UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO
PROGRAMA DE DOCTORADO EN CIENCIAS BIOMÉDICAS
INSTITUTO DE ECOLOGÍA

Evolución adaptativa de la resistencia contra herbívoros en

Datura stramonium

QUE PARA OPTAR POR EL GRADO DE:

DOCTOR EN CIENCIAS

PRESENTA:

Ivan Mijail De la Cruz Arguello

Tutor principal: Dr. Juan Núñez Farfán
Instituto de Ecología, UNAM

Tutor: Dr. Daniel Piñero Dalmau
Instituto de Ecología, UNAM

Tutor: Dr. Pedro L. Valverde Padilla
Departamento de Biología, UAM

Sinodales
Dr. Angélica Cibrián Jaramillo, Laboratorio Nacional de Genómica de la Biodiversidad, CINVESTAV
Dr. Ek del Val de Gortari, Escuela Nacional de Estudios Superiores campus Morelia, UNAM
Dr. Juan S. Núñez Fárfan, Instituto de Ecología, UNAM
Dr. Ken Oyama Nakagawa, Escuela Nacional de Estudios Superiores campus Morelia, UNAM
Dr. Luis E. Eguiarte Fruns, Instituto de Ecología, UNAM

Ciudad Universitaria, CDMX, CD. MX, Agosto 2020

Dedicatoria

Esta tesis es para mi amada madre y amado padre, para mis amados hermanos y para mi perrihijo Gandalf….Gracias por esta bella vida que me brindan y por todo su amor incondicional y cariño el cual es recíproco de mi parte. Ustedes son mi gasolina para aventurarme a muchas cosas en esta vida…Los amo

"As Forest Gump's mama said; life is like a box of chocolates you never know what you are gonna get……"

Agradecimientos institucionales

Al programa de Doctorado en Ciencias Biomédicas de la Universidad Nacional Autónoma de México por darme la oportunidad de realizar mis estudios de doctorado.

Al Instituto de Ecología de la UNAM, a la UNAM por si misma y al laboratorio de Genética Ecológica y Evolución por brindarme todo lo necesario para realizarme académica, profesional y personalmente.

Al Consejo Nacional de Ciencia y Tecnología por la beca de manutención para realizar mis estudios doctorales (beca #283799).

Este trabajo fue financiado por el Consejo Nacional de Ciencia y Tecnología a través de la convocatoria Fronteras de la Ciencia (beca #1527) y por el PAPIIT, UNAM (PAPIIT-DGAPA AG200717).

Al programa de apoyo para estudios de posgrado, PAEP, por las becas otorgadas para realizar trabajo de campo, un congreso, un curso en la ciudad de Copenhague y una estancia de investigación en el Instituto Max Planck para la Ecología Química.

Al laboratorio Nacional de Salud de la FES-Iztacala, UNAM a cargo del profesor Cesar Matero Flores y el M. en C. Luis Barbo por brindarme todo el apoyo para los análisis en el HPLC-MS-TOF.

A la Sociedad Max Planck, así como al Departamento de Ecología Molecular del Instituto Max Planck para la Ecología Química, a cargo del profesor Ian Thomas Baldwin por brindarme una beca y el apoyo para realizar una estancia de investigación en su grupo.

A la Agencia Nacional de Educación Finlandesa (Finnish National Agency for Education; EDUFI), por la beca para terminar mis estudios doctorales y terminar de investigar y escribir esta tesis en Finlandia (aplicación #KM-19-11065).

A la Universidad de Helsinki y su Unidad de Investigación en Genética Ecológica (EGRU) a cargo del profesor Juha Merilä por albergarme y apoyarme en todo momento para terminar está tesis.

Al grupo de trabajo "Popgen" de la Universidad de Copenhague por otorgarme una beca para asistir al curso "Summer course in analyses of genotyping and NGS data in medical and population genetics 2019".

A la Comisión Nacional para el Uso de la Biodiversidad (CONABIO) y al Laboratorio Nacional de Síntesis Ecológica, ENES campus Morelia, UNAM por brindarme acceso a sus respectivos equipos computacionales para los análisis bioinformáticos de esta tésis.

A las doctoras Ek del Val de Gortari, Ángelica Cibrián Jaramillo y a los doctores Luis Eguiarte Fruns, Juan Núñez Farfán y Ken Oyama Nakagawa, miembros de mi sínodo de examen de grado por sus valiosas sugerencias, comentarios y correcciones a esta tesis que ayudaron a mejorarla, muchas gracias.

A mi tutor principal del Instituto de Ecología, profesor Juan Núñez Farfán y a mis tutores profesores Daniel Piñero y Pedro Luis Valverde por apoyarme siempre en todo.

Agradecimientos personales

Esta tesis no pudo haber sido posible sin toda la ayuda en todos los sentidos de mi adorada y amada mamá y mi gran amado papá. Ustedes son parte esencial de que haya acabo esta tesis. Su amor, su apoyo emocional e incondicional, su apoyo económico y la felicidad que me brindan todos los días hizo posible que terminara esta tesis. ¡Gracias a la fundación De la Cruz-Arguello que me ha salvado de muchas! Siempre han apoyado mis ideas, muchas veces sin estar 100 % de acuerdo, pero me apoyan y se los agradezco eternamente, espero que esto sea también de su felicidad, verme cumplir mis metas (nuestras metas) y una pequeñísisisima parte de como retribuirles todo lo que han invertido y hecho por mí. A mi tío Ranulfo, por toda su ayuda, cariño y por sus consejos durante mi crecimiento personal desde siempre.

A mi tutor principal profesor Juan Núñez Farfán, te agradezco profundamente haberme aceptado en el laboratorio y brindarme todo el apoyo científico, personal y económico para realizar mi doctorado. Gracias por tenerme confianza y por introducirme al mundo de las *Datura* y de las interacciones planta-herbívoros. ¡De verdad, muchísimas gracias por todo docc! Espero seguir colaborando contigo y seguir recolectando antigüedades en Teotihuacán y sino pues me las guardas.

A la M. en C. Laura L. Cruz y a la Dra. Rosalinda Tapia por su esfuerzo y apoyo en campo, laboratorio, logística, amistad. No tengo duda, que sin su ayuda tampoco hubiésemos sacado esta tesis adelante. Estaré eternamente agradecido con ustedes.

A los profesores Daniel Piñero y P. L. Valverde por siempre atender con entusiasmo los tutorales, por su amabilidad, respeto y por toda la orientación que me dieron durante mi formación como estudiante de doctorado, ¡gracias!

A L.V.L. gracias por todo.

Dr. Asis Hallab, alias Kaliman, gracias por interesarte en el proyecto y aceptar ser mi mentor de bioinformática, yo sé que querías enseñarme todo lo que sabes y a veces era muy difícil. Pero sí aprendí y también gracias por empujarme a resolver análisis que parecían imposibles para mí. Gracias por confiar en mí y bueno también la pasamos muy bien divirtiéndonos y sacrificando personas a Tlaloc desde la torre de la CONABIO para que corrieran bien los análisis.

A mis hermanos que, aunque lejos, siempre están ahí pendientes de mi desarrollo y de mi bienestar. Por todas esas platicas por WhatsApp bien divertidas. Los amo mucho y extraño siempre.

A mis amigas y amigos con quienes comparto aventuras locas e interesantes, siempre ha sido demasiado bueno olvidarme del mundo de la ciencia un rato y bueno que diversión; Yesica, Daniel, Tavo, León, Mau, Nene, Chon. Esas pachangas de veras….

A la señora Antonieta y al señor Toribio, por brindarnos sus parcelas en Teotihuacán y Ticumán, respectivamente, para llevar a cabo los experimentos de campo.

Al Dr. Juha Merilä por su interés, amabilidad y apoyo para investigar *Datura* en su lab de sticklebacks.

A todo el grupo de Genética Ecológica (EGRU), los lab. meetings fueron demasiado enriquecedores.

Muchas personas, maestros, amigos, compañeros que me han enseñado, ayudado o han compartido un rato conmigo, muchas gracias. Dr. Ulises Rosas, gracias por las pocas, pero creativas y enriquecedoras charlas que tuvimos para este proyecto.

Finalmente, a mis compañeros y amigos del Laboratorio de Genética Ecológica y Evolución, UNAM; Kalil, Sabi, Lulú, Maried, Píli, Memo, Adán, Diana, Eunice, Marisol, Citlali Ivan Camargo, Jesus Torres, Adri, Gallo. Por las charlas científicas, salidas recreativas, otras charlas nada científicas y algun@s de ustedes me han brindado su amistad, que bueno conocerl@s.

Índice

Resumen

Dilucidar como las plantas terrestres (*i. e.,* embriofitas) se defienden del ataque de sus enemigos naturales representa aún hoy un reto para los biólogos evolutivos y ecólogos. Más aún, desentrañar las bases genéticas de la defensa de las plantas contra sus herbívoros es un tema relativamente poco desarrollado en el campo de las interacciones planta-herbívoros. En esta tesis, se estudia la evolución adaptativa de la resistencia contra herbívoros en la herbácea anual *Datura stramonium* (Solanaceae), empleando las herramientas de distintas disciplinas tales como la ecología, evolución, genética cuantitativa, genómica y química analítica. En el capitulo I, se proporciona un marco teórico sobre las bases moleculares, genéticas y químicas de las defensas de las plantas terrestres contra sus enemigos naturales. En el capitulo II, se estudia la adaptación local de 21 defensas putativas químicas y tricomas foliares de *D. stramonium* mediante un experimento de jardín común bajo condiciones ambientales controladas y usando un enfoque de genética cuantitativa. Por otro lado, en el capitulo III, se estudia a que grado la evolución de las defensas químicas de *D. stramonium* han sido impulsadas por la selección natural impuesta por sus herbívoros especialistas y analizamos sí la resistencia de las plantas contra sus herbívoros tiene una base genética. Finalmente, en el capitulo IV, se investiga la evolución adaptativa de la defensa en contra de los enemigos naturales en el genoma de *D. stramonium*. Lo resultados obtenidos a lo largo de los cuatro capítulos de esta investigación aportan evidencia química, genética/genómica y ecológica de la adaptación local en plantas de *D. stramonium* para enfrentar a sus distintas especies de herbívoros que varían geográficamente. Los resultados encontrados a lo largo de este estudio están en concordancia con la teoría del mosaico geográfico coevolutivo, la cual postula que, dado que la comunidad de herbívoros cambia a lo largo de la distribución de las plantas, se espera que la fuerza y dirección de la selección natural sobre la defensa de las plantas sea distintas entre las poblaciones, resultando en un mosaico coevolutivo entre las plantas y sus enemigos naturales.

Abstract

Elucidating how land plants (*i. e.,* bryophytes) defend themselves from the attack of their natural enemies is still a challenge for evolutionary biologists and ecologists. Furthermore, disentangling the genetic basis of plant defenses against their herbivores is a relatively underdeveloped topic in the field of plant-herbivore interactions. In this thesis, using tools from different disciplines such as ecology, evolution, quantitative genetics, genomics, and analytical chemistry, I studied the adaptive evolution of resistance against herbivores in the annual herb *Datura stramonium* (Solanaceae). In Chapter I, a theoretical framework is provided on the molecular, genetic, and chemical bases of the plant defenses. In Chapter II, it is studied the concentration levels of 21 tropane alkaloids and foliar trichomes of *D. stramonium* to elucidate local adaptation by means of a common garden experiment under controlled environmental conditions and using a quantitative genetic approach. In chapter III, it is studied to what extent the evolution of chemical defenses of *D. stramonium* has been driven by natural selection imposed by its specialist herbivores and I also investigate whether plant resistance has a genetic basis. Finally, in Chapter IV, the adaptive evolution of defense against natural enemies in the *D. stramonium* genome is investigated. The results obtained in this research provide chemical, genetic/genomic, and ecological evidence of local adaptation in *D. stramonium* plants to face their different species of herbivores that differ geographically. The results found throughout this study are in agreement with the coevolutionary geographic mosaic theory, which hypothesizes that, since the herbivore community changes throughout the distribution of plants, it is expected that the strength and direction of natural selection on plant defenses are different between populations, resulting in a coevolutionary mosaic between plants and their natural enemies.

