UNIVERSIDAD NACIONAL DE MORENO

DEPARTAMENTO DE CIENCIAS APLICADAS Y TECNOLOGÍA

Código: PICYDT-CAyT-05-2022

"IDENTIFICACIÓN COMPUTACIONAL DE RECEPTORES DE TIPO QUINASAS ASOCIADAS A RESPUESTA DE DEFENSA A PATÓGENOS EN EL PANGENOMA DE GIRASOL"

Director: FILIPPI, Carla Valeria

Co Directora: LIA, Veronica

Integrantes: RIVAROLA, Máximo Lisandro; AGUIRRE, Natalia; PANIEGO, Norma; VILLALBA, Pamela; RIVAS, Juan Gabriel; GONZALEZ, Yesica; GRAMAJO, Leila; TOLENTINO VASQUEZ, Miguel Ángel.

Año: 2025

Identificación del proyecto

Tipo de proyecto y año de convocatoria:	PICyDT VII UNM-2021
Nombre completo del proyecto:	"Identificación computacional de receptores de tipo quinasas asociadas a respuesta de defensa a patógenos en el pangenoma de girasol"
Director/a:	Carla Filippi
Lineamiento prioritario[1]	Bioprocesos y aplicaciones biotecnológicas
Fecha de inicio:	Agosto 2022
Fecha de finalización:	Julio 2024
Unidad de localización: Departamento/centro/ Programa	Programa Académico para la Investigación e Innovación en Biotecnología - Departamento de Ciencias Aplicadas y Tecnología Investigación e Innovación en Biotecnología

[1] Según Resolución CS 326/17 Líneas de investigación científica y desarrollo tecnológico prioritarias 2016-21/ó Res. R 449/18 Lineamientos estratégicos generales de Investigación y transferencia 2019/21 del CEDET

Av Bartolomé Mitre 1891, Moreno, prov. de Buenos Aires
Oficina A 103 - 1° Piso - Edificio Histórico
Tel: (0237) 466-7186/1529/4530
Internos: 134/350
Sec.investigación@unm.edu.ar

Resumen: *a(máx. 300 palabras)*	Los estreses bióticos reducen el rendimiento de los cultivos, constituyendo una amenaza a la seguridad alimentaria mundial, por lo que la obtención de cultivares resistentes a enfermedades es uno de los principales objetivos del fitomejoramiento. La resistencia de amplio espectro es un rasgo deseable porque confiere resistencia contra más de una especie de patógeno o contra la mayoría de las razas o cepas del mismo patógeno. Muchos de los loci que confieren resistencia de amplio espectro codifican para receptores de reconocimiento de patrones (PRR), entre ellos receptores de tipo quinasa (RLK) y las proteínas receptoras (RLP). Estos receptores pertenecen a grandes familias de genes/proteínas cuya clasificación computacional es dificultosa. En este proyecto, llevamos a cabo un estudio integral y a nivel pangenoma de las proteínas PRR. Aunque las subfamilias de RLKs se han estudiado previamente en otras plantas, no se ha realizado ningún estudio exhaustivo sobre esta familia de genes en girasol (H*elianthus annuus*), segunda oleaginosa en importancia en Argentina. Recientemente, y a partir de una colaboración con la Dra. Natalia Aguirre, observamos que tampoco se han caracterizado en Eucaliptus, especie forestal de gran implantación en el país. En este Proyecto nos propusimos con éxito desarrollar estrategias para la identificación a gran escala y caracterización de estas proteínas, en especies no modelo. Se trabajó también con la exploración de bases de datos publicas y trabajo sobre meta data para la selección de datos para estudios posteriores de variantes estructurales, de presencia ausencia, y de perfiles de expresión.
Palabras claves:	PRR, RLK, RLP, mejoramiento

Parte I

1. Introducción y objetivos *(mínimo 1 página- máximo 2 páginas)*

- Realizar una presentación general del estudio (tema/problema) y una justificación de su relevancia (motivos para estudiarlo, aportes potenciales).
- Indicar el objetivo general de la investigación y los interrogantes efectivamente trabajados en el proyecto.

Las plantas son organismos sésiles, es decir que pasan la mayor parte de su vida sujetos a un sustrato, sin posibilidad de desplazamiento. A lo largo de la evolución han desarrollado múltiples mecanismos biológicos y bioquímicos para censar las condiciones ambientales cambiantes, de modo de poder reaccionar rápidamente ante los retos que les impone el ambiente [1], ya sea de tipo abiótico (sequía, salinidad, entre otros), como bióticos (patógenos fúngicos, bacterias, virus, insectos).

Desde el punto de vista de la resistencia a estreses bióticos, a través de un proceso de coevolución en respuesta a los diferentes patógenos con los que se enfrentan, las plantas han desarrollado un complejo mecanismo de defensa [2]. La primera línea de defensa consiste en el reconocimiento de determinadas moléculas altamente conservadas en los microorganismos, conocidas como patrones moleculares asociados a patógenos o microbios (PAMPs o MAMPs). Entre ellos, encontramos a la flagelina bacteriana y la quitina fúngica. Este reconocimiento se produce a través de receptores transmembrana en las células vegetales, denominados receptores de reconocimiento de patrones (PRR), y desencadena una respuesta de inmunidad inducida por PAMP (PTI). La PTI implica la transmisión de señales dentro de la célula a través de cascadas de fosforilación, incluidas las proteínas quinasas activadas por mitógenos (MAPK) y otras quinasas [3]. Las proteínas preponderantes en esta primera capa de defensa son los RLK (receptores de tipo kinasa) y los RLP (proteínas receptoras). Si un patógeno logra superar esta primera línea de defensa, se enfrenta a una segunda línea, gobernada por genes R y sus productos: las proteínas R, principalmente de tipo NBS-LRR. A diferencia de los PRRs, las proteínas R reconocen moléculas patógeno-específicas, denominadas efectores (genes de avirulencia), que no están conservadas entre especies. Este reconocimiento de tipo gen a gen genera una respuesta hipersensible con la consecuente generación de especies reactivas de oxígeno y muerte celular localizada [4].

La introgresión de genes R es una práctica común en el mejoramiento genético de especies comerciales. Sin embargo, dado la naturaleza del mecanismo de resistencia conferido por dichos genes (de tipo gen a gen), suelen proveer una resistencia monogenética robusta mientras dura, pero de nula utilidad una vez superada, contra una única especie o cepa. Por su parte, la inmunidad inducida por PTI (es decir, mediada por PRRs) suele ser de más amplio espectro, capaz de detectar distintas cepas y/o especies de patógenos, aunque lo suficientemente específicas y selectivas como para permitir a la plantas diferenciar los estímulos favorables (ejemplo: simbiosis) de perjudiciales (ejemplo: patógeno) [5]. De esta forma, los

[2] Se solicita brindar información detallada en los campos que componen esta Parte I, ya que será publicada en el Repositorio online de la UNM. Esto permitirá difundir de manera amplia la investigación, sus resultados y visibilizar la labor de los miembros del equipo de investigación.

Av Bartolomé Mitre 1891, Moreno, prov. de Buenos Aires
Oficina A 103 - 1° Piso - Edificio Histórico
Tel: (0237) 466-7186/1529/4530
Internos: 134/350
Sec.investigación@unm.edu.ar

PRRs son blancos ideales para el mejoramiento, ya sea por introgresión, mejoramiento asistido por marcadores moleculares, o estrategias que involucren edición y transgénesis.

