rítmico en ND. Esto podría significar que al aumentar las horas de luz del fotoperiodo también se aumenta la regulación rítmica del transcriptoma en plantas superiores, pero sin alcanzar los niveles de las

microalgas. Cuando organizamos los genes rítmicos en grupos, basándonos en la clasificación que se muestra en la Figura 23, se observa de nuevo una distribución en forma de agujas del reloj (Figura 25B). Esto parece indicar que la organización básica de los ritmos diarios en plantas es independiente del fotoperiodo. Sin embargo, al construir una red de coexpresión agrupada compuesta por los genes en la intersección entre LD y ND (5143 genes, púrpura), se observa un patrón circadiano más difuso, lo que sugiere una alteración en la expresión rítmica de un número sustancial de genes debido al fotoperiodo extendido (Figura 25C).



5 Existe un nivel alto de conservación entre las dos microalgas y un nivel moderado entre éstas y *Arabidopsis*

Con el objetivo de identificar ortólogos con alta probabilidad de ser funcionales entre las tres especies, los datos de expresión génica fueron integrados con los resultados obtenidos previamente usando el método MBBH. En la literatura pueden encontrarse análisis que han utilizado aproximaciones similares basadas en la integración de similitud de secuencia, perfiles de expresión y patrones de coexpresión (Das et al., 2016; Romero-Campero et al., 2013). Así se asumió que dos potenciales genes ortólogos periódicos según MBBH de dos especies diferentes exhibían un patrón diario conservado si ambos presentaban el máximo y el mínimo en el mismo intervalo (mismo *cluster* circadiano) o cuando el coeficiente de correlación de Pearson entre sus perfiles de expresión era mayor que 0,98. Por tanto, estos dos genes fueron llamados expresólogos (Das et al., 2016).

De acuerdo a este criterio, aproximadamente el 34% de los genes rítmicos de *Arabidopsis* presentaron un expresólogo en *Ostreococcus* o *Chlamydomonas*. Esto indica un alto nivel de conservación a pesar de la gran distancia evolutiva entre las angiospermas y las microalgas, y presumiblemente persiste a lo largo de todo el linaje vegetal. Curiosamente, sólo el grupo de *Arabidopsis* que une a los genes rítmicos con el máximo al Atardecer estaba enriquecido en expresólogos de *Ostreococcus* y *Chlamydomonas* (p-valor de $4,84 \times 10^{-5}$ según el test exacto de Fisher), lo que sugiere que la mayoría de esos genes conservados están altamente regulados por las transiciones luz/oscuridad en cuanto a su expresión génica. En *Ostreococcus*, el 83,62% de los genes diarios rítmicos presentaron un expresólogo en *Chlamydomonas* pero sólo el 36,1% tenían uno en *Arabidopsis*. En *Chlamydomonas*, para el 52,71% de los genes rítmicos se encontró expresólogo en *Ostreococcus* mientras que sólo un 19,05% tenían un expresólogo en *Arabidopsis*, lo que indica una mayor divergencia evolutiva entre estas especies que con *Ostreococcus*.

Para estudiar la conservación de los patrones rítmicos diarios de los diferentes grupos entre las tres especies, más allá de la comparación entre los perfiles de

expresión génica individuales, se calculó el estadístico de preservación compuesta Zsummary como se define en Langfelder et al., 2011 (Figura 26). Un valor de Zsummary menor que 2 indica que no hay conservación, un valor entre 2 y 10 implica una conservación moderada, mientras que un valor mayor que 10 evidencia un alto nivel de conservación. Para cada grupo de genes rítmico se calculó el Zsummary para las seis comparaciones posibles (*Arabidopsis* vs. *Ostreococcus*, *Arabidopsis* vs. *Chlamydomonas*, *Ostreococcus* vs. *Arabidopsis*, *Ostreococcus* vs. *Chlamydomonas*, *Chlamydomonas* vs. *Arabidopsis* y *Chlamydomonas* vs. *Ostreococcus*) y se representó gráficamente la media y la desviación estándar correspondientes. En la Figura 26 se puede observar que los grupos con el máximo al Amanecer presentan un nivel moderado de conservación en el patrón de coexpresión rítmica. De hecho, el grupo con mayor nivel de conservación se observó en aquel que alcanza el máximo al Atardecer y el mínimo al Amanecer ($Z_{summary} > 10$). Esto sugiere que evolutivamente se ha conservado mayoritariamente la expresión génica con una gran influencia de las transiciones luz/oscuridad y oscuridad/luz, de nuevo señalando la importancia de estas transiciones en la fisiología vegetal.

Para corroborar la conservación de los patrones de coexpresión en las tres especies diferentes, se escogieron dos vías centrales diariamente reguladas en *Arabidopsis*, según nuestro análisis (Tabla 11). Para estos subgrupos génicos, denominados reguladores clave en fotoperiodo/reloj (Figura 26B) y enzimas centrales en el metabolismo del almidón/azúcar (Figura 26C), se construyeron genes de coexpresión génica más pequeñas. En estas redes, en las que las aristas de color rojo indican correlación positiva y las de color azul correlación negativa, se puede observar una conservación general (Figura 26B,C). La red fotoperiodo/reloj reveló una conservación alta entre los tres subgrupos. Sin embargo, se observó una conservación mayor entre *Arabidopsis* y *Ostreococcus* que con *Chlamydomonas*, indicando de nuevo que en esta microalga el núcleo del reloj circadiano ha divergido ligeramente con respecto a *Arabidopsis* (Figura 26B). Por el contrario, los genes del metabolismo del almidón regulados diariamente presentaron una alta conservación entre *Ostreococcus* y *Chlamydomonas*, sugiriendo que, durante el curso de la evolución, se han añadido pasos regulatorios adicionales en el control de la síntesis

de almidón en plantas superiores (Figura 26C). En general, esta visualización de *subclusters* nos permitió comparar diferentes procesos biológicos que mostraron correlación en nuestro análisis general, indicando además los niveles de conservación y divergencia entre las especies y constituyendo herramientas eficientes para iniciar el estudio sobre la evolución de procesos particulares.

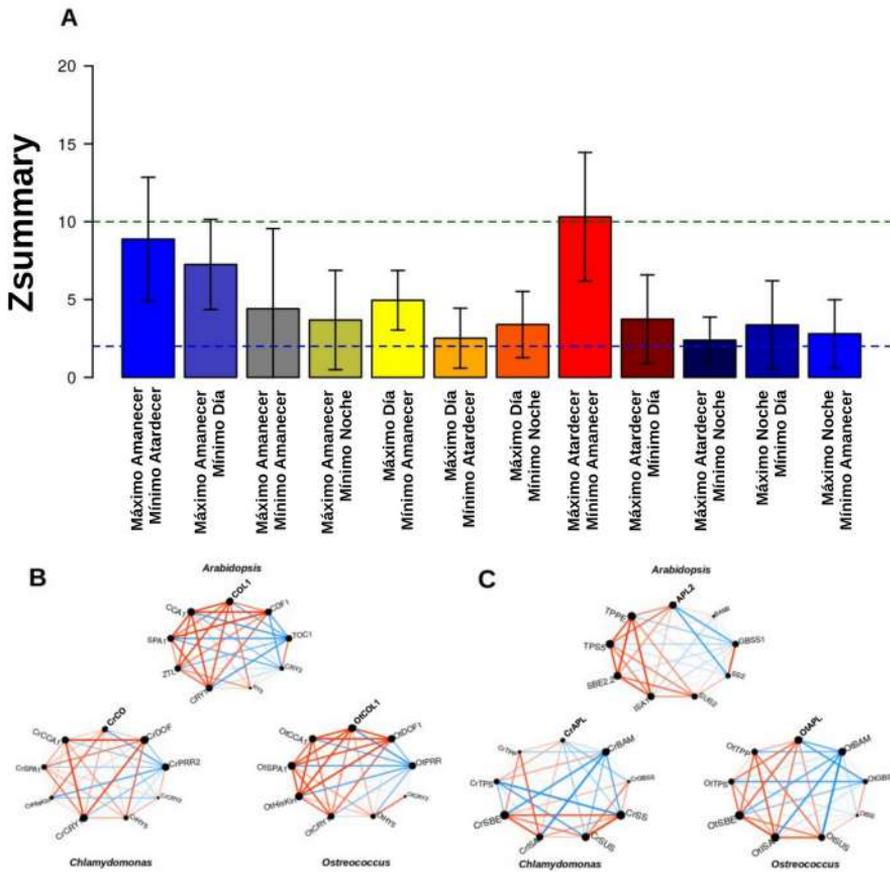


Figura 26. Conservación de patrones circadianos en *Arabidopsis*, *Chlamydomonas* y *Ostreococcus*.

(A) Gráfico de barras representando la media y la desviación estándar del índice de conservación Zsummary entre las diferentes especies para cada grupo génico. Sólo los grupos con el pico de expresión al Amanecer o al Atardecer están conservados significativamente entre las tres especies. (B) El patrón de coexpresión de reguladores del fotoperiodo y del reloj presenta un alto nivel de conservación entre las tres especies, especialmente entre *Arabidopsis* y *Ostreococcus*. Las aristas rojas indican correlación positiva entre los perfiles de expresión génica y las aristas azules correlación negativa. El grosor de la arista representa el valor absoluto de la correlación. (C) El patrón de coexpresión de enzimas clave del metabolismo del almidón y azúcar presenta un alto nivel de conservación entre las tres especies, de nuevo con pequeñas diferencias en *Chlamydomonas*.

Apoyando la veracidad de esta aproximación, se analizaron los promotores de los genes que mostraron un alto nivel de conservación en similitud de secuencia (Figura 27A,B a la izquierda) y en los patrones de coexpresión (Figura 27 A,B centro). Como se observa en la Figura 27A,B a la derecha, los promotores de los ortólogos *COLs* y *GBSSs* mostraron también conservación en los sitios de unión de factores de transcripción identificados. Estos resultados sugieren, que a lo largo de la historia evolutiva, grupos completos de genes han evolucionado sincronizadamente para mantener la regulación de estas redes funcionales y revelan porqué estas redes son tan resistentes al cambio a pesar del largo tiempo evolutivo.

Finalmente, para validar alternativamente la conservación de estos patrones rítmicos y la capacidad de predecir ortólogos en plantas y algas, se representó la expresión génica normalizada del gen *bZIP* de *Arabidopsis HY5* y de sus potenciales ortólogos detectados en este estudio en *Chlamydomonas* y *Ostreococcus* a partir del análisis del microarray en ND (Figura 27C). Estos datos provenientes de los microarrays se compararon con los obtenidos a partir de experimentos de QPCR a lo largo de 24 h en LD (Figura 27D). Ambos perfiles mostraron una similitud considerable, con la excepción de que en los perfiles de expresión en LD se observa una prolongación en la expresión hasta las últimas horas de luz, momento en que los niveles de mRNA empiezan a bajar de nuevo. Por tanto, la expresión génica de *HY5* y sus ortólogos mostraron un claro patrón de conservación, como se mostró en resultados previos (Figura 24), indicando que es altamente probable que estos tres genes sean ortólogos funcionales o expresólogos.

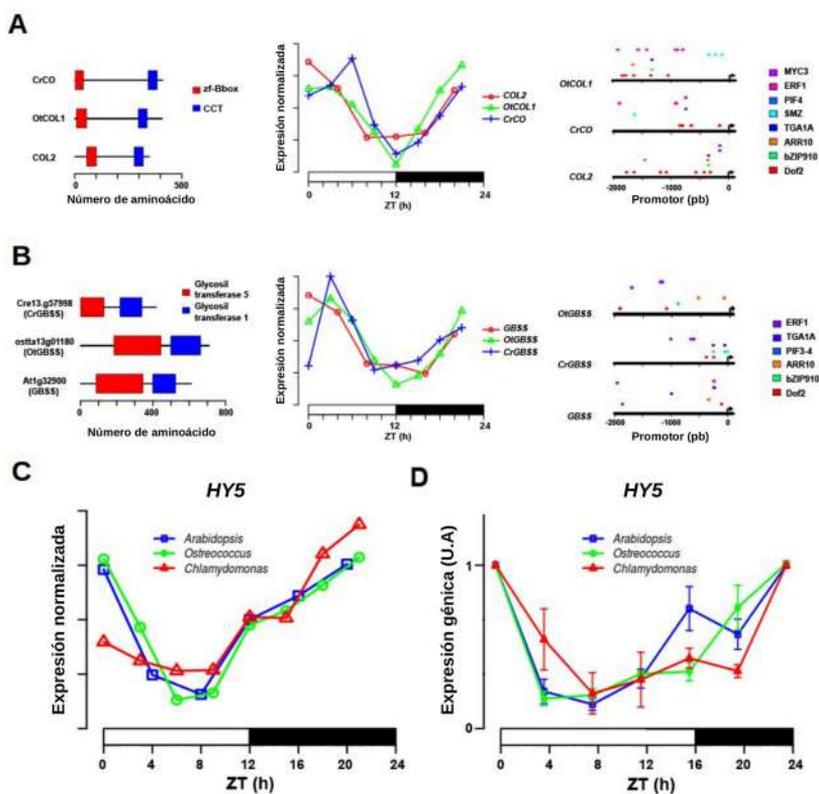


Figura 27. Factores de transcripción potencialmente ortólogos que muestran conservación en estructura de proteína, sitios de unión en el promotor y perfil de expresión en algas y plantas superiores. (A) Estructura proteica (a la izquierda), expresión durante un día neutro de 24 h (centro) y sitios de unión de factores de transcripción en el promotor del gen *COL2* de *Arabidopsis* y sus ortólogos. (B) Mismo análisis para el promotor del gen *GBSS* de *Arabidopsis* y sus potenciales ortólogos. (C) Expresión normalizada en un fotoperiodo ND proveniente de datos de microarray y RNAseq para el gen *HY5* de *Arabidopsis* (azul), *CrHY5* de *Chlamydomonas* (rojo) y *OtHY5* de *Ostreococcus* (verde). (D) Niveles de expresión en LD (Unidades Arbitrarias) de *HY5* (azul) en *Arabidopsis*, *CrHY5* (rojo) en *Chlamydomonas* y *OtHY5* (verde) en *Ostreococcus* en experimentos de QPCR. El patrón periódico de *HY5* se conserva en las tres especies a pesar de cambiar el fotoperiodo. Se muestra una escala de tiempo *Zeitgeber Time* (ZT) en horas (h). Para cada punto temporal se analizaron tres réplicas biológicas y tres réplicas técnicas, indicando el error estándar de la media.

Como se desprende de este capítulo, la comparación de redes de coexpresión génica entre especies es una herramienta útil para entender la evolución y la aparición de nuevos fenotipos y comportamientos. No obstante, existen limitaciones en el uso de las redes de coexpresión para inferir asociaciones funcionales. Genes que presentan el mismo patrón de expresión no tienen por qué tener funciones similares, o pueden no tener una interacción directa. Además, no siempre se tienen en cuenta regulaciones posteriores a nivel postranscripcional y postraduccional (Uygun et al., 2016; Ovens et al., 2021). Por otra parte, uno de los factores más importantes que regulan la expresión génica es la acción de factores de transcripción, por ello, la construcción de redes transcripcionales captura otro nivel de complejidad que nos permite estudiar cómo determinadas proteínas regulan la expresión génica de procesos biológicos clave. Este punto lo desarrollaremos en el siguiente capítulo.



The background features a complex network of nodes and lines in shades of blue and yellow, resembling a molecular or biological structure. The nodes are of varying sizes and are connected by thin lines, creating a web-like pattern. The overall aesthetic is scientific and modern.

Capítulo 2

**ATTRACTOR, una
herramienta para el
estudio del reloj
circadiano en
Arabidopsis.**

1 Flujo de trabajo computacional

En este trabajo se desarrolló un flujo de trabajo computacional automático para la construcción de redes transcripcionales a partir de datos de ChIPseq de factores de transcripción. Este flujo de trabajo se diseñó para trabajar en torres de servidores formadas por superordenadores empleando el sistema de colas *Sun Grid Engine* (<http://gridscheduler.sourceforge.net/>).

El archivo de entrada consiste en una serie de parámetros que especifica la dirección web para descargar el genoma de referencia y los datos crudos de ChIP-seq. Este archivo también contiene información sobre el diseño experimental (número de réplicas y muestras ChIP o control) y otros detalles técnicos como el espacio de trabajo.

Este flujo de trabajo se puede dividir en tres etapas (Figura 28). La primera fase es secuencial y genera la estructura de trabajo que contiene subcarpetas donde se almacenarán el genoma, la anotación, las muestras y los resultados. En esta fase se descargan el genoma de referencia y su anotación, y se generan los índices necesarios. La segunda etapa se realiza en paralelo para cada factor de transcripción incluido. Se descargan las muestras necesarias (ChIP y control) y se mapean contra el genoma de referencia simultáneamente. En este punto es necesaria una sincronización, ya que el alineamiento de muestras ChIP y control debe terminar antes de pasar al paso de determinación de picos o *peak calling*. Este paso de sincronización se lleva a cabo usando una estructura de pizarras (Mullineux, 1991). El alineamiento de las lecturas (ChIP y control) al genoma de referencia se lleva a cabo en paralelo, simultáneamente y de forma independiente, por ello se usa un archivo pizarra en el que ambos procesos pueden escribir y leer información. Cuando uno de los procesos termina su tarea, escribe un mensaje en el archivo y además lee la información para comprobar si el otro proceso ha terminado. Cuando la información de la pizarra indica que ambos alineamientos han terminado, entonces se lanza el paso de determinación de picos. Esta etapa paralela para cada factor de transcripción termina con la generación de un archivo con los potenciales genes diana. En este momento se necesita otro paso de sincronización para asegurar que todos los

ATTRACTOR, una herramienta para el estudio del reloj circadiano en Arabidopsis.

procesos que resultan en los genes diana de cada factor de transcripción han terminado antes de continuar con la tercera y última fase. Finalmente, esta última etapa es secuencial e integra todos los factores de transcripción y sus dianas para generar la red transcripcional.

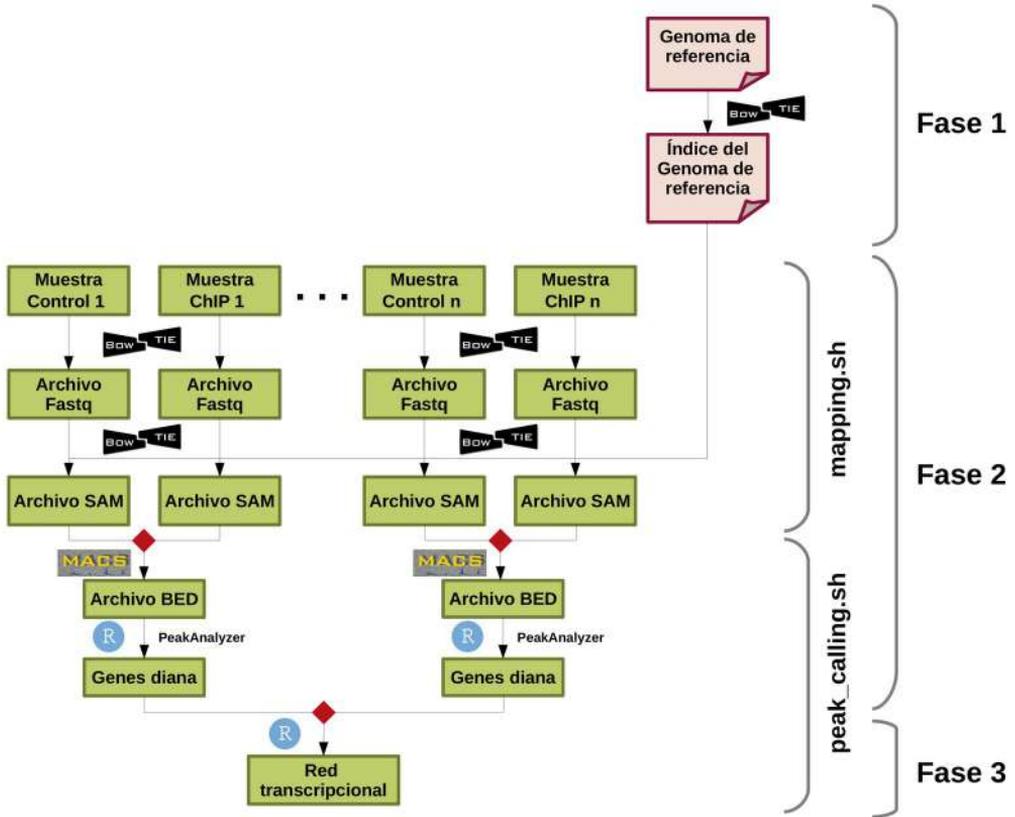


Figura 28. Flujo de trabajo computacional automático para construir redes transcripcionales a partir de datos brutos de ChIP-seq de factores de transcripción. En la primera fase se crea el índice del genoma de referencia. La segunda fase comprende el alineamiento de todas las muestras al genoma de referencia, la determinación de picos y la identificación de genes diana para cada factor de transcripción. Por último, en la tercera fase se integran todas estas dianas para generar la red transcripcional.

2 ATTRACTOR integra reguladores circadianos centrales y señalizadores lumínicos

ATTRACTOR consiste en una red transcripcional que integra datos transcriptómicos recolectados en ciclos de 24 h en condiciones ND (12:12) y datos de ChIP-seq de factores de transcripción involucrados en reloj circadiano y señalización lumínica. Esta red transcripcional está compuesta por 5788 nodos o genes con perfiles rítmicos diarios y 14529 aristas que representan regulaciones transcripcionales. Los factores de transcripción y reguladores transcripcionales que se incluyen en la red son los factores MYB del bucle central del reloj circadiano CCA1 y LHY, los integrantes de la familia de PSEUDO-RESPONSE REGULATORS TOC1, PRR5, PRR7 y PRR9; y los componentes del bucle de la tarde LUX, ELF3 y ELF4. Con el objetivo de analizar la regulación conjunta del reloj con la señalización por luz, se incluyeron los sensores lumínicos (y reguladores transcripcionales) PHYA, PHYB y CRY2; los factores de transcripción de la familia PHYTOCHROME-INTERACTING FACTOR PIF3, PIF4 y PIF5; y por último el factor de transcripción que interacciona con fitocromos FHY1 (Far-red elongated Hypocotyl 1) (Figura 29).

Esta red se construyó filtrando la red de coexpresión génica de *Arabidopsis* generada en el capítulo 1 de Resultados añadiendo regulaciones transcripcionales provenientes del análisis de los datos de ChIP-seq para los factores de transcripción incluidos. Por tanto, se pueden observar los efectos globales de expresión promovidos por la regulación de los factores de transcripción sobre sus dianas.

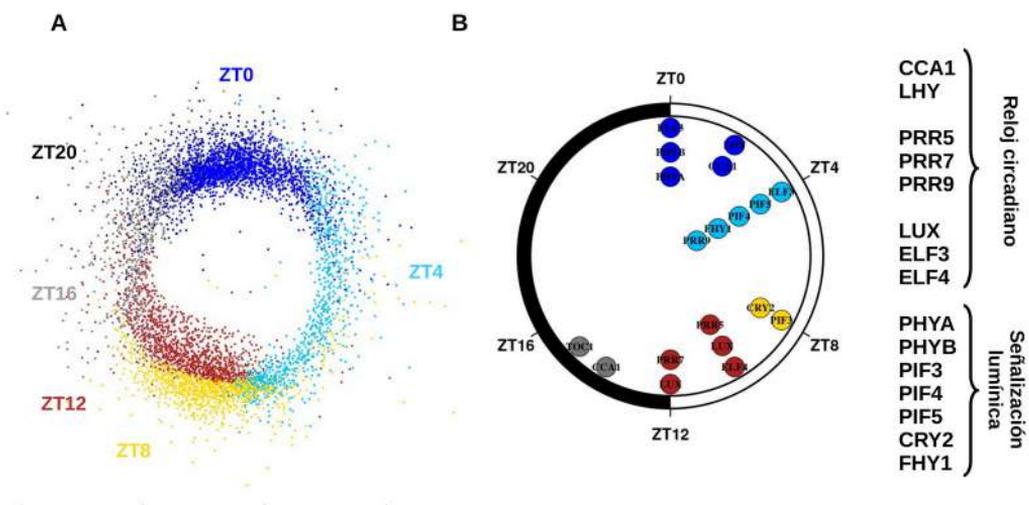


Figura 29. Red transcripcional ATTRACTOR formada por factores de transcripción involucrados en el reloj circadiano y en la percepción lumínica de *Arabidopsis*. (A) Representación gráfica de la red transcripcional en la que los factores de transcripción se localizan en el centro, dependiendo de la hora del día en la que se realizó el experimento de ChIPseq. Los genes diana periódicos se ubican alrededor, en el momento del día donde tienen su punto máximo de expresión. Se usó un código para colorear los nodos: azul oscuro para genes con el máximo de expresión a ZT0, celeste para ZT4, amarillo para ZT8, rojo para ZT12, gris para ZT16 y negro para ZT20. (B) Factores de transcripción incluidos en el análisis y con cuyos genes diana se construyó la red transcripcional. En el gráfico en forma de reloj se sitúan en su correspondiente hora donde muestran su máximo de expresión.

3 Herramienta web ATTRACTOR

Para la exploración de esta red transcripcional por la comunidad científica se generó una herramienta web mediante la librería de R *Shiny*, alojada en el enlace <https://greennetwork.us.es/ATTRACTOR/>. Se trata de una herramienta para el análisis del control transcripcional compartido por el reloj circadiano y la señalización lumínica sobre genes con expresión periódica. La estructura del código de la aplicación web consiste, como se describe en el apartado de Materiales, de una sección *server* y una sección *ui* (*user interface*) en la que se especifica la apariencia de la web, como los distintos paneles o pestañas.

Por una parte, ATTRACTOR permite explorar la regulación coordinada de varios factores de transcripción sobre un gen individual y el efecto observado en su perfil de expresión génica (Figura 30). En primer lugar se escoge un gen incluido en la red (p.e *GIGANTEA*) y los factores de transcripción deseados para explorar la regulación sobre el gen de interés. En la primera pestaña, “*Clock Visualizer*”, se usa una representación circular de un reloj para visualizar la regulación de los factores de transcripción sobre el gen elegido (Figura 30A). El gen se ubica en el centro del reloj en el punto temporal en el que su expresión alcanza el máximo y de nuevo se usa el código de color empleado en la Figura 29. Los factores de transcripción aparecen en la periferia, ubicados en el punto temporal en el que se realizó el experimento de ChIP-seq correspondiente. Se han representado con color rojo los potenciales represores, en verde los potenciales activadores, y gris los factores de transcripción sin un efecto claro sobre la expresión. En la siguiente pestaña, “*Expression Visualizer*”, se representa el perfil de expresión del gen elegido y de nuevo se representan los factores de transcripción ubicados en su correspondiente punto temporal (Figura 30B). Por último, en la pestaña “*Peak Visualizer*” se pueden explorar la unión de los factores de transcripción sobre el gen de interés, sus perfiles de unión, sus picos y los motivos de DNA identificados (Figura 30C). En este visualizador se pueden especificar los motivos de DNA a buscar en los sitios de unión, la puntuación mínima para la búsqueda de motivos, así como otros parámetros como el tamaño del promotor o la longitud del fragmento 3' UTR a analizar.

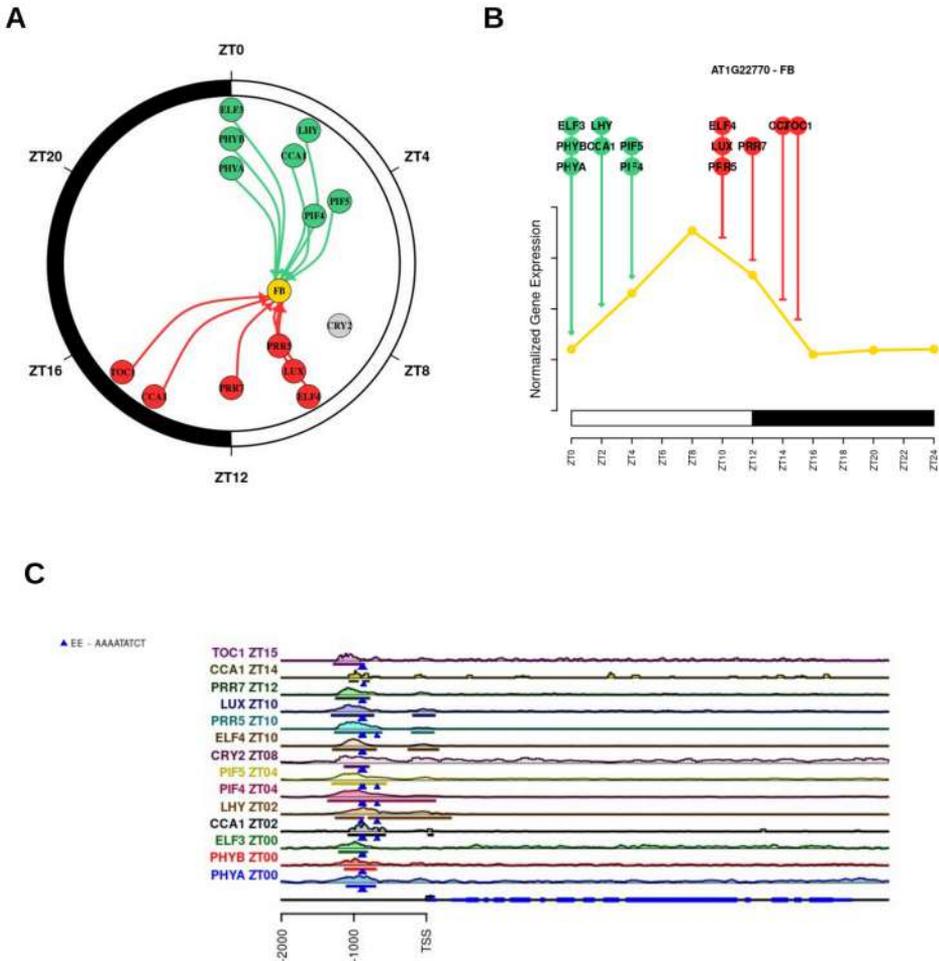
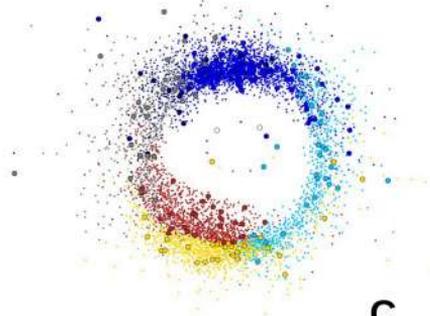


Figura 30. Ejemplo de análisis de un gen individual en ATTRACTOR (imágenes extraídas de la aplicación web). (A) Pestaña “Clock Visualizer” en la que se representa el ciclo circular de un día neutro (12:12). El gen escogido se localiza en la parte central, ligeramente orientado hacia el punto temporal en el que tiene su máximo de expresión. En la periferia se ubican los factores de transcripción elegidos y se traza una arista hasta el gen si dicha proteína se une a su promotor. Se usan aristas rojas para un efecto observado de represión, aristas verdes para activación y grises en caso de que no haya un cambio en la expresión génica tras la unión del factor de transcripción. (B) Pestaña “Expression Visualizer”. Se representa la expresión génica normalizada para el gen escogido y los factores de transcripción en el punto temporal correspondiente. (C) Pestaña “Peak Visualizer”. Se representa el *locus* del gen y para cada factor de transcripción el perfil de unión y los picos. El usuario escoge los motivos de DNA a buscar en los sitios de unión.

Por otra parte, ATTRACTOR también permite el análisis de la regulación de varios factores de transcripción sobre genes diana en común. En esta sección se pueden seleccionar varios reguladores transcripcionales de interés (PHYA, PHYB y PRR5) y también de forma opcional, genes con un perfil de expresión específico (“*peak ZT0*”). De esta forma, la aplicación selecciona los genes en la intersección de las dianas de los factores de transcripción y el conjunto de genes con el patrón de expresión indicado. En la primera pestaña, “*Network Visualization*”, se representa la red con los genes en cuestión indicados, siendo posible además visualizar las aristas (Figura 31A). En la siguiente pestaña, “*Gene Table*”, aparece una tabla interactiva con los genes seleccionados e información sobre ellos como la descripción, los reguladores en la red o el perfil de expresión, así como links que conducen a bases de datos. Esta tabla puede ser explorada por el usuario y descargada (Figura 31B). Posteriormente, en la siguiente pestaña se muestra la significancia del solapamiento. Aquí se muestra el diagrama de Venn para los conjuntos de genes seleccionados y se indican el p-valor y el enriquecimiento de la intersección (Figura 31C). En la siguiente pestaña “*Functional Enrichment*” es posible llevar a cabo un análisis de enriquecimiento en términos de ontología génica y en vías KEGG, con múltiples formas de visualización. En la Figura 31D se muestra la tabla en la que aparecen los términos GO enriquecidos para el conjunto de genes. De nuevo esta tabla puede ser explorada y descargada. Uno de los múltiples gráficos de visualización de los términos GO se muestra en la Figura 30E. De forma similar se realiza el análisis de enriquecimiento en vías KEGG. Asimismo, la identificación de motivos de DNA enriquecidos en los sitios de unión en los promotores de los genes se realiza en la pestaña “*TFBS Enrichment*”. Aquí el usuario puede ajustar parámetros tales como el grupo de genes empleado como universo en el análisis, la longitud del promotor y la puntuación mínima para la búsqueda de motivos (Figura 31F). El resultado es una tabla que enumera los motivos de DNA encontrados con información como el nombre, el identificador, el logo, el p-valor, el q-valor, el enriquecimiento y los genes que en cuyos promotores aparece el correspondiente motivo. Esta tabla igualmente puede ser descargada, explorada y contiene enlaces que conducen a la base de datos JASPAR. Por último, en la pestaña “*Data Retrieval*”, el usuario puede descargar una serie de datos descriptivos de los factores de transcripción seleccionados.