Introducción general

Defensas de las plantas contra herbívoros

Las estrategias defensivas que son desplegadas por las plantas terrestres (*i. e.*, embriofitas) para enfrentar el ataque de sus herbívoros pueden estudiarse desde los puntos de vista ontogénicos, fisiológicos, genéticos y epigéneticos, así como considerando las interacciones ecológicas con otras especies de plantas e insectos (War *et al.* 2012; Schuman & Baldwin 2016). Los primeros desarrollos conceptuales en la comprensión de la biología de la defensa de las plantas incluyen la propuesta de Dethier quien estudió las diversas formas en las cuales distintas especies de insectos pueden responder a los cambios fisiológicos dentro de las plantas (Dethier 1954); la hipótesis de Frankel, quien mencionó que los compuestos secundarios de las plantas funcionan como defensas adaptativas contra la herbivoría (Frankel 1959); y la conclusión de Ehrlich & Raven, quienes sugirieron de que la coevolución de defensa y contradefensa entre las plantas y los herbívoros resultó en la radiación adaptativa de angiospermas e insectos herbívoros (Ehrlich & Raven 1964).

Se ha estimado que las plantas terrestres y los insectos han vivido juntos por más de 350 millones años (Gatehouse 2002). Esto ha dado como resultado un arreglo interacciones complejas de defensa y contra defensa entre las plantas y sus insectos herbívoros (Wu & Baldwin 2010). Esta carrera armamentista entre plantas-herbívoros ha resultado en el desarrollo de una variedad de respuestas defensivas por parte de las plantas embriofitas (Kessler & Baldwin 2002; Wu & Baldwin 2010; War *et al.* 2012). Estas respuestas defensivas incluyen la producción de barreras físicas tales como espinas, tricomas y cutículas, así como metabolitos secundarios especializados de defensa que son tóxicos y/o reducen la digestibilidad de los herbívoros (Kessler & Baldwin 2002; Wu and Baldwin 2010; War *et al.* 2012). De hecho, evidencia robusta señala la importancia crucial de los metabolitos secundarios en las interacciones planta-insectos, planta-patógenos y en la comunicación planta-planta (Bennett and Wallsgrove 1994; Ueda *et al.* 2012; Schuman & Baldwin 2018).

Los rasgos de las plantas que confieren resistencia (*i. e.*, rasgos que tienen un efecto negativo directo o indirecto contra sus enemigos naturales; Núñez-Farfán *et al.* 2007) a plagas de insectos pueden ser clasificados de acuerdo con la manera en la cual estos son empleados (Howe & Jander 2008). Algunos rasgos son expresados constitutivamente (expresados continuamente) bajo el control de distintos programas moleculares y del desarrollo, conectados entre si y altamente específicos, independientes del nivel de amenaza de los herbívoros o si estos

están presentes o no (Howe & Jander 2008). En contraste, otras defensas de las plantas son inducidas y otras defensas son únicamente expresadas en respuesta a la herbivoría y en el sitio dañado (Bostock 2005). Las defensas inducidas de las plantas son una clase de sistema defensivo complejo que puede reconocer moléculas externas o señales de los tejidos dañados (Howe & Jander 2008; Verhage *et al.* 2010). Por ejemplo, las plantas perciben que están siendo dañadas a partir de reconocer señales químicas (moléculas), liberadas por los insectos (elicitores asociados a la herbivoría) en sus secreciones orales al momento de alimentarse del tejido de la planta (Xu *et al.* 2015). Consecuentemente, las plantas producen hormonas claves (acido jasmónico y sus derivativos) responsables de la activación de las respuestas químicas defensivas en contra de los herbívoros (Xu *et al.* 2015).

Las defensas también pueden ser desplegadas indirectamente como compuestos orgánicos volátiles (VOCs) o volátiles verdes de las hojas (GLVs) (Kessler & Baldwin 2002). Estos compuestos son liberados después del daño por los herbívoros y permiten atraer parasitoides y depredadores para los insectos herbívoros o pueden repeler la ovoposición de las plagas (Kessler & Baldwin 2002). Asimismo, las plantas también son capaces de tolerar el daño ejercido por los herbívoros (Rosenthal & Kotanen 1994; Strauss & Agrawal 1999). La tolerancia de las plantas a la herbivoría refleja el grado en el que una planta puede mantener su éxito reproductivo a través de la capacidad de reasignar recursos nutrimentales para el crecimiento, reverdecimiento y reproducción después del daño causado por los herbívoros (Rosenthal & Kotanen 1994; Strauss & Agrawal 1999).

Por otra parte, para un insecto herbívoro, una planta es más que una simple comida, es un estilo de vida (Berenbaum 1990). Algunas especies de insectos llevan cada una de las etapas de su ciclo de vida en su planta hospedera de la cual pueden alimentarse, escapar de sus depredadores, pasar el invierno o usarla en el apareamiento y ovoposición (Berenbaum 1990). Particularmente, los insectos con ámbitos hogareños cortos deben coordinar su ciclo de vida con el de su planta hospedera (Berenbaum 1990), ya que en el caso contrario no únicamente pierden su alimento sino todos los otros beneficios que una planta puede ofrecerles; por lo tanto, debe haber una presión selectiva muy fuerte en los herbívoros para poder adaptarse a las características de su planta hospedera (Berenbaum 1990). Sin embargo, muchas de estas peculiaridades en la planta hospedera, son el resultado de adaptaciones de la planta a la misma presión selectiva ejercida por los herbívoros (Berenbaum 1990; Karban & Baldwin 1997).

Históricamente, el estudio de las interacciones entre insectos y plantas ha abarcado una amplia gama de preguntas biológicas desde el nivel molecular hasta el ecosistema, todo unido por la biología evolutiva (Giron *et al.* 2018). Este campo de investigación ha sido revolucionado

recientemente por las nuevas tecnologías y enfoques analíticos, incluida la secuenciación de próxima generación (por sus siglas en inglés; NGS) y la tecnología de edición de genes (*e. g.,* CRISPR-Cas9) (Dyer *et al.* 2018; Giron *et al.* 2018). También se han realizado avances en análisis químicos de alta resolución como en la espectrometría de masas (Dyer *et al.* 2018; Giron *et al.* 2018). Actualmente, es posible integrar diversas disciplinas tales como la biología molecular, genómica, química, genética cuantitativa y ecología para el estudio de las interacciones planta-herbívoros en condiciones ambientales controladas y/o en entornos naturales (Giron *et al.* 2018). Esto ha permitido lograr una comprensión más completa de las redes ecológicas complejas, la dinámica fisiológica, ecológica y evolutiva de estas interacciones, y la base genética de los rasgos, así como probar hipótesis que antes eran difíciles de probar (Dyer *et al.* 2018). De esta manera la integración de varias disciplinas es fundamental para entender mejor la evolución adaptativa de la resistencia de las plantas contra sus enemigos naturales (Dyer *et al.* 2018; Giron *et al.* 2018).

En esta investigación, se estudia la evolución adaptativa de la resistencia contra los herbívoros especialistas en la herbácea anual *Datura stramonium* (Asteridae; Solanales; Solanaceae), usando distintas herramientas de la genética/genómica, ecología evolutiva, genética cuantitativa y química analítica para aportar conocimiento nuevo y de frontera al campo de las interacciones planta-herbívoro. De tal manera, que en este estudio se busca responder a dos preguntas básicas, pero aún relevantes en el campo de la ecología evolutiva de las interacciones planta-herbívoros; (1) ¿cómo las plantas de *D. stramonium* se defienden contra sus distintos herbívoros?, (2) ¿a que grado la evolución adaptativa de los rasgos defensivos ha sido promovida por la selección natural impuesta por los distintos herbívoros de *D. stramonium*?

Conceptos y métodos generales claves empleados en este estudio

Para responder a las dos preguntas antes mencionadas, se usaron distintas metodologías las cuales se describen en esta sección.

Metabólomica. La metabólomica cuantitativa de las plantas es una herramienta que ayuda a mejorar nuestro entendimiento de la bioquímica y el metabolismo de las plantas a través de obtener medidas confiables de la concentración de metabolitos que ocurren en distintos niveles de las muestras de plantas a analizar (Jorge *et al.* 2016). En este estudio se utilizó la cromatografía liquida de tiempo de vuelo acoplada a espectrometría de masas (por sus siglas en inglés; HPLC-TOF-MS, Fig. 1), este equipo posibilita la identificación de compuestos químicos conocidos o desconocidos que pueden ser identificados a partir de su formula química, de su masa/carga

(m/z) y patrones característicos isotópicos; tiempos de retención y fragmentación de la molécula (Jorge *et al.* 2016). Cuando los extractos de las plantas son inyectados al HPLC-TOF-MS, estos pasan por una columna que se encarga de separar los iones y estos iones son acelerados por un campo eléctrico (Ferrer & Thurman 2008) (Fig. 1). Esta aceleración da como resultado a un ion que posea la misma energía cinética que cualquier otro ion que tenga la misma carga (masa/carga; Ferrer & Thurman 2008). La velocidad del ion depende de la relación masa-carga; los iones más pesados con la misma carga alcanzan velocidades más bajas, mientras que los iones con carga más alta aumentan en velocidad (Ferrer & Thurman 2008). Se mide el tiempo que le toma al ion alcanzar un detector a una distancia conocida. Este tiempo dependerá de la velocidad del ion y, por lo tanto, es una medida de su relación masa-carga (Ferrer & Thurman 2008). A partir de esta relación y parámetros experimentales conocidos (*i. e.*, tiempo de retención, fragmentación de la molécula), se puede identificar el ion (Ferrer & Thurman 2008) (Fig. 1).

Una de las ventajas del HPLC-TOF-MS es que el volumen de muestra requerido para los análisis es muy pequeño (< 1g; Jorge *et al.* 2016) y la identificación de los compuestos es altamente precisa (Jorge *et al.* 2016). Asimismo, este equipo resulta ser muy útil cuándo los estándares de los compuestos no se encuentran disponibles (Jorge *et al.* 2016) (Fig. 1).

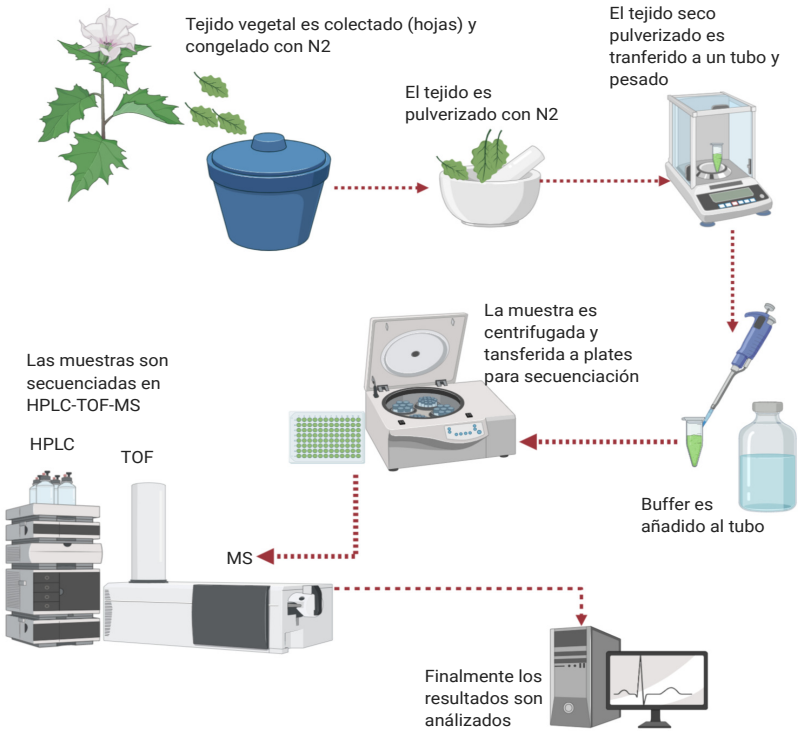


Fig. 1. Procedimiento general de extracción, secuenciación y análisis de los metabolitos secundarios (*e. g.*, alcaloides) utilizado en este estudio. N2 = nitrógeno liquido.