Contar con herramientas computacionales para la predicción y el análisis de los genes de resistencia, emerge como un componente clave para asistir a mejoradores en la identificación de genes de resistencia candidatos, que pueden ser útiles para la mejora de cultivos [6]. En los últimos años, se han publicado múltiples estudios y herramientas para identificar computacionalmente proteínas de resistencia citoplasmática (principalmente NBS-LRR) en diferentes especies vegetales [6-8]. En el marco del proyecto PICyDT 2018 UNM-R 251/19, nuestro grupo de trabajo ha establecido una rutina para la identificación y caracterización de genes de resistencia citoplasmáticos (genes R) en girasol (Tolentino, tesina de grado [9]). Para esto, aplicamos una estrategia mixta de identificación a partir del genoma ya anotado, y anotación de novo del genoma.

Debido a la diversidad de los dominios de receptores extracelulares, que los hace más difíciles de caracterizar en comparación con las proteínas R citoplasmáticas, los esfuerzos para identificar y caracterizar computacionalmente los PRRs han sido limitados [10]. Resultados preliminares de nuestro trabajo [9], dan cuenta de la incongruencia entre las distintas herramientas públicas disponibles para la predicción de RLKs (DRAGO2 [6], RRGpredictor [7] y RGAugury [8]. Por su parte, estos predictores casi no muestran variabilidad a la hora de predecir genes R citoplasmáticos. Resulta entonces preciso profundizar en la exploración y diseño de estrategias para la identificación y caracterización inequívoca de RLKs en plantas.

En paralelo, resulta evidente que los ensamblajes de referencia única representan sólo una pequeña fracción del espacio genómico de toda la especie. Dentro de las especies, los genomas varían tanto en el contenido de los genes (por ejemplo, genes duplicados en tándem, variantes de número copia dispersos por todo el genoma y presencia/ ausencia de genes) como en las porciones repetitivas del genoma [11]. De esta forma, es necesario caracterizar esta variación estructural dentro de una especie, a través de los estudios de pangenoma, para tener el repertorio completo de genes de dicha especie.

Haciendo uso de las herramientas y rutinas de trabajo establecidas previamente, la información generada en los últimos años por el grupo de Genómica de Girasol con sede en IABIMO (Instituto Nacional de Tecnología Agropecuaria, INTA - CONICET), el genoma de girasol [12] y datos públicos de secuenciación de nueva generación (NGS; next generation sequencing) de líneas de girasol, este proyecto propone los siguientes **objetivos**:

- Identificar el repertorio completo de proteínas quinasas (i.e el kinoma) en girasol. Realizar una caracterización in silico de esta familia en términos de: localización cromosómica, estructura génica, eventos de duplicación, expansión, localización subcelular, entre otros.
- Realizar un estudio a nivel pangenoma de quinasas: presencia/ausencia, polimorfismos, variación en el número de copias y/o variantes estructurales.
- Realizar un estudio filogenético extendido del kinoma, identificar los subclados de RLK en girasol.
- Estudiar la participación de las RLK en relación a una respuesta de defensa.

Av Bartolomé Mitre 1891, Moreno, prov. de Buenos Aires
Oficina A 103 - 1° Piso - Edificio Histórico
Tel: (0237) 466-7186/1529/4530
Internos: 134/350
Sec.investigación@unm.edu.ar

Por su parte, y a raíz de una colaboración con el grupo genómica forestal de IABIMO, se ampliaron los objetivos para trabajar en paralelo con otra especie no modelo, de gran implantación en el país: *Eucaliptus grandis*

La **hipótesis** que sustenta a este proyecto es que las proteínas RLK tienen un rol preponderante en los procesos de defensa de las plantas frente a patógenos. De esta forma, el desarrollo de herramientas genómicas y bioinformáticas que permitan su identificación inequívoca, acoplado a estudios a nivel transcriptoma (en condiciones de estrés) y pangenoma, son claves para acelerar los procesos de mejoramiento de cultivos, con vistas a la resistencia a enfermedades.

2. Marco de referencia *(min. 2 páginas- máx. 5 páginas)*

Describir en qué campo (temático, disciplinar) se inserta la investigación, indicando:
- estudios antecedentes (propios o no) sobre el tema, avances y áreas de discusión.
- marco teórico o encuadre de referencia de la investigación: con qué enfoque, conceptos, dimensiones o modelos se abordó el tema/problema.

El repertorio de proteínas quinasas, conocido como "kinoma" (término acuñado por Manning et al., 2002 [1]), describe el catálogo de proteínas quinasas en un genoma. Estudios recientes han observado que el kinoma de plantas con flor, es significativamente mayor que el kinoma de otras eucariotas. Esta gran variación entre organismos se debe principalmente a la expansión y contracción de unas pocas familias: más del 60% del kinoma pertenece a la familia RLK [2, 3]. Las RLK median la comunicación célula a célula, ya sea uniéndose a ligandos extracelulares o formando complejos heteroméricos que median la señalización intracelular [4]. Las RLK pertenecen a una gran familia genética monofilética, caracterizadas por poseer un dominio transmembrana (TM), un dominio extracelular amino-terminal variable y un dominio citoplasmático de tipo serina/treonina quinasa conservado en la región carboxilo-terminal [5]. Los dominios extracelulares desempeñan papeles importantes en el reconocimiento de estímulos ambientales y, según sus características, pueden utilizarse para clasificar las RLKs [6]. Recientemente, y usando una clasificación basada en dominios extracelulares se han identificado más de 21 clases estructurales en RLKs en la planta modelo *Arabidopsis thaliana,* siendo la categoría más abundante la que contiene repeticiones ricas en leucina (RLK-LRRs) [7]. Por su parte, el análisis filogenético de las RLKs de Arabidopsis utilizando los dominios quinasa, y la comparación estructural de sus dominios extracelulares permitió la identificación de más de 40 subfamilias [7].

Las funciones que cumplen las RLK en plantas son muy diversas, abarcando desde el crecimiento y el desarrollo [8-9], a las respuestas a estreses bióticos y abióticos [10-12]. En cuanto a las interacciones planta-patógeno, las RLK juegan un papel esencial en las respuestas de defensa mediante el reconocimiento de patrones moleculares conservados asociados a patógenos o microbios (PAMPs/MAMPs), como la flagelina y el factor de elongación EF-Tu [13]. Por su parte, las proteínas RLK que contienen LRR permiten el reconocimiento de patógenos porque su plasticidad estructural les permite unirse a distintos tipos de ligandos, ya sean proteínas, péptidos o lípidos [14].