ATTRACTOR, una herramienta para el estudio del reloj circadiano en Arabidopsis.

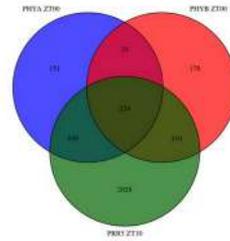
A



B

AGI ID	Gene Name	Gene Description	Regulators	Expression Peak Time	Expression Trough Time
AT2G11730	HY3	phytochrome B	LHY, PRR5, PHYA, PHYB, PHYE, PIF4, PIF5	ZTNA	ZTNA
AT2G27670	FHY1	far-red elongated hypocotyl 1	PRR5, PHYA, PHYB, FHY1	ZTNA	ZTNA
AT2G42190	PTHQ3455-SF4 - EXPRESSED PROTEIN (1 of 2)		LHY, PRR5, PHYA, PHYE, PIF4, PIF5	ZT4	ZT12
AT2G41900	OXS2	OXC-type zinc finger protein with ADM repeat domain	CCA1, LHY, PRR5, PHYA, PHYB, FHY1, PIF4, PIF5, ELF3	ZT12	ZT20
AT2G40100	ATHSPR02	ortholog of sugar beet H51 PR0-1.2	LHY, TOC1, PRR5, PHYA, PHYB, CRY2, PIF4, PIF5	ZT0	ZT12

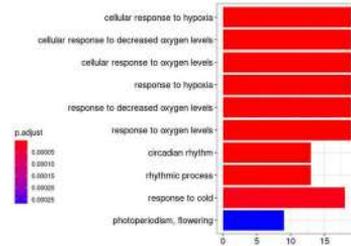
C



D

GO ID	Description	p-value	q-value	Enrichment (Target Ratio: 95 Refs)	Genes
00001406	cellular response to hypoxia	1.52537e-08	4.98245e-06	4.6 (16/205, 100/4912)	AT2G48880, AT2G54143, AT1G23710, AT1G01720, AT1G28480, AT1G13280, AT2G25240, AT1G20770, AT1G04880, AT1G23580, AT1G27300, AT1G03470, AT1G03880, AT1G04700, AT1G04410, AT1G03400, AT1G03440
00002504	cellular response to decreased oxygen levels	1.811225e-08	4.98245e-06	4.05 (19/205, 101/4912)	AT2G48880, AT2G54143, AT1G23710, AT1G01720, AT1G28480, AT1G13280, AT2G25240, AT1G20770, AT1G04880, AT1G23580, AT1G27300, AT1G03470, AT1G03880, AT1G04700, AT1G04410, AT1G03400, AT1G03440
00001402	cellular response to oxygen levels	1.811225e-08	4.98245e-06	4.05 (19/205, 101/4912)	AT2G48880, AT2G54143, AT1G23710, AT1G01720, AT1G28480, AT1G13280, AT2G25240, AT1G20770, AT1G04880, AT1G23580, AT1G27300, AT1G03470, AT1G03880, AT1G04700, AT1G04410, AT1G03400, AT1G03440

E



F

DNA motifs	Motif ID	DNA logo	P-values	Q-values	Enrichments	Genes
MYC4	MAD569.1		1.745677e-08	7.890460e-07	2.011658	AT2037670, AT2541430, AT2045880, AT2046830, AT2005790, AT1023710, AT1004660, AT2542790, AT2547180, AT1030640, AT1096570, AT1072820, AT1032640, AT1090100, AT1001720, AT1072450, AT1068440, AT1069570, AT1019110, AT3009000, AT3002140, AT3011670, AT3007390, AT3002120, AT3002990, AT3016050, AT3017040, AT2028890, AT1030110, AT3029975, AT4001020, AT4001720, AT4023050, AT4024240, AT4027410, AT4020240, AT3047160, AT3047470, AT3049530, AT3055430, AT3055980, AT3057880, AT3059060, AT3060330, AT3062700,

Figura 31. Ejemplo en ATTRACTOR del análisis de varios factores de transcripción. Se seleccionaron los factores de transcripción o reguladores transcripcionales PHYA, PHYB y PRR5 (imágenes extraídas de la aplicación web). (A) Imagen de la red transcripcional ATTRACTOR donde se destacan las dianas compartidas por los factores de transcripción seleccionados. **(B)** Tabla que enumera este grupo de genes y que provee información detallada sobre ellos. **(C)** Diagrama de Venn con la intersección de los genes diana de las proteínas escogidas. También se indican el p-valor y enriquecimiento del solapamiento. **(D)** Tabla con los términos GO enriquecidos para los genes diana en común. **(E)** Una de las múltiples formas de representar el análisis de enriquecimiento en términos GO de los genes en la intersección. **(F)** Tabla con la identificación de motivos enriquecidos en los sitios de unión de los factores de transcripción.

4 El reloj circadiano en *Arabidopsis* es robusto a ataques aleatorios y sensible a ataques dirigidos

Para analizar las propiedades del programa transcripcional controlado por el reloj circadiano y la percepción lumínica, se llevó a cabo un análisis topológico de la red transcripcional. La distribución del grado de los nodos se ajustó a una ley potencial negativa por lo que se determinó que ATTRACTOR constituía una red libre de escala. Estas redes muestran un alto grado de tolerancia a ataques aleatorios y una extrema fragilidad contra ataques dirigidos. Son resistentes a fallos aleatorios ya que la mayoría de sus nodos están poco conectados y su desaparición no afecta sustancialmente al sistema. Sin embargo, son frágiles a cambios dirigidos a nodos altamente conectados o *hubs* (Barabási y Albert, 1999; Yook et al., 2002; Albert et al., 2000). Esto implica que la arquitectura de estos sistemas, el reloj circadiano y la señalización lumínica en *Arabidopsis*, tienen las mismas características aparentemente opuestas, robustez y fragilidad, como ya se describió en estudios previos (Stelling et al., 2004).

Por otra parte, se determinó el camino medio mínimo de la red transcripcional obteniéndose el valor 2,089, es decir, para ir de un nodo aleatorio a otro hay que recorrer 2 aristas (Figura 32). Esto, sugiere que ATTRACTOR es una red de mundo pequeño, lo que se manifiesta en una rápida y eficiente propagación de información dentro de la red. Para verificar que la red era de mundo pequeño se generaron 10000 redes aleatorias con las mismas propiedades topológicas y se compararon sus

ATTRACTOR, una herramienta para el estudio del reloj circadiano en Arabidopsis.

coeficientes de agrupamiento con el de ATTRACTOR, verificándose que esta red transcripcional tiene un coeficiente significativamente mayor.

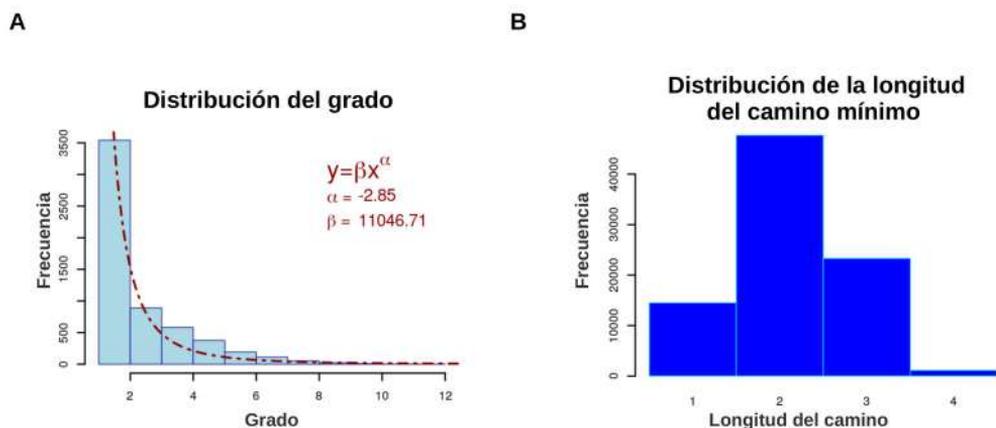


Figura 32. Propiedades topológicas de la red transcripcional ATTRACTOR. (A) Distribución del grado de los nodos. Se trata de una red libre de escala puesto que la distribución se ajusta a una ley potencial negativa. **(B)** Distribución de la longitud del camino mínimo entre dos nodos aleatorios.

5 El reloj es especialmente activo a ZT0

Para discernir en qué momento del día el programa transcripcional compartido por la señalización lumínica y el reloj era especialmente activo se identificaron los genes que presentaron una mayor conexión en la red. Para ello se identificaron los *hubs* como el 5% de los nodos con mayor grado o número de vecinos, obteniéndose un total de 199 genes.

Resultado de su representación en la red, se observó cómo estos genes se localizaban mayoritariamente en el grupo de genes que muestran un máximo de expresión a ZT0 (Figura 33A). Para verificarlo, se analizó el solapamiento de todos los grupos de genes con los nodos concentradores. Únicamente el grupo de genes con el máximo a ZT0 presentó un solapamiento significativo con los *hubs* de acuerdo al test exacto de Fisher con un p-valor de $5,822 \times 10^{-05}$ (Figura 33B). Por lo tanto, parece que el momento del día que aparentemente está más regulado por el reloj y por el sistema de señalización lumínica es a primera hora del día.

ATTRACTOR, una herramienta para el estudio del reloj circadiano en Arabidopsis.

A continuación, se realizó un análisis de enriquecimiento en términos de ontología génica sobre estos *hubs* encontrándose términos no redundantes como “*immune system process*”, “*transcription*”, “*multicellular organism development*”, “*response to organic substance*” o “*growth*” (Figura 33C). Por tanto, se puede afirmar que éstos son procesos biológicos altamente regulados por los ciclos diurnos.

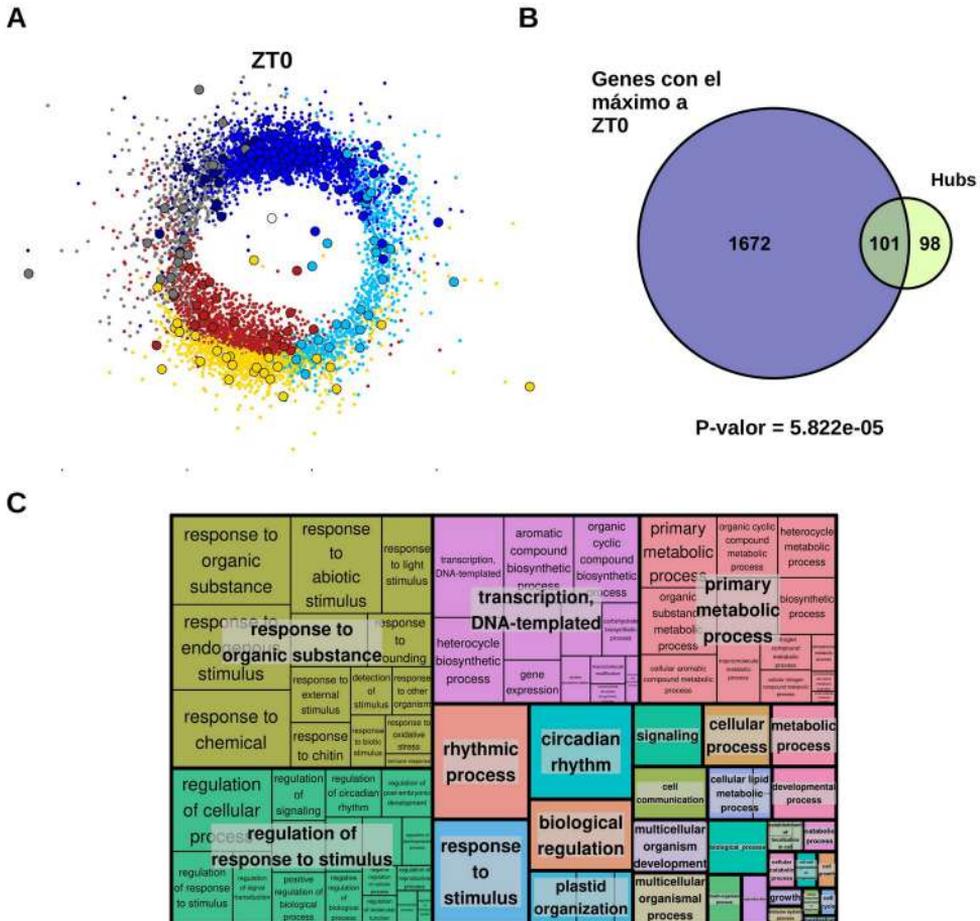


Figura 33. Genes más regulados o *hubs* en ATTRACTOR. (A) Representación de la red transcripcional donde se destacan los *hubs*. Se observa una mayor acumulación a ZT0. (B) Solapamiento significativo entre los *hubs* y los genes de la red que tienen su máximo de expresión a ZT0. (C) Resumen de los términos GO enriquecidos para los *hubs*. Cada rectángulo del *treemap* representa el $-\log_{10}(p\text{-valor})$ para el correspondiente término GO.

Siguiendo este criterio también puede deducirse que uno de los procesos más controlados por el reloj circadiano y el sistema de señalización lumínica es el sistema inmune. Esto concuerda con el hecho de que la apertura y cierre de estomas es un proceso cíclico y que supone una puerta de entrada para patógenos. De hecho, se ha descrito en múltiples estudios que el reloj circadiano regula el sistema de defensa contra microorganismos invasores (Sauerbrunn y Schlaich, 2004; Roden y Ingle, 2009). Por ejemplo, bajo condiciones constantes, *Arabidopsis* muestra una susceptibilidad cíclica frente a la infección por *Pseudomonas syringae* (Bhardwaj et al., 2011).

Otro de los procesos específicos en los que están implicados los *hubs* de la red es la respuesta a hormonas, en las que se incluyen auxinas, ácido abscísico, ácido jásmonico, brasinoesteroides o giberelinas. Aquí se encuentran genes que se inducen por la presencia de auxinas como *INDOLE-3-ACETIC ACID INDUCIBLE 2* (*IAA2*), *PINOID* (*PID*) y receptores de auxinas como *TRANSPORT INHIBITOR RESPONSE1* (*TIR1*). También se expresan a ZT0 genes periódicos como *ATAF1*, cuya expresión responde a los niveles de ácido abscísico o (*PYRABACTIN RESISTANCE 1-LIKE 5*) *PYL5*, que codifica un sensor de esta hormona. También se encontraron en este grupo proteínas relacionadas con los brasinoesteroides como el receptor *BRI1-ASSOCIATED RECEPTOR KINASE* (*BAK1*) o *BRASSINOSTEROID-INSENSITIVE 2* (*BIN2*), que actúa mediando las vías de señalización de auxinas y brasinoesteroides. Por tanto, parece que el reloj modula la síntesis de receptores hormonales, lo que permitiría a la planta anticiparse a la presencia de fitohormonas. Todas estas relaciones entre el reloj circadiano y las vías hormonales en *Arabidopsis* han sido descritas en recientes revisiones (Atamian y Harmer, 2016; Singh y Mas, 2018).

Por último, según nuestros análisis, el crecimiento en *Arabidopsis* también está estrechamente regulado por el reloj circadiano y la señalización lumínica. Efectivamente, se ha descrito un mecanismo de coincidencia externa por el cual la elongación está controlada por estímulos externos (señalización lumínica) e internos (reloj circadiano) (Nozue et al., 2007; Niwa et al., 2009; Kunihiro et al., 2011). Por otra parte, recientemente se ha demostrado el papel de los PRRs como represores del

crecimiento ejerciendo un papel antagonista al de los PIFs para regular la expresión génica (Martín et al., 2018).

6 Diferentes motivos de red regulan procesos fisiológicos esenciales

Para identificar patrones no aleatorios en la red transcripcional que presentaran funciones específicas se llevó a cabo una identificación de motivos de red. Como se describió anteriormente, un motivo de red es un patrón básico o subgrafo que aparece de forma recurrente en redes biológicas, mucho más frecuentemente que en redes aleatorias. Para identificar motivos de redes en ATTRACTOR generamos 10000 redes aleatorias con las mismas propiedades topológicas y comparamos la frecuencia en que aparecía cada subgrafo en nuestra red transcripcional con la frecuencia en las redes aleatorias, estimándose así un p-valor para cada uno.

El motivo de tres nodos más complejo que se identificó fue la combinación de tres genes que se retroalimentan entre sí y que tienen salida múltiple (Figura 34A). Su alta frecuencia en redes biológicas parece indicar que cumplen con funciones concretas, lo que propiciaría su conservación a lo largo de la evolución. Por tanto, se espera que regulen procesos biológicos importantes. Siguiendo esta idea, se exploraron diferentes motivos en la red compuestos por factores de transcripción y se encontró uno de triple retroalimentación con salida múltiple constituido por CCA1, PIF5 y PRR5 (Figura 34B). Empleando la aplicación web, se observó que los genes diana regulados por estos tres factores de transcripción estaban involucrados en procesos específicos como respuesta a sequía y respuesta a frío (Figura 34C). En este grupo se encontraron genes como *DREB2A* (*DEHYDRATION-RESPONSIVE ELEMENT BINDING PROTEIN 2*), *ERD7* (*EARLY-RESPONSIVE TO DEHYDRATION 7*) y *GRP7* (*GLYCINE RICH PROTEIN 7*).

ATTRACTOR, una herramienta para el estudio del reloj circadiano en Arabidopsis.

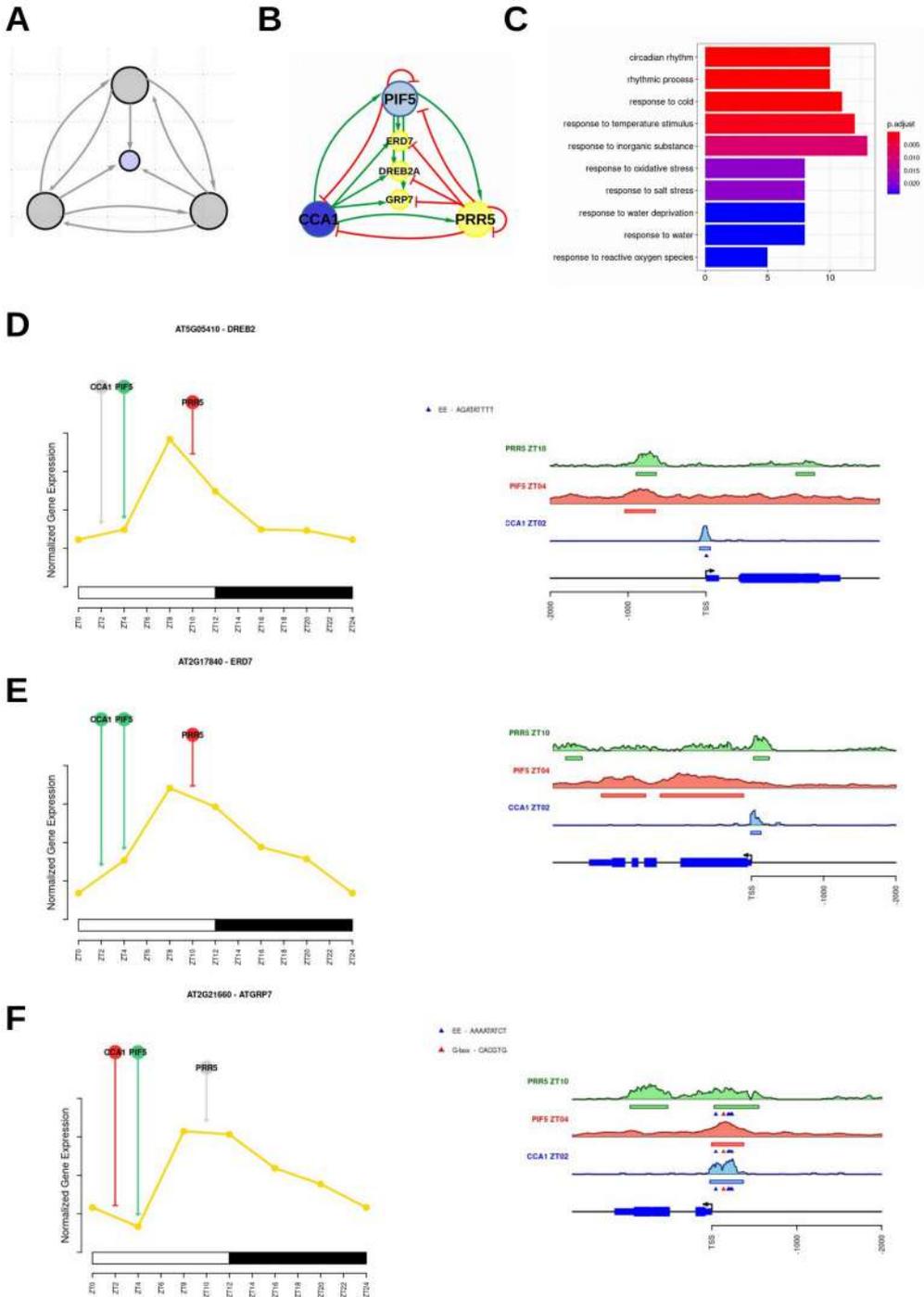
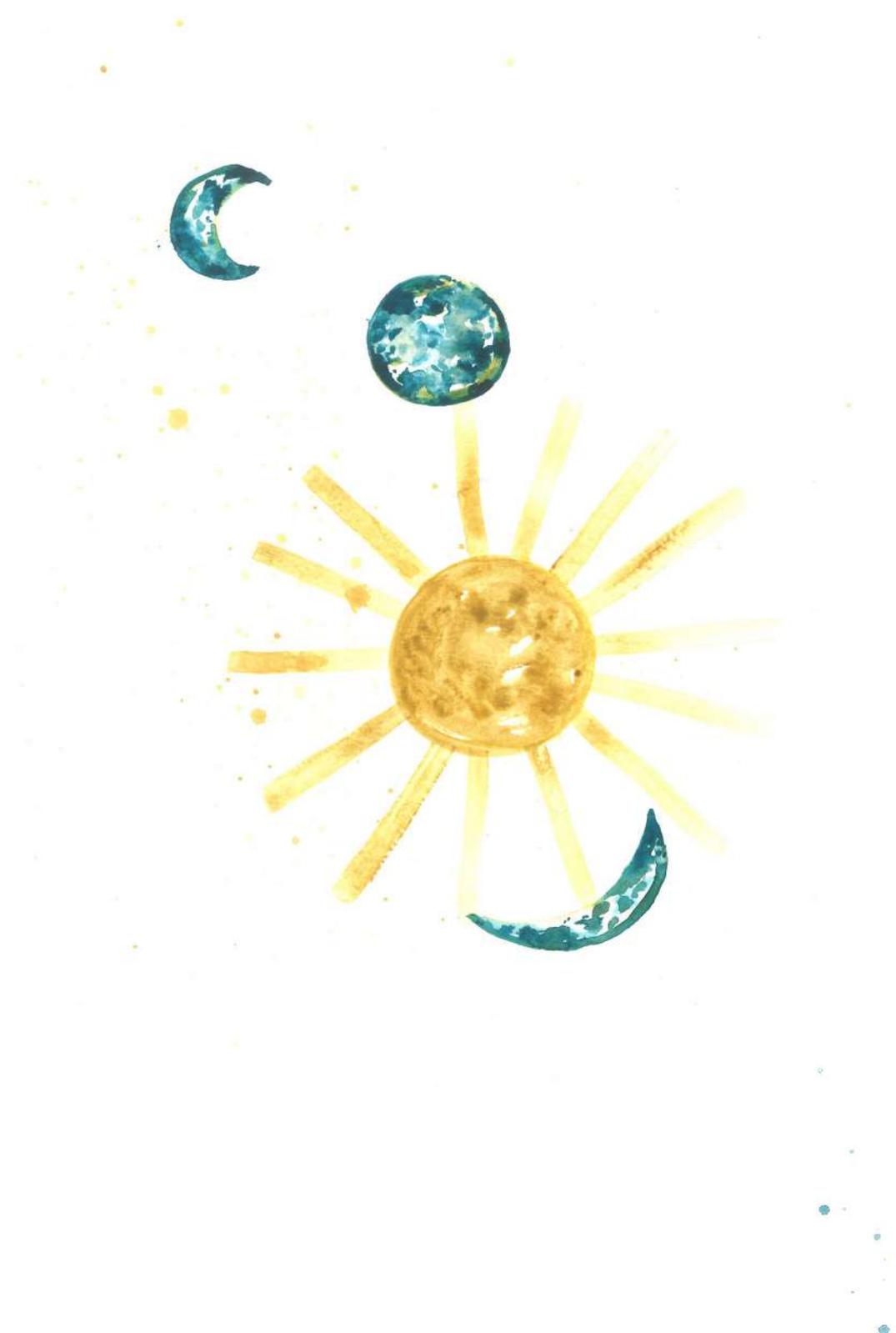


Figura 34. El motivo de red formado por CCA1, PIF5 y PRR5 modula procesos específicos. (A) Motivo de red encontrado en ATTRACTOR, triple retroalimentación con salida múltiple. (B) Uno de estos motivos, constituido por CCA1, PIF5 y PRR5. Se representan tres de los genes salida, *DREB2*, *ERD7* y *GRP7*. (C) Términos GO enriquecidos para los genes diana compartidos por los tres factores de transcripción, como respuesta a sequía o a frío. (D) Regulación del gen *DREB2A* por los tres factores de transcripción. (E) Mismo análisis para *ERD7*. (F) Mismo análisis para *GRP7*.

Al explorar en ATTRACTOR la regulación de los genes *DREB2A* y *ERD7* por los factores de transcripción estudiados observamos que su expresión empieza a aumentar por la mañana hasta por la tarde, alcanzando el mínimo durante la noche (Figura 34D-E izquierda). Esto podría relacionarse con el hecho de que su expresión se necesita en las horas centrales del día, cuando hay menos humedad, ya que estos genes están implicados en estrés por sequía y alta salinidad (Sakuma et al., 2006; Kiyosue et al., 1994). Por otra parte, la expresión de *GRP7* muestra un patrón similar (Figura 34F izquierda). Esta proteína tiene un efecto en la apertura de estomas bajo condiciones de estrés hídrico y en el cierre de estomas bajo condiciones de estrés por frío (Kim et al., 2008). También se observó que estos factores de transcripción se unen a las mismas regiones en algunos de estos promotores (*GRP7*), mientras que en otros se asocian en regiones distintas (*DREB2A* y *ERD7*) en las que se encontraron G-boxes y cajas EE (Evening Element) (Figura 34D-E derecha). Por tanto, el motivo de red compuesto por CCA1, PIF5 y PRR5, regula la expresión de estos genes ejerciendo una acción coordinada a lo largo del día, lo que permite modular procesos periódicos en *Arabidopsis*.

Mediante este ejemplo hemos mostrado que ATTRACTOR permite explorar cómo cooperan el sistema del reloj circadiano y la señalización lumínica para regular procesos cruciales en el desarrollo vegetal, en lugar de estudiar individualmente cada factor de transcripción. Es conocido que uno de los procesos más importantes regulados por estos sistemas es la transición floral fotoperiódica y el posterior desarrollo floral. Por ello, se decidió integrar en la red transcripcional datos de factores de transcripción involucrados en floración, con el objetivo de investigar posibles relaciones no exploradas aún entre ambos procesos.



Capítulo 3

Control del reloj circadiano por la vía fotoperiódica en *Arabidopsis*



1 Existe una alta interconexión entre el reloj circadiano y la vía fotoperiódica en *Arabidopsis*

Con el objetivo de investigar la interacción entre el reloj circadiano y la señalización mediada por luz con la floración fotoperiódica se analizaron datos publicados de ChIP-seq de 33 factores de transcripción involucrados en señalización lumínica, reloj circadiano, fotoperiodo y floración (Anexo 5) y posteriormente se construyó una red transcripcional llamada CircadianFloralNet (Figura 35A). Esta red está compuesta por 20601 nodos y 89377 aristas, conteniendo así el 75% del genoma de *Arabidopsis*. A partir de la red principal también se generó la red formada sólo por los factores de transcripción analizados y las interacciones entre ellos. Esta subred, denominada CircadianFloralTFNet, conforma el núcleo regulador de CircadianFloralNet, y captura el programa transcripcional que gobierna la interacción entre el reloj circadiano y la vía fotoperiódica de la floración (Figura 35B).

CircadianFloralNet es una red libre de escala puesto que su distribución del grado de los nodos sigue una ley potencial negativa de acuerdo al test de Kolmogorov-Smirnov. Además, es una red de mundo pequeño, lo que implica que que la distancia entre dos nodos al azar es más pequeña que la esperada para una red aleatoria de las mismas características. Estas propiedades topológicas indican que los programas transcripcionales que controlan el reloj circadiano y la vía fotoperiódica de la floración están altamente conectados (Figura 35C).

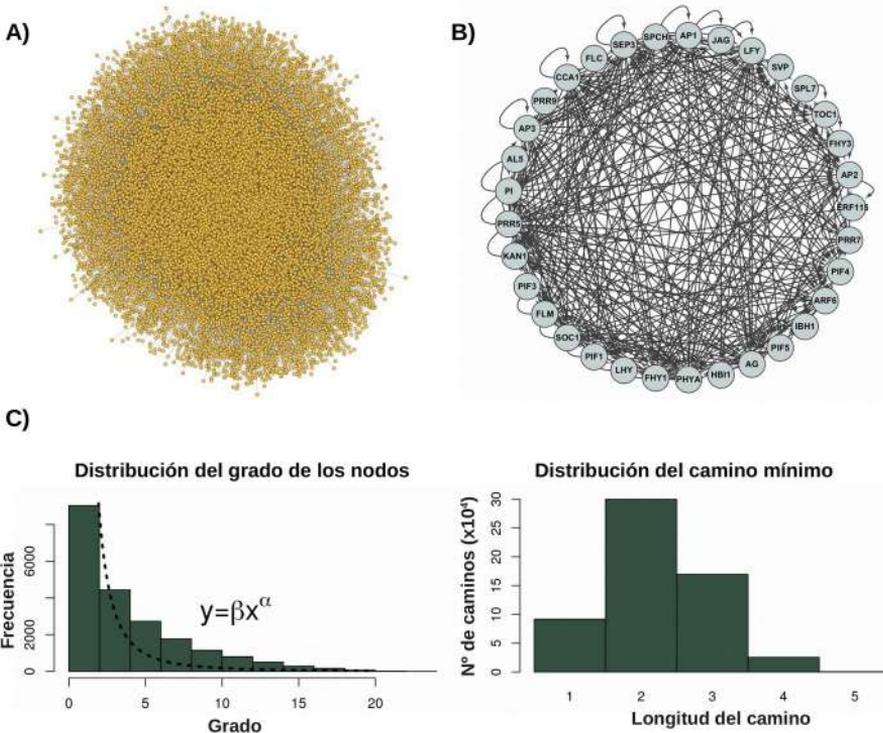


Figura 35. Red transcripcional CricadianFloralnet que integra datos de ChIP-seq provenientes de 33 factores de transcripción involucrados en ritmo circadiano, señalización lumínica, fotoperiodo y desarrollo floral. (A) Visualización gráfica de la red transcripcional. Está compuesta por 20601 nodos y 89377 aristas. **(B)** Red transcripcional inducida, CricadianFloralTFNet, compuesta únicamente por los 33 factores de transcripción analizados en este estudio y las interacciones entre ellos. Esta subred presenta 33 nodos y 341 aristas y constituye el núcleo regulador de la red global. **(C)** Análisis topológico de la red transcripcional. A la izquierda se representa la distribución del grado de los nodos, observándose que se ajusta a una ley potencial negativa, lo que indica que se trata de una red libre de escala. A la derecha, se muestra la distribución del camino mínimo entre dos nodos aleatorios de la red. Se trata de una red de mundo pequeño ya que la media del camino mínimo es significativamente menor que en redes aleatorias.