Secuenciación Masiva de DNA. La secuenciación de próxima generación se trata de un tipo de secuenciación masiva paralela que permite obtener información genómica de cientos de miles de moléculas de DNA en un solo ensayo (McCombie *et al.* 2019). En esta tesis se emplearon dos tipos de técnicas de secuenciación masiva del DNA; "double digestion restriction site associated DNA-sequencing" (por sus siglas en inglés, ddRadseq; Peterson *et al.* 2012) así como "Whole Genome sequencing" (por sus siglas en inglés, WGS; McCombie *et al.* 2019). ddRadseq involucra la digestión de dos enzimas de restricción y secuenciación de regiones adyacentes a los sitios de restricción o reconocimiento de las enzimas (Peterson *et al.* 2012). Esta técnica permite obtener la misma cantidad de subconjuntos de fragmentos de DNA a lo largo del genoma para cientos o miles de individuos, brindando cientos o miles de marcadores moleculares (por sus siglas en inglés, single-nucleotide polymorphisms; SNPs) (Peterson *et al.* 2012). Por otro lado, WGS es un método que permite secuenciar genomas completos (McCombie *et al.* 2019). La diferencia entre ddRadseq y WGS es que el primer método busca obtener una secuenciación reducida del genoma (~15% del genoma como máximo), mientras que el segundo llega a cubrir todo el genoma (Peterson *et al.* 2012; McCombie *et al.* 2019). Existen distintas plataformas de secuenciación masiva, unas como Illumina Hiseq, son capaces únicamente de secuenciar fragmentos de DNA de hasta 300 pb, mientras que otras son capaces de secuenciar hasta 20,000 pb (PacBio) (Peterson *et al.* 2012; McCombie *et al.* 2019) (Fig. 2).

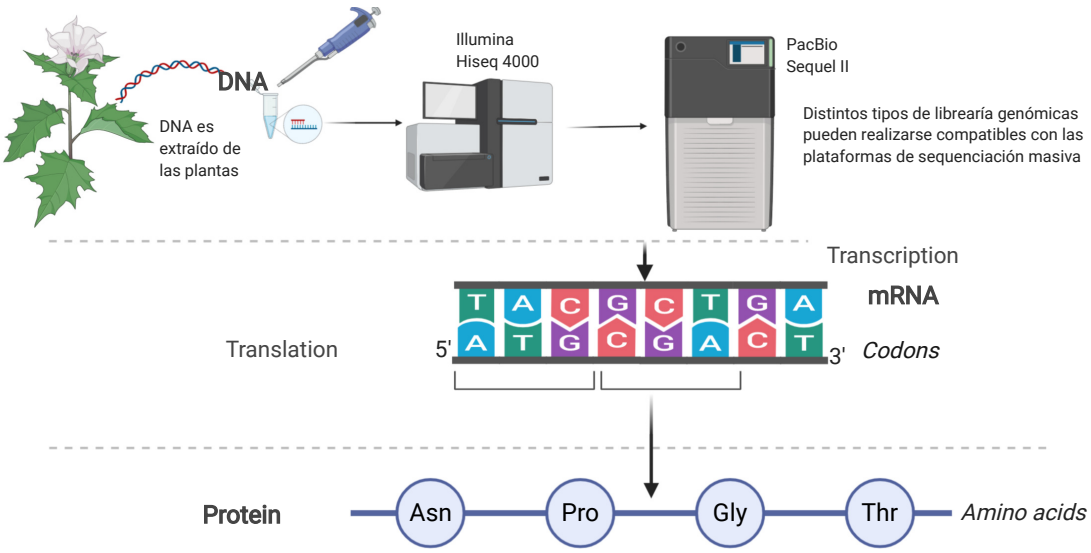


Fig. 2. Extracción y secuenciación del DNA.

Comparaciones Q_{ST}-F_{ST}. La mayoría de las especies eucariotas están divididas dentro de muchas subpoblaciones parcialmente aisladas y dependiendo de la magnitud de la selección natural, deriva génica, migración y mutación, estas subpoblaciones pueden llegar a diferenciarse; tanto genéticamente como fenotípicamente a través del tiempo (Wright 1965; Holsinger & Weir 2009; Leinonen *et al.* 2013). A nivel genético, el grado de diferenciación genética puede medirse usando el estadístico F_{ST} (Wright 1965; Holsinger & Weir 2009). El cual es una medida estandarizada de diferenciación genética entre las poblaciones para un locus genético (Leinonen *et al.* 2013). Los valores toman de 0 a 1; altos valores de F_{ST} indican un alto grado de diferenciación entre las poblaciones (Wright 1965). Esta medida es parte de los estadísticos *F* que han sido usados para analizar la estructura genética de las poblaciones (Wright 1965). Otros estadísticos *F* incluyen; F_{IS} (coeficiente de endogamia; valores positivos indican mayor endogamia, valores negativos excesos de heterocigotos; Wright 1965) y F_{IT}, el cual es el coeficiente de endogamia general de un individuo relativo al total de la población (Wright 1965; Holsinger & Weir 2009).

Por otra parte, el análogo genético de F_{ST} para caracteres cuantitativos es Q_{ST}, el cual mide la cantidad de varianza genética entre poblaciones relativa al total de varianza genética en un rasgo (en vez de en un locus como F_{ST}; Spitze 1993; Merilä & Crnokrak 2001; Leinonen *et al.* 2013). Debido a que Q_{ST} requiere datos genéticos cuantitativos de múltiples poblaciones; experimentos de cruzas para generar múltiples familias génicas, los cuales pueden llegar a demandar grandes cantidades de tiempo y esfuerzo, es posible usar una alternativa, P_{ST}, el cual tiene como base únicamente datos fenotípicos (ver DeFaveri & Merilä 2013). P_{ST} se vuelve más confiable cuándo los datos fenotípicos son medidos en condiciones controladas (*e. g.*, experimentos de jardín común en invernadero; Leinonen *et al.* 2013).

El valor de Q_{ST} para un fenotipo cuantitativo neutral que tiene una base genética aditiva es esperado en ser igual a F_{ST} para un locus neutral (Leinonen *et al.* 2013). De tal manera que sí F_{ST} es medida a partir de marcadores moleculares neutrales (*e. g.*, microsatellites), puede ser usada como una expectación nula para el grado de divergencia poblacional debido a la deriva y migración (Merilä & Crnokrak 2001; Leinonen *et al.* 2013). En los casos en el cual $Q_{ST}/P_{ST} \approx F_{ST}$, la inferencia es que la divergencia del rasgo entre subpoblaciones podría haber sido derivada únicamente por deriva génica (Merilä & Crnokrak 2001). Sí $Q_{ST}/P_{ST} > F_{ST}$, la divergencia del rasgo excede la neutralidad esperada y se infiere que esta divergencia pudo haber sido causada por selección direccional (Merilä & Crnokrak 2001). Sí $Q_{ST}/P_{ST} < F_{ST}$, la divergencia del rasgo entre poblaciones es menor que la esperada únicamente por deriva génica, entonces este patrón sugiere selección uniforme o estabilizadora a lo largo de las poblaciones (Merilä & Crnokrak 2001).

Identidad por Descendencia (por sus siglas en inglés; IBD). Un segmento de ADN es idéntico por estado (Identidad por Estado*;* IBS) en dos o más individuos si tienen secuencias de nucleótidos idénticas en este segmento (Thomson 2013). Un segmento de IBS es idéntico por descendencia (IBD) en dos o más individuos si lo han heredado de un ancestro común, es decir, el segmento tiene el mismo origen ancestral en estos individuos (Thomson 2013). Todos los individuos en una población finita están relacionados si se remontan al tiempo suficiente y, por lo tanto, compartirán segmentos de sus genomas (IBD) (Thomson 2013). El gen de IBD se rastrea a través de meiosis ancestrales y se define en relación con los fundadores de un pedigrí, o para un cierto tiempo ú origen mutacional en la coalescencia de un conjunto de genes existentes en una población (Thomson 2013).

Experimentos de jardín común y de transplantes recíprocos. Dilucidar adaptación y las bases genéticas de los rasgos adaptativos puede resultar ser difícil de estudiar debido a que la evidencia genética de la adaptación es casi siempre enmascarada por los efectos pervasivos de fenómenos evolutivos tales como la deriva génica, plasticidad fenotípica, historia demográfica compleja y arquitectura genética compleja (Kawecki & Ebert 2004; Blanquart 2013; Villemereuil *et al.* 2016). En el caso particular de adaptación local, los biólogos evolutivos han desarrollado herramientas para superar estos retos y los experimentos de jardín común y de transplantes recíprocos son unas de ellas (Kawecki & Ebert 2004; Blanquart 2013; Villemereuil *et al.* 2016).

En el caso especifico de los experimentos de jardín común, lo que se busca es controlar los efectos de la plasticidad fenotípica debido a las interacciones genotipo-ambiente (Kawakami *et al.* 2011; Brachi *et al.* 2013). Por ejemplo, plantas de distintas poblaciones pueden ser crecidas en un ambiente común con condiciones ambientales controladas (Villemereuil *et al.* 2016). De tal manera que usando las herramientas de genética cuantitativa como las comparaciones P_{ST}-F_{ST} (*ver arriba*), es posible dilucidar adaptación local de fenotipos complejos a lo largo de varias poblaciones sin los efectos del medio ambiente correspondiente a esas poblaciones ya que la divergencia fenotípica observada entre poblaciones se debe exclusivamente a la genética de los individuos (Kawakami *et al.* 2011; Gonda *et al.* 2011; Brachi *et al.* 2013; Villemereuil *et al.* 2016).

Por otro lado, en los experimentos de transplantes recíprocos se busca mover a los fenotipos entre ambientes (Leimu & Fischer 2008; Hereford 2009; Blanquart 2013), de tal manera que algún componente del fitness es medido para cada fenotipo bajo condiciones en las cuales esos fenotipos coinciden con su ambiente de origen (fenotipos residentes), así como bajo condiciones en las cuales esos fenotipos son expuestos a otros ambientes de otras poblaciones (fenotipos inmigrantes) (Kawecki & Ebert 2004; Hereford 2009; Blanquart 2013). La existencia de una interacción significativa entre el fitness y el ambiente sugiere la presencia de adaptación

local, ya que se espera que cada fenotipo se desempeñe mejor en su ambiente nativo (*i. e.*, cuando es residente) que en el ambiente foráneo (Kawecki & Ebert 2004; Blanquart 2013).

Objetivo general

El objetivo general de esta tesis es determinar experimentalmente la evolución adaptativa de la resistencia contra herbívoros en *Datura stramonium* (Asteridae; Solanales; Solanaceae) usando las herramientas de la genómica, ecología evolutiva, bioinformática, genética cuantitativa y química analítica.

Objetivos particulares

I. Proporcionar un marco teórico que permita comprender la evolución por selección natural de los rasgos de defensa en las plantas terrestres (embriofitas) contra sus enemigos naturales desde los puntos de vista de la genética y de la ecología química (Capitulo I).