El girasol constituye la segunda oleaginosa en importancia en Argentina. Si bien en los últimos 60 años los esfuerzos para el fitomejoramiento proporcionaron un suministro continuo de cultivares con características de rendimiento y calidad mejoradas, aún existe una gran brecha entre rendimiento real y potencial, debido principalmente a plagas y enfermedades [15]. Como se mencionara previamente, la introgresión de genes es una práctica común en el mejoramiento genético de especies comerciales. Girasol es un cultivo pionero en este aspecto, con más de 70 años de historia de introgresiones de resistencias, principalmente provenientes de genotipos silvestres [16]. En los últimos años, las tecnologías de genotipado y secuenciación masiva han permitido estudiar la composición de esas fuentes de resistencia, para la identificación de los genes causales. Estudios de mapeo de loci de carácter cuantitativo (QTL) y de asociación (GWAS) han permitido la identificación de regiones del genoma significativamente asociadas con procesos de resistencia, algunas de ellas conteniendo RLKs. A modo de ejemplo, el estudio de caracterización genética y fisiológica de la resistencia del girasol proporcionada por el gen OrDeb2 (de origen silvestre) contra razas altamente virulentas de *Orobanche cumana,* permitió la identificación de cinco genes codificantes para RLK [17]. En esta línea, estudios públicos y de nuestro grupo de trabajo ([18], [19], Filippi et al, manuscrito en prep) han permitido la identificación de genes de resistencia citoplasmáticos para la especie. No obstante, dada la abundancia de proteínas quinasas, su diversidad de función, y la complejidad del genoma de esta especie (~3.6 GigaBases, con >85% de regiones repetitivas), la predicción de RLK con putativa función en procesos de defensa constituye un desafío aún no abordado en girasol, aunque se cuenta con antecedentes en Arabidopsis, arroz, Populus, tomate, citrus, entre otros [20-25]. Estos estudios en diversas especies han demostrado que diferentes genomas contienen diferentes RLKs, y que hay una variación en el contenido de RLKs entre diferentes líneas o cultivares. Esta unión del repertorio completo de genes y regiones de una especie, considerando múltiples individuos o cultivares, se conoce como “pangenoma”.

Estudios recientes de nuestro grupo de trabajo [26] han evidenciado una estrecha base genética en los materiales actualmente usados a nivel global para el mejoramiento de la especie. Sin embargo, aunque existe estrecha diversidad genética, diferentes niveles de resistencia entre cultivares de girasol a las distintas plagas que afectan al cultivo han sido reportados, entre ellos a *Verticillium dahliae [27], Sclerotinia sclerotiorum [28], Phomopsis helianthi* [29], entre otras. En línea con esto, y dada la reciente disponibilidad de datos de secuenciación a genoma completo de cultivares y parientes silvestres de girasol, se vuelve factible la identificación, caracterización y comparación de las familias de genes de resistencia entre cultivares con diferentes niveles de resistencia (es decir, a nivel pangenoma), y no solo la exploración de un único genoma de referencia.

Por su parte, en Argentina, Eucalyptus es una especie exótica forestal implantada que tiene alta demanda para productos de madera, papel y pulpa, entre otros. A pesar del éxito que han mostrado como especies de plantación, estos árboles sucumben a plagas y patógenos, lo cual desafía y pone en riesgo su

Av Bartolomé Mitre 1891, Moreno, prov. de Buenos Aires
Oficina A 103 - 1° Piso - Edificio Histórico
Tel: (0237) 466-7186/1529/4530
Internos: 134/350
Sec.investigación@unm.edu.ar

sanidad. En este sentido, la identificación de genes que codifican para proteínas implicadas en resistencia es clave para proveer herramientas para el mejoramiento asistido.

Este proyecto se propone no solo la identificación y caracterización del kinoma en girasol, con énfasis en RLKs, sino también la conducción de un estudio a nivel pangenoma de las variables de presencia/ausencia y de los niveles de diversidad y combinación de las RLK en la especie mediante la exploración de datos de secuenciación públicos. En paralelo, nos propusimos la identificación y caracterización de PRR en una especie forestal de relevancia para el país: *Eucaliptus grandis.* Esperamos que este estudio permita aumentar la comprensión de los procesos de defensa del huésped, así como la generación de capacidades para el aprovechamiento de datos biológicos disponibles en bases de datos públicas.

3. Métodos y técnicas ***(min. 2 páginas- máx. 4 páginas)***

Indicar el trabajo de campo, documental y/o de laboratorio realizado, la forma de recolección de datos y sus fuentes. Al respecto, describir los métodos, técnicas, instrumentos y materiales utilizados para indagar el problema de investigación. Explicitar las unidades de análisis, los criterios de selección de muestras o casos. Indicar asimismo las formas de procesamiento y análisis de los datos recolectados.

a) Estudio del kinoma en girasol. Caracterización

Para identificar el kinoma, se descargó el proteoma de girasol (disponible en sunflowergenome.org). HMMER fue usado para la identificación de dominios. Mediante scripts propios, se retuvieron todas aquellas proteínas que contenían al menos un dominio quinasa típico (PFAM: PF00069, PF07714, PF08488; Interpro: IPR000719, IPR001245, IPR008271, IPR011009, IPR017441, IPR020635, IPR021820, IPR022126; SUPERFAMILY: SSF56112; SMART: SM00219, SM00220; Prosite: PS00107, PS00108, PS50011; Gene3D: G3DSA:3.30.200.20).

CD-HIT se utilizó para descartar secuencias redundantes, reteniendo en cada caso la más larga. Los dominios quinasa de las secuencias remanentes fueron alineados para confirmar la presencia de los mismos y descartar pseudogenes, tal cual está descrito en Lehti-Shiu y Shiu [1]. Solo se retuvieron aquellas secuencias cuyos dominios quinasa cubrían al menos el 50% del dominio descrito en la base de datos PFAM.

La localización cromosómica de las quinasas fue extraída del genoma anotado. Se determinaron posibles regiones cromosómicas homólogas y eventos de duplicación utilizando MCScanX [2]. Los genes adyacentes con un máximo de una interrupción génica se consideraron genes duplicados en tándem.

Las tasas de sustitución sinónima (Ks) y no sinónima (Ka) se calcularon utilizando la función 'add_ka_y_ks_to_collinearity.pl' de MCScanX. La organización génica (número de intrones y exones, largo, etc.) se extrajo del archivo GFF para la especie, utilizando scripts propios. Otras propiedades

Av Bartolomé Mitre 1891, Moreno, prov. de Buenos Aires
Oficina A 103 - 1° Piso - Edificio Histórico
Tel: (0237) 466-7186/1529/4530
Internos: 134/350
Sec.investigación@unm.edu.ar

físicas como el pI (punto isoeléctrico) teórico y el peso molecular de las quinasas se calcularon utilizando seqinR [3].

Para determinar las proteínas quinasa transmembrana (TM), se hizo una predicción de péptidos señal N-terminal y dominios TM utilizando SignalP [4] y TMHMM [5], respectivamente. La localización subcelular de las quinasas se realizó con TargetP [6].

Finalmente, y siguiendo a [7], el enfoque de identificación para determinar qué quinasas eran RLK siguió esta lógica (operadores lógicos: Y, O, y NO): ["presencia de péptido señal" Y "hélice transmembrana (al menos una)" Y "dominio/s de quinasa"] Y ["dominio/s extracelular/es (al menos uno de estos): LRR O L-Lectina O C-Lectina O G-Lectina O LysM O PR5K O TNFR NO WAK O Malectina O EGF o Stress-Antifung"] NO [dominios "NB-ARC"].

b) Identificación y caracterización de quinasas a nivel pangenoma

Se descargaron los datos de pangenoma publicados por Hubner y colaboradores [8], generado sobre más de 400 accesiones de girasol (incluyendo líneas endocriadas y especies de Helianthus silvestre). Sobre el pan-proteoma, se repitió el flujo de trabajo establecido en el objetivo anterior de este plan de Tesis, para la identificación del repertorio completo de quinasas en el pangenoma de girasol. Dado que dicho artículo [8] también contenía datos de variantes puntuales (SNPs, polimorfismo de nucleótido simple e INDELS, pequeñas inserciones/deleciones), se recuperaron los datos de variantes que colocalizaban con las quinasas identificadas (más/menos 2,000 pares de bases río arriba y abajo, para determinar variantes que afectaban también regiones regulatorias), permitiendo de esta manera detectar la estructura y polimorfismo de los mismos. Se identificaron también las variantes de presencia/ausencia y las variaciones en el número de copias de quinasas entre accesiones de girasol. Los datos de polimorfismos de materiales cultivados y silvestres fueron contrastados, para evaluar la contribución de quinasas, y específicamente de RLK, del germoplasma silvestre al cultivo. Este constituyó el primer análisis a nivel pangenoma de quinasas en girasol.