Para identificar componentes no aleatorios relevantes en las interacciones entre el reloj circadiano y la floración se llevó a cabo una identificación de motivos de red en la subred CircadianFloralTFNet. Como se ha descrito anteriormente, un motivo de red es un subgrafo que aparece significativamente con más frecuencia en una red de interés comparado con redes aleatorias con las mismas propiedades topológicas. Por tanto, se evaluó la significancia estadística de todos los posibles subgrafos de uno, dos o tres nodos generando 10000 grafos aleatorios con las mismas propiedades

topológicas que CircadianFloralTFNet. Estos motivos identificados se muestran en la Figura 36A. Uno de los más significativos y comunes fue el de “doble retroalimentación con salida” (p -valor $< 10^{-5}$). Esto indica una alta conexión existente dentro de la respuesta transcripcional que integra el reloj circadiano y la floración en *Arabidopsis*.

A)

Motivo	Diagrama	Apariciones en CFnet	Apariciones en grafos aleatorios (media \pm d.e)	p-valor
Autorregulación		24	10.32 \pm 2.63	$<10^{-5}$
Bucle de retroalimentación		62	51.67 \pm 5.03	1.65×10^{-2}
Bucle de retroalimentación con autorregulación		56	5.025 \pm 3.264	$<10^{-5}$
Regulador con salida múltiple		494	358 \pm 29.96	$<10^{-5}$
Bucle de retroalimentación con salida		235	73.93 \pm 8.53	$<10^{-5}$
Bucle de anteroalimentación/retroalimentación		187	66.64 \pm 12.58	$<10^{-5}$
Triple bucle de retroalimentación		50	4.98 \pm 2.61	$<10^{-5}$

B)



Figura 36. Motivos de red encontrados en la red transcripcional inducida. (A) Tabla resumen de los motivos más representativos encontrados en la subred CircadianFloralTFNet. Se muestra un diagrama ilustrativo, las apariciones en la red, la media y la desviación estándar de las apariciones en las 10000 redes generadas aleatoriamente y el p -valor estimado. **(B)** Gráfico de la regulación de PRR5 sobre CO identificada en la red transcripcional (rojo, línea continua), y la posible regulación de CO sobre PRR5 (verde, línea discontinua).

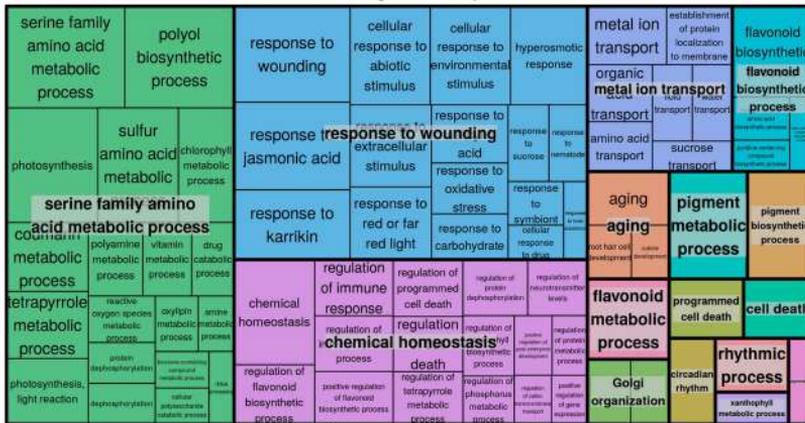
De acuerdo con la conectividad identificada en nuestra red transcripcional y a estudios previos (Suárez-López et al., 2001; Imaizumi et al., 2005; Sawa et al., 2007), la expresión de *CO* presenta una alta dependencia del reloj circadiano. Concretamente *PRR5* se une a elementos cis del promotor de *CO* para regular su expresión y esta conexión puede también observarse en CircadianFloralNet. Además, un número sustancial de genes de la familia *CO-like* (*COL*) están regulados por *PRR5* y otros factores de transcripción centrales en el reloj circadiano, como *CCA1* y *PRR7*. La alta regulación de genes *COL* por parte del reloj se puede examinar en la subred presentada en el Anexo 6. Teniendo en cuenta que la doble retroalimentación se encontró de forma significativa en CircadianFLoralNet y que *PRR5* regula a *CO*, es previsible que *CO* también regule a *PRR5* (Figura 36B). Esta conexión no podía existir en la red puesto que no se disponía de datos de ChIP-seq de *CO*. Así se predijo que una de las interacciones entre la transición floral y el reloj circadiano podría estar conformada por *CO* y *PRR5*.

2 CONSTANS altera el reloj circadiano bajo condiciones de Día Largo

Para analizar en qué procesos biológicos además de la floración podría estar implicado *CO*, incluyendo un posible papel sobre el reloj circadiano en *Arabidopsis*, se analizaron datos de RNAseq en plantas 35S:*CO* y Col-0 a ZT16 en condiciones LD generados previamente en el laboratorio. Se encontraron 540 genes activados en plantas sobreexpresoras estableciendo un *fold-change* de 1,5. De acuerdo al análisis de enriquecimiento en términos de ontología génica, se encontró que este grupo de genes están implicados en varios procesos biológicos como respuesta a estímulos, procesos de desarrollo y sorprendentemente, reloj circadiano (Figura 37A).

Concretamente, varios genes que codifican proteínas reguladoras del reloj circadiano mostraron una mayor expresión en las plantas 35S:*CO* (Figura 37B), como *GI*, *PSEUDO-RESPONSE REGULATORS 5/7/9*, los *REVEILLES RVE1* y *RVE8*, *LNK1*, *LNK2*, *LHY* y *CCA1*. Curiosamente, también mostraron una sobreexpresión los genes *CDF1*, *CDF3* y *CDF6*, que codifican represores transcripcionales de *CO* y *FT*.

A



B

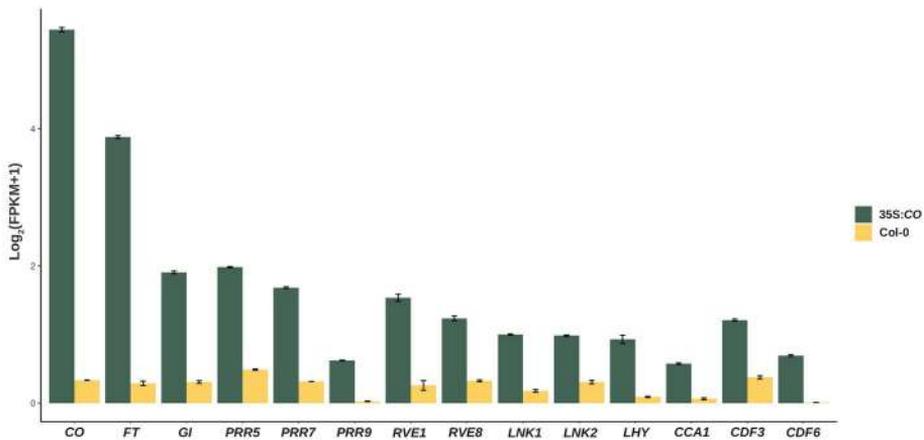


Figura 37. Efecto de la sobreexpresión de CO sobre el transcriptoma de Arabidopsis. (A) Treemap que resume los términos GO enriquecidos para los genes sobreexpresados en plantas 35S:CO. Se encontraron los términos específicos “*rhythmic process*” y “*circadian rhythm*” entre otros. Cada rectángulo del *treemap* representa el $-\log_{10}(p\text{-valor})$ para el correspondiente término GO. **(B) Expresión génica** normalizada de genes clave en el reloj circadiano que mostraron sobreexpresión: *GI*, *PRR5*, *PRR7*, *PRR9*, *RVE1*, *RVE8*, *LNK1*, *LNK2*, *LHY*, *CCA1*. Además, tres genes de la familia *CDF*; *CDF1*, *CDF3* y *CDF6* presentaron mayor expresión en el sobreexpresor de CO. Se incluyen como controles la expresión de CO y de *FT*.

Teniendo en cuenta que CO podía afectar al reloj, se analizaron los patrones circadianos regulados por éste. Para ello se emplearon los datos de RNAseq previamente analizados en plantas silvestres bajo condiciones LD en un periodo de 24

h (GSE43865; Rugnone et al., 2013) y se estimó la expresión génica en los puntos ZT2, ZT6, ZT10, ZT14, ZT18 y ZT22. Posteriormente, se identificaron los genes que presentaban un patrón rítmico o circadiano de expresión y fueron clasificados en *clusters* dependiendo de la hora donde tuvieron su nivel máximo y mínimo de expresión, tal y como se describió en el capítulo 1 de Resultados.

Para conocer los patrones circadianos afectados por CO, los genes activados en 35S:CO a ZT16 fueron solapados con los grupos de genes circadianos y se calculó la significancia estadística. Únicamente seis grupos circadianos presentaron una intersección significativa de acuerdo al test exacto de Fisher, como se muestra en el Anexo 7. De esos seis, cinco presentaron un máximo de expresión en la mañana o durante el día (ZT2 y ZT6), por lo que los unimos en un único grupo y de nuevo determinamos una intersección significativa de este grupo de genes con los genes activados en 35S:CO a ZT16, en concordancia con estudios previos (Anexo 8) (Gnesutta et al., 2017). Este resultado indica que CO por la tarde en LD activa principalmente genes que bajo condiciones de ND tienen su máximo por la mañana o durante el día, alterando sus patrones de expresión.

Para confirmar el efecto del fotoperiodo en el reloj circadiano, se decidió evaluar si CO era capaz de alterar aspectos fisiológicos del ritmo circadiano en *Arabidopsis*. Está descrito que en plantas uno de los comportamientos rítmicos más conocidos y caracterizados es el movimiento periódico diario de las hojas. Un desajuste en este movimiento indica alteraciones en el reloj circadiano, por lo que se decidió medir este fenotipo en plantas Col-0, *co-10* y 35S:CO. Así, se cultivaron las plántulas de *Arabidopsis* durante 8 días en condiciones LD y se transfirieron a una cámara con luz y temperatura constante, para eliminar cualquier estímulo de entrada del reloj circadiano (*free-running conditions*). Se monitorizó el movimiento de las dos primeras hojas y se estimaron parámetros circadianos como el periodo (el tiempo requerido para completar un ciclo circadiano), y la fase (el punto temporal en el que las hojas alcanzan el punto más alto). Como se muestra en la Figura 38A, y su representación en el gráfico *boxplot* (Figura 38B), las plantas 35S:CO mostraron un periodo significativamente más corto que las plantas silvestres (Col-0), mientras que el mutante *co-10* exhibió un periodo más largo.

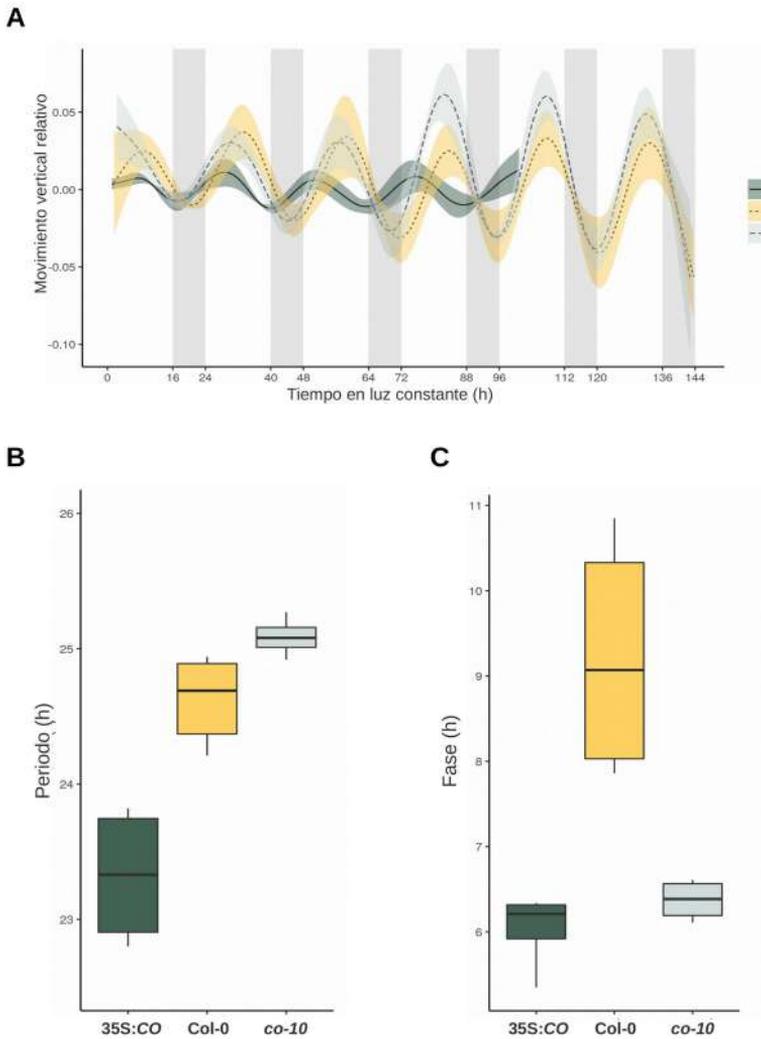


Figura 38. Efecto de la sobreexpresión y la mutación de CO sobre el movimiento de rítmico de hojas en Arabidopsis. (A) Movimiento relativo de las dos primeras hojas y los cotiledones en función del tiempo para las líneas Col-0, 35S:CO y co-10. El sombreado gris representa las horas correspondientes a la noche en condiciones de luz continua. La posición vertical de las hojas se suavizó mediante ajuste por polinomios (ajuste por *splines*) con 12 grados de libertad. Se representa la media \pm desviación estándar de 6 réplicas biológicas. (B) *Boxplot* mostrando el periodo obtenido del movimiento vertical para cada línea. Las plantas 35S:CO mostraron un periodo significativamente más corto mientras que las plantas mutantes co-10 mostraron un periodo alargado. (C) *Boxplot* que representa la fase (punto temporal en el que se alcanza al punto vertical más alto) para cada línea. Las plantas sobreexpresoras y mutantes mostraron una fase previa a las plantas Col-0. La fase está normalizada a “tiempo circadiano”, es decir, se ajusta el periodo a 24 h y se recalcula la fase.

Curiosamente, aunque también estaba alterado frente al silvestre, este comportamiento no se conservó en la fase del ritmo de movimiento de hojas, ya que 35S:CO y *co-10* mostraron una fase más corta comparada con las plantas Col-0 (Figura 38C). En cualquier caso, el cambio en el ritmo periódico del movimiento de las hojas en los mutantes de CO constituye una evidencia fisiológica de la participación de CO en la regulación del reloj circadiano.

Estos resultados indican que CO no es sólo una salida del reloj circadiano, sino que podría ejercer una regulación de retroalimentación sobre genes centrales del reloj y sobre patrones circadianos de expresión génica.

3 CONSTANS se une a promotores de genes centrales del reloj circadiano.

De acuerdo a la significancia estadística de los motivos de red encontrados en la red transcripcional se dedujo que CO podría regular la expresión de *PRR5* uniéndose a su promotor, lo que provocaría la alteración en el ritmo circadiano observado en las plantas mutantes y sobreexpresoras de CO. Presumiblemente, de acuerdo a nuestros datos de expresión, podría ocurrir que CO también estuviera regulando a otros genes clave del reloj que aparecieron sobreexpresados en plantas 35S:CO. Para averiguar esto y determinar los sitios de unión de CO a nivel de genoma completo se llevó a cabo un experimento de ChIPseq empleando las plantas 35S:CO y *co-10* (control). Específicamente se determinaron los picos o sitios de unión al genoma utilizando las muestras IP e input (DNA genómico) para cada línea y posteriormente se eliminaron los picos que aparecían en las dos líneas, considerándose falsos positivos.

Se determinaron 2417 genes diana siguiendo el criterio del gen más cercano al pico cadena abajo con un límite de 2000 pb. Se comparó el conjunto de dianas de CO con los genes activados en plantas 35S:CO en condiciones LD y se observó un solapamiento significativo con un p-valor de $1,006943 \times 10^{-139}$. De los 540 genes sobreexpresados en 35S:CO en el experimento de RNAseq, 318 eran dianas directas de CO. Los 222 restantes pueden considerarse genes activados por CO indirectamente. Además, se identificaron 2100 genes diana que no mostraron

sobreexpresión (Figura 39A). Es frecuente en factores de transcripción observar su unión a promotores *in vivo* sin afectar a los niveles de expresión de los genes diana, lo que sugiere que son necesarios en estas regiones la combinación de otros factores de transcripción para activar o reprimir la expresión génica. Por otra parte, se analizó la localización de los sitios de unión de CO a lo largo del genoma, determinándose que se asocia mayoritariamente en la región de los promotores más cercana al TSS con un 64.87%. Un 3,48% de los picos se encontraban en regiones 3'UTR, lo que puede indicar otros mecanismos de regulación génica como la existencia de bucles en la cromatina. Este tipo de mecanismo se ha descrito anteriormente en el homólogo *SP5G* de *FT* en tomate, cuya región 3'UTR regula su expresión génica (Zhang et al., 2018).

Se ha descrito que CO regula la expresión génica del florígeno *FT* a través de la unión a su promotor. Se han publicado numerosos estudios que detallan cómo el trímero de los factores NF-Ys se une a sitios CCAAT y puede estabilizar y reclutar a CO al promotor de *FT* en sitios CORE, con la posibilidad de formar bucles de cromatina que aproximen estos dos elementos del DNA (Tiwari et al., 2010; Cao et al., 2014; Wenkel et al., 2006). No obstante, en estudios más recientes se describe que la sustitución en el trímero de NF-YA por CO otorgaría especificidad de unión al DNA al complejo para unir elementos CCACA en lugar de CCAAT (Gnesutta et al., 2017). En este trabajo, para investigar las secuencias de DNA asociadas a CO se realizó en primer lugar una identificación de motivos *de novo* sobre los sitios de unión de CO, en la que las secuencias más significativas se detallan en la Figura 39C. Como se puede observar, los tres primeros elementos más significativos contienen la secuencia CCACA descrita en (Gnesutta et al., 2017), en tercer lugar una secuencia similar al CORE (Tiwari et al., 2010) pero que también contiene una caja CCACA (TGTGG) y curiosamente una caja G-box (CACGTG) para factores de transcripción BHLH o bZIP. De esta forma se obtuvieron las matrices de pesos posicionales o PWM (*position weight matrix*) de estos elementos con la intención de incluirlas en la base de datos JASPAR. Para simplificar, en la búsqueda de motivos de DNA sobre los sitios de unión analizados se incluyeron los elementos G-box (CACGTG), CCACA y los sitios CORE1

y CORE2, (Figura 40 y 43B) ya que la variabilidad observada en nuestros resultados es lo que se espera en experimentos masivos en el que se analizan miles de picos.

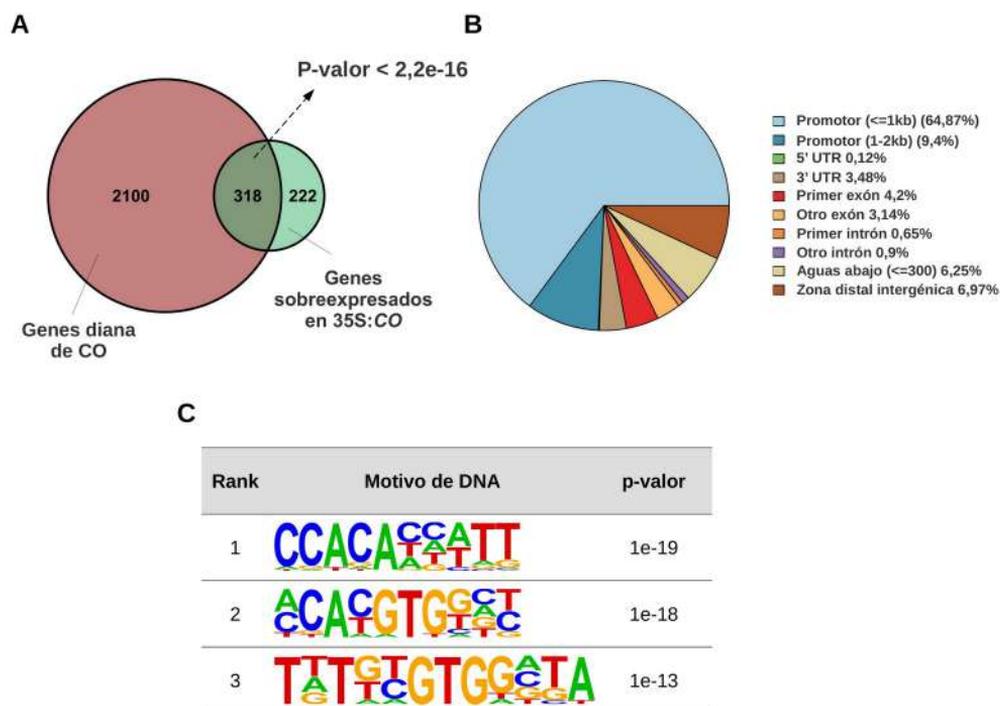


Figura 39. Anotación de los sitios de unión (picos) de CO al genoma de *Arabidopsis*. (A) Intersección entre los genes diana de CO y los genes activados en plantas 35S:CO. (B) Localización de los picos a lo largo del genoma. Se observa una acumulación en la región de los promotores más cercana al TSS. (C) Motivos de DNA enriquecidos en los sitios de unión de CO al DNA. Encontramos tres secuencias que contienen una G-box (CACGTG) y un elemento CCACA y en tercer lugar una secuencia que contiene la caja CCACA (TGTGG) y es similar al CORE descrito en el promotor de *FT*.

Como se esperaba, se observó que CO se une al promotor de *FT* en el sitio correspondiente a los elementos CORE1/2. También se encontraron sitios de unión en las dos zonas descritas en (Cao et al., 2014) que participan en bucles (proximal y distal). Confirmando nuestra hipótesis, se identificaron sitios de unión en los promotores de genes clave del reloj circadiano, incluyendo *PRR5*, *GI*, *RVE1*, *RVE8*, *LNK1* y *LNK2*, entre otros. Por otra parte, también se encontraron zonas enriquecidas en los promotores de algunos genes de la familia *CYCLING DOF FACTOR*, como *CDF1* y *CDF3* (Figura 40).

Control del reloj circadiano por la vía fotoperiódica en Arabidopsis

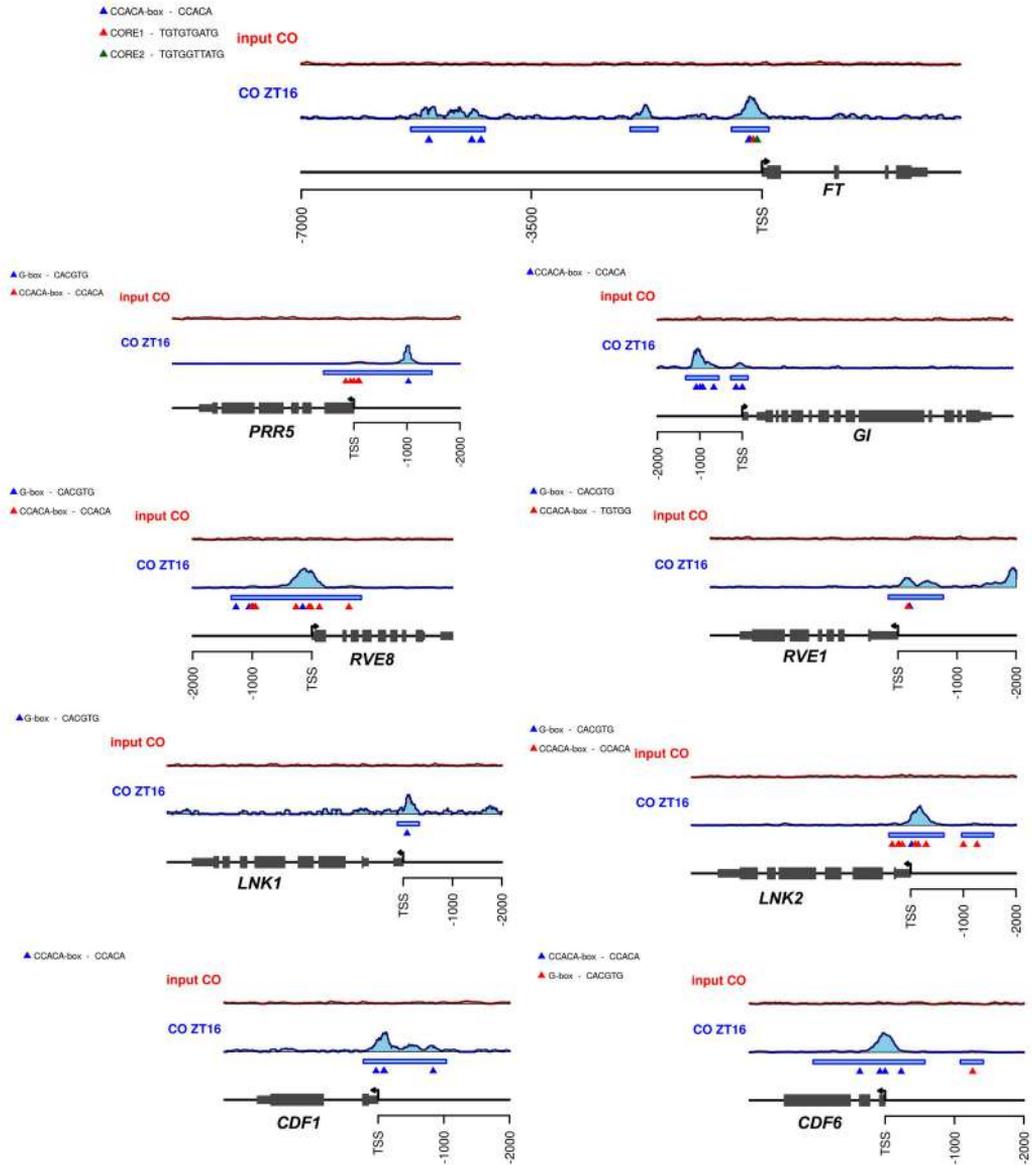


Figura 40. Unión de CO en promotores de genes involucrados en floración y reloj circadiano. Sitios de unión de CO en promotores de genes clave involucrados en floración (*FT*), reloj circadiano (*PRR5*, *GI*, *RVE1*, *RVE8*, *LNK1* y *LNK2*) y genes salida del reloj como *CDF1* y *CDF3*. Se indican los motivos encontrados con un umbral mínimo de puntuación del 95%.

Por tanto, CO se asocia a regiones que contienen elementos G-box en los promotores de genes circadianos o en genes que son salida del reloj, pero en otros

Control del reloj circadiano por la vía fotoperiódica en *Arabidopsis*

como en el de *FT* se asocia a en regiones que contienen secuencias CCACA y CORE, lo que indica un mecanismo independiente del complejo NF-CO. Estos resultados apoyan la existencia de una función no explorada de CO, alternativa a la vía fotoperiódica y al módulo CO-FT, regulando ritmos circadianos en *Arabidopsis* en condiciones LD.

De forma adicional se verificó la unión de CO a algunos sitios de estos promotores mediante ChIP-QPCR empleando las plantas 35S:CO y *co-10* (control). Se utilizaron cebadores específicos para la región del promotor de *Gf* en la que se encuentran cajas CCAAT y una secuencia similar una G-box CACGTA, la región G-box del promotor de *PRR5*, y la zona CORE1/2 del promotor de *FT* indicadas en la Figura 41A. Estos experimentos verificaron que CO se une a estos promotores en condiciones LD (ZT16), ya que estos amplicones presentaban un enriquecimiento significativamente mayor en 35S:CO que en las plantas control (Figura 41B).

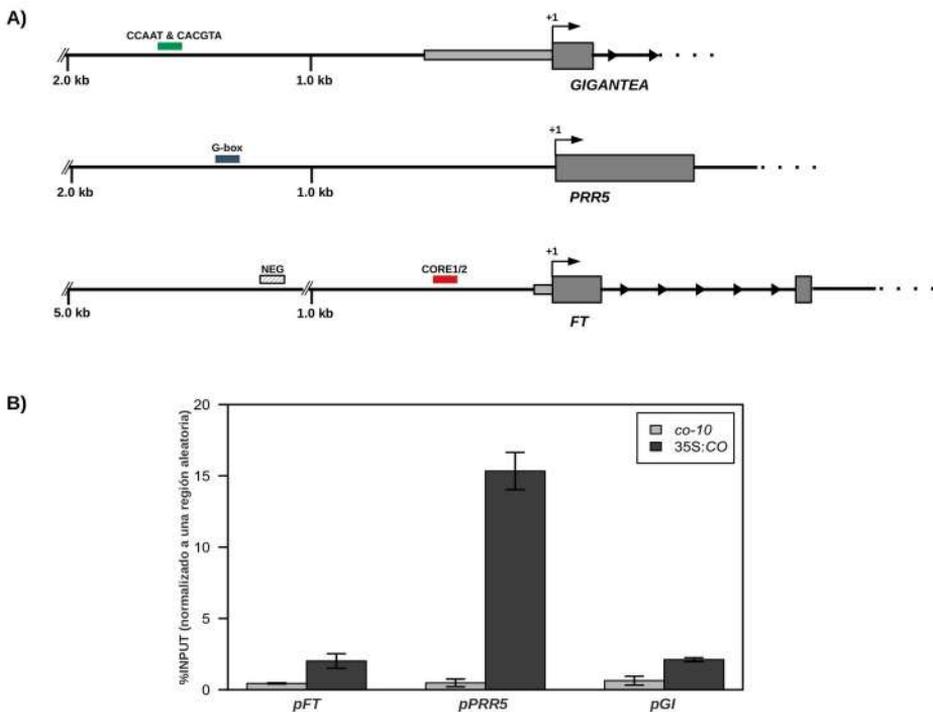


Figura 41. ChIP-QPCR de CO. (A) Amplicones analizados en los experimentos de QPCR. **(B)** Enriquecimiento obtenido sobre las regiones amplificadas en las plantas 35S:CO y *co-10*. Los resultados se normalizaron al enriquecimiento obtenido en una región aleatoria (NEG) del promotor de *FT*.

Una vez obtenidos los genes diana de CO, se integraron en la red transcripcional previamente generada. Como se había predicho, CO y PRR5 conformaron un bucle de retroalimentación en el que se regulan mutuamente. Además, un análisis más detallado reveló que se trata de un bucle de retroalimentación con salida múltiple, como se muestra en la Figura 42B, con una serie de genes diana en común. Como se ha descrito anteriormente, la familia de proteínas COLs y PRRs comparten un dominio CCT en su extremo carboxilo que contiene una señal de localización nuclear y un motivo de unión al DNA (Valverde, 2011). Esto sugiere que CO y estos factores de transcripción podrían compartir sitios de unión al DNA, y por lo tanto genes diana en común, como se ha propuesto en estudios recientes (Shen et al., 2020). Para evaluar esta hipótesis se analizó la significancia de la intersección entre los genes diana de CO con los genes diana de TOC1, PRR5, PRR7 y PRR9. Las dianas de TOC1, PRR5, PRR7 y PRR9 solaparon significativamente con las de CO con los p-valores $9,067828 \times 10^{-44}$, $1,552583 \times 10^{-249}$, $5,649396 \times 10^{-119}$ y $1,359 \times 10^{-2}$, respectivamente, de acuerdo al test exacto de Fisher (Figura 42A). PRR5 y CO se coexpresan en condiciones LD, y el hecho de que CO regule un número significativo de dianas de PRR5 apoya la existencia de un bucle de retroalimentación con salida múltiple, en el que los reguladores son CO y PRR5, constituyendo una de las conexiones entre los programas transcripcionales de la vía fotoperiódica y el reloj circadiano (Figura 42B). Como se observa en este motivo de red, algunos genes relevantes que son salida del bucle de retroalimentación son *Gl*, *PRR7*, *BETA-AMYLASE 3 (BAM9)* y el mismo *PRR5*. Para apoyar este resultado y validar alternativamente los datos de RNAseq, se midió la expresión génica de algunos de estas dianas compartidas por CO y PRR5 mediante experimentos de RT-QPCR durante un día largo completo. Así se determinó que la expresión de *PRR5*, *PRR7*, *Gl* y *BAM9* era mayor en plantas sobreexpresoras de CO, específicamente a ZT16, cuando se aumenta significativamente la estabilidad de CO, aunque también se observa cierto efecto a primera hora de la mañana (Figura 42C).