II. Demostrar adaptación local de las defensas putativas químicas de *Datura stramonium* mediante un experimento de jardín común en condiciones ambientales controladas, usando herramientas de genética cuantitativa (*i. e.*, P_{ST} *vs.* Q_{ST}) y química analítica (HPLC-TOF-MS) (Capitulo II).

III. Demostrar a que grado la evolución de las defensas químicas de *Datura stramonium* han sido impulsadas por la selección natural impuesta por sus herbívoros especialistas. Para este fin, se usó un experimento de trasplantes recíprocos, genómica y bioinformática (ddRadseq usada para calcular IBD en plantas de *Datura stramonium*) así como química analítica (HPLC-TOF-MS) (Capitulo III).

IV. Demostrar la evolución adaptativa de la defensa en contra de los enemigos naturales en el genoma de *Datura stramonium.* Para este fin, se usó genómica y bioinformática (WGS; ensamble y anotación del genoma, así como genómica comparativa con otras 11 especies de Solanáceas) así como química analítica (HPLC-TOF-MS) (Capitulo VI).

Sistema de estudio

Datura stramonium (Asteridae; Solanales; Solanaceae) es una hierba anual cosmopolita de origen tropical, que se reproduce ampliamente en todo el mundo (Núñez-Farfán 1994). Algunas plantas pueden alcanzar hasta 2 m de altura y se propaga exclusivamente por semillas (Núñez-Farfán 1994). En México, esta planta herbácea habita en sitios abiertos, cultivados y perturbados (Núñez-Farfán 1994). Esta hierba se ramifica dicotómicamente, cada bifurcación del tallo dicotómico tiene un brote floral (Núñez-Farfán 1994). Sus flores son tubulares, de color violáceo o blanco. Esta especie produce flores hermafroditas con variación en la distancia antera-estigma la cual que se ha demostrado que influye en la tasa de cruzamiento (Motten & Antonovics 1992; Motten & Stone 2000). De hecho, *D. stramonium* produce la mayoría de sus semillas por autofecundación (81.3-91.7%; Motten & Antonovics, 1992), pero la tasa de producción de semillas por cruzamiento puede variar dentro y entre poblaciones (Motten & Antonovics 1992; Motten & Stone 2000).

El fruto de *D. stramonium* es una cápsula espinosa dehiscente con cuatro válvulas y muchas semillas obscuras y reniformes (Núñez-Farfán 1994). Los alcaloides tropano y los tricomas de las hojas son componentes de la resistencia a la herbivoría y patógenos en *D. stramonium* (Shonle y Bergelson 2000; Valverde *et al.* 2001, Castillo *et al.* 2013, 2014, De-la-Cruz *et al.* 2020), mientras que también se ha observado la capacidad de tolerar el daño por herbívoros a través de invertir recursos en el crecimiento o producción de hojas (Fornoni & Núñez-Farfán 2000). De hecho, se ha reportado que, en algunas poblaciones de *Datura*, las plantas pueden utilizar de manera simultanea estrategias mixtas como la tolerancia y la resistencia para enfrentar a sus herbívoros (Valverde *et al.* 2003). Se ha detectado selección natural sobre su resistencia, tolerancia y en especifico sobre dos alcaloides tropanos y tricomas foliares (Castillo *et al.* 2014; Miranda-Pérez *et al.* 2016), variando, la fuerza y dirección de la selección entre poblaciones (Castillo *et al.* 2014). Casi todas las poblaciones de *D. stramonium* están infestadas principalmente por los escarabajos folívoros *Lema daturaphila* y *Epitrix párvula*, ambos pertenecientes a la familia Chrysomelidae, así como por el escarabajo depredador de semillas *Trichobaris soror* (Curculionidae) (Castillo *et al.* 2013). Otros herbívoros generalistas de *D. stramonium* comprenden *Sphenarium purpurascens* (Pyrgomorphidae) y *Manduca sp.* (Sphingidae), así como especies del género Helicoverpa (Noctuidae) y de la familia Saturniidae (Núñez-Farfán 1994).

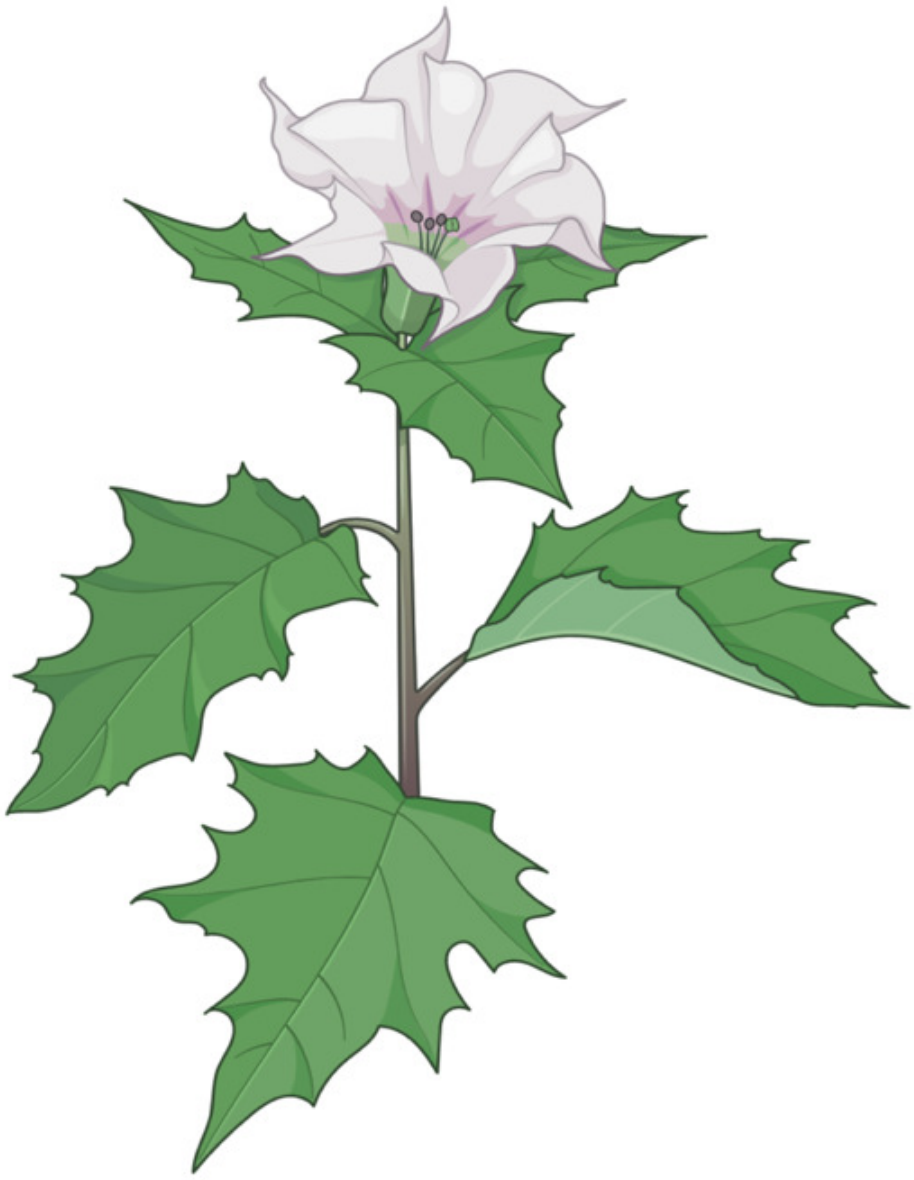


Fig. 3. *Datura stramonium*

Estructura de la tesis

Capitulo I

What do we know about the genetic basis of plant defensive responses to herbivores? A minireview

I. M. De-la-Cruz, S. Velázquez-Márquez, J. Núñez-Farfán

Publicado en el libro Evolutionary Ecology of Plant-Herbivore Interactions editado por Springer Nature.

Síntesis. En este capitulo, presento una revisión sobre que se conoce sobre las bases genéticas de las respuestas defensivas de las plantas a los herbívoros. En esta revisión me enfoqué en algunos ejemplos fascinantes que describen las bases genéticas de la defensa en plantas; (1) Se discute como los herbívoros pueden manejar los rasgos defensivos en las plantas, (2) Enfatizo en las distintas respuestas genéticas que las plantas despliegan en función a sus distintos estreses ambientales (bióticos y abióticos), (3) Discuto el papel de los elicitores en la interacción planta-herbívoro. (4) Destaco la importancia de las fitohormonas, factores de transcripción y mecanismos epigenéticos que regulan las respuestas genéticas defensivas. (5) Propongo algunas direcciones futuras en el estudio de la interacción planta-herbívoro a la luz de la genómica.

Contribuciones. I. M. De-la-Cruz concibió la idea, desarrollo la investigación, diseño las figuras y escribió el manuscrito. J. Núñez-Farfán, S. Velázquez-Márquez contribuyeron en la edición del texto y figuras.

Capitulo II

Evolutionary response to herbivory: population differentiation in microsatellite loci, tropane alkaloids and leaf trichome density in *Datura stramonium*

Publicado en Arthropod-Plant Interactions.

I. M. De-la-Cruz, L. L. Cruz, L. Martínez-García, P. L. Valverde, C. M. Flores-Ortiz, L. B. Hernández-Portilla, J. Núñez-Farfán

Síntesis. En el capítulo número dos se estudiaron plantas de dos poblaciones de *D. stramonium* (Teotihuacán, Estado de México y Ticumán, Morelos), crecidas bajo condiciones ambientales controladas mediante un experimento de jardín común. Se midieron defensas putativas químicas (21 tropano alcaloides) así como los tricomas foliares de cada individuo. Asimismo, se analizó la diferenciación genética entre poblaciones usando loci neutrales (microsatellites). Esto permitió detectar adaptación local en los caracteres putativos químicos de esta especie. Para

lograr esto, comparamos la magnitud de diferenciación entre poblaciones en la defensa de las plantas en contra de sus herbívoros (selección) y en loci neutrales (deriva génica); P_{ST} *vs.* F_{ST}.

Contribuciones. I. M. De-la-Cruz, J. Núñez-Farfán concibieron la idea. I. M. De-la-Cruz, L. L. Cruz, L. Martínez-García realizaron el experimento. C. M. Flores-Ortiz, L. B. Hernández-Portilla brindaron acceso y soporte técnico para el HPLC-TOF-MS. I. M. De-la-Cruz realizó los análisis. I. M. De-la-Cruz, P. L. Valverde, J. Núñez-Farfán escribieron el manuscrito. J. Núñez-Farfán financió el estudio.

Capitulo III

Local adaptation driven by herbivore-imposed natural selection in *Datura stramonium*:

genetic and chemical evidence

I. M. De-la-Cruz, J. Merilä, P. L. Valverde, C. M. Flores-Ortiz, J. Núñez-Farfán

Síntesis. En el capitulo número tres, se presentan los resultados de los experimentos en campo de transplantes recíprocos. Para este estudio se produjo una progenie de generación F_2 derivada de la cruza de dos parentales contrastantes en sus niveles de alcaloides, los cuales fueron seleccionados a partir del experimento descrito en el capitulo II (un parental de Teotihuacán y otro parental de Ticumán). Esta progenie fue trasplantada en las localidades de los parentales (Teotihuacán y Ticumán). Se obtuvo la identidad por descendencia entre cada planta F_2 y cada parental (Teotihuacán/Ticumán) y se analizaron los siete alcaloides más abundantes de *D. stramonium*. También, se midieron los niveles de infestación de los herbívoros especialistas en cada planta en ambas localidades. Se hicieron análisis de selección fenotípica sobre los alcaloides, sobre la resistencia y sobre la IBD. De tal manera que detectamos cuales son los alcaloides que confieren resistencia a *D. stramonium*, específicamente que alcaloides están funcionando como defensa para cierto tipo de herbívoros, pero no para otros. Detectamos cual es el herbívoro que ejerce mayor presión selectiva sobre las plantas, así como pudimos observar como la selección natural maneja la resistencia química en *D. stramonium* cuando es atacada por múltiples herbívoros especialistas y, como las plantas están adaptadas localmente a estas presiones selectivas por parte de los herbívoros. Más aún, pudimos detectar las bases genéticas de la resistencia contra los herbívoros especialistas en *D. stramonium.*

Contribuciones. I. M. De-la-Cruz, J. Núñez-Farfán concibieron la idea. I. M. De-la-Cruz, J. Núñez-Farfán, P. L. Valverde realizaron los experimentos. C. M. Flores-Ortiz brindó acceso y soporte técnico para el HPLC-TOF-MS. I. M. De-la-Cruz, J. Merilä realizaron los análisis. I. M.