c) Estudio filogenético extendido del kinoma

Para la clasificación del kinoma, nos basamos en las HMMs de las diferentes subfamilias reportadas por Lehti-Shiu y Shiu [1], basadas en secuencias aminoacídicas de dominios quinasa obtenidas de 21 especies de plantas, los cuales fueron analizados conjuntamente con los dominios quinasa identificados en el punto a) en girasol. Para optimizar los conjuntos de datos para los análisis evolutivos, se utilizó la herramienta "Decrease Redundancy", disponible en ExPaSy (www.expasy.org), para eliminar las secuencias idénticas o lejanamente relacionadas (parámetros: 99% de similitud máxima y 30% de similitud mínima). Los análisis filogenéticos se realizaron mediante el método de máxima verosimilitud, implementado en PhyML [9]. Se probaron doce modelos evolutivos diferentes (JTT, LG, DCMut, MtREV,

Av Bartolomé Mitre 1891, Moreno, prov. de Buenos Aires
Oficina A 103 - 1° Piso - Edificio Histórico
Tel: (0237) 466-7186/1529/4530
Internos: 134/350
Sec.investigación@unm.edu.ar

MtMam, MtArt, Dayhoff, WAG, RtREV, CpREV, Blosum62 y VT) utilizando el software ProtTest [10]. El modelo evolutivo que mejor se ajustó a los datos se determinó en base al criterio de información de Akaike. Los valores de soporte de los árboles se estimaron utilizando la prueba de razón de verosimilitud (LRT, PhyML). Los árboles se visualizaron y editaron con R. Se identificaron los subclados de RLK.

d) Estudio de la participación de RLK en respuesta a estrés en girasol

Para esto, se buscó en bases de datos públicas (ENA, European Nucleotide Archive; SRA, Sequence Read Archive; GEO, Gene Expression Omnibus) set de datos crudos de secuenciación de ARN (RNA-seq) en girasol. Los criterios para la selección de estos set de datos incluyeron: i) al menos tres réplicas por cada tratamiento; ii) que no se hubieran generado sobre el genotipo del genoma de referencia (para poder identificar también transcritos de genes ausentes en este genotipo, pero presentes en el pangenoma), iii) que existiera el manuscrito asociado, iv) largo mínimo de lecturas de NGS: 70 pb, v) lecturas pareadas. Cabe destacar que se consideraron datos de RNA-seq de distintos contextos: estrés biótico (como se previó), pero se estudió la inclusión de datos de respuesta a estrés abiótico, y (dependiendo de la disponibilidad), de diversos estadios de desarrollo y/o tejidos.
Brevemente, la estrategia de análisis fue: luego de la inspección por calidad (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/), y limpieza [11], las lecturas fueron mapeadas contra el transcriptoma de referencia (unido al pan-transcriptoma) usando Bowtie2 ([12], considerando solo lecturas de mapeo único, con <2 bases desapareadas). La matriz de conteos se generó utilizando Salmon [13]. La identificación de genes de expresión diferencial (DEG; differentially expressed genes) entre tratamientos se realizó utilizando DEseq2 [14], considerando DEG aquellos genes con q-valor < 0.05 y log2FC > 1. Se evaluó la intersección de las listas de todas las quinasas, y específicamente de las RLK, con las listas de DEG, buscando asociación entre los roles predichos, los perfiles de expresión y los contextos. Los análisis estadísticos se realizaron con R.

f) Línea de trabajo derivada: estudio de PRR en *Eucaliptus*

Por medio de una colaboración con la Dra Natalia Aguirre, se llevó adelante un trabajo de identificación y caracterización de PRR en *Eucalyptus grandis*. Utilizando el proteoma de *E. grandis* (V2.0) se realizó una búsqueda de PRR putativos mediante dos herramientas públicas (DRAGO3 y RRGPredictor) y se compararon con los predichos para esta especie por Ngou et al. (quienes usaron para ello una estrategia propia, DOI:10.1038/s41477-022-01260-5). Adicionalmente, a fin de validar su rol como potenciales PRR, se hizo un estudio de meta-análisis para determinar cuántos de ellos contaban con evidencia de

Av Bartolomé Mitre 1891, Moreno, prov. de Buenos Aires
Oficina A 103 - 1° Piso - Edificio Histórico
Tel: (0237) 466-7186/1529/4530
Internos: 134/350
Sec.investigación@unm.edu.ar

participación en procesos de defensa (PRR-DEG, o PRR diferencialmente expresados en respuesta a patógenos). Para esto, se colectaron todos los estudios de RNAseq, en estrés biótico, publicados a la fecha en Eucalyptus (5 artículos), y se recuperaron ciertos datos: carácter evaluado (enfermedad-patógeno: avispa de la agalla -*Leptocybe invasa*-; cancro -*Chrysoporthe austroafricana y Calonectria pseudoreteaudii*- y roya -*Austropuccinia psidii*-), versión del genoma de *E. grandis* usada como referencia (V1.0 o V2.0), log2 de la tasa de cambio (log2FC), p-valor ajustado (p-adj), entre otros. Para aquellos trabajos que usaron la V1.0 del genoma, se hizo la interconversión a V2.0 mediante una estrategia basada en BLAST. Adicionalmente, se analizó la región promotora (i.e. 1500 pares de bases río arriba) de estos PRR-DEGs para identificar elementos reguladores cis (CRE) asociados con la respuesta a estreses bióticos.

4. Resultados y discusión *(min. 5 páginas- máx. 15 páginas)*

Desarrollar los resultados, en relación a los objetivos del proyecto, especificando (de ser posible) los siguientes aspectos:
- nuevos conocimientos obtenidos sobre los casos o unidades bajo estudio.
- avances en materia de conocimiento científico sobre el tema bajo estudio, formulación de enfoques originales e innovadores (modelos, conceptos, etc.).
- Contribuciones para la resolución de problemas específicos y/o formulación de herramientas de intervención, diseño o mejora de productos y procesos.

Por último, desarrollar las conclusiones y reflexiones finales a las que se llegó luego de la investigación, en relación a los interrogantes y objetivos planteados.

Se presentan los resultados principales, discriminados por especie bajo estudio:

GIRASOL

Mediante la comparación del proteoma de referencia de girasol con un modelo oculto de Markov (HMM) específico para quinasas eucariotas, identificamos un total de 2,197 quinasas. Estas se clasificaron en 20 clases, basada en la clasificación de Lehti Shiu, siendo RLK la clase más abundante. La distribución en cromosomas mostró que el cromosoma 11 es el que más quinasas acumula (Figura 1).