Control del reloj circadiano por la vía fotoperiódica en Arabidopsis

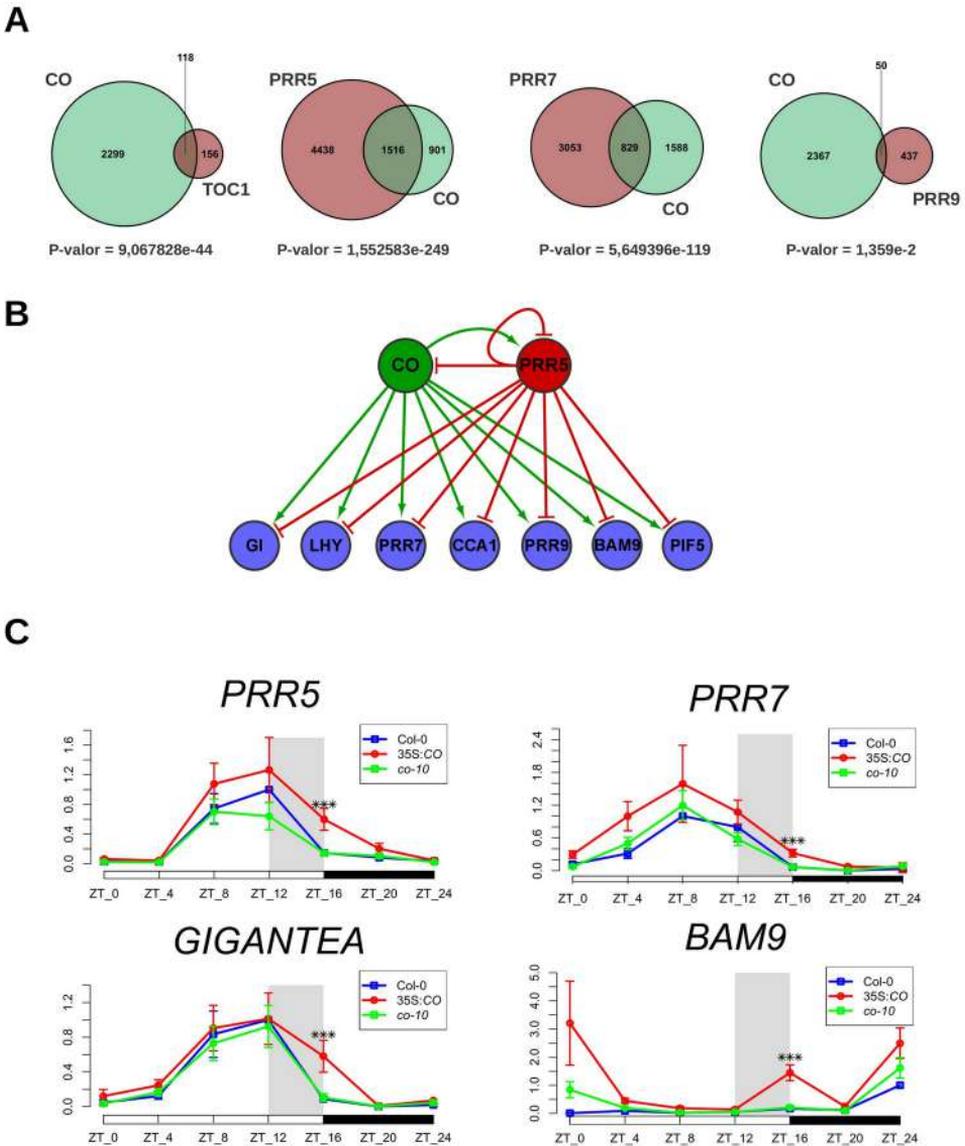


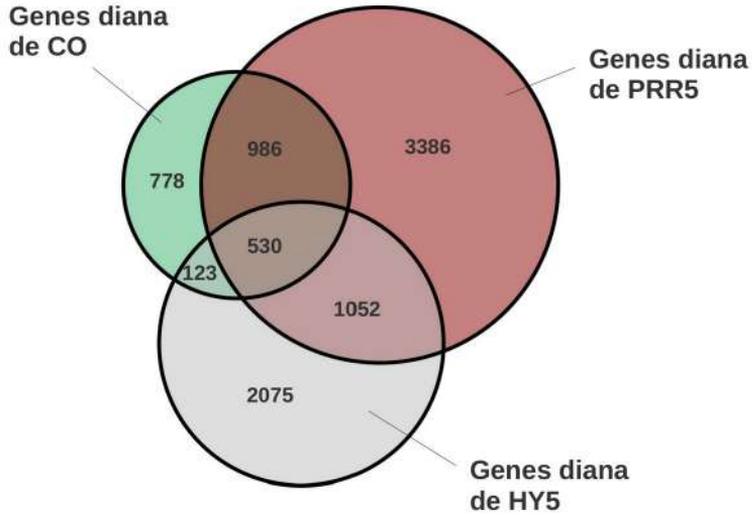
Figura 42. CO y PRR5 comparten un número significativo de dianas en común. (A) Solapamiento significativo entre los genes diana de CO y los genes diana de los PRRs TOC1, PRR5, PRR7 y PRR9. **(B)** Motivo de DNA de doble retroalimentación con salida encontrado en CircadianFloralNet, en el que los reguladores son CO y PRR5 afectando a la expresión de una serie de genes en común. **(C)** Expresión génica de algunos de los genes salida del motivo de red formado por CO y PRR5. Se muestra el perfil de expresión medido por RT-QPCR en condiciones LD de los genes *PRR5*, *PRR7*, *GI* y *BAM9* en las plantas 35S:CO (rojo), Col-0 (azul) y *co-10* (verde). En gris se indica el periodo de la ventana de coincidencia, momento en el que CO es estable en LD.

4 CO, PRRs y HY5 comparten sitios de unión al genoma en los promotores de genes regulados por el reloj

Anteriormente se había descrito que el motivo G-box de DNA participa en la regulación transcripcional ejercida por la familia de los PRRs, sugiriendo que su unión al DNA es indirecta, a través de otros factores de transcripción (Liu et al., 2016). En condiciones de día corto, los PRR9/7/5 y los factores BHLH tipo PIF podrían interactuar y unirse a las mismas regiones de promotores de genes relacionados con el crecimiento, como *CDF5* (Martín et al., 2018). En este estudio proponemos al factor de transcripción bZIP HY5 como intermediario en la unión de PRR5 y CO a promotores de genes relacionados con el reloj. Este tipo de factores de transcripción bZIP, generalmente presentan una unión directa al DNA (Chattopadhyay et al., 1998; Abbas et al., 2014; Binkert et al., 2014). HY5 se acumula en condiciones de luz (Osterlund et al., 2000), uno de sus motivos de unión al DNA es la G-box (Lee et al., 2007; Young et al., 2008; Zhang et al., 2011; Binkert et al., 2014), su degradación por proteasoma está mediada por COP1 (Osterlund et al., 2000) y por otra parte se ha descrito que CO y HY5 interactúan con el mismo remodelador de la cromatina, PICKLE (Jing et al., 2013, 2019). Estos indicios junto con resultados previos sugerían que HY5 era una de las proteínas que participaba en la unión al DNA.

El análisis a nivel de genoma completo de los datos de ChIP-seq de PRR5 (Nakamichi et al., 2012), ChIP-on-chip de HY5 (Lee et al., 2007) y las dianas de CO obtenidas de nuestro experimento de ChIP-seq, mostró un solapamiento significativo entre los genes diana de PRR5, HY5 y CO, con un p-valor < 0,001 y un enriquecimiento de 14,76 según el test realizado con la librería *SuperExactTest*. (Figura 43A). Posteriormente, para verificar que no sólo regulan a genes en común, sino que ocupan las mismas regiones en los promotores diana, se midió la significancia del solapamiento entre los sitios de unión de CO, HY5 y PRR5 a través de tests de permutaciones, obteniéndose p-valores significativos (Anexo 9). En la Figura 43B se representa el promotor de *LHY* uno de los genes co-regulados por CO, HY5 y PRR5.

A



P-valor < 0,001

B

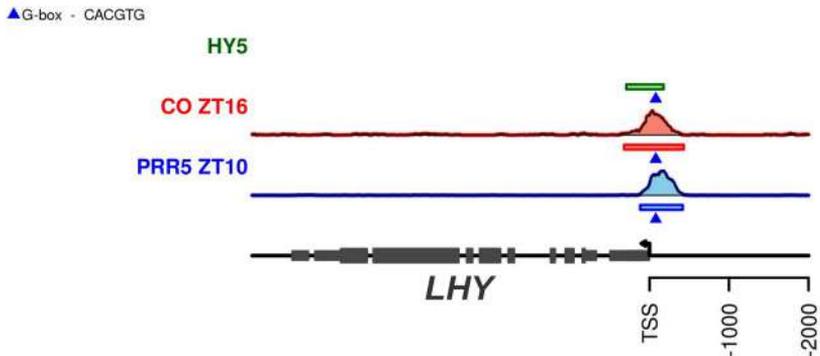


Figura 43. CO, PRR5 y HY5 podrían asociarse en el genoma para regular genes diana en común. (A) Solapamiento significativo de los genes diana de CO, PRR5 y HY5. **(B)** Solapamiento significativo de los sitios de unión de CO, HY5 y los PRRs. Se muestra el locus *LHY*.

Para demostrar la unión de HY5 a estas regiones del DNA en las mismas condiciones que CO (10 DAG, LD), llevamos a cabo un experimento de ChIP en plantas *35S:GFP:HY5 (hy5-211)* y observamos un enriquecimiento significativo de los amplicones analizados comparado con la planta control Col-0 (Figura 44). Estos

resultados indican que HY5 se une a los mismos sitios que CO en estos promotores del reloj circadiano, pero no en los sitios CORE asociados a CO en el promotor de *FT*, apuntando a que ésta es una vía regulación diferente llevada a cabo por CO relacionada con el ritmo circadiano, independiente de la vía fotoperiódica de la floración.

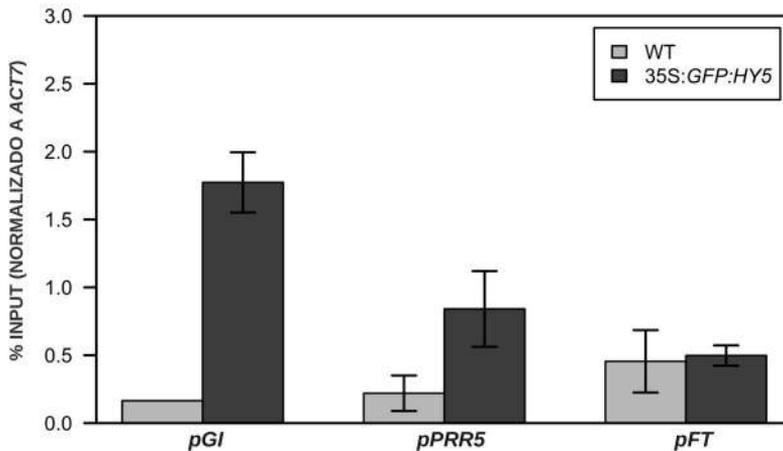


Figura 44. ChIP-QPCR de HY5. Se muestra el enriquecimiento en los amplicones indicados tras los experimentos de ChIP-QPCR en plantas Col-0 y 35S::GFP::HY5. Se observa que HY5 se une a regiones en los promotores de *GI* y *PRR5* pero no al sitio CORE1/2 localizado en el promotor de *FT*. El enriquecimiento de cada región se normalizó al obtenido en la región control del TSS de *ACT7*.

5 CO, PRR5 y HY5 interaccionan *in vivo*

En estudios previos se ha observado la interacción entre CO y algunos miembros de la familia de los PRRs y que esta unión estabiliza a CO promoviendo la floración en condiciones LD (Hayama et al., 2017). Sin embargo, el hecho de que CO y PRR5 se unan a las mismas regiones G-box en algunos promotores y que estas dos proteínas tengan efectos opuestos sobre la transcripción, sugiere que CO podría eliminar gradualmente a PRR5 del DNA, ocupar estas regiones y activar la transcripción, promoviendo un cambio de fase en la expresión de los genes diana. Por tanto, podría existir un complejo transitorio entre HY5, CO y PRR5 durante la tarde en condiciones LD. Para investigar este complejo, se realizaron experimentos de interacción proteína-proteína en planta.

En primer lugar, se llevó a cabo un ensayo BiFC en *Nicotiana* entre CO, HY5 y PRR5 por pares, observando interacción en los tres casos. En todos ellos se observaba una clara interacción entre los factores dos a dos, aunque mostraban diferentes patrones de distribución (Figura 45A). CO-HY5 mostraban una interacción nuclear homogénea, CO-PRR5 mostraban una distribución en gránulos nucleares intensos, mientras que PRR-HY5 se distribuían en gránulos de menor intensidad. Para investigar más profundamente la posibilidad de un complejo de 3 proteínas, medimos su colocalización en el núcleo de células de *Nicotiana*. Para ello, la proteína fluorescente amarilla (YFP) fue reconstituida fusionando cada una de sus partes a HY5 y a PRR5 y por otra parte se coinfiltró CO fusionada a la proteína fluorescente cian (CO:CFP) con el fin de poder detectar ambos fluoróforos independientemente. De esta manera pudo detectarse una colocalización nuclear clara comparado con el control (Figura 45B). De manera similar a lo observado en las interacciones dos a dos, la colocalización se realizaba en gránulos nucleares donde se había ya identificado a CO con anterioridad.

Por otra parte, con el objetivo de confirmar que existía una interacción estrecha y no sólo colocalización entre los tres factores de transcripción, se realizaron ensayos de FRET con los dos fluoróforos empleados anteriormente: YFP (reconstituida con HY5 y PRR5 fusionadas a cada una de las partes de la proteína fluorescente) y CFP fusionada a CO. Como se observa en la Figura 45C, los valores de N_{fret} para el complejo CO-HY5-PRR5 son significativamente mayores que los obtenidos en los correspondientes controles.

Dado que la interacción entre CO y PRR5 ha sido demostrada con anterioridad, resultaba de interés mostrar la interacción física entre CO y HY5. Ya que CO y HY5 parecen interactuar en el núcleo y que se unen a las mismas regiones en algunos promotores, investigamos si formaban un complejo para unirse al DNA. Para ello, los complejos de proteínas inmunoprecipitados resultantes de experimentos de ChIP fueron analizados mediante inmunoblot usando anticuerpos específicos anti-HY5. El ensayo mostró la presencia de la proteína HY5 en la muestra de plantas 35S:CO y no en el control *co-10* lo que apoyaba la interacción entre ambas proteínas y que esta unión está asociada al DNA (Figura 45C).

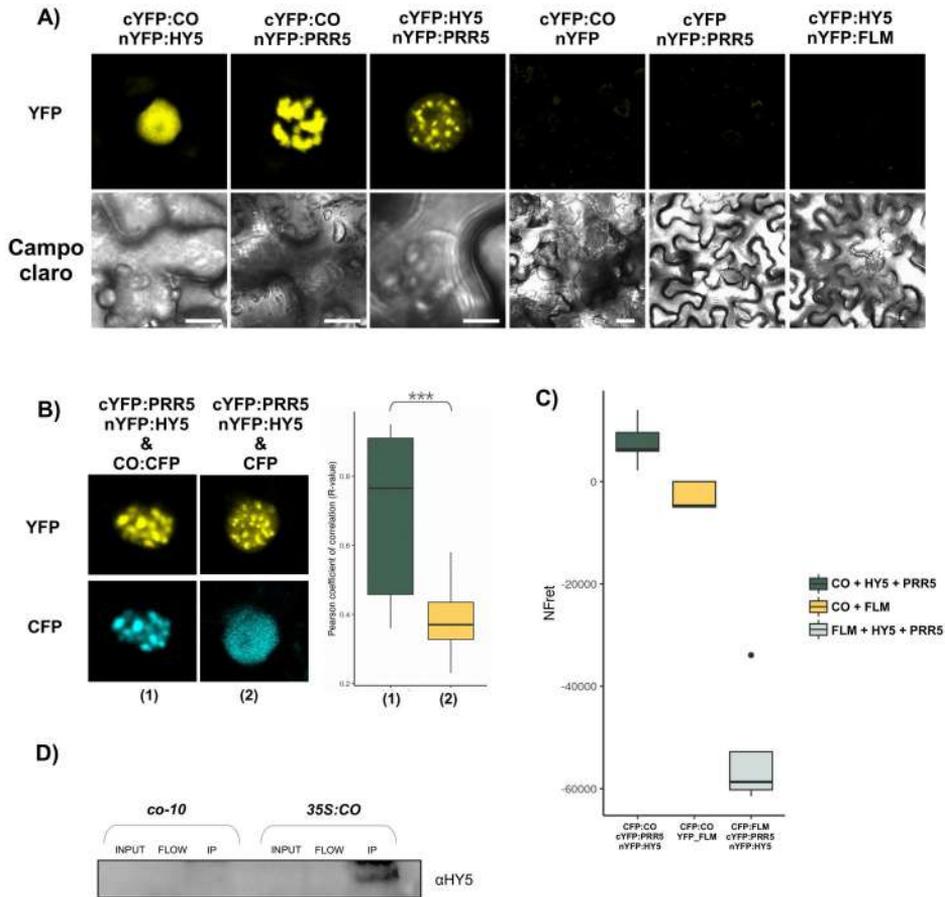
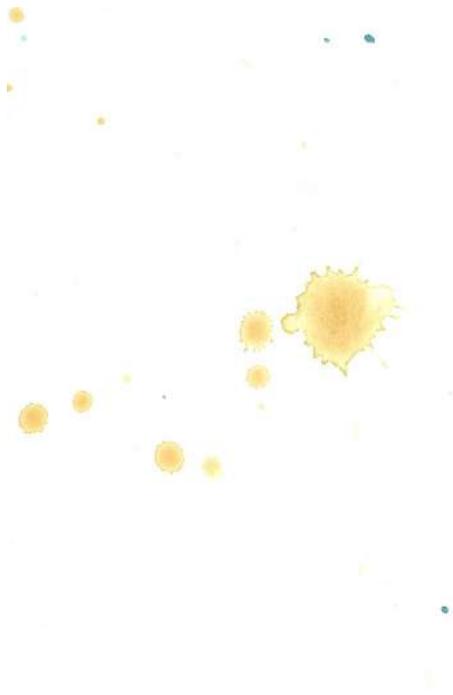


Figura 45. Interacción de las proteínas CO, HY5 y PRR5. (A) Núcleos de *Nicotiana* observados al microscopio confocal donde se observa interacción mediante BiFC por pares entre las tres proteínas. La barra de tamaño corresponde a 10 micras. (B) Experimento de colocalización de las tres proteínas. El complejo se localiza en *speckles* cuando se encuentran presentes los tres factores de transcripción. (C) Valores de Nfret para cada combinación. Se observa una interacción significativamente mayor en la combinación CO-HY5-PRR5 comparada con los controles en los que se incluye el factor de transcripción FLM. (D) ChIP-WB. La inmunoprecipitación de la cromatina empleando el anticuerpo anti-CO se analizó por inmunoblot con el anticuerpo anti-HY5. Se observa una banda correspondiente a HY5 en 35S:CO y no el control *co-10*.

Estos resultados indican que estas tres proteínas podrían formar un complejo de unión al DNA, en el que HY5 actuaría como mediador uniéndose directamente con el elemento G-box.



Discusión



Con la acumulación de datos transcriptómicos y filogenómicos, se hace más evidente la complejidad existente en la evolución de los patrones de expresión génica en eucariotas. Existe un patrón específico que controla el transcriptoma bajo ciclos de luz y oscuridad (Millar, 2016). En organismos fotosintéticos, este control se hace especialmente importante debido a la dependencia de la luz solar para llevar a cabo los principales procesos fisiológicos, así, se puede identificar un mecanismo periódico de expresión génica presente ya en las microalgas (Mittag et al., 2005; Corellou et al., 2009). En esta tesis se ha empleado una aproximación basada en Biología Molecular de Sistemas para investigar esta dependencia del fotoperiodo y para encontrar mecanismos controlados ancestralmente por ritmos diarios y aquellos que han adquirido una nueva regulación. Para llevar a cabo dicho estudio, se ha generado la herramienta MBBH, que ha permitido identificar ortólogos en especies evolutivamente distantes, como microalgas y plantas. La combinación con redes de coexpresión génica nos ha permitido entender cómo diferentes procesos han conservado una regulación periódica a lo largo de la historia evolutiva de la línea vegetal. Además, se ha desarrollado una herramienta web llamada CircadiaNet que permite el estudio de la evolución de genes rítmicos desde algas a plantas superiores de una forma similar a herramientas previas que comparan perfiles de expresión diurnos entre monocotiledóneas y dicotiledóneas (Mockler et al., 2007).

Por otra parte, a través de la construcción y el análisis de redes transcripcionales se ha explorado la regulación coordinada de la señalización lumínica y el reloj interno autónomo en *Arabidopsis*. Se ha construido una herramienta web interactiva para estudiar esta regulación sobre genes individuales y analizar el efecto de varios factores de transcripción sobre grupos de genes con expresión rítmica. Estas herramientas junto con flujos de trabajo, bases de datos y medios web previamente desarrollados en el laboratorio (Romero-Campero et al., 2013, 2016; Romero-Losada et al., 2021), además de aportarnos datos inéditos sobre la evolución y la composición de las redes de expresión de algas y plantas, constituyen recursos útiles para la comunidad científica.

Por último, de nuevo a través de la integración de datos en redes génicas y el análisis de múltiples datos ómicos se ha explorado una conexión desconocida entre la

vía fotoperiódica y el reloj circadiano, constituida por PRR5 y CO. Así CO adquiere una función novedosa afectando al reloj circadiano y a vías controladas por éste. Por tanto, la monitorización de la longitud del día ejercería una señalización retrógrada hacia el reloj circadiano para proveer información estacional en el control de los procesos diarios.

1 El reloj circadiano regula más procesos transcripcionales en algas que en plantas

Usando las herramientas generadas, el algoritmo MBBH y las redes de coexpresión génica, confirmamos que en las microalgas picoeucariotas primitivas como *Ostreococcus*, prácticamente todo el transcriptoma está gobernado por mecanismos rítmicos diarios (Monnier et al., 2010). Otras microalgas han reducido la dependencia de regulación del transcriptoma por el reloj, pero el número de genes controlados por los ciclos diurnos sigue siendo sustancialmente numeroso. Éste es el caso de *Chlamydomonas* (Harris, 2001), que puede llevar a cabo algunos procesos fisiológicos complejos con independencia de los ciclos luz/oscuridad, controlados en parte por otros factores como estímulos bióticos externos. Siguiendo esta tendencia, *Arabidopsis* ha reducido la dependencia de los ciclos diurnos sobre su expresión génica a sólo un tercio de su transcriptoma, aun así, sigue siendo significativamente mayor que en mamíferos, en los que sólo un 10-15% de los genes exhiben un control periódico (Lowrey y Takahashi, 2011). No obstante, se ha descrito que cuando se combinan con el fotoperiodo otras señales externas como la temperatura, casi el 90% de la expresión génica de *Arabidopsis* podría seguir un control periódico (Michael et al., 2008). Por tanto, las plantas parecen tener la capacidad de integrar distintas señales externas para regular periódicamente la expresión génica, lo que no ocurre en microalgas. La ausencia en microalgas de los genes del bucle de la tarde GI y ZTL (Figura 4) puede explicar esta diferencia, ya que estas proteínas permiten mantener un periodo constante en un amplio intervalo de temperatura (Edwards et al., 2005; Kim et al., 2020).

Teniendo en cuenta que el curso de la evolución ha hecho más complejo el reloj en *Arabidopsis* (Figura 3, Figura 4), podría intuirse que ejerce un mayor control

periódico del transcriptoma. Sin embargo, ocurre lo contrario, la regulación de la expresión génica ha perdido mucha de la dependencia de procesos diurnos que tenía en algas. Esto podría significar que se adquirió un ajuste más fino y preciso del reloj para que algunos procesos nuevos fueran independientes de las transiciones luz/oscuridad en plantas superiores. Tras la terrestreización, las plantas tuvieron que adaptarse a un nuevo ecosistema, probablemente más cambiante que el ecosistema acuático. Aunque algunos eventos pueden predecirse y anticiparse, como la llegada de patógenos o la cantidad de humedad en el ambiente, otros son inesperados, como inundaciones, sequías, predación por herbívoros o competencia con otras plantas. No obstante, se necesitarían más experimentos circadianos en tejidos y estadios de desarrollo diferentes para profundizar en esta aparente paradoja.

2 Algas y plantas presentan una alta conservación en la expresión génica en las transiciones luz/oscuridad, a pesar de la gran distancia evolutiva que las separa

El análisis de agrupamiento de genes periódicos realizado en este estudio ha mostrado que las redes de coexpresión en los tres organismos presentan la misma distribución de reloj, revelando la naturaleza asociativa y temporal de este sistema para coordinar los diferentes procesos fisiológicos, como el metabolismo del almidón o la ribogénesis (Figura 22A-C). En microalgas, los grupos de genes que mostraron su máxima expresión en las transiciones día/noche fueron los más abundantes, mientras que en *Arabidopsis*, los genes ortólogos se expresaron principalmente durante los periodos de oscuridad o de luz. Esto parece indicar que *Arabidopsis* ha desarrollado un reloj que permite predecir el amanecer y el atardecer de forma más eficiente que las microalgas, lo que mejora la anticipación de estas transiciones. Además, al realizar el análisis de conservación global entre grupos de genes, sólo aquellos con máximo y mínimo en el amanecer o en el atardecer mostraron una alta conservación entre las especies analizadas (Figura 26A). Esto indica una diversificación en el control temporal de la expresión génica durante la evolución, a excepción de determinados momentos del día importantes para la fisiología vegetal.

Una de las propiedades clave del reloj circadiano en todos los organismos es la anticipación de determinados eventos periódicos, como los ciclos de luz y oscuridad. De esta forma cualquier organismo está preparado antes de que ocurra una señal externa predecible. Antes del amanecer suceden cambios en el transcriptoma que permiten la expresión de genes necesarios posteriormente, como los de las enzimas encargadas de la síntesis de pigmentos fotosintéticos o de la maquinaria de los fotosistemas. Lo mismo ocurre al atardecer, como preparación para la noche, inhibiendo prematuramente la expresión de los genes que regulan los procesos fotosintéticos o activando los encargados de la apertura estomática. Por ello, tiene sentido que los genes que se expresan mayoritariamente en las transiciones hayan conservado su perfil de expresión a lo largo de la evolución.

3 Amplificación y diversificación génica de microalgas hacia plantas superiores

Como resultado de la identificación de ortólogos realizada en esta tesis, se ha observado una amplificación a lo largo de la evolución vegetal, especialmente en familias de factores de transcripción, como los CO-likes o los PSEUDO-RESPONSE REGULATORS.

Como se ha mostrado anteriormente (Figura 20D, Figura 21D), los PRRs son una familia de factores de transcripción que aparecieron en microalgas y han evolucionado hasta una familia más numerosa en *Arabidopsis*. Los PRRs de espermatofitas son similares en número a los COLs y de forma similar han sufrido una amplificación a partir de una única copia génica en *Ostreococcus* (*OtPRR*, *ostta13g01820*) y dos en *Chlamydomonas* (*CrPPR1*, *Cre02g.094150*; *CrPPR2*, *Cre16.g676421*) a una familia multigénica en *Arabidopsis*. En plantas estas proteínas contienen un dominio N-terminal PRR y un dominio CCT carboxilo-terminal, sin embargo, las proteínas de *Ostreococcus* y *Chlamydomonas* presentan un potencial motivo fosfoceptor DDK, que no está presente en *Arabidopsis*. Esto sugiere que las proteínas de algas aún retienen parte del ancestral mecanismo señalizador de fosforelay, mientras que en plantas superiores evolucionó a un “Pseudo” *Response Regulator* (Mizuno y Nakamichi, 2005; Satbhai et al., 2011). Se necesitarían análisis

más extensos empleando plantas primitivas como hepáticas, helechos y angiospermas son necesarios para investigar la desaparición del dominio RR. No obstante, nuestro estudio parece indicar que la expansión de esta familia y su diversificación en el patrón de expresión ha permitido una mayor complejidad y plasticidad en el reloj de *Arabidopsis*.

Los COL constituyen otra familia de factores de transcripción que ha sufrido una amplificación y diversificación génica (Figura 20B, Figura 21B). En *Ostreococcus* se encontraron dos potenciales ortólogos (*OtCOL1*, ostta04g03620 y *OtCOL2*, ostta09g01510) que muestran un dominio N-terminal B-box y un CCT C-terminal. Por otro lado, en *Chlamydomonas* sólo se encontró un miembro de la familia CO-like que ha sido ampliamente estudiado (Valverde, 2011). Otros estudios han investigado esta familia de factores de transcripción en el musgo *Physcomitrella patens*, encontrándose cinco miembros (Zobell et al., 2005), pero la rama de CO parece estar ausente. Más tarde en la evolución encontramos 17 miembros en *A. thaliana* donde parece haber ocurrido una amplificación y diversificación en los patrones de expresión (Figura 21B). Como se ha demostrado en diferentes análisis, CrCO está implicado en múltiples procesos como mantenimiento de ritmos biológicos, ciclo celular, contenido lipídico o fotoprotección, sin embargo, la floración fotoperiódica no es uno de ellos. El hecho de que CrCO pueda rescatar el fenotipo de floración tardía en el mutante *co-10* en *Arabidopsis* (Serrano et al., 2009) indica que durante la evolución vegetal estas familias de factores de transcripción crecieron y divergieron reteniendo parte de la función ancestral, pero adquirieron en plantas superiores nuevas funciones. Esta expansión también se ha observado en otras plantas de la familia de las Brassicaceas como *Brassica napus*, donde se ha observado una amplificación y diversificación similar de los genes *COL* (Chen et al., 2021).

4 Características del reloj circadiano en *Arabidopsis*

Tras el análisis topológico de la red transcripcional ATTRACTOR (Figura 32), se determinó que se trataba de una red libre de escala, puesto que la distribución del grado de los nodos seguía una ley potencial negativa. Es decir, la mayoría de genes

Discusión

tienen pocas conexiones mientras que un grupo pequeño presenta un alto grado de conexión, lo que indica la existencia de un grupo de genes clave cuya expresión está estrictamente regulada. Además, la distribución del camino mínimo entre nodos reveló que constituyen una red de mundo pequeño, específicamente de mundo “ultrapequeño”, ya que este camino es menor de lo esperado. Hay una relación evidente entre esta característica y el concepto libre de escala, ya que la existencia de *hubs* o elementos altamente regulados reducen las distancias entre nodos drásticamente.

Estas características confieren unas propiedades concretas a los sistemas biológicos. Así, los programas transcripcionales gobernados por el reloj y la percepción lumínica en *Arabidopsis* son robustos frente a fallos aleatorios y frágiles frente a mutaciones dirigidas a *hubs*, como ya se describió en estudios anteriores (Stelling et al., 2004). El hecho de que el camino entre dos nodos aleatorios sea corto implica rapidez en la propagación de la información dentro de la red y por tanto que las señales lumínicas recibidas por la planta pueden rápidamente regular procesos específicos.

Como se desarrolla en el segundo capítulo de esta tesis, los genes más controlados por el reloj circadiano en *Arabidopsis* se expresan a primera hora de la mañana. Esto sugiere que éste es un momento clave para la fisiología vegetal y que está estrictamente regulado. Durante el amanecer ocurren cambios severos en las condiciones ambientales que afectan a la fisiología y al desarrollo de los organismos vegetales como la llegada de patógenos, cambios en la humedad, luz, etc. Por tanto, tiene sentido que el reloj circadiano de *Arabidopsis* regule finamente estos genes en este lapso de tiempo, permitiendo anticiparse y prepararse para el periodo del día. De nuevo esto se relaciona con la robustez del sistema, ya que si se elimina la función de alguno de estos factores de transcripción, la expresión de los genes matinales estaría asegurada al estar controlados por muchas otras proteínas.

El modelo Barabasi-Albert es un modelo diseñado para capturar los mecanismos responsables de la emergencia de la propiedad libre de escala en redes teóricas, sin embargo, presenta una serie de limitaciones para explicar algunas

características observadas en redes reales (Barabási y Albert, 1999). En dicho modelo, cuando se añade un nodo, se conecta al nodo existente de mayor grado, de esta forma el nodo más antiguo siempre será el más conectado. Si nos fijamos en los factores de transcripción presentes en la red transcripcional, observamos que algunos de los más recientes en la evolución presentan un mayor número de conexiones (genes diana), como los PHYs o los PIFs. Esto no puede explicarse aplicando el modelo Barabási-Albert, por lo que se han desarrollado otros modelos para describir los mecanismos que gobiernan la evolución de las redes biológicas, que incluyen otros parámetros como la aparición de conexiones internas, eliminación de nodos y aristas y el “*fitness*” de un nodo, como el modelo Bianconi-Barabási (Bianconi y Barabási, 2000b, 2000a). La característica “*fitness*” se puede definir como una propiedad intrínseca de cada nodo que puede otorgar una conexión preferencial, no sólo dependiente del momento de su aparición. Este concepto explica cómo estos factores de transcripción han sido beneficiados por la presión selectiva para acumular un mayor número de genes diana a pesar de su reciente aparición, y podría ser incluido en futuros análisis más profundos.