De-la-Cruz, J. Merilä, J. Núñez-Farfán escribieron el manuscrito. J. Núñez-Farfán financió el estudio.

Capitulo IV

Genomic signatures of the evolution of defence against its natural enemies in the poisonous and medicinal plant *Datura stramonium* (Solanaceae)

I. M. De-la-Cruz, A. Hallab, U. Olivares, R. Tapia-López, S. Velázquez-Márquez, D. Piñero, K. Oyama, B. Usadel, J. Núñez-Farfán

Síntesis. En el capitulo cuatro, se presentan los dos primeros genomas ensamblados y anotados de *D. stramonium* así como se hace un análisis extensivo de genómica comparativa con 11 especies de Solanáceas y se investiga a detalle la evolución de ocho genes involucrados en la producción de tropano alcaloides. Estos genomas corresponden a los abuelos parentales que fueron seleccionados para producir la progenie de generación F_2 utilizada en el capitulo III de esta tesis. El genoma de *D. stramonium*, es un genoma de tamaño grande y complejo, compuesto mayormente por elementos repetidos. Se obtuvieron hallazgos importantes tales como genes expandidos, positivamente seleccionados y con divergencia fisicoquímica en *D. stramonium* que revelan la evolución y adaptación de la defensa en contra de sus enemigos naturales. Estos genes pertenecen a la familia de genes *R* (genes de resistencia). Asimismo, los genes positivamente seleccionados y expandidos o con divergencia fisicoquímica están involucrados en la producción de metabolitos secundarios involucrados en defensa en contra de patógenos como herbívoros, bacterias, virus, hongos. Esta investigación también revela la diferenciación en la arquitectura de dominio en varios de los genes involucrados en la ruta de los alcaloides tropanos entre ambos parentales de *D. stramonium*, los cuales fueron seleccionados por su alta diferenciación en la producción de alcaloides.

Contribuciones. I. M. De-la-Cruz, A. Hallab concibieron la idea. I. M. De-la-Cruz, R. Tapia-López, realizaron los experimentos. I. M. De-la-Cruz, A. Hallab realizaron los análisis. U. Olivares, S. Velázquez-Márquez, D. Piñero, K. Oyama, B. Usadel, J. Núñez-Farfán brindaron asistencia en logística, financiación, metodología. I. M. De-la-Cruz, J. Núñez-Farfán, A. Hallab escribieron el manuscrito. J. Núñez-Farfán financió el estudio.

Referencias

Bennett, R. N., & Wallsgrove, R. M. (1994). Secondary metabolites in plant defence mechanisms. *New Phytologist*, *127*(4), 617–633.

Blanquart, F., Kaltz, O., Nuismer, S. L., & Gandon, S. (2013). A practical guide to measuring local adaptation. *Ecology Letters*, *16*(9), 1195–1205.

Berenbaum, M. R. (1990). Coevolution between herbivorous insects and plants: tempo and orchestration. In F. Gilbert (Ed.), *Insect Life Cycles* (pp. 87–99). Springer.

Bostock, R. M. (2005). Signal crosstalk and induced resistance: Straddling the line between cost and benefit. *Annual Review of Phytopathology*, *43*(1), 545–580.

Brachi, B., Villoutreix, R., Faure, N., Hautekèete, N., Piquot, Y., Pauwels, M., Roby, D., Cuguen, J., Bergelson, J., & Roux, F. (2013). Investigation of the geographical scale of adaptive phenological variation and its underlying genetics in *Arabidopsis thaliana*. *Molecular Ecology*, *22*(16), 4222–4240.

Castillo, G., Cruz, L. L., Hernández-Cumplido, J., Oyama, K., Flores-Ortiz, C. M., Fornoni, J., Valverde, P. L., & Núñez-Farfán, J. (2013). Geographic association and temporal variation of chemical and physical defense and leaf damage in *Datura stramonium*. *Ecological Research*, *28*(4), 663–672.

Castillo, G., Cruz, L. L., Tapia-López, R., Olmedo-Vicente, E., Carmona, D., Anaya-Lang, A. L., Fornoni, J., Andraca-Gómez, G., Valverde, P. L., & Núñez-Farfán, J. (2014). Selection mosaic exerted by specialist and generalist herbivores on chemical and physical defense of *Datura stramonium*. *PLOS ONE*, *9*(7), e102478.

Villemereuil, P., Gaggiotti, O. E., Mouterde, M., & Till-Bottraud, I. (2016). Common garden experiments in the genomic era: New perspectives and opportunities. *Heredity*, *116*(3), 249–254.

DeFaveri, J., & Merilä, J. (2014). Local adaptation to salinity in the three-spined stickleback? *Journal of Evolutionary Biology*, *27*(2), 290–302.

De-la-Cruz, I. M., Cruz, L. L., Martínez-García, L., Valverde, P. L., Flores-Ortiz, C. M., Hernández-Portilla, L. B., & Núñez-Farfán, J. (2020). Evolutionary response to herbivory: Population

differentiation in microsatellite loci, tropane alkaloids and leaf trichome density in *Datura stramonium*. *Arthropod-Plant Interactions*, *14*(1), 21–30.

Dethier, V. G. (1954). Evolution of feeding preferences in phytophagous insects. *Evolution*, *8*(1), 33–54.

Dyer, L. A., Philbin, C. S., Ochsenrider, K. M., Richards, L. A., Massad, T. J., Smilanich, A. M., Forister, M. L., Parchman, T. L., Galland, L. M., Hurtado, P. J., Espeset, A. E., Glassmire, A. E., Harrison, J. G., Mo, C., Yoon, S., Pardikes, N. A., Muchoney, N. D., Jahner, J. P., Slinn, H. L., … Jeffrey, C. S. (2018). Modern approaches to study plant–insect interactions in chemical ecology. *Nature Reviews Chemistry*, *2*(6), 50–64.

Ehrlich, P. R., & Raven, P. H. (1964). Butterflies and plants: a study in coevolution. *Evolution*, *18*(4), 586–608.

Ferrer, I., & Thurman, E. M. (2008). Existence and use of accurate mass databases and elemental composition tools for target and non-target analyses. In *Liquid Chromatography Time-of-Flight Mass Spectrometry* (pp. 57–70). John Wiley & Sons, Ltd.

Fraenkel, G. S. (1959). The raison d'Être of secondary plant substances: these odd chemicals arose as a means of protecting plants from insects and now guide insects to food. *Science*, *129*(3361), 1466–1470.

Fornoni, J., & Núñez-Farfán, J. (2000). Evolutionary ecology of *Datura stramonium*: genetic variation and costs for tolerance to defoliation. *Evolution*, *54*(3), 789–797.

Gatehouse, J. A. (2002). Plant resistance towards insect herbivores: A dynamic interaction. *New Phytologist*, *156*(2), 145–169.

Giron, D., Dubreuil, G., Bennett, A., Dedeine, F., Dicke, M., Dyer, L. A., Erb, M., Harris, M. O., Huguet, E., Kaloshian, I., Kawakita, A., Lopez-Vaamonde, C., Palmer, T. M., Petanidou, T., Poulsen, M., Sallé, A., Simon, J.-C., Terblanche, J. S., Thiéry, D., … Pincebourde, S. (2018). Promises and challenges in insect–plant interactions. *Entomologia Experimentalis et Applicata*, *166*(5), 319–343.

Gonda, A., Herczeg, G., & Merilä, J. (2011). Population variation in brain size of nine-spined sticklebacks (*Pungitius pungitius*)-Local adaptation or environmentally induced variation? *BMC Evolutionary Biology*, *11*(1), 75.

Hereford, J., Elle, A. E. E., & Geber, E. M. A. (2009). A quantitative survey of local adaptation and fitness trade-offs. *The American Naturalist*, *173*(5), 579–588.

Howe, G. A., & Jander, G. (2008). Plant immunity to insect herbivores. *Annual Review of Plant Biology*, *59*, 41–66.

Jorge, T. F., Mata, A. T., & António, C. (2016). Mass spectrometry as a quantitative tool in plant metabolomics. *Philosophical Transactions. Series A, Mathematical, Physical, and Engineering Sciences*, *374*(2079).

Karban, R., & Baldwin I. T. (1997) Induced responses to herbivory. University of Chicago Press, Chicago, IL.

Kawakami, T., Morgan, T. J., Nippert, J. B., Ocheltree, T. W., Keith, R., Dhakal, P., & Ungerer, M. C. (2011). Natural selection drives clinal life history patterns in the perennial sunflower species, Helianthus maximiliani. *Molecular Ecology*, *20*(11), 2318–2328.

Kawecki, T. J., & Ebert, D. (2004). Conceptual issues in local adaptation. *Ecology Letters*, *7*(12), 1225–1241

Kessler, A., & Baldwin, I. T. (2002). Plant responses to insect herbivory: the emerging molecular analysis. *Annual Review of Plant Biology*, *53*(1), 299–328.

Leimu, R., & Fischer, M. (2008). A meta-analysis of local adaptation in plants. *PLOS ONE*, *3*(12), e4010.

Leinonen, T., McCairns, R. J. S., O'Hara, R. B., & Merilä, J. (2013). Q_{ST}-F_{ST} comparisons: Evolutionary and ecological insights from genomic heterogeneity. *Nature Reviews. Genetics*, *14*(3), 179–190.

McCombie, W. R., McPherson, J. D., & Mardis, E. R. (2019). Next-Generation Sequencing technologies. *Cold Spring Harbor Perspectives in Medicine*, *9*(11), a036798.

Merilä, J., & Crnokrak, P. (2001). Comparison of genetic differentiation at marker loci and quantitative traits. *Journal of Evolutionary Biology*, *14*(6), 892–903.

Miranda-Pérez, A., Castillo, G., Hernández-Cumplido, J., Valverde, P. L., Borbolla, M., Cruz, L. L., Tapia-López, R., Fornoni, J., Flores-Ortiz, C. M., & Núñez-Farfán, J. (2016). Natural selection drives chemical resistance of *Datura stramonium*. *PeerJ*, *4*, e1898.

Motten, A. F., & Stone, J. L. (2000). Heritability of stigma position and the effect of stigma-anther separation on outcrossing in a predominantly self-fertilizing weed, *Datura stramonium* (Solanaceae). *American Journal of Botany*, *87*(3), 339–347.

Motten, A. F., & Antonovics J. (1992). Determinants of outcrossing rate in a predominantly self-fertilizing weed, *Datura stramonium* (Solanaceae). *American Journal of Botany* 79: 419–427.

Núñez-Farfán, J., & Dirzo, R. (1994). Evolutionary ecology of *Datura Stramonium* L. in Central Mexico: natural selection for resistance to herbivorous insects. *Evolution*, *48*(2), 423–436.

Núñez-Farfán, J., Fornoni, J., & Valverde, P. L. (2007). The evolution of resistance and tolerance to herbivores. *Annual Review of Ecology, Evolution, and Systematics*, *38*(1), 541–566.