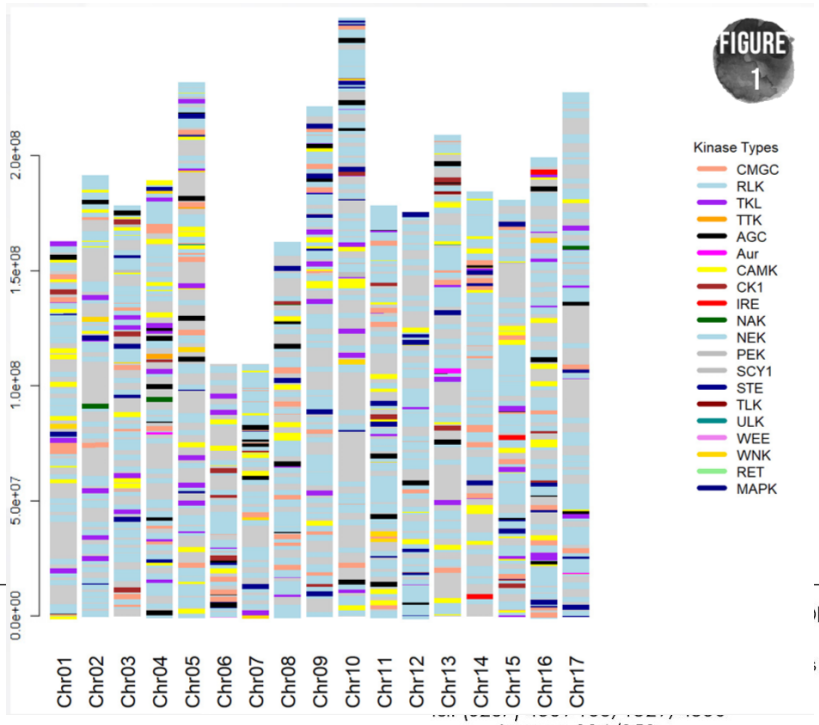


Figura 1. Representación esquemática de los 17 cromosomas de girasol y la distribución de las quinasas predichas. Dichas quinasas se colorearon en base a la clasificación de Lehti Shiu.

Los análisis filogenéticos basados en sus dominios quinasa revelaron la presencia de dos clados principales: quinasas receptoras y quinasas solubles, cada uno subdividido a su vez en subclados funcionales más específicos (figura 2A). Esto se verifica también al realizar la predicción de localización subcelular de estas quinasas (Figura 2B).

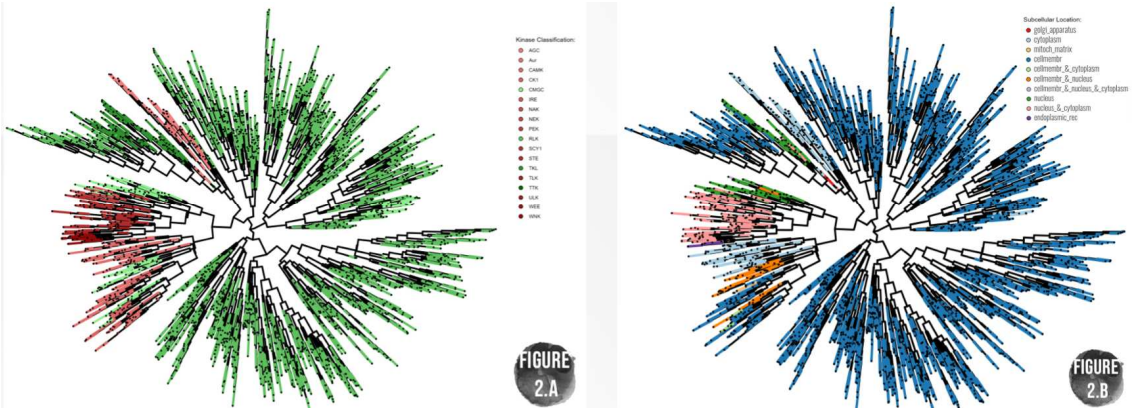


Figura 2. Árbol circular basado en distancias, para las 2197 kinasas predichas. A) quinasas coloreadas en dos grupos mayoritarios: solubles (gamas de rojo) y receptoras (gamas de verde). Las quinasas receptoras (i.e. ancladas a membrana) incluyen las CMGC, TKL, TTK, y las RLK. B) quinasas coloreadas de acuerdo a su localización subcelular predicha. El grupo mayoritario, azul, corresponde a las quinasas ancladas a membrana. Esto se asocia con su clasificación observada en A) como quinasas receptoras (gamas de verde).

Para profundizar en la caracterización de las quinasas receptoras, desarrollamos una estrategia novedosa que permite la identificación precisa de quinasas tipo receptor (RLKs) basándonos en combinaciones específicas de dominios. Esta estrategia se resume mediante la siguiente función lógica: ["presencia de un péptido señal" O "al menos un dominio transmembrana"] Y ["dominio quinasa"] Y ["dominio extracelular (ECD): LRR O L-Lectina O C-Lectina O G-Lectina O LysM O PR5K O TNFR O WAK O Malectina O EGF"] Y ["ECD>30 aminoácidos"] NO ["NB-ARC"].

A través de esta estrategia, logramos identificar un total de 723 RLKs en girasol, las cuales fueron clasificadas en 20 clases funcionales basándonos en su homología con las RLKs de *Arabidopsis thaliana*. Además, las predicciones y análisis de bloques sinténicos y eventos de duplicación de RLKs revelaron un enriquecimiento significativo en duplicaciones segmentales (n=906). Paralelamente, la comparación de las tasas evolutivas, medidas a través de Ka (sustituciones no sinónimas), Ks (sustituciones sinónimas) y

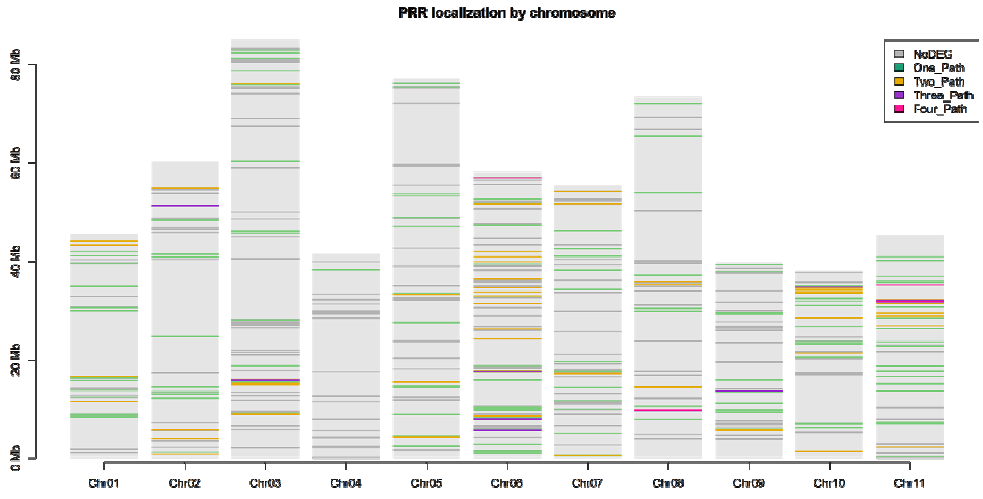
Av Bartolomé Mitre 1891, Moreno, prov. de Buenos Aires
Oficina A 103 - 1° Piso - Edificio Histórico
Tel: (0237) 466-7186/1529/4530
Internos: 134/350
Sec.investigación@unm.edu.ar

su cociente (Ka/Ks), indicó que las RLKs están bajo una fuerte selección purificadora (promedio Ka/Ks=0.26), lo que sugiere una alta conservación funcional de estas proteínas en el girasol.