5 Fotoperiodo y reloj circadiano en *Arabidopsis*

5.1 Contribución de CO al reloj circadiano

Una perspectiva basada en ciencia de redes es esencial para comprender sistemas complejos interconectados (Barabási, 2015). Los programas transcripcionales que gobiernan el reloj circadiano y la floración mostraron estar altamente relacionados en nuestra red transcripcional CircadianFloralNet de acuerdo a sus propiedades de libre de escala y mundo pequeño. En el campo de la teoría de grafos, los motivos de red son particularmente útiles ya que pueden explicar la aparición de funciones emergentes y cómo se controlan procesos biológicos. Los motivos de red son subgrafos que aparecen de forma significativa más frecuentemente en una red de interés que en redes aleatorias (Milo et al., 2002; Shen-Orr et al., 2002; Stone et al., 2019). A pesar de la gran complejidad de las redes génicas, existe un cierto nivel de simplicidad en los componentes que las forman, que están presentes en todas las

Discusión

redes y especies (Alon, 2007b). Se ha debatido ampliamente sobre la validez de analizar estos motivos de red puesto que son parte de redes más grandes y más complejas (Mellis y Raj, 2015). De hecho, CircadianFloralNet representa todas las posibles interacciones en multitud de condiciones, tejidos, estadios de crecimiento y puntos temporales, por tanto, también puede resultar útil tener en cuenta redes aisladas más pequeñas. Siguiendo esta línea de razonamiento, el motivo de red formado por CO y PRR5 actuaría en condiciones específicas, durante la tarde en condiciones LD.

En los últimos años, se ha ido abandonando la idea del reloj circadiano como un reloj molecular simple y se ha constituido como un sistema regulatorio de gran complejidad que modula múltiples procesos, como vías hormonales, metabólicas y de respuesta a estrés, ejerciendo algunas de ellas una regulación *feedback* hacia el reloj (Sanchez y Kay, 2016). En efecto, se han identificado varios factores de transcripción que regulan diversos procesos fisiológicos y que contribuyen al funcionamiento del reloj circadiano (McClung, 2014).

La función de CO está altamente regulada en *Arabidopsis*, a nivel transcripcional y postraduccional, lo que indica que la vía fotoperiódica ha sido sometida a una alta presión selectiva (Romero y Valverde, 2009). No se han descrito funciones alternativas a la floración fotoperiódica para CO, aunque estudios recientes han sugerido que podría estar implicado en el control de varios genes circadianos (Gnesutta et al., 2017). En este estudio se muestra que CO altera el reloj circadiano regulando genes clave de este sistema, lo que revela una función no explorada. Nuestra investigación demuestra, por tanto, que existe una regulación más compleja entre el reloj circadiano y la vía fotoperiódica.

Los PRRs han sido caracterizados como factores de transcripción que reprimen la expresión de sus genes diana. De acuerdo a nuestro análisis de redes basado en datos de ChIP-seq, PRR5 se une al promotor de CO (Figura 36B). Sin embargo, los PRRs generan un efecto positivo global sobre la vía fotoperiódica y el módulo regulador CO-FT, ya que estabilizan la proteína CO (Hayama et al., 2017) y reprimen la expresión de *CDF1*, lo que conlleva la activación de CO y FT (Nakamichi

et al., 2007). Por tanto, la activación de la expresión de *PRR5* por CO (Figura 37B, Figura 39A) podría establecer un bucle de retroalimentación positivo entre el reloj circadiano y la vía fotoperiódica. Además, los *PRRs* se expresan secuencialmente durante el día, por lo que CO ejerce un efecto positivo sobre su expresión para alargar su patrón de expresión, adaptando el reloj a condiciones de día largo (Figura 45). De esta manera, se establecería un sistema de retroalimentación vespertino entre el reloj circadiano y el sistema que mide el fotoperiodo para ajustar en cada momento del año el reloj a la longitud del día y las estaciones.

Por otra parte, nuestro análisis revela que la expresión de *GI* también está regulada por CO (Figura 37B, Figura 39A). Ya que *GI* ejerce un papel activador sobre CO (Imaizumi et al., 2005; Sawa et al., 2007; Song et al., 2014), esto podría constituir de nuevo otro bucle de retroalimentación positiva para mejorar y mantener la respuesta fotoperiódica durante la tarde en condiciones LD.

La familia REVEILLE (*RVE*) incluye también a reguladores del reloj circadiano en *Arabidopsis*. Dos de sus miembros, *RVE1* y *RVE8*, también aparecen como genes regulados por CO en nuestro análisis (Figura 37B, Figura 39A). Estos factores de transcripción presentan un dominio MYB al igual que *CCA1/LHY* pero son activadores de la expresión génica (Farinas y Mas, 2011). Específicamente, *RVE8* forma un bucle de retroalimentación negativo con *PRR5*, de forma similar a CO (Rawat et al., 2011). Por su parte, la expresión de *RVE1* también está inhibida por *PRR9*, *PRR7* y *PRR5* a través de la unión a su promotor (Liu et al., 2013, 2016). Así, CO estaría activando estos dos genes del reloj, desplazando su perfil de expresión máxima hacia la tarde, ya que en condiciones de día neutro *RVE1* y *RVE8* presentan un máximo de expresión a primera hora de la mañana.

Otros genes que mostraron una mayor expresión génica en la línea 35S:CO (Figura 37B) y que son dianas de CO en el experimento de ChIP-seq fueron *LNK1* y *LNK2* (Figura 39A). Estas proteínas desempeñan un papel en la regulación de la expresión génica durante las primeras horas del día y en el control de los ritmos diarios y estacionales en *Arabidopsis*. Su perfil de expresión presenta un máximo por la mañana aunque realizan su función por la tarde, modificando la expresión de

Discusión

algunos genes implicados en el reloj y en el tiempo de floración, como el florigeno *FT* (Rugnone et al., 2013). De hecho, el doble mutante *Ink1-ink2*, presenta un periodo más largo en el movimiento de sus hojas, al igual que *co-10*. Por tanto, de nuevo *CO* está activando genes que se expresan por la mañana, aunque en este caso sus proteínas están presentes por la tarde. Esto podría significar un refuerzo en la señal para transmitir información estacional al reloj circadiano (Figura 46).

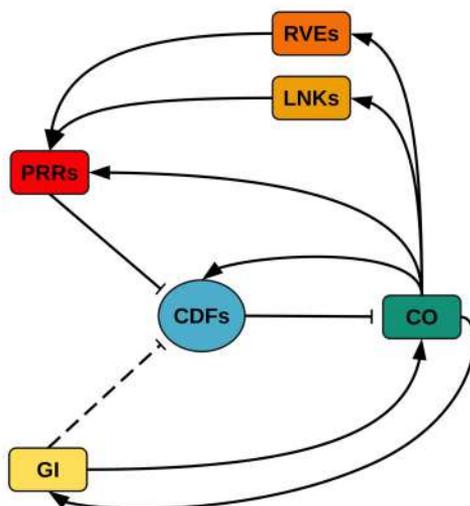


Figura 46. Modelo de la regulación de *CO* sobre el reloj circadiano. *CO* activa la expresión de varios genes del oscilador en *Arabidopsis*, los *PRRs* (principalmente *PRR5*), *RVE1* y *RVE8*, *LNK1* y *LNK2* y *GI*. *PRR5* y *GI* promueven el descenso de los niveles de *CDF1/3/6*, que son represores de la expresión de *CO*. Por tanto, *CO* formaría un bucle de retroalimentación positiva con *PRR5* y *GI* en cuanto a que estas proteínas promueven un efecto positivo global sobre *CO*. Por otra parte, nuestros análisis revelan que *CO* regula la expresión de *RVE1*, *RVE8*, *LNK1* y *LNK2*, activando su expresión a lo largo de la tarde. Esto constituiría un *input* de la vía fotoperiódica al reloj circadiano, a través de *CO*, aportando información estacional. Las líneas continuas indican regulación transcripcional mientras que las discontinuas señalan regulación postraducciona.

Sorprendentemente, según nuestros análisis *CO* también activaría a tres de los miembros de la familia *CDFs* (*CDF1*, *CDF3* y *CDF6*) uniéndose a su promotor (Figura 37B, Figura 39A). Se ha demostrado que estos factores de transcripción reprimen transcripcionalmente a *CO* (Corrales et al., 2017; Krahmer et al., 2019; Imaizumi et al., 2005; Sawa et al., 2007; Nakamichi et al., 2007). De confirmarse, este hecho constituiría un bucle de retroalimentación negativo que habría que explorar con

más profundidad. No obstante, resulta de interés observar en los promotores de estos genes la casi total ausencia de elementos G-box (excepto uno en una región distal del promotor de *CDF6*), lo que sugiere que en este caso, CO estaría regulando su expresión a través del complejo NF-CO y no, como en el caso de los otros genes del reloj regulados por CO, a través de un sitio de unión de factores de transcripción como HY5. Todas estas regulaciones de CO sobre el reloj se ilustran en la Figura 46.

El efecto opuesto sobre la expresión génica de CO y PRR5 viene apoyado por el hecho de que los mutantes *prp5* y *co-10* muestran fenotipos contrarios en el movimiento rítmico de sus hojas. Mientras que en el mutante *prp5* se acorta el ciclo con respecto a la planta silvestre, *co-10* exhibe un periodo más largo (Anexo 10). Los mutantes del reloj *cca1-lhy-20*, *toc1-101*, *prp5-1* (reanalizado en esta tesis), y el triple mutante *prp5-1prp7-3prp9-1* presentan un acortamiento del ciclo circadiano o en el último caso, arritmia (Zhang et al., 2013; Kikis et al., 2005; Michael et al., 2003; Greenham et al., 2015). Por el contrario, el mutante de las proteínas ZTL y FKF1 *ztl-4fkf1-2*, al igual que *co-10*, exhibe un retraso en el ciclo del movimiento de hojas (Baudry et al., 2010), lo que podría indicar que CO podría actuar en el reloj de forma similar a la de estas dos proteínas. Sin embargo, el fenotipo severo del reloj observado en el mutante *ztl-4fkf1-2* se debe a la falta de degradación de TOC1 y PRR5 y no a la regulación transcripcional sobre sus genes diana. En realidad, se necesitaría una caracterización más profunda de todos los elementos del regulador central del reloj circadiano en diferentes fotoperiodos y en el mutante *co-10* para poder delimitar en profundidad el efecto de la participación de CO en el reloj.

De cualquier forma, de nuestros análisis resulta indudable que existe un efecto de CO sobre la integración de señales estacionales en el *core* del reloj circadiano, confirmando la idea ya establecida de un sistema complejo de bucles interconectados, en lugar de una estructura unidireccional compuesto por entrada-oscilador-salida.

5.2 Procesos co-regulados por CO y PRR5

CO, además de regular genes centrales del reloj como *PRR5* o *GI*, también podría controlar la expresión de genes diana del reloj, especialmente un número

Discusión

significativo de dianas de PRR5 que son reguladores transcripcionales. Este grupo de genes estarían presumiblemente reprimidos por PRR5 y activados por CO, por lo que se produciría una modificación drástica de la expresión génica de múltiples genes, propio de un cambio de fase transcripcional (Figura 47). De hecho, el paso de la fase vegetativa a la fase reproductiva requiere de cambios masivos en la expresión génica, por lo que CO podría promover la activación de un grupo de genes que anteriormente estaban reprimidos. Asimismo, esto podría no estar limitado al momento de la transición floral, sino que se trataría de una función de CO en condiciones LD.

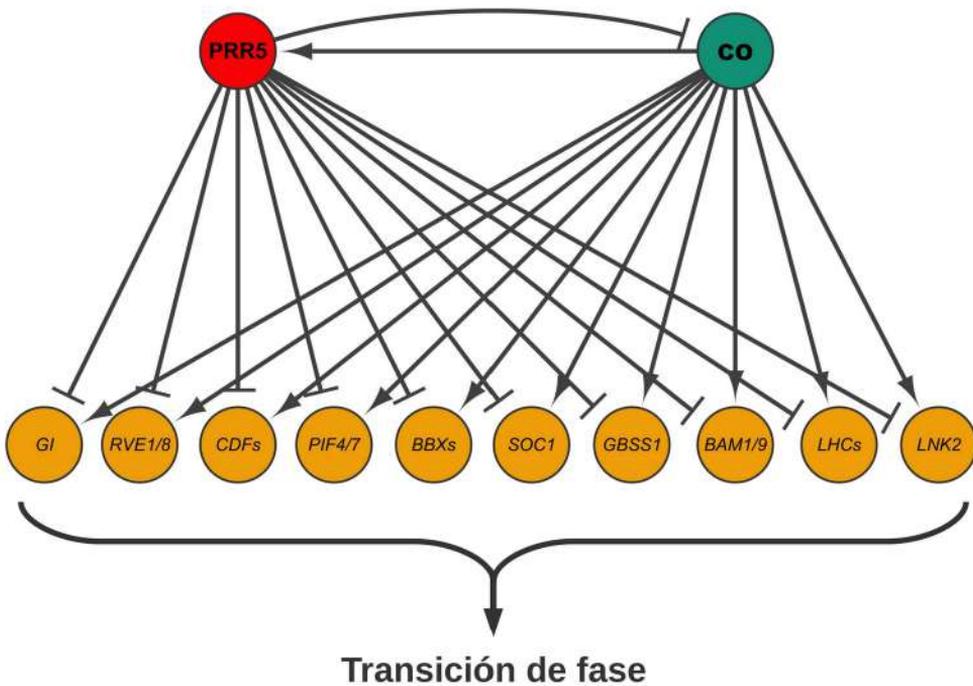


Figura 47. Motivo de red ampliado formado por CO y PRR5. Estos factores de transcripción regulan un número significativo de dianas en común que estarían reprimidas por PRR5 y activadas por CO, produciéndose una transición de fase en su transcripción.

Como se ha señalado, CO comparte un número significativo de dianas con PRR5, no obstante, también lo hace, aunque en menor grado con el resto de la familia de los PRRs (Figura 42A). CONSTANS y PRR5 se coexpresan en LD, acumulándose su proteína por la tarde, por lo que estos dos factores de transcripción coexisten en el

tiempo y el espacio, pudiendo competir por sus potenciales dianas. Sin embargo, la familia de los CO-likes es también numerosa, y presentan funciones solapantes con CO, además de roles alternativos y una diversificación en su patrón de expresión diaria (Figura 21B). Esto sugiere la posibilidad de que las familias de los PRRs y los CO-likes hayan coevolucionado desde microalgas a plantas superiores y que el mismo caso de CO-PRR5 se pueda observar con otros miembros de los CO-likes y los PRRs (Figura 21B, D). Para verificar esta hipótesis son necesarios más experimentos ómicos con el resto de factores de transcripción CO-like.

Trabajos previos han descrito un mecanismo similar por el cual los PRRs se unen al DNA ocupando sitios G-box, y eliminando durante el día secuencialmente a la familia de factores de transcripción PIFs (Martín et al., 2018). No obstante, esto ocurre en condiciones de día corto y permite la represión de genes relacionados con el crecimiento, que son activados por los PIFs durante la noche e inhibidos por los PRRs durante las horas de luz. El mecanismo propuesto en nuestra investigación ocurriría en condiciones LD, momento en que ocurre la acumulación de la proteína, existiendo una competencia entre CO y PRR5 para regular genes relacionados con otros procesos.

Tras el análisis de los genes diana en común entre CO y PRR5 se determinó que estaban involucrados en procesos específicos como ritmo circadiano (discutido anteriormente), respuesta inmune, fotosíntesis u homeostasis del hierro. Pasaremos a discutir alguno de estos aspectos.

El papel del reloj circadiano en la inmunidad innata vegetal ha sido ampliamente estudiado en *Arabidopsis* (Sauerbrunn y Schlaich, 2004; Wang et al., 2011a; Bhardwaj et al., 2011; Zhang et al., 2013; Wang et al., 2011b). Se ha descrito previamente que el gen *PCC1* está implicado en defensa contra patógenos y se regula por el reloj circadiano, concretamente por CCA1 (Sauerbrunn y Schlaich, 2004). Posteriormente, se describió que el transportador de fosfato PHT4 es una pieza clave en la defensa basal y también está regulado por el reloj (Wang et al., 2011a). Otros estudios han identificado otros genes bajo control circadiano, concretamente por CCA1/LHY, que participan en la defensa contra patógenos y permiten a la planta

Discusión

anticiparse al amanecer, cuando normalmente los patógenos dispersan sus esporas (Wang et al., 2011b; Bhardwaj et al., 2011; Zhang et al., 2013). Aunque no se ha identificado previamente un papel directo de los PRRs ni CO en la inmunidad de *Arabidopsis*, sí se ha observado una relación entre el fotoperiodo y la respuesta inmune. En este sentido se ha descrito que plantas crecidas en fotoperiodo de día largo mostraron una mayor resistencia a infecciones que en día corto (Cecchini et al., 1998), sugiriendo que GI jugaba un papel importante en esta regulación (Cecchini et al., 2002), así como una función para el módulo DET1/COP1-PIF4 en el equilibrio defensa/crecimiento en *Arabidopsis* (Gangappa y Kumar, 2018). Entre los genes controlados por CO y PRR5 podemos encontrar proteínas involucradas en la resistencia contra patógenos de plantas como *Pseudomonas syringae*, *Botrytis cinerea*, *Fusarium oxysporum* o *Peronospora parasitica*. A la luz de nuestros resultados y dado que la inmunidad vegetal es un proceso afectado tanto por el reloj circadiano como por el fotoperiodo, es posible que el módulo CO/PRR5 también participe en el control de los recursos que la planta dedica a la defensa contra patógenos.

Otros de los procesos biológicos enriquecidos, tanto en los experimentos de expresión génica masiva como en ChIPseq, entre los genes diana de CO y PRR5 es la fotosíntesis. Observando los datos experimentales en detalle encontramos una parte significativa de genes pertenecientes al complejo “Light Harvesting” (LHC), que reciben la luz gracias a los pigmentos asociados y la transfieren a los fotosistemas I y II. La superfamilia de proteínas que forma el LHC incluye las proteínas de unión a clorofila a/b (CAB), las ELIP (*early light-induced proteins*), las HLIP (*high-light-induced proteins*) y las PSBS. Este tipo de proteínas antena tienen una característica intrínseca de fotoprotección, reduciendo los efectos adversos del estrés por alta intensidad lumínica y protegiendo a los pigmentos fotosintéticos de la fotooxidación (Montané y Kloppstech, 2000; Dittami et al., 2010). Recientemente, se ha descrito un papel de fotoprotección para CrCO en *Chlamydomonas*, regulando genes que codifican proteínas PSB y LIGHT-HARVESTING COMPLEX STRESS-RELATED PROTEINS (LHCSRs) (Gabilly et al., 2019; Tokutsu et al., 2019). Entre los genes regulados por CO y PRR5 encontramos 15 LHCs. Entre ellos se encuentra CAB1, un

gen descrito en multitud de estudios como regulado por el reloj. También aparecen otros genes implicados en fotoprotección como *PGR5* (Munekage et al., 2002). Esto indica que este mecanismo de fotoprotección activado por CO ha podido ser conservado durante la evolución desde microalgas hasta plantas superiores.

Por último, sorprendentemente encontramos que algunos de los genes regulados por CO/PRR5 participan en la homeostasis del hierro. Varios autores han descrito que la captación y la homeostasis de hierro es un proceso regulado por el reloj circadiano, ya que es crítico para la supervivencia y la fisiología de la planta. Además, la disponibilidad de hierro puede afectar al reloj, constituyendo una señal retrógrada (Chen et al., 2013; Hong et al., 2013; Salomé et al., 2013). Uno de los genes dentro de este grupo es *FERRITIN1* (*FER1*), que codifica una proteína representativa de la homeostasis del hierro en *Arabidopsis* cuya expresión aumenta en presencia de este ión, participando en su almacenamiento efectivo (Gaymard et al., 1996; Petit et al., 2001). Además su expresión génica está regulada por el reloj circadiano a través de *CCA1/LHY* (Hong et al., 2013). Por tanto PRR5 y CO también podrían regular este gen conjuntamente y participar en la homeostasis de hierro en *Arabidopsis*, lo que necesitaría de una mayor experimentación para asegurar esta observación.

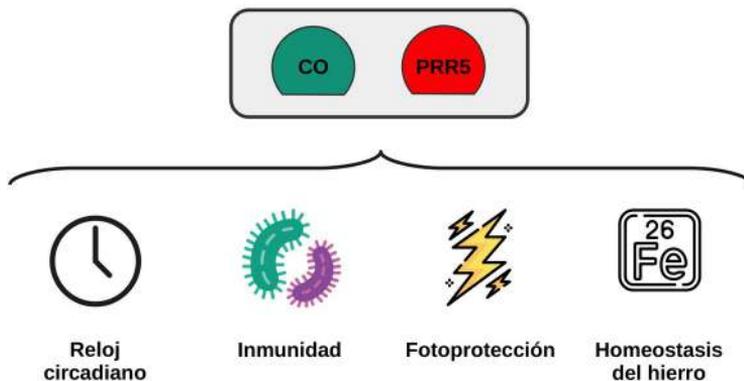


Figura 48. Procesos regulados por CO y PRR5. Los genes diana en común están involucrados en reloj circadiano, inmunidad vegetal, fotoprotección y homeostasis del hierro.

5.3 Interacción de CO con el DNA

Desde el descubrimiento de CO ha existido controversia en cuanto a si actúa como un clásico factor de transcripción, uniéndose directamente al DNA o como un regulador transcripcional o co-activador (Blackman y Michaels, 2010). En principio se postuló el modelo de reemplazo por el cual CO sustituía a la proteína NF-YA en el trímero NF-YA/NF-YB/NF-YC y se unía a secuencias CCAAT (Wenkel et al., 2006; Siefers et al., 2009; Li et al., 2011). Además, otros autores señalaron el motivo CORE (TGTG(N2-3)ATG) como elemento al que se une CO en el DNA y que era suficiente para su efecto sobre la expresión de *FT* (Tiwari et al., 2010). Posteriormente, mediante experimentos de captura de la conformación de los cromosomas (3C) se han descrito dos bucles en el promotor de *FT* que aproximan secuencias CCAAT a la región próxima al TSS donde se encuentran los sitios CORE1 y 2. Estos autores mostraron que la región distal (-5.3 kb) contactaba con el sitio CORE1/2 y esto era dependiente de la interacción de CO con el trímero NF-Ys. Además, se discutía si el complejo CO/NF-YB/NF-YC sería capaz de unirse a las secuencias CCAAT ya que CO carece de los aminoácidos necesarios para su unión a esta secuencia, que sí están presentes en NF-YA. Por ello se postuló el modelo de reclutamiento, en el que el trímero NF-YA/NF-YB/NF-YC recluta y estabiliza a CO en el promotor de *FT* (Cao et al., 2014). Recientemente, un estudio que incluye datos masivos, describe que la sustitución de NF-YA por CO en el trímero otorga especificidad de secuencia diana, reconociendo un nuevo elemento CCACA, en lugar de CCAAT (Gnesutta et al., 2017). Estos modelos se representan en la Figura 49. Incluso otros autores han demostrado que CO puede asociarse en complejos represores junto a TOPLESS (Graeff et al., 2016).

Curiosamente, nuestro análisis de identificación *de novo* de motivos de DNA encontró enriquecimiento en los sitios de unión de CO en una secuencia que contiene el elemento CCACA, pero no al CCAAT (Wenkel et al., 2006) (Figura 39). Esto al mismo tiempo que otorga veracidad a nuestro análisis, confirma la especificidad de diana otorgada por CO al trímero NF-CO (Gnesutta et al., 2017). Por otra parte, también se encontró significativamente una secuencia consenso similar al motivo CORE TGTG(N2-3)ATG encontrado en el promotor de *FT*, y que es reconocido por CO para unirse y activar la expresión de este gen (Tiwari et al., 2010). Sin embargo, el

motivo que hemos detectado en este estudio puede reconstituir esta secuencia consenso excepto en la última posición, en la que la adenina aparece con mayor probabilidad (Figura 39). Es necesario además tener en cuenta que esta secuencia contiene la caja CCACA (TGTGG).

De hecho, en este trabajo demostramos que CO es capaz de unirse a elementos G-box en el DNA, probablemente a través de un factor de transcripción bZIP, como HY5, descrito como una proteína involucrada en multitud de procesos relacionados con la regulación por luz (Gangappa y Botto, 2016). En condiciones LD, CO desplazaría a PRR5 de estos promotores, activando la expresión de los genes diana (Figura 49D). De cualquier forma, la coexistencia de elementos CCACA y G-box en algunos sitios de unión de CO (Figura 40) sugiere que en estos promotores CO podría estar regulando la expresión formando diferentes complejos.

La ausencia de elementos G-box en los sitios de unión del promotor de *FT* indica que CO no se une en estas regiones a través de HY5 (Figura 43), lo que podría explicar en parte el fenotipo de floración temprana del mutante *hy5*. Como se observa en el Anexo 11, *hy5* presenta un tiempo de floración significativamente menor al de la planta silvestre, mientras que *hyh* no exhibe un cambio significativo. De nuevo, el doble mutante *hy5/hyh* muestra un tiempo de floración menor, al igual que el mutante simple *hy5*, por lo que se deduce que la contribución viene de la ausencia de la proteína HY5. A falta del factor HY5, CO no podría unirse de forma tan eficiente al resto de sus sitios de unión que contienen G-boxes, por lo que habría una mayor cantidad de proteína CO disponible para activar la expresión de *FT*.

Discusión

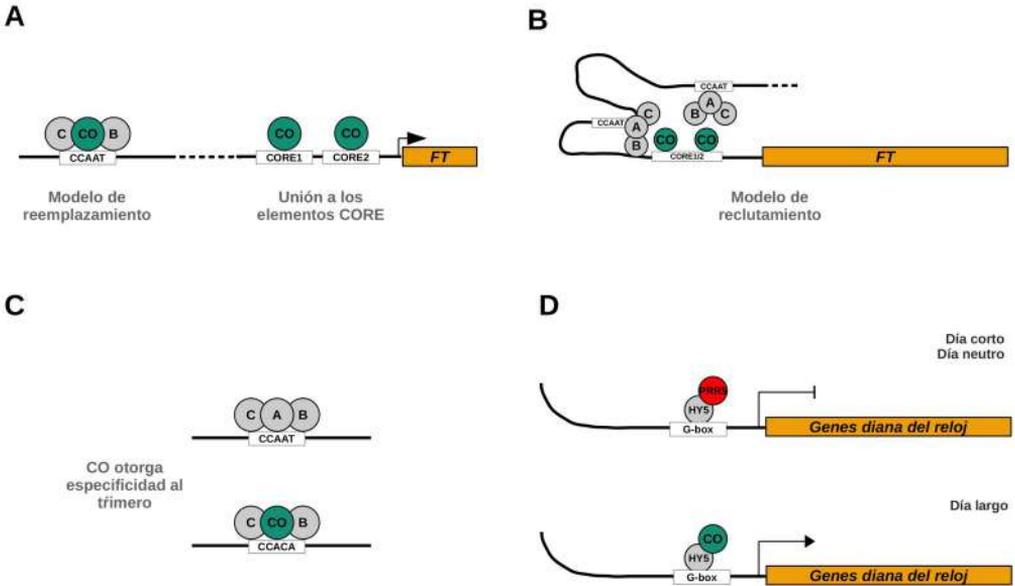


Figura 49. Posibles interacciones de CO con el DNA. (A) Modelo de reemplazamiento por el cual CO sustituye a NF-YA del trímero y se une a regiones CCAAT, además se asocia en elementos CORE en el promotor de *FT* (Wenkel et al., 2006; Tiwari et al., 2010). (B) Modelo de reclutamiento de CO postulado por Cao et al., 2014 en el que bucles en el promotor de *FT* mantienen a los complejos NF-Y próximos a los sitios CORE. La unión de CO al sitio CORE1/2 se estabiliza a través de la interacción con los NF-Y. (C) Modelo descrito por Gnesutta et al., 2017, en el que CO otorga especificidad al sustituir a NF-YA en el trímero CO/NF-YB/NF-YC, pasando a reconocer elementos CCACA. (D) Modelo de la unión de CO al motivo G-box para regular genes del reloj. CO se asociaría a sitios previamente ocupados por PRR5 para activar genes diana en común. El factor de transcripción bZIP HY5 podría participar en la unión al motivo G-box.

En conclusión, nuestras observaciones muestran cierta plasticidad en la función de CO como regulador transcripcional. Así, dependiendo de las proteínas con las que se asocie, CO exhibirá distintos efectos sobre la transcripción y una afinidad por diferentes sitios de unión al DNA. Además, se demuestra su habilidad para regular el núcleo del reloj circadiano durante la tarde en condiciones LD, proporcionando información estacional necesaria para el desarrollo vegetal.





Conclusiones

1. Los eventos de amplificación y diversificación génica desde microalgas hasta plantas superiores han permitido una mayor plasticidad y una menor dependencia de los ciclos diurnos en la fisiología de las plantas.
2. A pesar de la gran distancia evolutiva entre las microalgas y las plantas superiores, se observa una conservación en los patrones de expresión génica que controlan procesos clave como la respuesta fotoperiódica, el reloj circadiano o el metabolismo del azúcar y el almidón, especialmente en las transiciones luz/oscuridad.
3. Los sistemas transcripcionales del reloj circadiano y la percepción luminosa han evolucionado hacia una mayor complejidad en *Arabidopsis*. Este programa transcripcional es robusto a mutaciones que afecten a genes poco conectados y frágil frente a la eliminación de *hubs*.
4. Los genes con una mayor regulación por parte del reloj circadiano se expresan durante las primeras horas de la mañana y controlan procesos clave como la respuesta inmune, respuesta a hormonas o crecimiento.
5. Existe una señalización retrógrada desde la vía fotoperiódica hacia el oscilador central en *Arabidopsis*, ejercida por CO, que regula genes centrales del reloj circadiano activando su expresión en condiciones de día largo.
6. Una de las conexiones entre el reloj circadiano y la vía fotoperiódica estaría formada por CO y PRR5, que constituyen un motivo de doble retroalimentación con salida múltiple. CO y PRR5 regulan la expresión de una serie de genes diana en común a través de la unión a elementos G-box en sus promotores, ejerciendo un papel antagonista. La unión de estos factores de transcripción a estas regiones podría estar mediada por HY5.

Conclusiones

1. Gene amplification and diversification throughout evolution have promoted a lesser dependency of diurnal cycles and a greater plasticity on higher plants physiology.
2. Despite the great evolutionary distance between microalgae and higher plants, a certain level of conservation is observed in the gene expression patterns that control key processes such as the photoperiodic response, the circadian clock or sugar and starch metabolism, particularly in the day/night transitions.
3. The transcriptional systems that govern the circadian clock and light perception have evolved to reach a greater complexity in *Arabidopsis*. This transcriptional program is robust to mutations that affect poorly connected genes and fragile against the elimination of hubs.
4. Genes with a higher dependence by the circadian clock are expressed during the early morning hours and control essential processes such as the immune response, hormone signalling or growth.
5. A retrograde signal from the photoperiodic pathway to the central oscillator in *Arabidopsis* is exerted by CO, which regulates circadian clock central genes, activating their expression under long-day conditions.
6. One of the connections between the circadian clock and the photoperiodic pathway would be established by CO and PRR5, which constitute a double feedback motif with multiple outputs. CO and PRR5 regulate the expression of a set of common target genes through binding to G-box elements in their promoters, exerting an antagonistic role. The binding of these transcription factors to these regions could be mediated by HY5.



The background of the page is a watercolor-style illustration. It features a white background with scattered splatters and blotches of light blue, teal, and yellow. The splatters are of various sizes and shapes, some resembling ink blots or watercolor washes. The overall effect is artistic and textured.