Peterson, B. K., Weber, J. N., Kay, E. H., Fisher, H. S., & Hoekstra, H. E. (2012). Double digest RADseq: an inexpensive method for *de novo* SNP discovery and genotyping in model and non-model species. *PLOS ONE*, *7*(5), e37135.

Rosenthal, J. P., & Kotanen, P. M. (1994). Terrestrial plant tolerance to herbivory. *Trends in Ecology & Evolution*, *9*(4), 145–148.

Schuman, M. C., & Baldwin, I. T. (2018). Field studies reveal functions of chemical mediators in plant interactions. *Chemical Society Reviews*, *47*(14), 5338–5353.

Shonle, I., & Bergelson, J. (2000). Evolutionary ecology of the tropane alkaloids of *Datura Stramonium* L. (solanaceae). *Evolution*, *54*(3), 778–788.

Spitze, K. (1993). Population structure in *Daphnia obtusa*: Quantitative genetic and allozymic variation. *Genetics*, *135*(2), 367–374.

Strauss, S. Y., Agrawal, A. A., Strauss, S. Y., & Agrawal, A. A. (1999). The ecology and evolution of plant tolerance to herbivory. *Trends in Ecology & Evolution*, *14*(5), 179–185.

Ueda, H., Kikuta, Y., & Matsuda, K. (2012). Plant communication: Mediated by individual or blended VOCs? *Plant Signaling & Behavior*, *7*(2), 222.

Valverde, P. L., Fornoni, J., & Núñez-Farfán, J. (2003). Evolutionary ecology of *Datura stramonium*: Equal plant fitness benefits of growth and resistance against herbivory. *Journal of Evolutionary Biology*, *16*(1), 127–137.

Verhage, A., van Wees, S. C. M., & Pieterse, C. M. J. (2010). Plant Immunity: It's the hormones talking, but what do they say? *Plant Physiology*, *154*(2), 536–540.

Wright, S. (1965). The interpretation of population structure by *F*-statistics with special regard to systems of mating. *Evolution*, *19*(3), 395–420.

Wu, J., & Baldwin, I. T. (2010). New insights into plant responses to the attack from insect herbivores. *Annual Review of Genetics*, *44*(1), 1–24.

Xu, S., Zhou, W., Pottinger, S., & Baldwin, I. T. (2015). Herbivore associated elicitor-induced defences are highly specific among closely related *Nicotiana* species. *BMC Plant Biology*, *15*(1), 2.

CHAPTER 1

Review article

2020

WHAT DO WE KNOW ABOUT THE GENETIC BASIS OF PLANT DEFENSIVE RESPONSES TO HERBIVORES? A MINIREVIEW

This manuscript is published in ***Evolutionary Ecology of Plant-Herbivore Interaction Edited by Springer Nature*** https://doi.org/10.1007/978-3-030-46012-9_16

Format for the book Evolutionary Ecology of Plant-Herbivore Interactions Edited by Springer Nature. Eds. Juan Núñez Fárfan and Pedro Luis Valverde Padilla

Chapter 16

What Do We Know About the Genetic Basis of Plant Defensive Responses to Herbivores? A Minireview

Ivan M. De-la-Cruz[1], Sabina Velázquez-Márquez[1] and Juan Núñez-Farfán[1*]

Abstract. Terrestrial plants are frequently attacked by herbivores and pathogens (bacteria, virus, fungi, oomycete) and have evolved constitutive and induced defences to prevent/diminish fitness costs. Here, we review recent progress in the study of the defence genes in plants. The sophisticated signalling network of plant defence responses is elicited and driven by both herbivore-induced factors (*e. g.*, elicitors, effectors, and wounding) and plant signalling (*e. g.*, phytohormone and plant volatiles) in response to arthropod factors. Genome-wide data offer many advantages over sparser sets of genetic markers. It is now possible to detect selection across the genome and detect if those selected genes are associated with the herbivory. Genomic tools are now allowing genome-wide studies, and recent theoretical advances can help to design research strategies that combine genomics and field experiments to examine the genetics of local adaptation (cf. Savolainen *et al*. 2013). Plant and arthropod genomics provide many opportunities to understand the plant immunity to arthropod herbivores. Also, it will provide new insights into basic mechanisms of chemical communication and plant-animal coevolution and may also facilitate new approaches to crop protection and improvement.

Keywords. Crosstalk, herbivory, local adaptation, phytohormonal, plant defences, quantitative trait loci.

I. M. De-la-Cruz • S. Velázquez-Márquez • J. Núñez-Farfán (✉)
[1]Departamento de Ecología Evolutiva, Instituto de Ecología,
Universidad Nacional Autónoma de México (UNAM),
Ciudad de México
*Corresponding author: farfan@unam.mx

The evolutionary history of terrestrial plants (*i. e.*, Embryophyte clade) and their arthropod associates are inextricably linked (Smith and Clement, 2012), since plants are the food source of nearly one million or more insect species from diverse taxonomic groups (Howe and Jander, 2008). Because plants are sessile organisms and have no chance to escape from the attack of herbivores, they evolved particular strategies to defend themselves (Mithöfer and Boland, 2012). Likewise, phytophagous arthropods have evolved ways to overcome plant defences (Futuyma and Agrawal, 2009). This coevolutionary relationship is based on land plants and herbivores continually co-adapting to changing environments and biotic pressures (Futuyma and Agrawal, 2009).

Terrestrial plants traits that confer resistance to insect pests can be classified according to the manner in which they are deployed (Howe and Jander, 2008). Some traits are expressed constitutively under the control of hard-wired developmental programs, irrespective of the herbivore threat level and if the herbivores are present or not (Howe and Jander, 2008) (Figure 1 and 2). In contrast, other plants defences, including some toxins, are induced, while other defences are only expressed in response to herbivory and at the site of tissue damage. In many cases, defences are produced systemically in undamaged tissues (Bostock, 2005) (Figure 1 and 2).

Terrestrial plant induced defences are a sophisticated defence system that can recognize the non-self-molecules or signals from damaged cells, much like the animals, activating the plant's immune response against herbivores (Hare, 2010; Howe and Jander, 2008; Verhage *et al*. 2010). Induced defence in terrestrial plants is mediated in general, by the recognition of specific cues, for instance herbivore-associated molecular patterns (HAMPs) in their oral secretions, followed by the elicitation of complex signalling networks, involving mitogen- activated protein kinase (MAPK) cascades, as well as signalling via the Jasmonic acid, salycilic acid (JA, SA), and ethylene pathways (Campos *et al*. 2014) (Figure 1, 2 and 3). This signalling, in turn, leads to a reconfiguration of the transcriptome and proteome,

as well as the biosynthesis of defence chemicals (Wu and Baldwin, 2010) (Figure 2). Induced and constitutive defence can be displayed directly as physical barriers (tissue toughness, plant pubescence, and glandular and nonglandular trichomes) or allelochemicals in plant tissues exhibiting antifeedant, toxic, or repellent effects on herbivores (Mithöfer and Boland, 2012) (Figure 1, 2 and 3).

Terrestrial plant defenses also can be displayed indirectly as volatile organic compounds (VOCs) and green leaf volatiles (GLVs) (Figure 1). These compounds are released by herbivore-damaged plants that attract arthropod predators and parasitoids, or that may repel oviposition of pest arthropods (Baldwin *et al*. 2006; Kessler and Baldwin, 2001, 2002). For example, HAMPs are specific plant indirect defence responses to specific herbivore-derived elicitors in oral or ovipositor secretions, that facilitate indirect defences against herbivores (Mithöfer and Boland, 2012) (Figure 1 and 2). The most-studied HAMPs are insect fatty acid-plant amino acid conjugates from lepidoptera larvae (Halitschke *et al*. 2001; Kessler and Baldwin, 2002; Schmelz *et al*. 2012).

In summary, constitutive and induced defences in terrestrial plants are regulated by complex physiological, biochemical and molecular processes, such as chemical signal cascades involving jasmonic acid (JA), salicylic acid (SA), ethylene, abscisic acid, and gibberellic acid that result in a downstream production of direct and indirect defences (Smith and Clement, 2011) (Figure 2). Despite the wealth of information on plant defence to natural enemies, our understanding of plants communication with plant neighbours, symbionts, pathogens, herbivores, and herbivores' natural enemies, above and below ground, is still limited (War *et al*. 2012). Much of the understanding of the molecular mechanisms and evolutionary origins of immune recognition in plants derives mainly from studies of plant-pathogen (bacteria, virus, fungi, oomycete) interactions (Howe and Jander, 2008; Jones and Dangl, 2006) and several other studies of plant-herbivore interactions in model plants (Howe and Jander, 2008). In particular, there is little information about the genetic basis of plant defence traits for insect herbivores (Figure 3).

In this review we focus on some fascinating examples of the genetic basis of defence in terrestrial plants. We highlight the recent advances in the understanding of some defence genes and (1) discuss how herbivores can drive the genetic evolution of defence traits in plants; 2) we emphasize the different gene responses to a combination of stressors, such as drought + herbivory, and the genetic architecture of defence traits (*i. e*., whether few or many genes do explain a high fraction of phenotypic variance in plant defence); 3) we also discuss on the role of elicitors in the plant-herbivore interaction; 4) and on the role of the phytohormones, transcription factors and epigenetic mechanisms that regulate the defence genetic responses; 5) finally, we point out some future directions in the study of plant-herbivore interactions in the light of genomics.

Herbivores as Selective Agents on Defensive Genes

Geographic analyses of genetic variation in several plant species indicate clear genetic signals of local adaptation (Linhart and Grant, 1996) caused by spatial differences in selection (Züst *et al*. 2012; Johnson, 2018). Insects use different feeding strategies to obtain nutrients from all above and belowground plant parts (Howe and Jander, 2008). Although all phytophagous insects inflict mechanical damage on plant tissues, the quantity and quality of injury varies greatly depending on the feeding tactic (Howe and Jander, 2008). Despite evidence indicates that climate and soil variability can exert strong local selective pressures and play essential roles in shaping large-scale plant genetic patterns (Hancock *et al*. 2011), there is less direct evidence that biotic forces, such as herbivory or competition, can lead to the maintenance of genetic variation across broad geographic scales (Howe and Jander, 2008).

In *Arabidopsis thaliana* (Brassicaceae) *GS-ELONG* locus constitutes an insect resistance QTL caused by variation in glucosinolate quantity, quality or both (Kroymann *et al*. 2003). Züst *et al*. (2012) studied alleles of the *GS-ELONG* and *GS-AOP* loci that mechanistically determine the accumulation and structure of aliphatic glucosinolates. Aliphatic glucosinolates are the first defence in *A. thaliana,* and the *GS-ELONG* locus regulates the carbon side-chain elongation (*3C* or *4C*) (Kroymann *et al*. 2003), whereas the *GS-AOP* locus modifies the functional group of the biologically active glucosinolate side chain (*ALK*, *OH*, or *NULL*). The combination of these alleles yields six distinct chemotypes present in natural populations in varying proportions (Züst *et al*. 2012). Züst *et al*. (2012) mapped the geographic variation in the abundance of six *A. thaliana* chemotypes within Europe, from a set of 96 accessions with known chemical profiles, and found that the frequency of *3C* to *4C* chemotypes of *GS-ELONG* locus increased with latitude and longitude, and that the herbivores *Brevicoryne brassicae* (Aphididae) and *Lipaphis erysimi*

(Aphididae) drive these chemical patterns as both are abundant. This study is a strong evidence that the magnitude and direction of selection exerted by two specialist aphids on *GS-ELONG* drives the changes in the chemistry of *Arabidopsis* at least across Europe.