Para completar la caracterización del repertorio de quinasas en esta especie, estamos extendiendo el estudio al nivel del pangenoma. En este contexto, hemos realizado una anotación *de novo* de 714 nuevas quinasas que no estaban presentes en el genoma de referencia XRQ. Estas nuevas quinasas fueron identificadas a partir del pangenoma que construyó nuestro grupo utilizando datos de resecuenciación de 149 individuos. Este trabajo representa el primer esfuerzo integral y exhaustivo para la caracterización de las quinasas en la especie *Helianthus annuus,* proporcionando información fundamental sobre la diversidad y evolución de estas importantes proteínas. Dado que este trabajo forma parte del proyecto de Tesis de Maestria del Lic Tolentino, y atendiendo que los proyectos de formación de recursos humanos llevan tiempos diferentes, aún está en proceso de avance. En este sentido, estamos aún estudiando los perfiles de expresión diferencial para la identificación de RLK diferencialmente expresadas en respuesta a estrés a partir de datos de RNAseq públicos. Los scripts para estos análisis ya han sido desarrollados, pero el análisis integrado de los datos de transcriptoma y pangenoma, para la identificación de un catálogo de RLK promisorias para el mejoramiento, aún está en proceso.

EUCALIPTUS

De un total de 2835 genes predichos por al menos una de estas tres estrategias, 730 fueron reconocidos consistentemente por las tres, siendo entonces candidatos robustos a ser PRR. Estos 730 PRR putativos fueron caracterizados teniendo en cuenta su clasificación, distribución en cromosomas, presencia o no en clusters, entre otros. Se lograron localizar 663 genes entre los once cromosomas y 67 a nivel scaffolds, observándose que los cromosomas 6 y 4 son los que más y menos PRR acumulan (130 y 29 respectivamente), sin correlación al tamaño de los cromosomas. Se hizo una anotación funcional (GO, PFAM) y posterior análisis de enriquecimiento (figura 3).



Av Bartolomé Mitre 1891, Moreno, prov. de Buenos Aires
Oficina A 103 - 1° Piso - Edificio Histórico
Tel: (0237) 466-7186/1529/4530
Internos: 134/350
Sec.investigación@unm.edu.ar

Figura 3. Localización y distribución cromosómica de PRR y PRR-DEG en *E. grandis*. En gris PRR no DEG; en verde los PRR-DEG en respuesta a un patógeno, en amarillo PRR-DEG en respuesta a dos patógenos, en violeta PRR-DEG en respuesta a tres patógenos y en fucsia PRR-DEG en respuesta a 4 patógenos.

La exploración de la historia evolutiva de estos genes mediante el análisis de duplicación génica con la herramienta MCScanX, permitió identificar 122 genes duplicados segmentales (figura 4). Las tasas evolutivas fueron evaluadas a través de la relación Ka/Ks (siendo Ka 'sustituciones no sinónimas' y Ks 'sustituciones sinónimas'), lo que indicó que la mayoría de los PRR están bajo selección purificadora, con un rango de Ka/Ks entre 0.051 y 0.597. Esto sugiere que muchos de estos genes desempeñarían funciones conservadas en la especie.

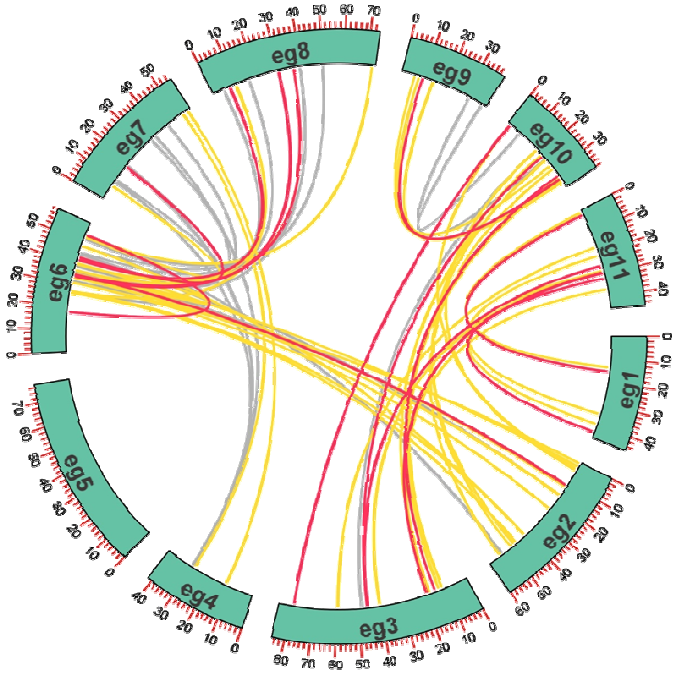


Figura 4: Localización cromosómica y colinealidad de genes PRR en *E. grandis*. En gris se indican los pares de genes duplicados que no fueron DEGs, en amarillo los pares génicos donde al menos uno de ellos es DEG y en fucsia los pares duplicados donde ambos fueron DEG.

Para validar su rol potencial en la defensa contra patógenos, realizamos un meta-análisis de estudios transcriptómicos (RNA-seq), Identificando 283 genes diferencialmente expresados entre los 730 PRR candidatos (PRR-DEGs), de los cuales 16 destacaron por estar regulados diferencialmente frente a al menos tres patógenos en estudios independientes, y regulados al alza en al menos una condición, lo que sugiere que estos PRR proporcionarán respuesta defensiva de amplio espectro (Figuras 3, 4, y 5).

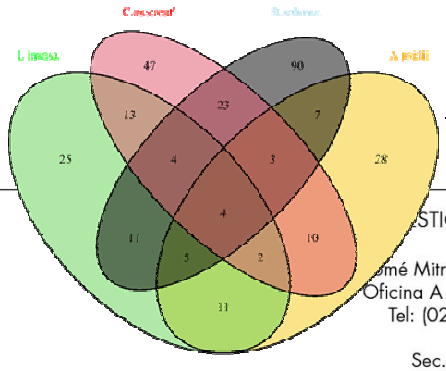


Figura 5: 283 PRR-DEG diferenciados por patógenos. Avispa de la agalla -*Leptocybe invasa*-; cancro -*Chrysoporthe austroafricana* y *Calonectria pseudoreteaudii*- y roya -*Austropuccinia psidii*-

Adicionalmente, el análisis de la región promotora (i.e. 1500 pares de bases río arriba) de estos 16 PRR-DEGs permitio identificar elementos reguladores cis (CRE) asociados con la respuesta a estreses bióticos. Entre ellos se encontraron motivos relacionados con regulación hormonal y respuesta al estrés, destacándose aquellos de respuesta a fitohormonas como auxina, metil jasmonato y ácido salicílico, que están vinculados con las respuestas de defensa en plantas (Figura 6).