Bibliografía

- Abbas, N., Maurya, J.P., Senapati, D., Gangappa, S.N., y Chattopadhyaya, S.** (2014). Arabidopsis CAM7 and HY5 physically interact and directly bind to the HY5 promoter to regulate its expression and thereby promote photomorphogenesis. *Plant Cell* **26**: 1036-1052.
- Albert, R., Jeong, H., y Barabási, A.L.** (2000). Error and attack tolerance of complex networks. *Nature* **406**: 378-382.
- Alexa, A. y Rahnenfuhrer, J.** (2021). topGO: Enrichment Analysis for Gene Ontology. R package version 2.44.0.
- Alon, U.** (2003). Biological networks: The tinkerer as an engineer. *Science* (80-.). **301**: 1866-1867.
- Alon, U.** (2007a). Network motifs: theory and experimental approaches. *Nat. Rev. Genet.* **8**: 450-61.
- Alon, U.** (2007b). Simplicity in biology. *Nature* **446**: 497.
- Ando, E., Ohnishi, M., Wang, Y., Matsushita, T., Watanabe, A., Hayashi, Y., Fujii, M., Ma, J.F., Inoue, S.I., y Kinoshita, T.** (2013). TWIN SISTER OF FT, GIGANTEA, and CONSTANS have a positive but indirect effect on blue light-induced Stomatal opening in Arabidopsis. *Plant Physiol.* **162**: 1529-1538.
- Aoki, K., Ogata, Y., y Shibata, D.** (2007). Approaches for extracting practical information from gene co-expression networks in plant biology. *Plant Cell Physiol.* **48**: 381-90.
- Atamian, H.S. y Harmer, S.L.** (2016). Circadian regulation of hormone signaling and plant physiology. *Plant Mol. Biol.* 2016 916 **91**: 691-702.
- Barabási, A.-L.** (2015). *Network Science*.
- Barabási, A.-L. y Oltvai, Z.N.** (2004). Network biology: understanding the cell's functional organization. *Nat. Rev. Genet.* 2004 52 **5**: 101-113.
- Barabási, A.L. y Albert, R.** (1999). Emergence of Scaling in Random Networks. *Science* (80-.). **286**: 509-512.
- Barak, S., Tobin, E.M., Green, R.M., Andronis, C., y Sugano, S.** (2000). All in good time: the Arabidopsis circadian clock. *Trends Plant Sci.* **5**: 517-522.
- Barbazuk, W.B., Emrich, S.J., Chen, H.D., Li, L., y Schnable, P.S.** (2007). SNP discovery via 454 transcriptome sequencing. *Plant J.* **51**: 910-918.
- Baudry, A., Ito, S., Song, Y.H., Strait, A.A., Kiba, T., Lu, S., Henriques, R., Prunedo-Paz, J.L., Chua, N.H., Tobin, E.M., Kay, S.A., y Imaizumi, T.** (2010). F-box proteins FKF1 and LKP2 act in concert with ZEITLUPE to control Arabidopsis clock progression. *Plant Cell* **22**: 606-622.
- Beel, B., Müller, N., Kottke, T., y Mittag, M.** (2012a). News about cryptochrome photoreceptors in algae. <https://doi.org/10.4161/psb.22870> **8**.
- Beel, B., Prager, K., Spexard, M., Sasso, S., Weiss, D., Müller, N., Heinicke, M., Dewez, D., Ikoma, D., Grossman, A.R., Kottke, T., y Mittag, M.** (2012b). A Flavin Binding Cryptochrome Photoreceptor Responds to Both Blue and Red Light in *Chlamydomonas reinhardtii*. *Plant Cell* **24**: 2992-3008.
- Belda-Palazón, B., Ruiz, L., Martí, E., Tárraga, S., Tiburcio, A.F., Culiáñez, F., Farràs, R., Carrasco, P., y Ferrando, A.** (2012). Aminopropyltransferases Involved in Polyamine Biosynthesis Localize Preferentially in the Nucleus of Plant Cells. *PLoS One* **7**: e46907.

Bibliografía

- Bernard, P. y Couturier, M.** (1992). Cell killing by the F plasmid CcdB protein involves poisoning of DNA-topoisomerase II complexes. *J. Mol. Biol.* **226**: 735-745.
- Bhardwaj, V., Meier, S., Petersen, L.N., Ingle, R.A., y Roden, L.C.** (2011). Defence Responses of *Arabidopsis thaliana* to Infection by *Pseudomonas syringae* Are Regulated by the Circadian Clock. *PLoS One* **6**: e26968.
- Bianconi, G. y Barabási, A.-L.** (2000a). Bose-Einstein condensation in complex networks. *Phys. Rev. Lett.* **86**: 5632-5635.
- Bianconi, G. y Barabási, A.L.** (2000b). Competition and multiscaling in evolving networks. *Struct. Dyn. Networks* **9781400841356**: 361-367.
- Binkert, M., Kozma-Bognár, L., Terecskei, K., De Veylder, L., Nagy, F., y Ulm, R.** (2014). UV-B-Responsive association of the *Arabidopsis* bZIP transcription factor ELONGATED HYPOCOTYL5 with target genes, including its own promoter. *Plant Cell* **26**: 4200-4213.
- Blackman, B.K. y Michaels, S.D.** (2010). Does CONSTANS act as a transcription factor or as a co-activator? The answer may be - yes. *New Phytol.* **187**: 1-3.
- Blanc-Mathieu, R. et al.** (2014). An improved genome of the model marine alga *Ostreococcus tauri* unfolds by assessing Illumina de novo assemblies. *BMC Genomics* **15**: 1103.
- Bläsing, O.E., Gibon, Y., Günther, M., Höhne, M., Morcuende, R., Osuna, D., Thimm, O., Usadel, B., Scheible, W.R., y Stitt, M.** (2005). Sugars and circadian regulation make major contributions to the global regulation of diurnal gene expression in *Arabidopsis*. *Plant Cell* **17**: 3257-3281.
- Blazquez, M.A., Green, R., Nilsson, O., Sussman, M.R., y Weigel, D.** (1998). Gibberellins Promote Flowering of *Arabidopsis* by Activating the LEAFY Promoter. *Plant Cell* **10**: 791.
- Böhlenius, H., Huang, T., Charbonnel-Campaa, L., Brunner, A.M., Jansson, S., Strauss, S.H., y Nilsson, O.** (2006). CO/FT Regulatory Module Controls Timing of Flowering and Seasonal Growth Cessation in Trees. *Science* (80-.). **312**: 1040-1043.
- Bowler, C., Benvenuto, G., Laflamme, P., Molino, D., Probst, A. V., Tariq, M., y Paszkowski, J.** (2004). Chromatin techniques for plant cells. *Plant J.* **39**: 776-789.
- BRUCE, V.G.** (1970). The Biological Clock in *Chlamydomonas reinhardi*. *J. Protozool.* **17**: 328-334.
- Byrne, T.E., Wells, M.R., y Johnson, C.H.** (1992). Circadian rhythms of chemotaxis to ammonium and of methylammonium uptake in *chlamydomonas*. *Plant Physiol.* **98**: 879-886.
- Cao, S., Kumimoto, R.W., Gnesutta, N., Calogero, A.M., Mantovani, R., y Holt, B.F.** (2014). A distal CCAAT/NUCLEAR FACTOR Y complex promotes chromatin looping at the FLOWERING LOCUS T promoter and regulates the timing of flowering in *Arabidopsis*. *Plant Cell* **26**: 1009-1017.
- Cecchini, E., Al-Kaff, N.S., Bannister, A., Giannakou, M.E., McCallum, D.G., Maule, A.J., Milner, J.J., y Covey, S.N.** (1998). Pathogenic interactions between variants of cauliflower mosaic virus and *Arabidopsis thaliana*. *J. Exp. Bot.* **49**: 731-737.
- Cecchini, E., Geri, C., Love, A.J., Coupland, G., Covey, S.N., y Milner, J.J.** (2002). Mutations that delay flowering in *Arabidopsis* de-couple symptom response from cauliflower mosaic virus accumulation during infection. *Mol. Plant Pathol.* **3**: 81-90.

- Cerutti, H., Johnson, A.M., Gillham, N.W., y Boynton, J.E.** (1997). Epigenetic silencing of a foreign gene in nuclear transformants of *Chlamydomonas*. *Plant Cell* **9**: 925-945.
- Chattopadhyay, S., Ang, L.H., Puente, P., Deng, X.W., y Wei, N.** (1998). Arabidopsis bZIP protein HY5 directly interacts with light-responsive promoters in mediating light control of gene expression. *Plant Cell* **10**: 673-683.
- Chen, Y., Zhou, R., Hu, Q., Wei, W., y Liu, J.** (2021). Conservation and Divergence of the CONSTANS-Like (COL) Genes Related to Flowering and Circadian Rhythm in *Brassica napus*. *Front. Plant Sci.* **0**: 2634.
- Chen, Y.Y., Wang, Y., Shin, L.J., Wu, J.F., Shanmugam, V., Tsednee, M., Lo, J.C., Chen, C.C., Wu, S.H., y Yeh, K.C.** (2013). Iron Is Involved in the Maintenance of Circadian Period Length in Arabidopsis. *Plant Physiol.* **161**: 1409-1420.
- Chen, Z.J.** (2013). Genomic and epigenetic insights into the molecular bases of heterosis. *Nat. Rev. Genet.* 2013 147 **14**: 471-482.
- Chomczynski, P. y Sacchi, N.** (1987). Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Anal. Biochem.* **162**: 156-159.
- Cohen, R. y Havlin, S.** (2003). Scale-Free Networks Are Ultrasmall. *Phys. Rev. Lett.* **90**: 4.
- Conti, L.** (2017). Hormonal control of the floral transition: Can one catch them all? *Dev. Biol.* **430**: 288-301.
- Corellou, F., Schwartz, C., Motta, J.-P., Djouani-Tahri, E.B., Sanchez, F., y Bouget, F.-Y.** (2009). Clocks in the green lineage: comparative functional analysis of the circadian architecture of the picoeukaryote *ostreococcus*. *Plant Cell* **21**: 3436-49.
- Corrales, A.R., Carrillo, L., Lasierra, P., Nebauer, S.G., Dominguez-Figueroa, J., Renau-Morata, B., Pollmann, S., Granell, A., Molina, R.V., Vicente-Carbajosa, J., y Medina, J.** (2017). Multifaceted role of cycling DOF factor 3 (CDF3) in the regulation of flowering time and abiotic stress responses in Arabidopsis. *Plant. Cell Environ.* **40**: 748-764.
- Courties, C., Vaquer, A., Troussellier, M., Lautier, J., Chrétiennot-Dinet, M.J., Neveux, J., Machado, C., y Claustre, H.** (1994). Smallest eukaryotic organism [6]. *Nature* **370**: 255.
- Csardi, G.N.T.** (2006). The igraph software package for complex network research. *Inter J Complex Syst* **1695**.
- Dalquen, D.A. y Dessimoz, C.** (2013). Bidirectional best hits miss many orthologs in duplication-rich clades such as plants and animals. *Genome Biol. Evol.* **5**: 1800-1806.
- Das, M., Haberer, G., Panda, A., Das Laha, S., Ghosh, T.C., y Schäffner, A.R.** (2016). Expression Pattern Similarities Support the Prediction of Orthologs Retaining Common Functions after Gene Duplication Events. *Plant Physiol.* **171**: 2343-2357.
- Das, P.M., Ramachandran, K., vanWert, J., y Singal, R.** (2018). Chromatin immunoprecipitation assay. <https://doi.org/10.2144/04376RV01> **37**: 961-969.
- Deng, X., Fan, X., Li, P., y Fei, X.** (2015). A photoperiod-regulating gene CONSTANS is correlated to lipid biosynthesis in *chlamydomonas reinhardtii*. *Biomed Res. Int.* **2015**.
- Derelle, E. et al.** (2006). Genome analysis of the smallest free-living eukaryote *Ostreococcus tauri* unveils many unique features. *Proc. Natl. Acad. Sci. U. S. A.* **103**: 11647-11652.

Bibliografía

- Dessimoz, C. et al.** (2012). Toward community standards in the quest for orthologs. *Bioinformatics* **28**: 900-904.
- Dittami, S.M., Michel, G., Collén, J., Boyen, C., y Tonon, T.** (2010). Chlorophyll-binding proteins revisited - a multigenic family of light-harvesting and stress proteins from a brown algal perspective. *BMC Evol. Biol.* **10**: 365.
- Dixon, L.E., Knox, K., Kozma-Bognar, L., Southern, M.M., Pokhilko, A., y Millar, A.J.** (2011). Temporal repression of core circadian genes is mediated through EARLY FLOWERING 3 in Arabidopsis. *Curr. Biol.* **21**: 120-125.
- Djouani-Tahri, E.-B.B., Christie, J.M., Sanchez-Ferandin, S., Sanchez, F., Bouget, F.Y.F.-Y., y Corellou, F.** (2011). A eukaryotic LOV-histidine kinase with circadian clock function in the picoalga *Ostreococcus*. **65**: 578-588.
- Dunlap, J.C.** (1999). Molecular bases for circadian clocks. *Cell* **96**: 271-290.
- Edwards, K.D., Lynn, J.R., Gyula, P., Nagy, F., y Millar, A.J.** (2005). Natural Allelic Variation in the Temperature-Compensation Mechanisms of the Arabidopsis thaliana Circadian Clock. *Genetics* **170**: 387.
- Endo, M., Tanigawa, Y., Murakami, T., Araki, T., y Nagatani, A.** (2013). Phytochrome-dependent late-flowering accelerates flowering through physical interactions with phytochrome B and CONSTANS. *Proc. Natl. Acad. Sci. U. S. A.* **110**: 18017-18022.
- Euler, L.** (1741). Solutio problematis ad geometriam situs pertinentis. *Comment. Acad. Sci. Petropolitanae*.
- Fankhauser, C. y Staiger, D.** (2002). Photoreceptors in Arabidopsis thaliana: light perception, signal transduction and entrainment of the endogenous clock. *Planta* **216**: 1-16.
- Farinas, B. y Mas, P.** (2011). Functional implication of the MYB transcription factor RVE8/LCL5 in the circadian control of histone acetylation. *Plant J.* **66**: 318-329.
- Feige, J.N., Sage, D., Wahli, W., Desvergne, B., y Gelman, L.** (2005). PixFRET, an ImageJ plug-in for FRET calculation that can accommodate variations in spectral bleed-throughs. *Microsc. Res. Tech.* **68**: 51-58.
- Fernández, M.B., Tossi, V., Lamattina, L., y Cassia, R.** (2016). A Comprehensive Phylogeny Reveals Functional Conservation of the UV-B Photoreceptor UVR8 from Green Algae to Higher Plants. *Front. Plant Sci.* **7**: 1698.
- Fornara, F., Panigrahi, K.C.S., Gissot, L., Sauerbrunn, N., Rühl, M., Jarillo, J.A., y Coupland, G.** (2009). Arabidopsis DOF Transcription Factors Act Redundantly to Reduce CONSTANS Expression and Are Essential for a Photoperiodic Flowering Response. *Dev. Cell* **17**: 75-86.
- Frazer, A.C., Perteua, G., Jaffe, A.E., Langmead, B., Salzberg, S.L., y Leek, J.T.** (2015). Ballgown bridges the gap between transcriptome assembly and expression analysis. *Nat. Biotechnol.* **33**: 243-246.
- Fried, M. y Crothers, D.M.** (1981). Equilibria and kinetics of lac repressor-operator interactions by polyacrylamide gel electrophoresis. *Nucleic Acids Res.* **9**: 6505-6525.
- Fujiwara, S., Wang, L., Han, L., Suh, S.S., Salomé, P.A., McClung, C.R., y Somers, D.E.** (2008). Post-translational Regulation of the Arabidopsis Circadian Clock through Selective Proteolysis and Phosphorylation of Pseudo-response Regulator Proteins. *J. Biol. Chem.* **283**: 23073-23083.

- Gabilly, S.T., Baker, C.R., Wakao, S., Crisanto, T., Guan, K., Bi, K., Guiet, E., Guadagno, C.R., y Niyogi, K.K.** (2019). Regulation of photoprotection gene expression in *Chlamydomonas* by a putative E3 ubiquitin ligase complex and a homolog of CONSTANS. *Proc. Natl. Acad. Sci. U. S. A.* **116**: 17556-17562.
- Gangappa, S.N. y Botto, J.F.** (2016). The Multifaceted Roles of HY5 in Plant Growth and Development. *Mol. Plant* **9**: 1353-1365.
- Gangappa, S.N. y Kumar, S.V.** (2018). DET1 and COP1 Modulate the Coordination of Growth and Immunity in Response to Key Seasonal Signals in Arabidopsis. *Cell Rep.* **25**: 29-37.e3.
- Garner, M.M. y Revzin, A.** (1981). A gel electrophoresis method for quantifying the binding of proteins to specific DNA regions: application to components of the *Escherichia coli* lactose operon regulatory system. *Nucleic Acids Res.* **9**: 3047-3060.
- Gates, A.J., Ke, Q., Varol, O., y Barabási, A.-L.** (2019). Nature's reach: narrow work has broad impact. *Nat.* 2021 5757781 **575**: 32-34.
- Gautier, L., Cope, L., Bolstad, B.M., y Irizarry, R.A.** (2004). affy—analysis of Affymetrix GeneChip data at the probe level. *Bioinformatics* **20**: 307-315.
- Gaymard, F., Boucherez, J., y Briat, J.F.** (1996). Characterization of a ferritin mRNA from *Arabidopsis thaliana* accumulated in response to iron through an oxidative pathway independent of abscisic acid. *Biochem. J.* **318 (Pt 1)**: 67-73.
- Gendron, J.M., Pruneda-Paz, J.L., Doherty, C.J., Gross, A.M., Kang, S.E., y Kay, S.A.** (2012). Arabidopsis circadian clock protein, TOC1, is a DNA-binding transcription factor. *Proc. Natl. Acad. Sci. U. S. A.* **109**: 3167-3172.
- Ginzinger, D.G.** (2002). Gene quantification using real-time quantitative PCR: An emerging technology hits the mainstream. *Exp. Hematol.* **30**: 503-512.
- Gnesutta, N., Kumimoto, R.W., Swain, S., Chiara, M., Siriwardana, C., Horner, D.S., Holt, B.F., y Mantovani, R.** (2017). CONSTANS imparts DNA sequence specificity to the histone fold NF-YB/NF-YC Dimer. *Plant Cell* **29**: 1516-1532.
- Goff, L.A., Trapnell, C., y Kelley, D.** (2014). CummeRbund: Visualization and Exploration of Cufflinks High-throughput Sequencing Data.
- Goto, K. y Johnson, C.H.** (1995). Is the cell division cycle gated by a circadian clock? The case of *Chlamydomonas reinhardtii*. *J. Cell Biol.* **129**: 1061-1069.
- Graeff, M., Straub, D., Eguen, T., Dolde, U., Rodrigues, V., Brandt, R., y Wenkel, S.** (2016). MicroProtein-Mediated Recruitment of CONSTANS into a TOPLESS Trimeric Complex Represses Flowering in Arabidopsis. *PLoS Genet.* **12**: e1005959.
- Greenham, K., Lou, P., Remsen, S.E., Farid, H., y McClung, C.R.** (2015). TRiP: Tracking Rhythms in Plants, an automated leaf movement analysis program for circadian period estimation. *Plant Methods* **11**.
- Hajdu, A., Ádám, É., Sheerin, D.J., Dobos, O., Bernula, P., Hiltbrunner, A., Kozma-Bognár, L., y Nagy, F.** (2015). High-level expression and phosphorylation of phytochrome B modulates flowering time in Arabidopsis. *Plant J.* **83**: 794-805.
- Hanahan, D.** (1983). Studies on transformation of *Escherichia coli* with plasmids. *J. Mol. Biol.* **166**: 557-580.

Bibliografía

- Harris, E.H.** (2001). CHLAMYDOMONAS AS A MODEL ORGANISM. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **52**: 363-406.
- Hartwell, L.H., Hopfield, J.J., Leibler, S., y Murray, A.W.** (1999). From molecular to modular cell biology. *Nat.* 1999 4026761 **402**: C47-C52.
- Hayama, R., Sarid-Krebs, L., Richter, R., Fernández, V., Jang, S., y Coupland, G.** (2017). PSEUDO RESPONSE REGULATORS stabilize CONSTANS protein to promote flowering in response to day length. *EMBO J.* **36**: 904-918.
- Hazen, S.P., Schultz, T.F., Pruneda-Paz, J.L., Borevitz, J.O., Ecker, J.R., y Kay, S.A.** (2005). LUX ARRHYTHMO encodes a Myb domain protein essential for circadian rhythms. *Proc. Natl. Acad. Sci.* **102**: 10387-10392.
- HEIJDE, M. et al.** (2010). Characterization of two members of the cryptochrome/photolyase family from *Ostreococcus tauri* provides insights into the origin and evolution of cryptochromes. *Plant. Cell Environ.* **33**: 1614-1626.
- Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., y Glass, C.K.** (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell* **38**: 576-89.
- Helfer, A., Nusinow, D.A., Chow, B.Y., Gehrke, A.R., Bulyk, M.L., y Kay, S.A.** (2011). LUX ARRHYTHMO encodes a nighttime repressor of circadian gene expression in the Arabidopsis core clock. *Curr. Biol.* **21**: 126-133.
- Herrero, E. et al.** (2012). EARLY FLOWERING4 Recruitment of EARLY FLOWERING3 in the Nucleus Sustains the Arabidopsis Circadian Clock. *Plant Cell* **24**: 428-443.
- Hoheisel, J.D.** (2006). Microarray technology: beyond transcript profiling and genotype analysis. *Nat. Rev. Genet.* 2006 73 **7**: 200-210.
- Hong, S., Kim, S.A., Guerinot, M. Lou, y Robertson McClung, C.** (2013). Reciprocal Interaction of the Circadian Clock with the Iron Homeostasis Network in Arabidopsis. *Plant Physiol.* **161**: 893.
- Huang, K., Merkle, T., y Beck, C.F.** (2002). Isolation and characterization of a Chlamydomonas gene that encodes a putative blue-light photoreceptor of the phototropin family. *Physiol. Plant.* **115**: 613-622.
- Huang, W., Pérez-García, P., Pokhilko, A., Millar, A.J., Antoshechkin, I., Riechmann, J.L., y Mas, P.** (2012). Mapping the core of the Arabidopsis circadian clock defines the network structure of the oscillator. *Science* (80-.). **335**: 75-79.
- Iliev, D., Voytsekh, O., Schmidt, E.M., Fiedler, M., Nykytenko, A., y Mittag, M.** (2006). A heteromeric RNA-binding protein is involved in maintaining acrophase and period of the circadian clock. *Plant Physiol.* **142**: 797-806.
- Imaizumi, T., Schultz, T.F., Harmon, F.G., Ho, L.A., y Kay, S.A.** (2005). Plant science: FKF1 F-box protein mediates cyclic degradation of a repressor of CONSTANS in Arabidopsis. *Science* (80-.). **309**: 293-297.
- Isabel Ortiz-Marchena, M., Albi, T., Lucas-Reina, E., Said, F.E., Romero-Campero, F.J., Cano, B., Teresa Ruiz, M., Romero, J.M., y Valverde, F.** (2014). Photoperiodic control of carbon distribution during the floral transition in Arabidopsis. *Plant Cell* **26**: 565-584.

- Jang, S., Marchal, V., Panigrahi, K.C.S., Wenkel, S., Soppe, W., Deng, X.W., Valverde, F., y Coupland, G.** (2008). Arabidopsis COP1 shapes the temporal pattern of CO accumulation conferring a photoperiodic flowering response. *EMBO J.* **27**: 1277-1288.
- Jin, J., Tian, F., Yang, D.C., Meng, Y.Q., Kong, L., Luo, J., y Gao, G.** (2017). PlantTFDB 4.0: Toward a central hub for transcription factors and regulatory interactions in plants. *Nucleic Acids Res.* **45**: D1040-D1045.
- Jing, Y., Guo, Q., Zha, P., y Lin, R.** (2019). The chromatin-remodelling factor PICKLE interacts with CONSTANS to promote flowering in Arabidopsis. *Plant. Cell Environ.* **42**: 2495-2507.
- Jing, Y., Zhang, D., Wang, X., Tang, W., Wang, W., Huai, J., Xu, G., Chen, D., Li, Y., y Lin, R.** (2013). Arabidopsis chromatin remodeling factor PICKLE interacts with transcription factor HY5 to regulate hypocotyl cell elongation. *Plant Cell* **25**: 242-256.
- Kamioka, M., Takao, S., Suzuki, T., Taki, K., Higashiyam, T., Kinoshita, T., y Nakamichi, N.** (2016). Direct repression of evening genes by CIRCADIAN CLOCK-ASSOCIATED1 in the Arabidopsis circadian clock. *Plant Cell* **28**: 696-711.
- Kaufmann, K., Muiño, J.M., Østerås, M., Farinelli, L., Krajewski, P., y Angenent, G.C.** (2010). Chromatin immunoprecipitation (ChIP) of plant transcription factors followed by sequencing (ChIP-SEQ) or hybridization to whole genome arrays (ChIP-CHIP). *Nat. Protoc.* 2010 53 **5**: 457-472.
- Kiba, T., Henriques, R., Sakakibara, H., y Chua, N.-H.** (2007). Targeted Degradation of PSEUDO-RESPONSE REGULATOR5 by an SCFZTL Complex Regulates Clock Function and Photomorphogenesis in Arabidopsis thaliana. *Plant Cell* **19**: 2516-2530.
- Kikis, E.A., Khanna, R., y Quail, P.H.** (2005). ELF4 is a phytochrome-regulated component of a negative-feedback loop involving the central oscillator components CCA1 and LHY. *Plant J.* **44**: 300-313.
- Kim, D., Paggi, J.M., Park, C., Bennett, C., y Salzberg, S.L.** (2019). Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat. Biotechnol.* 2019 378 **37**: 907-915.
- Kim, J.S., Jung, H.J., Lee, H.J., Kim, K.A., Goh, C.H., Woo, Y., Oh, S.H., Han, Y.S., y Kang, H.** (2008). Glycine-rich RNA-binding protein 7 affects abiotic stress responses by regulating stomata opening and closing in Arabidopsis thaliana. *Plant J.* **55**: 455-466.
- Kim, T.S., Wang, L., Kim, Y.J., y Somers, D.E.** (2020). Compensatory Mutations in GI and ZTL May Modulate Temperature Compensation in the Circadian Clock. *Plant Physiol.* **182**: 1130-1141.
- Kim, W.-Y., Fujiwara, S., Suh, S.-S., Kim, J., Kim, Y., Han, L., David, K., Putterill, J., Nam, H.G., y Somers, D.E.** (2007). ZEITLUPE is a circadian photoreceptor stabilized by GIGANTEA in blue light. *Nat.* 2007 4497160 **449**: 356-360.
- Kiyosue, T., Yamaguchi-Shinozaki, K., y Shinozaki, K.** (1994). Cloning of cDNAs for genes that are early-responsive to dehydration stress (ERDs) in Arabidopsis thaliana L.: identification of three ERDs as HSP cognate genes. *Plant Mol. Biol.* **25**: 791-798.
- Kleinberg, J.M.** (1997). Authoritative Sources in a Hyperlinked Environment *.
- Kloosterman, B., Abelenda, J.A., Gomez, M.D.M.C., Oortwijn, M., De Boer, J.M., Kowitzanich, K., Horvath, B.M., Van Eck, H.J., Smaczniak, C., Prat, S., Visser, R.G.F., y Bachem, C.W.B.** (2013). Naturally occurring allele diversity allows potato cultivation in northern latitudes. *Nature* **495**: 246-250.
- Kobourov, S.G.** (2012). Spring Embedders and Force Directed Graph Drawing Algorithms.

Bibliografía

- Kovaka, S., Zimin, A. V., Pertea, G.M., Razaghi, R., Salzberg, S.L., y Pertea, M.** (2019). Transcriptome assembly from long-read RNA-seq alignments with StringTie2. *Genome Biol.* 2019 201 **20**: 1-13.
- Krahmer, J., Goralogia, G.S., Kubota, A., Zardilis, A., Johnson, R.S., Song, Y.H., MacCoss, M.J., Le Bihan, T., Halliday, K.J., Imaizumi, T., y Millar, A.J.** (2019). Time-resolved interaction proteomics of the GIGANTEA protein under diurnal cycles in Arabidopsis. *FEBS Lett.* **593**: 319-338.
- Kunihiro, A., Yamashino, T., Nakamichi, N., Niwa, Y., Nakanishi, H., y Mizuno, T.** (2011). Phytochrome-interacting factor 4 and 5 (PIF4 and PIF5) activate the homeobox ATHB2 and auxin-inducible IAA29 genes in the coincidence mechanism underlying photoperiodic control of plant growth of Arabidopsis thaliana. *Plant Cell Physiol.* **52**: 1315-1329.
- Lagercrantz, U. y Axelsson, T.** (2000). Rapid Evolution of the Family of CONSTANS LIKE Genes in Plants. *Mol. Biol. Evol.* **17**: 1499-1507.
- Lamesch, P. et al.** (2012). The Arabidopsis Information Resource (TAIR): improved gene annotation and new tools. *Nucleic Acids Res.* **40**: D1202-10.
- Langfelder, P. y Horvath, S.** (2012). Fast R functions for robust correlations and hierarchical clustering. *J. Stat. Softw.* **46**: 1-17.
- Langfelder, P., Luo, R., Oldham, M.C., y Horvath, S.** (2011). Is my network module preserved and reproducible? *PLoS Comput. Biol.* **7**: 1001057.
- Langmead, B., Trapnell, C., Pop, M., y Salzberg, S.L.** (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**: R25.
- Lau, O.S. y Bergmann, D.C.** (2015). MOBE-CHIP: a large-scale chromatin immunoprecipitation assay for cell type-specific studies. *Plant J.* **84**: 443-50.
- Laubinger, S., Marchal, V., Gentilhomme, J., Wenkel, S., Adrian, J., Jang, S., Kulajta, C., Braun, H., Coupland, G., y Hoecker, U.** (2006). Arabidopsis SPA proteins regulate photoperiodic flowering and interact with the floral inducer CONSTANS to regulate its stability. *Development* **133**: 3213-3222.
- Lazaro, A., Valverde, F., Piñeiro, M., y Jarillo, J.A.** (2012). The Arabidopsis E3 ubiquitin ligase HOS1 negatively regulates CONSTANS abundance in the photoperiodic control of flowering. *Plant Cell* **24**: 982-999.
- Ledger, S., Strayer, C., Ashton, F., Kay, S.A., y Putterill, J.** (2001). Analysis of the function of two circadian-regulated CONSTANS-LIKE genes. *Plant J.* **26**: 15-22.
- Lee, J., He, K., Stolc, V., Lee, H., Figueroa, P., Gao, Y., Tongprasit, W., Zhao, H., Lee, I., y Deng, X.W.** (2007). Analysis of transcription factor HY5 genomic binding sites revealed its hierarchical role in light regulation of development. *Plant Cell* **19**: 731-49.
- León, R. y Fernández, E.** (2007). Nuclear transformation of eukaryotic microalgae: Historical overview, achievements and problems. *Adv. Exp. Med. Biol.* **616**: 1-11.
- Li, C., Distelfeld, A., Comis, A., y Dubcovsky, J.** (2011). Wheat flowering repressor VRN2 and promoter CO2 compete for interactions with NUCLEAR FACTOR-Y complexes. *Plant J.* **67**: 763-773.
- Li, F.W. y Mathews, S.** (2016). Evolutionary aspects of plant photoreceptors. *J. Plant Res.* **129**: 115-122.
- Li, Y., Pearl, S.A., y Jackson, S.A.** (2015). Gene Networks in Plant Biology: Approaches in Reconstruction and Analysis. *Trends Plant Sci.* **20**: 664-675.

- Lister, R., Gregory, B.D., y Ecker, J.R.** (2009). Next is now: new technologies for sequencing of genomes, transcriptomes, and beyond. *Curr. Opin. Plant Biol.* **12**: 107-118.
- Liu, J., Cheng, X., Liu, P., Li, D., Chen, T., Gu, X., y Sun, J.** (2017). MicroRNA319-regulated TCPs interact with FBHs and PFT1 to activate CO transcription and control flowering time in Arabidopsis. *PLOS Genet.* **13**: e1006833.
- Liu, L., White, M.J., y MacRae, T.H.** (1999). Transcription factors and their genes in higher plants. *Eur. J. Biochem.* **262**: 247-257.
- Liu, L.J., Zhang, Y.C., Li, Q.H., Sang, Y., Mao, J., Lian, H.L., Wang, L., y Yang, H.Q.** (2008). COP1-mediated ubiquitination of CONSTANS is implicated in cryptochrome regulation of flowering in Arabidopsis. *Plant Cell* **20**: 292-306.
- Liu, T., Carlsson, J., Takeuchi, T., Newton, L., y Farré, E.M.** (2013). Direct regulation of abiotic responses by the Arabidopsis circadian clock component PRR7. *Plant J.* **76**: 101-14.
- Liu, T.L., Newton, L., Liu, M.-J., Shiu, S.-H., y Farré, E.M.** (2016). A G-Box-Like Motif Is Necessary for Transcriptional Regulation by Circadian Pseudo-Response Regulators in Arabidopsis. *Plant Physiol.* **170**: 528-539.
- Locke, J.C.W., Kozma-Bognár, L., Gould, P.D., Fehér, B., Kevei, É., Nagy, F., Turner, M.S., Hall, A., y Millar, A.J.** (2006). Experimental validation of a predicted feedback loop in the multi-oscillator clock of Arabidopsis thaliana. *Mol. Syst. Biol.* **2**: 59.
- Locke, J.C.W., Southern, M.M., Kozma-Bognár, L., Hibberd, V., Brown, P.E., Turner, M.S., y Millar, A.J.** (2005). Extension of a genetic network model by iterative experimentation and mathematical analysis. *Mol. Syst. Biol.* **1**: 2005.0013.
- de Los Reyes, P., Romero-Campero, F.J., Teresa Ruiz, M., Romero, J.M., y Valverde, F.** (2017). Evolution of daily gene co-expression patterns from algae to plants. *Front. Plant Sci.* **8**: 1217.
- Lowrey, P.L. y Takahashi, J.S.** (2011). Genetics of Circadian Rhythms in Mammalian Model Organisms. *Adv. Genet.* **74**: 175-230.
- Lucas-Reina, E., Romero-Campero, F.J., Romero, J.M., y Valverde, F.** (2015). An evolutionarily conserved DOF-CONSTANS module controls plant photoperiodic signaling. *Plant Physiol.* **168**: 561-574.
- Martín, G. et al.** (2018). Circadian Waves of Transcriptional Repression Shape PIF-Regulated Photoperiod-Responsive Growth in Arabidopsis. *Curr. Biol.* **28**: 311-318.e5.
- Matsuo, T., Okamoto, K., Onai, K., Niwa, Y., Shimogawara, K., y Ishiura, M.** (2008). A systematic forward genetic analysis identified components of the Chlamydomonas circadian system. *Genes Dev.* **22**: 918-930.
- Matus, J.T., Ruggieri, V., Romero, F.J., Moretto, M., y Wong, D.C.J.** (2019). Status and Prospects of Systems Biology in Grapevine Research. *En (Springer, Cham)*, pp. 137-166.
- McClung, C.R.** (2014). Wheels within wheels: New transcriptional feedback loops in the Arabidopsis circadian clock. *F1000Prime Rep.* **6**.
- McClung, C.R., Salomé, P.A., y Michael, T.P.** (2002). The Arabidopsis Circadian System. *Arab. B.* **1**: e0044.