A GWAS (genome-wide analysis association) study of herbivory in *A. thaliana* studied the genetics of *Pieris rapae* (Pieridae)-host herbivory (Nallu *et al*. 2018), resulting in a total of 90 associated SNPs in linkage disequilibrium with 389 genes. A subset of 12 well-supported candidate genes contained three or more associated SNPs (single nucleotide polymorphisms). Eight of these genes were functionally validated using mutants. This validated gene set includes both well- known and novel defence genes. For instance, the cytochrome P450 gene *CYP79B2* is involved in the conversion of tryptophan to indole-3-acetaldoxime, a precursor of indole glucosinolates and indole-3-acetic acid. Indole glucosinolates are essential secondary metabolites used for defence by *Arabidopsis* and other species of Brassicaceae (Nallu *et al*. 2018). They also found that genes *PROPEP1* and *PROPEP3,* which belong to the *At*Pep (endogenous danger peptides) gene family, are associated with activation of danger or HAMPs (herbivore-associated molecular patterns) immunity in plants against both pathogen and herbivore attacks.

Rausher and Huang (2016) investigated for how long particular plant defensive genes have been involved in the coevolutionary process. They assessed the patterns of selection on the defensive gene threonine deaminase (*TD*) in several Solanaceae species. They found that a single copy of *TD* underwent two duplications near the base of the Solanaceae (Rausher and Huang, 2016). One copy retains the housekeeping function, whereas a second copy evolved defensive functions (*TD2*). According to the authors, *TD2* experienced adaptive substitutions for a period of 30-50 My (Rausher and Huang, 2016). They suggested that the most likely explanation for this result is fluctuating herbivore abundances: When herbivores are rare, relaxed selection increases the likelihood that slightly disadvantageous mutations will be fixed by drift; when herbivores are common, increased selection causes the evolution of compensatory adaptive mutations (Rausher and Huang, 2016).

Different Genes Responses to Biotic and Abiotic Stresses Combined, and the Genetic Architecture of the Defensive Traits

Davila-Olivas *et al*. (2017) studied the genetic architecture of plant responses before the attack of different herbivores (*Pieris rapae;* Pieridae and *Plutella xylostella;* Plutellidae), a pathogen (*Botrys cinereal*; Sclerotiniaceae) and drought. Results of this study show that 18 and 32 QTLs are linked to damage by *P. rapae* and *P. xylostella*, respectively. Revealing different genetic architecture in plant responses to different herbivore species (Mao *et al*. 2011).

Regarding gene responses to biotic and abiotic stresses combined, Davila Olivas *et al*. (2016) found, that QTLs 19 and 25 are associated to the response to combined stresses Drought + *Pieris* and *Botrytis* + *Pieris*, respectively. Interestingly, in the response to Drought + *Pieris*, QTL 19 (chromosome 5) contained the transcription factors MYC1 and AT5G50915. The two genes were induced by *P. rapae* infestation and slightly induced by drought (Davila Olivas *et al*. 2016). Natural variation in trichome density in *A. thaliana* is associated with genetic variation in MYC1 (Davila Olivas *et al*. 2016). Thus, this study found that different genetic components control resistance to the two caterpillars and there is limited overlap in the quantitative trait loci (QTLs) underlying resistance to combined stresses by drought plus *P. rapae* or *B. cinerea* plus *P. rapae*.

Thoen *et al*. (2016), investigated the genetic architecture underlying plant responses to 11 single stresses and several of their combinations by phenotyping 350 *Arabidopsis thaliana* accessions by GWA analyses. They found that stress responses that share phytohormonal signaling pathways also share genetic architecture underlying these responses. Thoen *et al*. (2016) also show that for the 30 most significant SNPs in their study, average quantitative trait locus (QTL) effect sizes were larger for dual stresses than for single stresses. Plants appear to deploy broad-spectrum defensive mechanisms influencing multiple traits in response to combined stresses (Thoen *et al*. 2016). The approach of this study provided an unprecedented comprehensive genetic analysis of how plants deal with a wide spectrum of stress conditions.

R Genes Family and *Mi-1.2* Gene are the Only Genes that have been Described in a Gene- for-Gene Interaction Between Plants and their Natural Enemies

In terrestrial plants, resistance (*R*) genes play a key role in their remarkable immune responses (Kourelis and van der Hoorn 2018). *R* alleles are usually dominant that provide full or partial resistance to one or more pathogens (Kourelis and van der Hoorn 2018). *R* genes exist in natural plant populations and have been used by humankind since early crop domestication (Kourelis and van der Hoorn 2018). Selection during domestication favored dominant *R* alleles providing full resistance, but recessive *R* alleles and *R* alleles that provide partial resistance may provide more durable resistance (Kourelis and van der Hoorn 2018). Most identified *R* alleles are polymorphic in plant populations, which led to their initial characterization and use in plant breeding programs. However, individual plants have up to a few hundred *R* gene analogs that make no identified contribution to resistance (Kourelis and van der Hoorn 2018). Many of these *R* gene analogs are also fixed in plant species and are thought to contribute to nonhost resistance (Schulze-Lefert and Panstruga, 2011).

The plant *R*-genes are the only ones that are involved in gene-for-gene interactions with pathogens (bacteria, virus, fungi), and they may undergo coevolutionary arms races in which plant specificity and pathogen virulence or insect infestation, continually adapt in response to each other (Bergelson *et al*. 2001). The *R*-genes evolution is shaped by natural selection for resistance to different insect species, but especially for aphids (Bergelson *et al*. 2001; Howe and Jander, 2008; Michelmore and Meyers, 1998; Smith and Boyko, 2006). Evidence points that the products of *R* genes mediate resistance to phloem-feeding insects in several monocot and dicot crop species (Smith and Boyko, 2006; Howe and Jander, 2008).

In tomato, the *Mi-1.2* gene provides resistance to some isolates of *Macrosiphum euphorbiae* (Aphididae) and *Bemisia tabaci* (Aleyrodidae), but not to *Myzus persicae* (Aphididae) (Nombela *et al*. 2003). *Mi-1.2* confers resistance to multiple species of arthropods and nematodes (Nombela *et al*. 2003). The LRR (leucine-rich repeat) region of *Mi-1.2* signals programmed cell death, and one model proposes a gene-for-gene interaction between *Mi-1.2* and aphid elicitors, similar to plant-pathogen interactions (Hwang and Williamson, 2003). Other studies suggest NBS-LRR (nucleotide binding site-leucine rich repeat) involvement in aphid resistance in other crops (see Smith and Clement, 2012).

Elicitors Induce Defensive Genes

Herbivore-induced defence responses are often specific-different herbivores induce different defence responses in plants - and their specificity is largely mediated by chemical cues (herbivore-associated elicitors: HAEs) in insect oral or oviposition secretions (Xu *et al*. 2015).

One remarkable study regarding HAEs was carried out by Lawrence *et al*. (2008). Adding regurgitant of Colorado potato beetle (CPB), *Leptinotarsa decemlineata* (Chrysomelidae), to wounded leaves of potato plants, elicits the expression of 73 genes when compared to leaves only wounded (Lawrence *et al*. 2008). An analysis of five differentially expressed genes between treatments found that genes involved on induction are related to secondary metabolism and stress. One induced gene encodes an aromatic amino acid decarboxylase, responsible for the synthesis of the precursor of 2-phenylethanol, which is recognized by the predator of *L. decemlineata* (*Perillus bioculatis*; Pentatomidae). Also, 3 out 16 type 1 and type 2 proteinase inhibitor clones duplicates present on the potato microarray (the TIGR 11,421 EST Solanaceae) were repressed by application of CPB regurgitant to wounded leaves. Given that proteinase inhibitors are known to interfere with the digestion of proteins in the insect midgut, repression of these proteinase inhibitors by CPB may inhibit this component of the plant's defence arsenal (Lawrence *et al*. 2008). Therefore, the authors suggest that beyond the wound response, CPB elicitors play a role in mediating the plant-insect interaction.

Xu *et al*. (2015) compared the responses of six closely related *Nicotiana* (Solanaceae) species to a synthetic elicitor, N-linolenoyl-glutamic acid and HAE of two insect herbivores (the Solanaceae specialist *Manduca sexta*; Sphingidae) and the generalist *Spodoptera littoralis*; Noctuidae). They found that HAE-induced defences are highly specific among closely related *Nicotiana*. This study is an example of how the specific responses to different HAEs

in terrestrial plants species are likely due to the perception by the plant of each specific component of the HAE (Xu *et al*. 2015).

Phytohormones as Regulators of Defensive Response

Herbivorous insects produce oral secretions containing compounds that elicit plant responses (Bonaventure *et al*. 2011; Stam *et al*. 2014) (Figure 2 and 3). The chemical nature of active compounds is remarkably diverse, including small organic compounds such as benzyl cyanide, fatty acid-amino acid conjugates, and proteins such as β-glucosidase (Stam *et al*. 2014). The recognition of herbivore elicitors by plant receptors initiates a cascade of responses, including changes in plasma membrane potential and activation of networks of MAP kinases and phytohormones (Stam *et al*. 2014) (Figure 2). In particular, this response to arthropod herbivory triggers reactive oxygen species and signal cascades involving jasmonic acid (JA), salicylic acid (SA), ethylene, abscisic acid, cytokinins, auxins and gibberellic acid that result in a downstream production of direct and indirect defence proteins such as *R* proteins (Smith and Clement, 2012; Kourelis and van der Hoorn 2018). Defence response gene up-regulation via JA and other pathways results in the production of many defence allelochemicals (Chen, 2008; Smith and Clement, 2012) (Figure 2). Less, however, is known on arthropod-induced expression of plant metabolism genes, but sparse evidence indicates that some of these genes are downregulated in the initial hours after the onset of arthropod herbivory and subsequently upregulated during ensuing days (Smith and Boyko, 2008; Smith and Clement, 2012) (Figure 3).

The expression of a gene is determined by the cis-acting DNA elements located in the vicinity of the gene and the trans-acting protein factors that interact with them (Signor and Sergey 2018) (Table 1). These cis-acting elements are concentrated in a relatively small promoter region of a few hundred nucleotides upstream of the transcriptional start site; other regulatory sequences are located at a distance of several thousands of nucleotides from the gene (Memelink, 2009). Several cis-acting elements in various gene promoters that mediate phytohormones (such as jasmonate) responsiveness have been identified. The most common jasmonate-responsive promoter sequences are the GCC motif and the G-box. Besides several other jasmonate-responsive promoter elements have been reported (Memelink, 2009).

The most studied phytohormone, jasmonic acid (JA) and its cyclic precursors and derivatives, are collectively referred to as jasmonates (JAs), and constitute a family of bioactive oxylipins that regulate plant responses to environmental and developmental cues (Wasternack, 2007) (Figure 2). These signalling molecules affect a variety of plant processes including fruit ripening (Creelman and Mullet, 1997), root elongation (Staswick *et al*. 1992), response to wounding (Zhang and Turner, 2008) and abiotic stresses, defence against insects (McConn *et al*. 1997) and necrotrophic pathogens (Thomma *et al*. 1999). Also, there is evidence that the jasmonates 12-oxo-phytodienoic acid (OPDA), JA, and methyl-jasmonic acid (MeJA) act as active signalling molecules to herbivory (Wasternack, 2007; Memelink, 2009) (Figure 2). Evidence on the role of jasmonates in plant-insect interactions derives from the analysis of mutants that fail to perceive JA/MeJA (Howe and Jander, 2008). Mutants that are defective in the *Coronatine insensitive* 1 (*COI1*) gene are impaired in all jasmonate-signalled processes and highly susceptible to a wide range of arthropod herbivores (reviewed in Howe and Jander, 2008) (Figure 2).