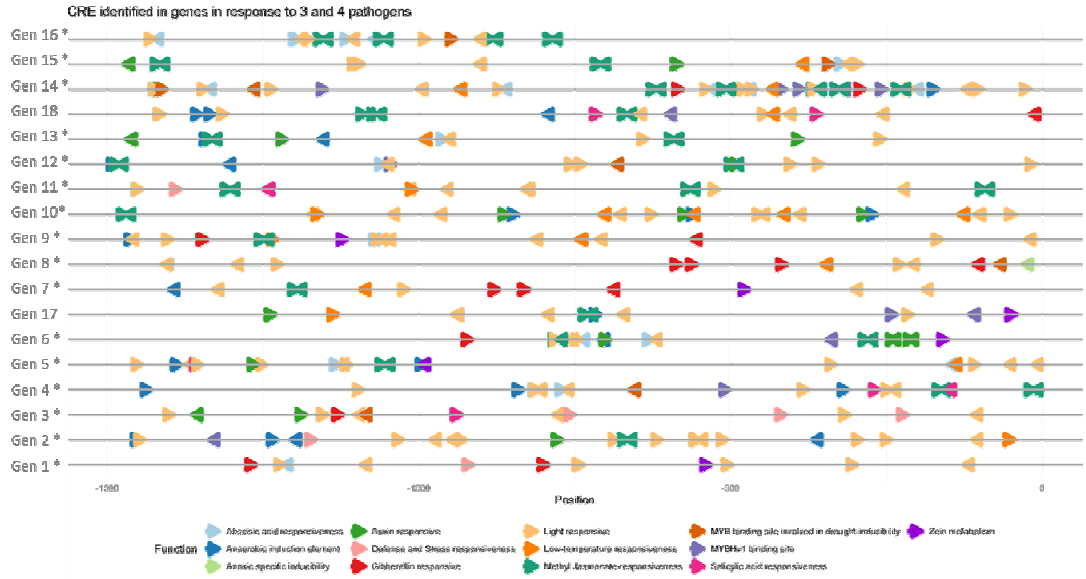


Figura 6: Predicción de elementos que actúan en cis en las regiones promotoras (1500 pb rio arriba) de genes PRR-DEGs que presentaron evidencia en estreses bióticos para 3 y 4 patógenos. Las flechas hacia la derecha indican que el elemento está ubicado en la hebra positiva, mientras que las flechas apuntando hacia la izquierda señalan su presencia en la hebra negativa.

Este trabajo contribuye a la caracterización funcional y evolutiva de PRRs en *E. grandis*, ofreciendo una base para el desarrollo de herramientas que permitan la identificación temprana de individuos resistentes a múltiples patógenos en programas de mejoramiento forestal.

5. Nuevos interrogantes y líneas de investigación a futuro

Consignar si la investigación hizo surgir nuevos interrogantes o si emergieron potenciales líneas de investigación a desarrollarse en el futuro a partir de los hallazgos.

En el marco del PICyDT “Identificación computacional de receptores de tipo quinasas asociadas a respuesta de defensa a patógenos en el pangenoma de girasol”, hemos identificado el repertorio completo de proteínas quinasas en girasol (i.e., el ‘quinoma’ de la especie) tanto a nivel genómico como

pangenómico (i.e., considerando múltiples individuos). Entre estas quinasas, nos hemos enfocado en la familia de RLK, observando que, de manera análoga a lo que sucede en los NLR, muchas RLK han experimentado duplicaciones, y estos genes duplicados tienden a estar organizados preferentemente en clústeres. Esta observación dio lugar a la hipótesis central del Trabajo Final de Investigación (TFI) que está llevando adelante la estudiante Leila Gramajo, según la cual ciertos TE están más próximos a las RLK de lo esperado por azar y podrían estar mediando la evolución y/o el silenciamiento de las RLK implicadas en la resistencia a enfermedades en girasol.

En paralelo, estamos escribiendo el manuscrito Gonzalez et al, derivado de la TFI de Yesica Gonzalez.

6. Bibliografía *(min. 2 página- máx. 4 páginas)*

Consignar los textos y fuentes utilizados en la redacción de los campos anteriores.

PLANTEAMIENTO DEL PROBLEMA

[1] Gimenez F et al. Worldwide research on plant defense against biotic stresses as improvement for sustainable agriculture. Sustainability, 2018, vol. 10, no 2, p. 391.

[2] Yuan, M, et al. Pattern-recognition receptors are required for NLR-mediated plant immunity. Nature, 2021, vol. 592, no 7852, p. 105-109.

[3] Romeis T. Protein kinases in the plant defence response. Current opinion in plant biology, 2001, vol. 4, no 5, p. 407-414.

[4] Lu, Y; Tsuda K. Intimate association of PRR-and NLR-mediated signaling in plant immunity. Molecular Plant-Microbe Interactions, 2021, vol. 34, no 1, p. 3-14.

[5] Li, W, et al. Exploiting broad-spectrum disease resistance in crops: from molecular dissection to breeding. Annual review of plant biology, 2020, vol. 71, p. 575-603.

[7] Silva, R; Micleli F. RRGPredictor, a set-theory-based tool for predicting pathogen-associated molecular pattern receptors (PRRs) and resistance (R) proteins from plants. Genomics, 2020, vol. 112, no 3, p. 2666-2676.

[8] Li, P et al. RGAugury: a pipeline for genome-wide prediction of resistance gene analogs (RGAs) in plants. BMC genomics, 2016, vol. 17, no 1, p. 1-10.

[9] Tolentino M. Busqueda y caracterización de genes R en girasol: herramientas para el mejoramiento del cultivo. Tesis para optar por el título de Licenciado en Biotecnología de la Universidad Nacional de Moreno. 2021

[10] Sekhwal MK, Li P, Lam I, Wang X, Cloutier S, You FM. Disease resistance gene analogs (RGAs) in plants. Int J Mol Sci. 2015;16:19248–90.

[11] Della Coletta R, et al. How the pan-genome is changing crop genomics and improvement. Genome biology, 2021, vol. 22, no 1, p. 1-19.

[12] Badouin, H et al. The sunflower genome provides insights into oil metabolism, flowering and Asterid evolution. Nature, 2017, vol. 546, no 7656, p. 148-152.

MARCO TEÓRICO

[1] Manning G, Whyte DB, Martinez R, Hunter T, Sudarsanam S. The protein kinase complement of the human genome. Science. 2002;298(5600):1912–34.

[2] Lehti-Shiu MD, Shiu SH. Diversity, classification and function of the plant protein kinase superfamily. Philos Trans R Soc Lond Ser B Biol Sci. 2012;367(1602):2619–39.

[3] Morillo SA, Tax FE. Functional analysis of receptor-like kinases in monocots and dicots. Curr Opin Plant Biol. 2006;9(5):460–9.

[4] Osakabe Y, Yamaguchi-Shinozaki K, Shinozaki K, Tran L-SP. Sensing the environment: key roles of membrane-localized kinases in plant perception and response to abiotic stress. J Exp Bot. 2013;64:445–58.

[5] Walker JC. Structure and function of the receptor-like protein kinases of higher plants. Plant Mol Biol. 1994;26:1599–609.

[6] Shiu SH, Bleecker AB. Plant receptor-like kinase gene family: diversity, function, and signaling. Sci STKE. 2001. doi:10.1126/stke.2001.113.re22.

[7] Shiu SH, Bleecker AB. Receptor-like kinases from Arabidopsis form a monophyletic gene family related to animal receptor kinases. Proc Natl Acad Sci U S A. 2001;98:10763–8.

[8] Osakabe Y, Maruyama K, Seki M, Satou M, Shinozaki K, Yamaguchi-shinozaki K. Leucine-Rich Repeat Receptor-Like Kinase1 Is a Key Membrane-Bound Regulator of Abscisic Acid Early Signaling in Arabidopsis. Plant Cell. 2005;17:1105–19.