Bibliografía

- Mellis, I.A. y Raj, A.** (2015). Half dozen of one, six billion of the other: What can small- and large-scale molecular systems biology learn from one another? *Genome Res.* **25**: 1466-1472.
- Merchant, S.S. et al.** (2007). The *Chlamydomonas* genome reveals the evolution of key animal and plant functions. *Science* **318**: 245-50.
- Mesarović, M.D.** (1968). Systems Theory and Biology—View of a Theoretician. En *Systems Theory and Biology* (Springer Berlin Heidelberg), pp. 59-87.
- Michael, T.P. et al.** (2008). Network Discovery Pipeline Elucidates Conserved Time-of-Day–Specific cis-Regulatory Modules. *PLOS Genet.* **4**: e14.
- Michael, T.P. y McClung, C.R.** (2003). Enhancer trapping reveals widespread circadian clock transcriptional control in *Arabidopsis*. *Plant Physiol.* **132**: 629-639.
- Michael, T.P., Salomé, P.A., Yu, H.J., Spencer, T.R., Sharp, E.L., McPeck, M.A., Alonso, J.M., Ecker, J.R., y McClung, C.R.** (2003). Enhanced Fitness Conferred by Naturally Occurring Variation in the Circadian Clock. *Science* (80-.). **302**: 1049-1053.
- Michaels, S.D. y Amasino, R.M.** (1999). FLOWERING LOCUS C encodes a novel MADS domain protein that acts as a repressor of flowering. *Plant Cell* **11**: 949-956.
- Millar, A.J.** (2016). The Intracellular Dynamics of Circadian Clocks Reach for the Light of Ecology and Evolution.
- Milo, R., Shen-Orr, S., Itzkovitz, S., Kashtan, N., Chklovskii, D., y Alon, U.** (2002). Network motifs: Simple building blocks of complex networks. *Science* (80-.). **298**: 824-827.
- Mittag, M., Kiaulehn, S., y Johnson, C.H.** (2005). The Circadian Clock in *Chlamydomonas reinhardtii*. What Is It For? What Is It Similar To? *Plant Physiol.* **137**: 399.
- Mizoguchi, T., Wright, L., Fujiwara, S., Cremer, F., Lee, K., Onouchi, H., Mouradov, A., Fowler, S., Kamada, H., Putterill, J., y Coupland, G.** (2005). Distinct roles of GIGANTEA in promoting flowering and regulating circadian rhythms in *Arabidopsis*. *Plant Cell* **17**: 2255-2270.
- Mizuno, T. y Nakamichi, N.** (2005). Pseudo-response regulators (PRRs) or true oscillator components (TOCs). *Plant Cell Physiol.* **46**: 677-685.
- Mockler, T.C., Michael, T.P., Priest, H.D., Shen, R., Sullivan, C.M., Givan, S.A., Mcentee, C., Kay, S.A., y Chory, J.** (2007). The DIURNAL project: DIURNAL and circadian expression profiling, model-based pattern matching, and promoter analysis. *Cold Spring Harb. Symp. Quant. Biol.* **72**: 353-363.
- Monnier, A., Liverani, S., Bouvet, R., Jesson, B., Smith, J.Q., Mosser, J., Corellou, F., y Bouget, F.Y.** (2010). Orchestrated transcription of biological processes in the marine picoeukaryote *Ostreococcus* exposed to light/dark cycles. *BMC Genomics* **11**.
- Montané, M.H. y Kloppstech, K.** (2000). The family of light-harvesting-related proteins (LHCs, ELIPs, HLIPs): was the harvesting of light their primary function? *Gene* **258**: 1-8.
- Mortazavi, A., Williams, B.A., McCue, K., Schaeffer, L., y Wold, B.** (2008). Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat. Methods* **5**: 621-628.
- Mullineux, G.** (1991). A blackboard structure for handling engineering design data. *Eng. Comput.* **7**: 185-195.

- Munekage, Y., Hojo, M., Meurer, J., Endo, T., Tasaka, M., y Shikanai, T.** (2002). PGR5 is involved in cyclic electron flow around photosystem I and is essential for photoprotection in Arabidopsis. *Cell* **110**: 361-371.
- Murashige, T. y Skoog, F.** (1962). A Revised Medium for Rapid Growth and Bio Assays with Tobacco Tissue Cultures. *Physiol. Plant.* **15**: 473-497.
- Nakagawa, T. et al.** (2007). Improved Gateway binary vectors: high-performance vectors for creation of fusion constructs in transgenic analysis of plants. *Biosci. Biotechnol. Biochem.* **71**: 2095-2100.
- Nakamichi, N., Kiba, T., Henriques, R., Mizuno, T., Chua, N.-H., y Sakakibara, H.** (2010). PSEUDO-RESPONSE REGULATORS 9, 7, and 5 are transcriptional repressors in the Arabidopsis circadian clock. *Plant Cell* **22**: 594-605.
- Nakamichi, N., Kiba, T., Kamioka, M., Suzuki, T., Yamashino, T., Higashiyama, T., Sakakibara, H., y Mizuno, T.** (2012). Transcriptional repressor PRR5 directly regulates clock-output pathways. *Proc. Natl. Acad. Sci. U. S. A.* **109**: 17123-8.
- Nakamichi, N., Kita, M., Niinuma, K., Ito, S., Yamashino, T., Mizoguchi, T., y Mizuno, T.** (2007). Arabidopsis Clock-Associated Pseudo-Response Regulators PRR9, PRR7 and PRR5 Coordinately and Positively Regulate Flowering Time Through the Canonical CONSTANS-Dependent Photoperiodic Pathway. *Plant Cell Physiol.* **48**: 822-832.
- Nakamura, S., Mano, S., Tanaka, Y., Ohnishi, M., Nakamori, C., Araki, M., Niwa, T., Nishimura, M., Kaminaka, H., Nakagawa, T., Sato, Y., y Ishiguro, S.** (2010). Gateway binary vectors with the bialaphos resistance gene, bar, as a selection marker for plant transformation. *Biosci. Biotechnol. Biochem.* **74**: 1315-1319.
- Nguyen, K.T., Park, J., Park, E., Lee, I., y Choi, G.** (2015). The Arabidopsis RING Domain Protein BOI Inhibits Flowering via CO-dependent and CO-independent Mechanisms. *Mol. Plant* **8**: 1725-1736.
- Nikaido, S.S. y Johnson, C.H.** (2007). Daily and Circadian Variation in Survival From Ultraviolet Radiation in *Chlamydomonas reinhardtii*. *Photochem. Photobiol.* **71**: 758-765.
- Niwa, Y., Yamashino, T., y Mizuno, T.** (2009). The circadian clock regulates the photoperiodic response of hypocotyl elongation through a coincidence mechanism in Arabidopsis thaliana. *Plant Cell Physiol.* **50**: 838-854.
- Nozue, K., Covington, M.F., Duek, P.D., Lorrain, S., Fankhauser, C., Harmer, S.L., y Maloof, J.N.** (2007). Rhythmic growth explained by coincidence between internal and external cues. *Nature* **448**: 358-361.
- Nusinow, D.A., Helfer, A., Hamilton, E.E., King, J.J., Imaizumi, T., Schultz, T.F., Farré, E.M., y Kay, S.A.** (2011). The ELF4-ELF3-LUX complex links the circadian clock to diurnal control of hypocotyl growth. *Nature* **475**: 398-404.
- Onai, K. y Ishiura, M.** (2005). PHYTOCLOCK 1 encoding a novel GARP protein essential for the Arabidopsis circadian clock. *Genes to Cells* **10**: 963-972.
- Osterlund, M.T., Hardtke, C.S., Ning, W., y Deng, X.W.** (2000). Targeted destabilization of HY5 during light-regulated development of Arabidopsis. *Nature* **405**: 462-466.
- Ovens, K., Eames, B.F., y McQuillan, I.** (2021). Comparative Analyses of Gene Co-expression Networks: Implementations and Applications in the Study of Evolution. *Front. Genet.* **0**: 1464.

Bibliografía

- Palenik, B. et al.** (2007). The tiny eukaryote *Ostreococcus* provides genomic insights into the paradox of plankton speciation. *Proc. Natl. Acad. Sci. U. S. A.* **104**: 7705-7710.
- Para, A., Farré, E.M., Imaizumi, T., Pruneda-Paz, J.L., Harmon, F.G., y Kay, S.A.** (2007). PRR3 Is a Vascular Regulator of TOC1 Stability in the *Arabidopsis* Circadian Clock. *Plant Cell* **19**: 3462.
- Parsons, R., Parsons, R., Garner, N., Oster, H., y Rawashdeh, O.** (2020). CircaCompare: A method to estimate and statistically support differences in mesor, amplitude and phase, between circadian rhythms. *Bioinformatics* **36**: 1208-1212.
- Pérez-García, P., Ma, Y., Yanovsky, M.J., y Mas, P.** (2015). Time-dependent sequestration of RVE8 by LNK proteins shapes the diurnal oscillation of anthocyanin biosynthesis. *Proc. Natl. Acad. Sci. U. S. A.* **112**: 5249-5253.
- Pertea, M., Kim, D., Pertea, G.M., Leek, J.T., y Salzberg, S.L.** (2016). Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. *Nat. Protoc.* **11**: 1650-1667.
- Petit, J.M., Van Wuytswinkel, O., Briat, J.F., y Lobréaux, S.** (2001). Characterization of an iron-dependent regulatory sequence involved in the transcriptional control of *AtFer1* and *ZmFer1* plant ferritin genes by iron. *J. Biol. Chem.* **276**: 5584-5590.
- Piñeiro, M. y Jarillo, J.A.** (2013). Ubiquitination in the control of photoperiodic flowering. *Plant Sci.* **198**: 98-109.
- Pokhilko, A., Fernández, A.P., Edwards, K.D., Southern, M.M., Halliday, K.J., y Millar, A.J.** (2012). The clock gene circuit in *Arabidopsis* includes a repressilator with additional feedback loops. *Mol. Syst. Biol.* **8**: 574.
- Pokhilko, A., Hodge, S.K., Stratford, K., Knox, K., Edwards, K.D., Thomson, A.W., Mizuno, T., y Millar, A.J.** (2010). Data assimilation constrains new connections and components in a complex, eukaryotic circadian clock model. *Mol. Syst. Biol.* **6**: 416.
- Portolés, S. y Más, P.** (2010). The Functional Interplay between Protein Kinase CK2 and CCA1 Transcriptional Activity Is Essential for Clock Temperature Compensation in *Arabidopsis*. *PLOS Genet.* **6**: e1001201.
- Pruneda-Paz, J.L., Breton, G., Para, A., y Kay, S.A.** (2009). A Functional Genomics Approach Reveals CHE as a Component of the *Arabidopsis* Circadian Clock. *Science* (80-.). **323**: 1481-1485.
- Putterill, J., Robson, F., Lee, K., Simon, R., y Coupland, G.** (1995). The *CONSTANS* gene of *Arabidopsis* promotes flowering and encodes a protein showing similarities to zinc finger transcription factors. *Cell* **80**: 847-857.
- Ral, J.P., Colleoni, C., Wattedled, F., Dauvillée, D., Nempont, C., Deschamps, P., Li, Z., Morell, M.K., Chibbar, R., Purton, S., D'Hulst, C., y Ball, S.G.** (2006). Circadian clock regulation of starch metabolism establishes GBSSI as a major contributor to amylopectin synthesis in *Chlamydomonas reinhardtii*. *Plant Physiol.* **142**: 305-317.
- Ravasz, E. y Barabási, A.-L.** (2003). Hierarchical organization in complex networks. *Phys. Rev. E* **67**: 026112.
- Raven, J.A., Beardall, J., Larkum, A.W.D., y Sánchez-Baracaldo, P.** (2013). Interactions of photosynthesis with genome size and function. *Philos. Trans. R. Soc. B Biol. Sci.* **368**.

- Rawat, R., Takahashi, N., Hsu, P.Y., Jones, M.A., Schwartz, J., Salemi, M.R., Phinney, B.S., y Harmer, S.L.** (2011). REVELLE8 and PSEUDO-REPONSE REGULATOR5 form a negative feedback loop within the arabidopsis circadian clock. *PLoS Genet.* **7**.
- Reisdorph, N.A. y Small, G.D.** (2004). The CPH1 Gene of *Chlamydomonas reinhardtii* Encodes Two Forms of Cryptochrome Whose Levels Are Controlled by Light-Induced Proteolysis. *Plant Physiol.* **134**: 1546-1554.
- Rensing, S.A.** (2014). Gene duplication as a driver of plant morphogenetic evolution. *Curr. Opin. Plant Biol.* **17**: 43-48.
- Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., y Smyth, G.K.** (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**: e47-e47.
- Rizzini, L., Favory, J.-J., Cloix, C., Faggionato, D., O'Hara, A., Kaiserli, E., Baumeister, R., Schäfer, E., Nagy, F., Jenkins, G.I., y Ulm, R.** (2011). Perception of UV-B by the Arabidopsis UVR8 Protein. *Science* (80-.). **332**: 103-106.
- Roberts, A., Trapnell, C., Donaghey, J., Rinn, J.L., y Pachter, L.** (2011). Improving RNA-Seq expression estimates by correcting for fragment bias. *Genome Biol.* **12**: R22.
- Rockwell, N.C., Duanmu, D., Martin, S.S., Bachy, C., Price, D.C., Bhattacharya, D., Worden, A.Z., y Lagarias, J.C.** (2014). Eukaryotic algal phytochromes span the visible spectrum. *Proc. Natl. Acad. Sci.* **111**: 3871-3876.
- Roden, L.C. y Ingle, R.A.** (2009). Lights, Rhythms, Infection: The Role of Light and the Circadian Clock in Determining the Outcome of Plant–Pathogen Interactions. *Plant Cell* **21**: 2546-2552.
- Romero-Campero, F.J., Lucas-Reina, E., Said, F.E., Romero, J.M., y Valverde, F.** (2013). A contribution to the study of plant development evolution based on gene co-expression networks. *Front. Plant Sci.* **4**: 291.
- Romero-Campero, F.J., Perez-Hurtado, I., Lucas-Reina, E., Romero, J.M., y Valverde, F.** (2016). ChlamyNET: A *Chlamydomonas* gene co-expression network reveals global properties of the transcriptome and the early setup of key co-expression patterns in the green lineage. *BMC Genomics* **17**: 227.
- Romero-Losada, A.B., Arvanitidou, C., De Los Reyes, P., García-González, M., Romero-Campero, F.J., y Romero, F.J.** (2021). ALGAEFUN with MARACAS, microALGAE FUNctional enrichment tool for MicroAlgae RnA-seq and Chip-seq Analysis. *bioRxiv*: 2021.08.14.456338.
- Romero, J.M. y Valverde, F.** (2009). Evolutionarily conserved photoperiod mechanisms in plants: When did plant photoperiodic signaling appear? *Plant Signal. Behav.* **4**: 642-644.
- Rosenfeld, N., Elowitz, M.B., y Alon, U.** (2002). Negative Autoregulation Speeds the Response Times of Transcription Networks. *J. Mol. Biol.* **323**: 785-793.
- Rugnone, M.L., Faigón Soverna, A., Sanchez, S.E., Schlaen, R.G., Hernando, C.E., Seymour, D.K., Mancini, E., Chernomoretz, A., Weigel, D., Más, P., y Yanovsky, M.J.** (2013). LNK genes integrate light and clock signaling networks at the core of the Arabidopsis oscillator. *Proc. Natl. Acad. Sci. U. S. A.* **110**: 12120-5.

Bibliografía

- Sakuma, Y., Maruyama, K., Osakabe, Y., Qin, F., Seki, M., Shinozaki, K., y Yamaguchi-Shinozaki, K.** (2006). Functional analysis of an Arabidopsis transcription factor, DREB2A, involved in drought-responsive gene expression. *Plant Cell* **18**: 1292-1309.
- Salmon-Divon, M., Dvinge, H., Tammoja, K., y Bertone, P.** (2010). PeakAnalyzer: Genome-wide annotation of chromatin binding and modification loci. *BMC Bioinforma.* 2010 111 **11**: 1-12.
- Salomé, P.A., Oliva, M., Weigel, D., y Krämer, U.** (2013). Circadian clock adjustment to plant iron status depends on chloroplast and phytochrome function. *EMBO J.* **32**: 511.
- Samach, A., Onouchi, H., Gold, S.E., Ditta, G.S., Schwarz-Sommer, Z., Yanofsky, M.F., y Coupland, G.** (2000). Distinct roles of constans target genes in reproductive development of Arabidopsis. *Science* (80-.). **288**: 1613-1616.
- Sambrook, J., Fritsch, E.F., y Maniatis, T.** (1989). *Molecular cloning: a laboratory manual*. Mol. cloning a Lab. manual.
- Sanchez, S.E. y Kay, S.A.** (2016). The Plant Circadian Clock: From a Simple Timekeeper to a Complex Developmental Manager. *Cold Spring Harb. Perspect. Biol.* **8**: a027748.
- Sanchez, S.E., Rugnone, M.L., y Kay, S.A.** (2020). Light Perception: A Matter of Time. *Mol. Plant* **13**: 363-385.
- Satbhai, S.B., Yamashino, T., Okada, R., Nomoto, Y., Mizuno, T., Tezuka, Y., Itoh, T., Tomita, M., Otsuki, S., y Aoki, S.** (2011). Pseudo-response regulator (PRR) homologues of the moss physcomitrella patens: Insights into the evolution of the prr family in land plants. *DNA Res.* **18**: 39-52.
- Sauerbrunn, N. y Schlaich, N.L.** (2004). PCC1: A merging point for pathogen defence and circadian signalling in Arabidopsis. *Planta* **218**: 552-561.
- Sawa, M., Nusinow, D.A., Kay, S.A., y Imaizumi, T.** (2007). FKF1 and GIGANTEA complex formation is required for day-length measurement in Arabidopsis. *Science* (80-.). **318**: 261-265.
- Schindelin, J. et al.** (2012). Fiji: An open-source platform for biological-image analysis. *Nat. Methods* **9**: 676-682.
- Schreiber, K.J., Bentham, A., Williams, S.J., Kobe, B., y Staskawicz, B.J.** (2016). Multiple Domain Associations within the Arabidopsis Immune Receptor RPP1 Regulate the Activation of Programmed Cell Death. *PLOS Pathog.* **12**: e1005769.
- Schulze, A. y Downward, J.** (2001). Navigating gene expression using microarrays - A technology review. *Nat. Cell Biol.* **3**.
- Serrano-Bueno, G., Romero-Campero, F.J., Lucas-Reina, E., Romero, J.M., y Valverde, F.** (2017). Evolution of photoperiod sensing in plants and algae. *Curr. Opin. Plant Biol.* **37**: 10-17.
- Serrano-Bueno, G., Said, F.E., los Reyes, P., Lucas-Reina, E.I., Ortiz-Marchena, M.I., Romero, J.M., y Valverde, F.** (2019). CONSTANS–FKBP12 interaction contributes to modulation of photoperiodic flowering in Arabidopsis. *Plant J.: tpj*.14590.
- Serrano, G., Herrera-Palau, R., Romero, J.M., Serrano, A., Coupland, G., y Valverde, F.** (2009). Chlamydomonas CONSTANS and the Evolution of Plant Photoperiodic Signaling. *Curr. Biol.* **19**: 359-368.

- Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., y Ideker, T.** (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**: 2498-504.
- Sharrock, R.A. y Quail, P.H.** (1989). Novel phytochrome sequences in *Arabidopsis thaliana*: structure, evolution, and differential expression of a plant regulatory photoreceptor family.
- Sheerin, D.J., Menon, C., Oven-Krockhaus, S. Zur, Enderle, B., Zhu, L., Johnen, P., Schleifenbaum, F., Stierhof, Y.D., Huq, E., y Hiltbrunner, A.** (2015). Light-activated phytochrome A and B interact with members of the SPA family to promote photomorphogenesis in *Arabidopsis* by reorganizing the COP1/SPA complex. *Plant Cell* **27**: 189-201.
- Shen-Orr, S.S., Milo, R., Mangan, S., y Alon, U.** (2002). Network motifs in the transcriptional regulation network of *Escherichia coli*. *Nat. Genet.* **31**: 64-68.
- Shen, C., Liu, H., Guan, Z., Yan, J., Zheng, T., Yan, W., Wu, C., Zhang, Q., Yin, P., y Xing, Y.** (2020). Structural Insight into DNA Recognition by CCT/NF-YB/YC Complexes in Plant Photoperiodic Flowering. *Plant Cell* **32**: 3469-3484.
- Sheth, B.P. y Thaker, V.S.** (2014). Plant systems biology: insights, advances and challenges. *Planta* 2014 2401 **240**: 33-54.
- Shim, J.S. y Jang, G.** (2020). Environmental Signal-Dependent Regulation of Flowering Time in Rice. *Int. J. Mol. Sci.* 2020, Vol. 21, Page 6155 **21**: 6155.
- Shyu, Y.J., Suarez, C.D., y Hu, C.D.** (2008). Visualization of ternary complexes in living cells by using a BiFC-based FRET assay. *Nat. Protoc.* **3**: 1693-1702.
- Siefers, N., Dang, K.K., Kumimoto, R.W., Bynum IV, W.E., Tayrose, G., y Holt, B.F.** (2009). Tissue-Specific Expression Patterns of *Arabidopsis* NF-Y Transcription Factors Suggest Potential for Extensive Combinatorial Complexity. *Plant Physiol.* **149**: 625-641.
- Simpson, G.G.** (2004). The autonomous pathway: epigenetic and post-transcriptional gene regulation in the control of *Arabidopsis* flowering time. *Curr. Opin. Plant Biol.* **7**: 570-574.
- Singh, M. y Mas, P.** (2018). A Functional Connection between the Circadian Clock and Hormonal Timing in *Arabidopsis*. *Genes (Basel)*. **9**: 567.
- Smoot, M.E., Ono, K., Ruscheinski, J., Wang, P.-L., y Ideker, T.** (2011). Cytoscape 2.8: new features for data integration and network visualization. *Bioinformatics* **27**: 431-2.
- Song, Y.H., Estrada, D.A., Johnson, R.S., Kim, S.K., Lee, S.Y., MacCoss, M.J., y Imaizumi, T.** (2014). Distinct roles of FKF1, GIGANTEA, and ZEITLUPE proteins in the regulation of CONSTANS stability in *Arabidopsis* photoperiodic flowering. *Proc. Natl. Acad. Sci. U. S. A.* **111**: 17672-17677.
- Song, Y.H., Smith, R.W., To, B.J., Millar, A.J., y Imaizumi, T.** (2012). FKF1 conveys timing information for CONSTANS stabilization in photoperiodic flowering. *Science (80-.)*. **336**: 1045-1049.
- Soy, J., Leivar, P., González-Schain, N., Martín, G., Diaz, C., Sentandreu, M., Al-Sady, B., Quail, P.H., y Monte, E.** (2016). Molecular convergence of clock and photosensory pathways through PIF3-TOC1 interaction and co-occupancy of target promoters. *Proc. Natl. Acad. Sci. U. S. A.* **113**: 4870-4875.
- Stelling, J., Gilles, E.D., y Doyle, F.J.** (2004). Robustness properties of circadian clock architectures. *Proc. Natl. Acad. Sci.* **101**: 13210-13215.

Bibliografía

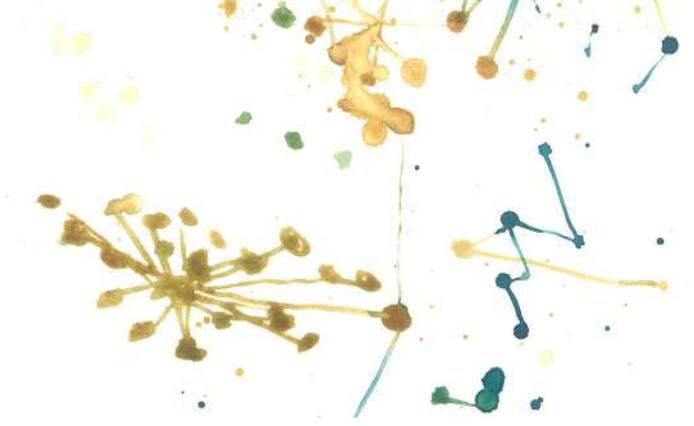
- Stone, L., Simberloff, D., y Artzy-Randrup, Y.** (2019). Network motifs and their origins. *PLOS Comput. Biol.* **15**: e1006749.
- Straley, S.C. y Bruce, V.G.** (1979). Stickiness to Glass. *Plant Physiol.* **63**: 1175-1181.
- Strayer, C., Oyama, T., Schultz, T.F., Raman, R., Somers, D.E., Mas, P., Panda, S., Kreps, J.A., y Kay, S.A.** (2000). Cloning of the Arabidopsis clock gene TOC1, an autoregulatory response regulator homolog. *Science* (80-.). **289**: 768-771.
- Suárez-López, P., Wheatley, K., Robson, F., Onouchi, H., Valverde, F., y Coupland, G.** (2001). CONSTANS mediates between the circadian clock and the control of flowering in Arabidopsis. *Nature* **410**: 1116-1120.
- Supek, F., Bošnjak, M., Škunca, N., y Šmuc, T.** (2011). REVIGO Summarizes and Visualizes Long Lists of Gene Ontology Terms. *PLoS One* **6**: e21800.
- Thaben, P.F. y Westermark, P.O.** (2014). Detecting rhythms in time series with RAIN. *J. Biol. Rhythms* **29**: 391-400.
- Tilbrook, K., Dubois, M., Crocco, C.D., Yin, R., Chappuis, R., Allorent, G., Schmid-Siebert, E., Goldschmidt-Clermont, M., y Ulma, R.** (2016). UV-B perception and acclimation in chlamydomonas reinhardtii. *Plant Cell* **28**: 966-983.
- Tiwari, S.B. et al.** (2010). The flowering time regulator CONSTANS is recruited to the FLOWERING LOCUS T promoter via a unique cis-element. *New Phytol.* **187**: 57-66.
- Tokutsu, R., Fujimura-Kamada, K., Matsuo, T., Yamasaki, T., y Minagawa, J.** (2019). The CONSTANS flowering complex controls the protective response of photosynthesis in the green alga Chlamydomonas. *Nat. Commun.* 2019 101 **10**: 1-10.
- Trapnell, C., Pachter, L., y Salzberg, S.L.** (2009). TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* **25**: 1105-11.
- Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D.R., Pimentel, H., Salzberg, S.L., Rinn, J.L., y Pachter, L.** (2012). Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat. Protoc.* **7**: 562-78.
- Trapnell, C., Williams, B.A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M.J., Salzberg, S.L., Wold, B.J., y Pachter, L.** (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.* **28**: 511-5.
- Uygun, S., Peng, C., Lehti-Shiu, M.D., Last, R.L., y Shiu, S.-H.** (2016). Utility and Limitations of Using Gene Expression Data to Identify Functional Associations. *PLOS Comput. Biol.* **12**: e1005244.
- Valverde, F.** (2011). CONSTANS and the evolutionary origin of photoperiodic timing of flowering. *J. Exp. Bot.* **62**: 2453-63.
- Valverde, F., Mouradov, A., Soppe, W., Ravenscroft, D., Samach, A., y Coupland, G.** (2004). Photoreceptor Regulation of CONSTANS Protein in Photoperiodic Flowering. *Science* (80-.). **303**: 1003-1006.
- Voinnet, O., Rivas, S., Mestre, P., y Baulcombe, D.** (2003). Retracted: An enhanced transient expression system in plants based on suppression of gene silencing by the p19 protein of tomato bushy stunt virus. *Plant J.* **33**: 949-956.

- Wall, M.E., Hlavacek, W.S., y Savageau, M.A.** (2004). Design of gene circuits: lessons from bacteria. *Nat. Rev. Genet.* 2004 51 **5**: 34-42.
- Waltenberger, H., Schneid, C., Grosch, J.O., Bareiss, A., y Mittag, M.** (2014). Identification of target mRNAs for the clock-controlled RNA-binding protein Chlamy 1 from *Chlamydomonas reinhardtii*. *Mol. Genet. Genomics* 2001 2651 **265**: 180-188.
- Wang, C.Q., Guthrie, C., Sarmast, M.K., y Dehesh, K.** (2014). BBX19 interacts with CONSTANS to repress FLOWERING LOCUS T transcription, defining a flowering time checkpoint in Arabidopsis. *Plant Cell* **26**: 3589-3602.
- Wang, G.Y., Shi, J.L., Ng, G., Battle, S.L., Zhang, C., y Lu, H.** (2011a). Circadian Clock-Regulated Phosphate Transporter PHT4;1 Plays an Important Role in Arabidopsis Defense. *Mol. Plant* **4**: 516-526.
- Wang, H., Pan, J., Li, Y., Lou, D., Hu, Y., y Yu, D.** (2016). The DELLA-CONSTANS transcription factor cascade integrates gibberellic acid and photoperiod signaling to regulate flowering. *Plant Physiol.* **172**: 479-488.
- Wang, M., Yuan, D., Gao, W., Li, Y., Tan, J., y Zhang, X.** (2013). A Comparative Genome Analysis of PME and PME1 Families Reveals the Evolution of Pectin Metabolism in Plant Cell Walls. *PLoS One* **8**: e72082.
- Wang, M., Zhao, Y., y Zhang, B.** (2015). Efficient Test and Visualization of Multi-Set Intersections. *Sci. Rep.* **5**: 16923.
- Wang, W., Barnaby, J.Y., Tada, Y., Li, H., Tör, M., Caldelari, D., Lee, D.U., Fu, X.D., y Dong, X.** (2011b). Timing of plant immune responses by a central circadian regulator. *Nat.* 2011 4707332 **470**: 110-114.
- Wang, Y., Wu, J.F., Nakamichi, N., Sakakibara, H., Nam, H.G., y Wu, S.H.** (2011c). LIGHT-REGULATED WD1 and PSEUDO-RESPONSE REGULATOR9 Form a positive feedback regulatory loop in the Arabidopsis circadian clock. *Plant Cell* **23**: 486-498.
- Wang, Z., Gerstein, M., y Snyder, M.** (2009). RNA-Seq: A revolutionary tool for transcriptomics. *Nat. Rev. Genet.* **10**: 57-63.
- Wang, Z.Y., Kenigsbuch, D., Sun, L., Harel, E., Ong, M.S., y Tobin, E.M.** (1997). A Myb-related transcription factor is involved in the phytochrome regulation of an Arabidopsis Lhcb gene. *Plant Cell* **9**: 491.
- Watson, B., Currier, T.C., Gordon, M.P., Chilton, M.D., y Nester, E.W.** (1975). Plasmid required for virulence of *Agrobacterium tumefaciens*. *J. Bacteriol.* **123**: 255.
- Wenkel, S., Turck, F., Singer, K., Gissot, L., Le Gourrierc, J., Samach, A., y Coupland, G.** (2006). CONSTANS and the CCAAT box binding complex share a functionally important domain and interact to regulate flowering of Arabidopsis. *Plant Cell* **18**: 2971-84.
- Wolf, Y.I. y Koonin, E. V** (2012). Tight Link between Orthologs and Bidirectional Best Hits in Bacterial and Archaeal Genome. *Genome Biol. Evol.* **4**: 1286–1294.
- Wu, D., Hu, Q., Yan, Z., Chen, W., Yan, C., Huang, X., Zhang, J., Yang, P., Deng, H., Wang, J., Deng, X., y Shi, Y.** (2012). Structural basis of ultraviolet-B perception by UVR8. *Nat.* 2012 4847393 **484**: 214-219.