[9] Pitorre D, Llauro C, Jobet E, Guilleminot J, Brizard J-P, Delseny M, Lasserre E. RLK7, a leucine-rich repeat receptor-like kinase, is required for proper germination speed and tolerance to oxidative stress in Arabidopsis thaliana. Planta. 2010;232:1339–53.

[10] Ouyang SQ, Liu YF, Liu P, Lei G, He SJ, Ma B, Zhang WK, Zhang JS, Chen SY. Receptor-like kinase OsSIK1 improves drought and salt stress tolerance in rice (Oryza sativa) plants. Plant J. 2010;62:316–29.

[11] de Lorenzo L, Merchan F, Laporte P, Thompson R, Clarke J, Sousa C, Crespi M. A novel plant leucine-rich repeat receptor kinase regulates the response of Medicago truncatula roots to salt stress. Plant Cell. 2009;21:668–80.

[12] Zipfel C. Pattern-recognition receptors in plant innate immunity. Curr Opin Immunol. 2008;20:10–6.

[13] Boller T, Felix G. A renaissance of elicitors: perception of microbe-associated molecular patterns and danger signals by pattern-recognition receptors. Annu Rev Plant Biol. 2009;60:379–406.

[14] Holt III BF, Mackey D, Dangl JL. Recognition of pathogens by plants. Curr Biol. 2000;10:R5–7.

[15] Gururani, M, et al. Plant disease resistance genes: current status and future directions. Physiological and molecular plant pathology, 2012, vol. 78, p. 51-65.

[16] Seiler, GJ.; et al. Utilization of sunflower crop wild relatives for cultivated sunflower improvement. Crop Science, 2017, vol. 57, no 3, p. 1083-1101.

[17] Fernandez-Aparicio, M et al. Genetic and physiological characterization of sunflower resistance provided by the wild-derived OrDeb2 gene against highly virulent races of Orobanche cumana Wallr. Theoretical and Applied Genetics, 2021, p. 1-25.

[18] Neupane, S, et al. Genome-wide identification of NBS-encoding resistance genes in sunflower (Helianthus annuus L.). Genes, 2018, vol. 9, no 8, p. 384.

[19] Tolentino M. Busqueda y caracterizacion de genes R en girasol: herramientas para el mejoramiento del cultivo. Tesis para optar por el título de Licenciado en Biotecnología de la Universidad Nacional de Moreno. 2021

[20] Diévart A, Gilbert N, Droc G, Attard A, Gourgues M, Guiderdoni E, Périn C. Leucine-rich repeat receptor kinases are sporadically distributed in eukaryotic genomes. BMC Evol Biol. 2011;11:367.

[21] Zan Y, Ji Y, Zhang Y, Yang S, Song Y, Wang J. Genome-wide identification, characterization and expression analysis of populus leucine-rich repeat receptor-like protein kinase genes. BMC Genomics. 2013;14:318.

[22] Sakamoto T, Deguchi M, Brustolini OJB, Santos AA, Silva FF, Fontes EPB. The tomato RLK superfamily: phylogeny and functional predictions about the role of the LRRII-RLK subfamily in antiviral defense. BMC Plant Biol. 2012;12:229.

[23] Shiu S, Karlowski WM, Pan R, Tzeng Y, Mayer KFX, Li W. Comparative Analysis of the Receptor-Like Kinase Family in Arabidopsis and Rice. Plant Cell. 2004;16:1220–34.

[24] Fischer I, Diévart A, Droc G, Dufayard J-F, Chantret N. Evolutionary dynamics of the Leucine-Rich Repeats Receptor-Like Kinase (LRR-RLK) subfamily in angiosperms. Plant Physiol. 2016;170:1595–610.

[25] Magalhaes DM., et al. LRR-RLK family from two Citrus species: genome-wide identification and evolutionary aspects. BMC genomics, 2016, vol. 17, no 1, p. 1-13.

[26] Filippi CV et al. Genetic diversity, population structure and linkage disequilibrium assessment among international sunflower breeding collections. Genes, 2020, vol. 11, no 3, p. 283.

[27] Montecchia, JF., et al. On-field phenotypic evaluation of sunflower populations for broad-spectrum resistance to Verticillium leaf mottle and wilt. Scientific reports, 2021, vol. 11, no 1, p. 1-14.

[28] Filippi CV, et al. Phenotyping sunflower genetic resources for sclerotinia head rot response: assessing variability for disease resistance breeding. Plant disease, 2017, vol. 101, no 11, p. 1941-1948.

[29] Pogoda CS., et al. Genetic loci underlying quantitative resistance to necrotrophic pathogens Sclerotinia and Diaporthe (Phomopsis), and correlated resistance to both pathogens. Theoretical and Applied Genetics, 2021, vol. 134, no 1, p. 249-259.

METODOLOGÍA

Av Bartolomé Mitre 1891, Moreno, prov. de Buenos Aires
Oficina A 103 - 1° Piso - Edificio Histórico
Tel: (0237) 466-7186/1529/4530
Internos: 134/350
Sec.investigación@unm.edu.ar

[1] Lehti-Shiu MD, Zou C, Shiu S-H. Origin, diversity, expansion history, and functional evolution of the plant receptor-like kinase/pelle family. In: Tax F, Kemmerling B, editors. Receptor-like kinases in plants: from development to defense. Berlin, Heidelberg: Springer Berlin Heidelberg; 2012. p. 1–22.

[2] Wang, Y et al. MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. Nucleic acids research, 2012, vol. 40, no 7, p. e49-e49.

[3] Charif, D; Lobry, JR. SeqinR 1.0-2: a contributed package to the R project for statistical computing devoted to biological sequences retrieval and analysis. En Structural approaches to sequence evolution. Springer, Berlin, Heidelberg, 2007. p. 207-232.

[4] Petersen, TN, et al. SignalP 4.0: discriminating signal peptides from transmembrane regions. Nature methods, 2011, vol. 8, no 10, p. 785-786.

[5] Krogh, A, et al. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. Journal of molecular biology, 2001, vol. 305, no 3, p. 567-580.

[6] Armenteros, JJA, et al. Detecting sequence signals in targeting peptides using deep learning. Life science alliance, 2019, vol. 2, no 5.

[7] Restrepo-Montoya D, et al. Computational identification of receptor-like kinases “RLK” and receptor-like proteins “RLP” in legumes. BMC genomics, 2020, vol. 21, no 1, p. 1-17.

[8] Hubner, S, et al. Sunflower pan-genome analysis shows that hybridization altered gene content and disease resistance. Nature plants, 2019, vol. 5, no 1, p. 54-62.

[9] Guindon, S, et al. PHYML Online—a web server for fast maximum likelihood-based phylogenetic inference. Nucleic acids research, 2005, vol. 33, no suppl_2, p. W557-W559.

[10] Darriba, Diego, et al. ProtTest 3: fast selection of best-fit models of protein evolution. Bioinformatics, 2011, vol. 27, no 8, p. 1164-1165.

[11] Bolger, Anthony M et al. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics (Oxford, England) vol. 30,15 (2014): 2114-20.

[12] Langmead, Ben, and Steven L Salzberg. Fast gapped-read alignment with Bowtie 2. Nature methods vol. 9,4 357-9. 4 Mar. 2012.

[13] Patro, Rob et al. Salmon provides fast and bias-aware quantification of transcript expression. Nature methods vol. 14,4 (2017): 417-419.

[14] Love, Michael I et al. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome biology vol. 15,12 (2014): 550.

Parte II

Dimensiones de cumplimiento del Plan de Trabajo