Bibliografía

- Wu, J.F., Wang, Y., y Wu, S.H.** (2008). Two new clock proteins, LWD1 and LWD2, regulate arabidopsis photoperiodic flowering. *Plant Physiol.* **148**: 948-959.
- Wu, T. et al.** (2021). clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *Innov.* **2**: 100141.
- Xia, Z. y Liu, Y.** (2001). Reliable and global measurement of fluorescence resonance energy transfer using fluorescence microscopes. *Biophys. J.* **81**: 2395.
- Xie, Q. et al.** (2014). LNK1 and LNK2 are transcriptional coactivators in the Arabidopsis circadian oscillator. *Plant Cell* **26**: 2843-2857.
- Xu, F., Li, T., Xu, P.-B., Li, L., Du, S.-S., Lian, H.-L., y Yang, H.-Q.** (2016). DELLA proteins physically interact with CONSTANS to regulate flowering under long days in *Arabidopsis*. *FEBS Lett.* **590**: 541-549.
- Yanovsky, M.J. y Kay, S.A.** (2002). Molecular basis of seasonal time measurement in Arabidopsis. *Nature* **419**: 308-312.
- Yook, S.H., Jeong, H., y Barabási, A.L.** (2002). Modeling the Internet's large-scale topology. *Proc. Natl. Acad. Sci.* **99**: 13382-13386.
- Young, H.S. et al.** (2008). DNA-binding study identifies C-box and hybrid C/G-box or C/A-box motifs as high-affinity binding sites for STF1 and long hypocotyl5 proteins. *Plant Physiol.* **146**: 1862-1877.
- Yu, G., Wang, L.-G., y He, Q.-Y.** (2015). ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. *Bioinformatics* **31**: 2382-2383.
- Zhang, C., Xie, Q., Anderson, R.G., Ng, G., Seitz, N.C., Peterson, T., McClung, C.R., McDowell, J.M., Kong, D., Kwak, J.M., y Lu, H.** (2013). Crosstalk between the Circadian Clock and Innate Immunity in Arabidopsis. *PLOS Pathog.* **9**: e1003370.
- Zhang, H., He, H., Wang, X., Wang, X., Yang, X., Li, L., y Deng, X.W.** (2011). Genome-wide mapping of the HY5-mediated genenetworks in Arabidopsis that involve both transcriptional and post-transcriptional regulation. *Plant J.* **65**: 346-358.
- Zhang, J.D., Ruschhaupt, M., y Biczok, R.** (2016). ddCt method for qRT-PCR data analysis.
- Zhang, S., Jiao, Z., Liu, L., Wang, K., Zhong, D., Li, S., Zhao, T., Xu, X., y Cui, X.** (2018). Enhancer-Promoter Interaction of SELF PRUNING 5G Shapes Photoperiod Adaptation. *Plant Physiol.* **178**: 1631.
- Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W., y Liu, X.S.** (2008). Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**: R137.
- Zhao, B., Schneid, C., Iliev, D., Schmidt, E.-M., Wagner, V., Wollnik, F., y Mittag, M.** (2004). The Circadian RNA-Binding Protein CHLAMY 1 Represents a Novel Type Heteromer of RNA Recognition Motif and Lysine Homology Domain-Containing Subunits. *Eukaryot. Cell* **3**: 815.
- Zhu, L.J., Gazin, C., Lawson, N.D., Pagès, H., Lin, S.M., Lapointe, D.S., y Green, M.R.** (2010). ChIPpeakAnno: a Bioconductor package to annotate ChIP-seq and ChIP-chip data. *BMC Bioinforma.* **2010** **111** **11**: 1-10.

- Zhu, T., Nevo, E., Sun, D., y Peng, J.** (2012). Phylogenetic analyses unravel the evolutionary history of *nac* proteins in plants. *Evolution* (N. Y). **66**: 1833-1848.
- Zielinski, T., Moore, A.M., Troup, E., Halliday, K.J., y Millar, A.J.** (2014). Strengths and Limitations of Period Estimation Methods for Circadian Data. *PLoS One* **9**: e96462.
- Zobell, O., Coupland, G., y Reiss, B.** (2005). The family of *CONSTANS*-like genes in *Physcomitrella patens*. *Plant Biol.* **7**: 266-275.
- Zones, J.M., Blaby, I.K., Merchant, S.S., y Umen, J.G.** (2015). High-resolution profiling of a synchronized diurnal transcriptome from *Chlamydomonas reinhardtii* reveals continuous cell and metabolic differentiation. *Plant Cell* **27**: 2743-2769.
- Zuo, Z., Liu, H., Liu, B., Liu, X., y Lin, C.** (2011). Blue light-dependent interaction of *CRY2* with *SPA1* regulates *COP1* activity and floral initiation in *arabidopsis*. *Curr. Biol.* **21**: 841-847.





Anexos

Anexo 1

Breve descripción de los datos transcriptómicos de *Arabidopsis* usados en este estudio.
Se incluyen los experimentos, condiciones e identificador de cada muestra.

Experimento	Condiciones	Muestra
Análisis <i>genome-wide</i> de la expresión génica en un ciclo día/noche (microarray) (GSE3416)	WT Día Neutro, ZT0	GSM77055
		GSM77056
		GSM77057
	WT Día Neutro, ZT4	GSM77058
		GSM77059
		GSM77060
	WT Día Neutro, ZT8	GSM77061
		GSM77062
		GSM77063
	WT Día Neutro, ZT12	GSM77064
		GSM77065
		GSM77066
	WT Día Neutro, ZT16	GSM77067
		GSM77068
		GSM77069
	WT Día Neutro, ZT20	GSM77070
		GSM77071
		GSM77072
Análisis <i>genome-wide</i> de los efectos de LINK1 y LINK2 en el transcriptoma de <i>Arabidopsis</i> (RNAseq) (GSE43865)	WT Día Largo, ZT2.	SRR653561
		SRR653562
		SRR653563
	WT Día Largo, ZT6.	SRR653564
		SRR653565
		SRR653566
	WT Día Largo, ZT10	SRR653567
		SRR653568
		SRR653569
	WT Día Largo, ZT14	SRR653570
		SRR653571

Anexos

		SRR653572
	WT Día Largo, ZT18	SRR653573
		SRR653574
		SRR653575
	WT Día Largo, ZT22	SRR653576
		SRR653577
SRR653578		
Análisis <i>genome-wide</i> de los efectos de CO en el transcriptoma de Arabidopsis (RNAseq)	Col-0 (WT), ZT16.	-
		-
	35S:CO, ZT16.	-
		-

Anexo 2

Breve descripción de los datos transcriptómicos de *Chlamydomonas* usados en este estudio. Se incluyen los experimentos, condiciones y e identificador de cada muestra.

Experimento	Condición	Muestra
Análisis <i>genome-wide</i> de la expresión génica de <i>Chlamydomonas</i> durante un día neutro (GSE71469)	Día Neutro 1, ZT1	GSM1835200
	Día Neutro 1, ZT2	GSM1835201
	Día Neutro 1, ZT3	GSM1835202
	Día Neutro 1, ZT4	GSM1835203
	Día Neutro 1, ZT5	GSM1835204
	Día Neutro 1, ZT6	GSM1835205
	Día Neutro 1, ZT7	GSM1835206
	Día Neutro 1, ZT8	GSM1835207
	Día Neutro 1, ZT9	GSM1835208
	Día Neutro 1, ZT10	GSM1835209
	Día Neutro 1, ZT11	GSM1835210
	Día Neutro 1, ZT11.5	GSM1835211
	Día Neutro 1, ZT12	GSM1835212
	Día Neutro 1, ZT12.5	GSM1835213
	Día Neutro 1, ZT13	GSM1835214
	Día Neutro 1, ZT13.5	GSM1835215
	Día Neutro 1, ZT14	GSM1835216
	Día Neutro 1, ZT14.5	GSM1835217
	Día Neutro 1, ZT15	GSM1835218
	Día Neutro 1, ZT16	GSM1835219
	Día Neutro 1, ZT17	GSM1835220
	Día Neutro 1, ZT18	GSM1835221
	Día Neutro 1, ZT19	GSM1835222
	Día Neutro 1, ZT20	GSM1835223
	Día Neutro 1, ZT21	GSM1835224
Día Neutro 1, ZT22	GSM1835225	
Día Neutro 1, ZT23	GSM1835226	
Día Neutro 1, ZT24	GSM1835227	
Día Neutro 2, ZT1	GSM1835228	

Anexos

	Día Neutro 2, ZT2	GSM1835229
	Día Neutro 2, ZT3	GSM1835230
	Día Neutro 2, ZT4	GSM1835231
	Día Neutro 2, ZT5	GSM1835232
	Día Neutro 2, ZT6	GSM1835233
	Día Neutro 2, ZT7	GSM1835234
	Día Neutro 2, ZT8	GSM1835235
	Día Neutro 2, ZT9	GSM1835236
	Día Neutro 2, ZT10	GSM1835237
	Día Neutro 2, ZT11	GSM1835238
	Día Neutro 2, ZT11.5	GSM1835239
	Día Neutro 2, ZT12	GSM1835240
	Día Neutro 2, ZT12.5	GSM1835241
	Día Neutro 2, ZT13	GSM1835242
	Día Neutro 2, ZT13.5	GSM1835243
	Día Neutro 2, ZT14	GSM1835244
	Día Neutro 2, ZT14.5	GSM1835245
	Día Neutro 2, ZT15	GSM1835246
	Día Neutro 2, ZT16	GSM1835247
	Día Neutro 2, ZT17	GSM1835248
	Día Neutro 2, ZT18	GSM1835249
	Día Neutro 2, ZT19	GSM1835250
	Día Neutro 2, ZT20	GSM1835251
	Día Neutro 2, ZT21	GSM1835252
	Día Neutro 2, ZT22	GSM1835253
	Día Neutro 2, ZT23	GSM1835254
	Día Neutro 2, ZT24	GSM1835255

Anexo 3

Breve descripción de los datos transcriptómicos de *Ostreococcus* usados en este estudio. Se incluyen los experimentos, condiciones e identificador de cada muestra.

Experimento	Condición	Muestra
Análisis <i>genome-wide</i> de la expresión génica a lo largo de un día neutro (microarray) (GSE16422)	Cinética 1, Día Neutro 1, ZT0	GSM412869
	Cinética 1, Día Neutro 1, ZT3	GSM412870
	Cinética 1, Día Neutro 1, ZT6	GSM412871
	Cinética 1, Día Neutro 1, ZT9	GSM412872
	Cinética 1, Día Neutro 1, ZT12	GSM412873
	Cinética 1, Día Neutro 1, ZT15	GSM412874
	Cinética 1, Día Neutro 1, ZT18	GSM412875
	Cinética 1, Día Neutro 1, ZT21	GSM412876
	Cinética 1, Día Neutro 2, ZT0	GSM412877
	Cinética 2, Día Neutro 1, ZT0	GSM412878
	Cinética 2, Día Neutro 1, ZT3	GSM412879
	Cinética 2, Día Neutro 1, ZT6	GSM412880
	Cinética 2, Día Neutro 1, ZT9	GSM412881
	Cinética 2, Día Neutro 1, ZT12	GSM412882
	Cinética 2, Día Neutro 1, ZT15	GSM412883
	Cinética 2, Día Neutro 1, ZT18	GSM412884
	Cinética 2, Día Neutro 1, ZT21	GSM412885
	Cinética 2, Día Neutro 2, ZT0	GSM412886
	Cinética 3, Día Neutro 1, ZT0	GSM412887
	Cinética 3, Día Neutro 1, ZT3	GSM412888
	Cinética 2, Día Neutro 1, ZT6	GSM412889
	Cinética 3, Día Neutro 1, ZT9	GSM412890
	Cinética 3, Día Neutro 1, ZT12	GSM412891
	Cinética 3, Día Neutro 1, ZT15	GSM412892
	Cinética 3, Día Neutro 1, ZT18	GSM412893
	Cinética 2, Día Neutro 1, ZT21	GSM412894
	Cinética 2, Día Neutro 2, ZT0	GSM412895

Anexo 4

Enriquecimiento en términos de ontología génica en los grupos génicos periódicos para *Ostreococcus tauri*.

	<i>Ostreococcus tauri</i>		
	Tamaño	Términos GO enriquecidos	Genes
Máx al amanecer / Mín al amanecer	438	<i>Metal ion transport</i> <i>Protein catabolic process</i> <i>DNA conformation change</i>	<i>Ot04g02990</i> <i>Ot11g00020</i> <i>Ot17g01710</i>
Máx al amanecer / Mín en el día	1370	<i>Translation</i> <i>RNA metabolic process</i> <i>Ribosome biogenesis</i>	<i>Ot02g03640</i> <i>Ot14g00400</i> <i>Ot04g00890</i>
Máx al amanecer / Mín al atardecer	1732	<i>Tetrapyrrole metabolic process</i> <i>Lipid metabolic process</i> <i>Single-organism carbohydrate catabolic process</i>	<i>Ot02g04440</i> <i>Ot11g00810</i> <i>Ot01g03020</i>
Máx al amanecer / Mín por la noche	769	<i>Nucleotide metabolic process</i> <i>Oxidation-reduction process</i>	<i>Ot05g04200</i> <i>Ot17g02140</i>
Máx durante el día / Mín al amanecer	946	<i>DNA metabolic process</i> <i>Carbohydrate biosynthetic process</i> <i>Proteolysis</i>	<i>Ot02g00960</i> <i>Ot11g01520</i> <i>Ot02g05670</i>
Máx durante el día / Mín en el día	137	<i>Protein localization to membrane</i> <i>Cellular modified amino acid metabolic process</i>	<i>Ot01g00420</i> <i>Ot17g02400</i>
Máx durante el día / Mín al atardecer	503	<i>Tetrapyrrole biosynthetic process</i> <i>Pyruvate metabolic process</i> <i>Lipid metabolic process</i>	<i>Ot02g04440</i> <i>Ot01g03020</i> <i>Ot07g03180</i>
Máx durante el día / Mín por la noche	1032	<i>Photosynthesis</i> <i>Carbohydrate metabolic process</i> <i>DNA replication</i>	<i>Ot02g07190</i> <i>Ot13g00970</i> <i>Ot01g02530</i>
Máx al atardecer / Mín al amanecer	1551	<i>DNA metabolic process</i> <i>Carbohydrate metabolic process</i> <i>Vesicle-mediated transport</i>	<i>Ot02g05840</i> <i>Ot11g01580</i> <i>Ot09g02760</i>
Máx al atardecer / Mín en el día	433	<i>Microtubule-based process</i> <i>Macromolecule modification</i> <i>Chromosome organization</i>	<i>Ot10g00670</i> <i>Ot06g00140</i> <i>Ot03g01750</i>
Máx al atardecer / Mín al atardecer	85	<i>Nucleobase-containing compound metabolic process</i>	<i>Ot10g00640</i>
Máx al atardecer / Mín por la noche	695	<i>Vesicle-mediated transport</i> <i>Catabolic process</i> <i>DNA replication</i>	<i>Ot08g01900</i> <i>Ot01g03070</i> <i>Ot01g02530</i>
Máx por la noche / Mín al amanecer	507	<i>RNA processing</i> <i>Phosphatidylinositol biosynthetic process</i> <i>cellular aromatic compound metabolic process</i>	<i>Ot11g03480</i> <i>Ot10g03600</i> <i>Ot12g02920</i>

Máx por la noche / Mín en el día	1140	<i>RNA metabolic process</i> <i>Ribosome biogenesis</i> <i>Vitamin metabolic process</i>	<i>Ot11g01040</i> <i>Ot08g03020</i> <i>Ot01g06150</i>
Máx por la noche / Mín al atardecer	647	<i>Gene expression</i> <i>Peptide biosynthetic process</i>	<i>Ot02g03000</i> <i>Ot11g03490</i>
Máx por la noche / Mín por la noche	62	<i>L-serine metabolic process</i>	<i>Ot12g03090</i>

Enriquecimiento en términos de ontología génica en los grupos génicos periódicos para *Chlamydomonas reinhardtii*.

	<i>Chlamydomonas reinhardtii</i>		
	Tamaño	Términos GO enriquecidos	Genes
Máx al amanecer / Mín al amanecer	1426	<i>Single-organism cellular localization</i> <i>Carboxylic acid metabolic process</i> <i>Cellular amino acid metabolic process</i>	<i>Cre16.g679200</i> <i>Cre16.g687350</i> <i>Cre01.g029250</i>
Máx al amanecer / Mín en el día	1801	<i>Ribosome biogenesis</i> <i>RNA metabolic process</i> <i>Phosphorylation</i>	<i>Cre13.g573050</i> <i>Cre01.g012350</i> <i>Cre10.g436100</i>
Máx al amanecer / Mín al atardecer	2869	<i>Cellular amino acid metabolic process</i> <i>Tetrapyrrole metabolic process</i> <i>Carbohydrate catabolic process</i>	<i>Cre12.g514050</i> <i>Cre06.g306300</i> <i>Cre01.g006950</i>
Máx al amanecer / Mín por la noche	1212	<i>Protein targeting to ER</i> <i>Cyclic nucleotide metabolic process</i>	<i>Cre16.g683950</i> <i>Cre02.g074150</i>
Máx durante el día / Mín al amanecer	1987	<i>Protein catabolic process</i> <i>Carbohydrate metabolic process</i> <i>Actin cytoskeleton organization</i>	<i>Cre12.g531100</i> <i>Cre02.g117500</i> <i>Cre16.g676050</i>
Máx durante el día / Mín en el día	256	<i>Peptide biosynthetic process</i> <i>Vitamin metabolic process</i>	<i>Cre10.g421600</i> <i>Cre02.g090150</i>
Máx durante el día / Mín al atardecer	962	<i>Photosynthesis</i> <i>Tetrapyrrole biosynthetic process</i> <i>Pyruvate metabolic process</i>	<i>Cre10.g425900</i> <i>Cre06.g306300</i> <i>Cre02.g093450</i>
Máx durante el día / Mín por la noche	1467	<i>Photosynthesis</i> <i>Protein catabolic process</i> <i>Carbohydrate derivative biosynthetic process</i>	<i>Cre16.g673650</i> <i>Cre10.g418100</i> <i>Cre12.g508700</i>
Máx al atardecer / Mín al amanecer	3751	<i>DNA metabolic process</i> <i>Carbohydrate derivative biosynthetic process</i> <i>Vesicle-mediated transport</i>	<i>Cre01.g017450</i> <i>Cre12.g508700</i> <i>Cre01.g029650</i>
Máx al atardecer / Mín en el día	1573	<i>Mitotic cell cycle</i> <i>Microtubule-based process</i> <i>Regulation of cell cycle</i>	<i>Cre02.g086650</i> <i>Cre10.g456350</i> <i>Cre17.g720350</i>

Anexos

Máx al atardecer / Mín al atardecer	308	<i>Peptide biosynthetic process</i> <i>Regulation of cellular metabolic process</i>	<i>Cre01.g011000</i> <i>Cre12.g489000</i>
Máx al atardecer / Mín por la noche	882	<i>Vesicle-mediated transport</i> <i>Protein catabolic process</i> <i>Chromatin organization</i>	<i>Cre02.g087551</i> <i>Cre10.g418100</i> <i>Cre16.g668200</i>
Máx por la noche / Mín al amanecer	1488	<i>Phosphorylation</i> <i>Phosphatidylinositol biosynthetic process</i> <i>Nucleotide biosynthetic process</i>	<i>Cre10.g423200</i> <i>Cre10.g441100</i> <i>Cre13.g606250</i>
Máx por la noche / Mín en el día	1479	<i>Protein metabolic process</i> <i>Regulation of cellular process</i> <i>Cellular carbohydrate metabolic process</i>	<i>Cre01.g035500</i> <i>Cre12.g489000</i> <i>Cre03.g195600</i>
Máx por la noche / Mín al atardecer	951	<i>Protein metabolic process</i> <i>Signal transduction</i>	<i>Cre01.g001400</i> <i>Cre13.g571200</i>
Máx por la noche / Mín por la noche	57	<i>Cellular respiration</i>	<i>Cre13.g568800</i>

Enriquecimiento en términos de ontología génica en los grupos génicos periódicos para *Arabidopsis thaliana*.

	<i>Arabidopsis thaliana</i>		
	Tamaño	Términos GO enriquecidos	Genes
Máx al amanecer / Mín al amanecer	0		
Máx al amanecer / Mín en el día	1209	<i>Phosphorylation</i> <i>Carbohydrate catabolic process</i> <i>Monocarboxylic acid metabolic process</i>	<i>AT5G19450</i> <i>AT4G17090</i> <i>AT4G25700</i>
Máx al amanecer / Mín al atardecer	811	<i>Response to auxin</i> <i>Carbohydrate biosynthetic process</i> <i>Fatty acid metabolic process</i>	<i>AT1G29440</i> <i>AT1G70290</i> <i>AT1G01120</i>
Máx al amanecer / Mín por la noche	132	<i>Regulation of transcription, DNA-templated</i> <i>Organic cyclic compound metabolic process</i>	<i>AT5G62430</i> <i>AT1G20330</i>
Máx durante el día / Mín al amanecer	983	<i>DNA replication</i> <i>Carbohydrate metabolic process</i> <i>Actin cytoskeleton organization</i>	<i>AT4G02060</i> <i>AT3G27300</i> <i>AT1G13180</i>
Máx durante el día / Mín en el día	127		
Máx durante el día / Mín al atardecer	222	<i>Tricarboxylic acid metabolic process</i> <i>Isoprenoid metabolic process</i>	<i>AT5G04950</i> <i>AT4G15560</i>
Máx durante el día / Mín por la noche	935	<i>Photosynthesis</i> <i>Ribosome biogenesis</i> <i>Tetrapyrrole biosynthetic process</i>	<i>AT4G10340</i> <i>AT5G16750</i> <i>AT4G18480</i>
Máx al atardecer / Mín al amanecer	909	<i>DNA metabolic process</i> <i>Carbohydrate catabolic process</i>	<i>AT2G42120</i> <i>AT2G32290</i>

		<i>Alpha-amino acid metabolic process</i>	<i>AT3G17820</i>
Máx al atardecer / Mín en el día	0		
Máx al atardecer / Mín al atardecer	0		
Máx al atardecer / Mín por la noche	196	<i>Peptide biosynthetic process</i>	<i>AT4G29430</i>
Máx por la noche / Mín al amanecer	122	<i>Hexose metabolic process</i>	<i>AT1G24280</i>
Máx por la noche / Mín en el día	1014	<i>Dephosphorylation Establishment of localization Autophagy</i>	<i>AT1G71860 AT1G08190 AT3G60640</i>
Máx por la noche / Mín al atardecer	177	<i>Hexose metabolic process Response to stimulus</i>	<i>AT5G35790 AT1G10470</i>
Máx por la noche / Mín por la noche	3		

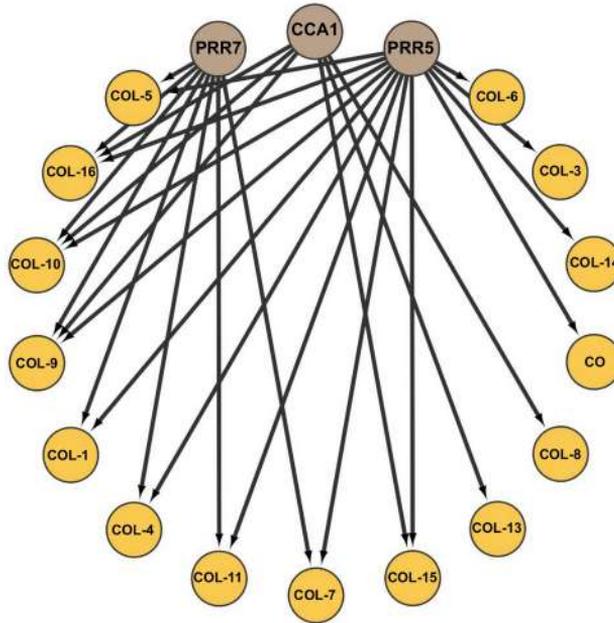
Anexo 5

Factores de transcripción que conforman CircadianFloralNet. Se indica el número de acceso de la base de datos GEO (NCBI) y el número de genes diana de acuerdo a nuestro análisis de ChIPseq.

Factor de transcripción (Número de acceso)	Número de genes diana
SOC1 – At2g45660 (GSE45846)	1328
PIF4 – At2g43010 (GSE43284)	1038
FHY1 – At2g37678 (GSE58083)	6539
APETALA1 – At1g69120 (GSE46987)	2420
SEPALLATA3 – At1g24260 (GSE46987)	5709
PHYA – At1g09570 (GSE48769)	7911
FLM – At1g77080 (GSE48082)	3627
SVP – At2g22540 (GSE33120)	149
APETALA3 - At3g54340 (GSE38358)	8501
PISTILLATA – At5g20240 (GSE38358)	6316
FLC – At5g10140 (*)	4
TOC1 – At5g61380 (GSE35952)	274
AGAMOUS – At4g18960 (GSE45938)	2295
LEAFY – At5g61850 (GSE24568)	4598
PIF4 – At2g43010 (GSE43284)	1038
ARF6 – At1g30330 (GSE51770)	1984
JAG – At1g68480 (GSE51537)	292
PIF5 – At3g59060 (GSE35059)	1789
PIF3 – At1g09530 (GSE39215)	229
PIF1 – At2g20180 (GSE43283)	3190
SPEECHLESS – At5g53210 (GSE57497)	6027
HBI1 – At2g18300 (GSE53099)	15
AL5 – At5g20510 (GSE56706)	24
KAN1 – At5g16560 (GSE48081)	6881
APETALA 2 – At4g36920 (GSE21301)	1572
ERF115 – At5g07310 (GSE48793)	1266
IBH1 – At4g30410 (GSE51120)	59

LHY – At1g01060 (GSE52175)	86
SPL7 – At5g18830 (GSE45213)	3
CCA1 – At2g46830 (GSE70533)	5485
PRR9 – At2g46790 (GSE71397)	137
PRR7 - At5g02810 (GSE49282)	1873
FHY3 – At3g22170 (GSE30713)	1802

Anexo 6



Regulación de genes *COL* por parte del reloj (*PRR7*, *CCA1* y *PRR5*) extraída de la red transcripcional CircadianFloralNet.

Anexo 7

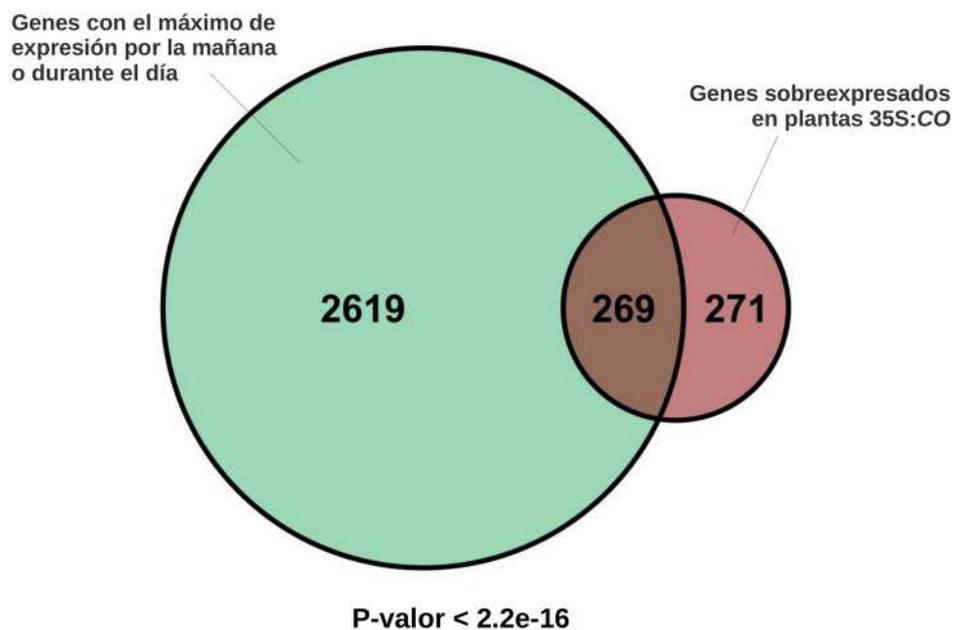
Significancia de la intersección del grupo de genes activados en plantas 35S:CO con cada uno de los grupos de genes rítmicos según el test Exacto de Fisher.

Grupo de genes rítmicos		P-valor
Mínimo	Máximo	
ZT2	ZT2	1
ZT2	ZT6	0.87978794356843
ZT2	ZT10	0.728875629932714
ZT2	ZT14	0.425515877393787
ZT2	ZT18	0.597387907743152
ZT2	ZT22	0.996297367944849
ZT6	ZT2	0.99895687750138
ZT6	ZT6	1
ZT6	ZT10	0.409542847200008
ZT6	ZT14	0.924317768531103
ZT6	ZT18	0.381233248084804
ZT6	ZT22	0.999526156817779
ZT10	ZT2	8.62512622797079e-50
ZT10	ZT6	0.409542847200008
ZT10	ZT10	1
ZT10	ZT14	1
ZT10	ZT18	0.984773510748382
ZT10	ZT22	0.033836115774578
ZT14	ZT2	4.07577907169486e-10
ZT14	ZT6	0.001725259976476
ZT14	ZT10	0.396098439911409
ZT14	ZT14	1
ZT14	ZT18	0.002300528315685
ZT14	ZT22	0.636419726937007
ZT18	ZT2	5.75790456681666e-33
ZT18	ZT6	6.6485038331489e-10
ZT18	ZT10	0.714552843024272

Anexos

ZT18	ZT14	2.14191104338971e-05
ZT18	ZT18	1
ZT18	ZT22	1
ZT22	ZT2	6.37446843169192e-12
ZT22	ZT6	0.019932003627689
ZT22	ZT10	0.435620176459361
ZT22	ZT14	0.012957689584519
ZT22	ZT18	1
ZT22	ZT22	1

Anexo 8



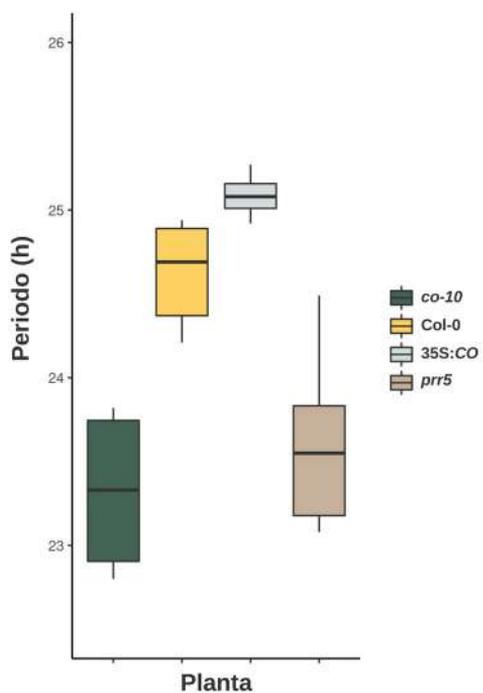
Solapamiento entre los genes activados en 35S:CO y la unión de los grupos de genes rítmicos que mostraron un solapamiento significativo.

Anexo 9

Resultado del test de permutaciones para probar la significancia estadística del solapamiento entre los picos de CO, HY5 y PRR5. Se realizaron 10000 permutaciones y en cada una de ellas se comparó el resultado con el número real de intersecciones, estimándose un p-valor.

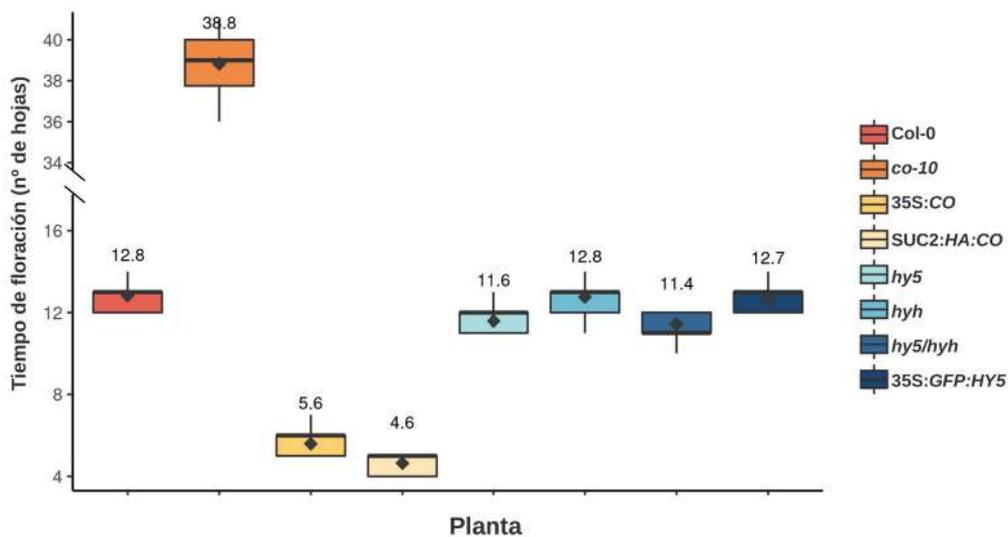
Factor de transcripción 1	Factor de transcripción 2	Número real de intersecciones	Número de intersecciones con picos aleatorios (media ± d.e)	p-valor	FDR
HY5	CO	496	91,827 ± 9,3266912523666	<0,0001	0,0001
PRR5	CO	2132	286,4514 ± 16,9317780091358	<0,0001	0,0001
PRR5	HY5	1603	478,6958 ± 24,5070871808995	<0,0001	0,0001

Anexo 10



Periodo del movimiento rítmico de las hojas obtenido a partir del movimiento vertical incluyendo las líneas *co-10*, *Col-0*, *35S:CO* y *prp5*.

Anexo 11



Tiempo de floración expresado en número de hojas de roseta en el momento de la floración.