Once herbivory occurs, JA is produced via the octadecanoid pathway. In *Arabidopsis*, the enzyme jasmonoyl isoleucine conjugate synthase 1 (JAR1) activates JA by conjugating it to the amino acid isoleucine (Ile) to form JA-Ile (Stam *et al*. 2014). Within the JA signalling pathway two branches have been identified which act antagonistically (Pieterse *et al*. 2012). The MYC2 branch positively regulates the expression of wound-inducible JA-responsive marker genes such as vegetative storage protein 2 (*VSP2*) and lipoxygenase 2 (*LOX2*) (Pieterse *et al*. 2012). In the ethylene (ET) response factor (ERF) branch of the JA pathway, JA and ET synergistically induce the expression of JA/ET-responsive transcription factors, including ERF1 and octadecanoid-responsive *Arabidopsis* 59 (ORA59), which positively regulate JA/ET- responsive genes such as plant defensin 1.2 (*PDF1.2*) (Pieterse *et al*. 2012; Stam *et al*. 2014). The ERF branch is mainly involved in induced defense against necrotrophic pathogens, whereas the MYC2 branch mediates defence against herbivorous insects (Stam *et al*. 2014).

Salicylic acid (SA), a benzoic acid derivative, is also an important phytohormone involved in regulation of plant defence (War *et al*. 2012). It is an essential endogenous plant growth regulator involved in a wide range of metabolic and physiological responses in plants, including defence and plant growth and development (Rivas-San and

Plasencia, 2011). Responses to SA depend on a regulatory protein called Non-Expressor of Pathogenesis-Related Genes1 (*NPR1*) (Rivas-San and Plasencia, 2011). The *NPR1* gene is activated through redox pathways by SA accumulation and is translocated to the nucleus (War *et al*. 2012). However, it does not bind to DNA directly, but acts through transcription factors (War *et al*. 2012) (Table 1). SA induces greater defence against insects that pierce or suck plants rather than Chewers (War *et al*. 2012). Moreover, production of reactive oxygen species (ROS) by SA pathway is thought to induce resistance in plants against insect pests (*e. g*., tomato and *H. armigera;* Peng *et al*. 2007).

Initially, it was thought that plant damage by arthropods' chewing mouthparts would elicit JA-based transcriptomes, and that arthropods with piercing-sucking mouthparts would induce JA-SA-based transcriptomes (Smith and Clement, 2012). Nonetheless, it has been demonstrated that JA-SA signalling and JA-SA cross-talk are induced by both types of herbivores' feeding habits (Smith and Boyko, 2006; Smith and Clement, 2011). Crosstalk between phytohormonal signalling pathways may permit herbivores to manipulate plant defences in their interest (Stam *et al*. 2014). In example, feeding by *Manduca sexta* caterpillars induced an ET burst and suppressed nicotine accumulation in tobacco plants (Kahl *et al*. 2000). It has been hypothesized that by activating the SA signalling pathway, phloem feeders suppress the JA-dependent defences to which phloem feeders are more sensitive (Stam *et al*. 2014; Zarate *et al*. 2007). Recent studies indicate an interference between SA with JA-inducible defences against chewing insects (*e. g*., Lu *et al*. 2014), although not always phloem-feeding insects interfere with defences induced by chewing herbivores, perhaps due to density effects or to differences between species (Stam *et al*. 2014).

Ethylene is another important phytohormone that plays an active role in plant defence against many insects (van Loon *et al*. 2006). Ethylene signalling pathway participates directly and indirectly on induced plant defence against herbivores and pathogens *(*van Loon *et al*. 2006). For instance, infestation by *Phytophthora alni* (Peronosporaceae) induces the emission of ethylene and the release of various volatiles in *Alnus glutinosa* (Betulaceae) leaves (War *et al*. 2012).

The hormone systemin (Pearce *et al*. 1991; Ryan and Pearce, 1998) plays a regulatory role in many aspects of the plant life, including growth, development, fertilization, and interactions with symbiotic organisms (Wang *et al*. 2018). It is an amino acid peptide derived from a larger precursor protein. It was proposed that systemin functions spreading signal that triggers the systemic defence responses observed in plants after wounding or attack by herbivores (Pearce *et al*. 1991). A leucine-rich repeat receptor kinase (LRR-RK) is identified as the systemin receptor 160 (SR160) (Torii, 2004). SR160 is a tomato homologue of Brassinosteroid Insensitive 1 (BRI1), which mediates the regulation of growth and development in response to the steroid hormone brassinolide (Torii, 2004). Want *et al*. (2018) demonstrated that the perception of systemin depends on a pair of distinct LRR-RKs (leucine-rich repeat receptor kinase) called SYR1 and SYR2. SYR1 acts as a genuine systemin receptor that binds systemin with high affinity and specificity and the authors showed that the presence of SYR1 is important for defence against insect herbivory (Wang *et al*. 2018).

Epigenetic Regulation in Response to Defence

Environmental factors may modify the plant's regulation of individual genes through different mechanisms (Figure 1 and 3). For instance, DNA methylation, lysine methylation in histones, histone acetylation, histone phosphorylate or through RNA interference or transposition of mobile elements: insulators, promoters, enhancers, transposons (see Ramirez-Prado *et al*. 2018).

For instance, a large portion of many plant genomes consists of transposable elements (TEs) (Sahebi *et al*. 2018). TEs affect the expression of not only nearby genes but also unlinked inserted genes (Sahebi *et al*. 2018). TEs can create new promoters, leading to novel expression patterns or alternative coding regions to generate alternate transcripts in plant species (Sahebi *et al*. 2018). TEs can also provide novel cis-acting regulatory elements that act as enhancers or inserts within original enhancers that are required for transcription (Sahebi *et al*. 2018). TEs are able to generate new genes and modify existing gene structures by duplicating, mobilizing and recombining gene fragments (Sahebi *et al*. 2018). Hence, TE insertions can not only act as simple mutagens but can also alter the elementary functions of the plant defensive genes (Ramirez-Prado *et al*. 2018; Sahebi *et al*. 2018).

Recent evidence demonstrates that plant defence gene expression also involves DNA methylation and histone modifications (Law and Jacobsen, 2010; Lämke and Bäurle, 2017; Ramirez-Prado *et al*. 2018). For instance, a major class of *R* proteins are the leucine rich containing proteins (NLR) immune receptors that mediate ethylene signaling as defence to various pathogens (Espinas *et al*. 2016). *NLR* genes often form gene clusters in the genome that contain repetitive sequences and TEs (Meyers *et al*. 2003). The repetitive nature of *NLR* gene clusters is thought to facilitate rapid expansion and sequence diversification of these genes, possibly by promoting unequal recombination (Friedman and Baker, 2007; Espinas *et al*. 2016). As we have mentioned above, it is well documented that TEs inserted in the promoter region often regulate neighboring genes in both animals and plants by changing their epigenetic states (Slotkin and Martienssen, 2007). A recent report shows that TEs in intronic regions can regulate *NLR* expression in *Arabidopsis* (Eulgem *et al*. 2007). *Arabidopsis RPP7* gene encodes a CC-NBS-LRR (N- terminal coiled-coil domain; CC, a central nucleotide-binding site; NBS, and a C-terminal leucine-rich repeat; LRR) class of NLR that confers resistance to downy mildew, *Hyaloperonospora arabidopsidis* (Peronosporaceae) (Eulgem *et al*. 2007).

Transcriptions Factors as Regulators of Defensive traits

Regulation of gene expression is given by transcription factors; they are protein sequences that specifically bind to cis-regulation DNA sequences and may have activities as corepressors or coactivators (Petrillo *et al*. 2014) (Table 1). The control phytohormonal and gene expression in plants crosstalk to herbivory result in transcriptional responses that have a degree of specificity (Stam *et al*. 2014). Transcriptional responses of the plant response depend on the feeding guild of the attacker and the phytohormonal signal signature that the attacker induces (Stam *et al*. 2014). Recent studies in *Nicotiana attenuata* showed that aphids suppressed more genes than chewing herbivores did, and aphids upregulated the expression of SA-dependent genes and suppressed the expression of JA-mediated genes (Heidel and Baldwin, 2004) (Table 1). This review lists the recent findings related to the responses defence function of plants' transcription factors and the regulations of expression (Table 1).

Conclusions and Future Directions in the Genomics Era

Plant's genotype determines not only constitutive plant traits, but also inducible plant responses, such as the production of metabolites or structural changes (Kessler and Baldwin 2002; Howe and Jander 2008; War *et al*. 2012; Ramirez-Prado *et al*. 2018). The extent to which constitutive or inducible traits affect plant-insect interactions influences the relative importance of the inducible and constitutive phenotypes concerning their impact on community dynamics (Bidart-Bouzat and Kliebenstein, 2011). In this review we discuss the relevant studies on the molecular basis of defensive traits in plants. These studies help in understanding plant defence genes and the role of natural selection in natural, and cultivated, plant terrestrial populations.

It is important to stress the need for more studies on the different genetic responses to different stresses to elucidate the physiological mechanisms activated by the plants. Progress in genetics, including functional genomics, genome-wide association studies and QTL studies, along with advances in analytical chemistry and metabolomics are rapidly aiding to our understanding of the mechanisms linking physiological responses to ecological interactions in plants (Figure 3). Genome-wide data offer many advantages over sparser sets of genetic markers (Savolainen *et al*. 2013). It is now possible to detect the signature of selection across the genome and if putative genes are associated with herbivory (Figure 3). The understanding of the genetic basis of plant defence traits is today important in the face of climate change, crop production and pest control (Figure 3). Likewise, genomic tools may help in designing research strategies to combine genomics and field experiments to examine the genetics of local adaptation (Savolainen *et al*. 2013) (Figure 3). Thus, we can tackle the analysis of phenotypic patterns generated by spatially varying selection, genetic mapping and the genetic architecture of defence adaptive traits (Savolainen *et al*. 2013).

The recently identified interactions among signalling pathways involved in plant growth with defence signalling networks and the role of the phytohormones provide a starting point to test hypotheses on the regulation of ontogenically driven defence responses (Stam *et al*. 2014). Fundamental research on crosstalk among growth hormones and defence responses in angiosperms has mostly been performed with model organisms, such as *A. thaliana* and *Solanum lycopersicum*. The elucidation of the genetic architecture of defence traits in other non-model plant species will be important to uncover the manifold evolutionary phenotypic route taken by interacting plants and herbivores (Figure 3).

Acknowledgements This work was financed by the CONACyT grant (#1527) and the PAPIIT grant (PAPIIT-DGAPA AG200717) to JNF. Ivan Mijail De la Cruz Arguello is a doctoral student from the "Programa de Doctorado en Ciencias Biomédicas, National Autonomous University of Mexico (UNAM)" and acknowledges the scholarship (#283799) granted by CONACyT, and S. Velázquez-Márquez thanks to CONACyT for the Postdoctoral Fellowship (#2019-000019-01NC). We thank R. Tapia-López for logistics support for the project.

References

Baldwin, I. T., Halitschke, R., Paschold, A., Dahl, C. C. von, and Preston, C. A. (2006). Volatile signaling in plant-plant interactions: "talking trees" in the genomics era. Science, 311(5762), 812–815.

Bergelson, J., Kreitman, M., Stahl, E. A., and Tian, D. (2001). Evolutionary dynamics of plant *R*-genes. Science, 292(5525), 2281– 2285.

Bidart-Bouzat, M. G., and Kliebenstein, D. (2011). An ecological genomic approach challenging the paradigm of differential plant responses to specialist versus generalist insect herbivores. Oecologia, 167(3), 677.

Bonaventure, G., VanDoorn, A., and Baldwin, I. T. (2011). Herbivore-associated elicitors: FAC signaling and metabolism. Trends in Plant Science, 16(6), 294–299.

Bostock, R. M. (2005). Signal crosstalk and induced resistance: straddling the line between cost and benefit. Annual Review of Phytopathology, 43(1), 545–580.

Campos, M. L., Kang, J.-H., and Howe, G. A. (2014). Jasmonate-triggered plant immunity. Journal of Chemical Ecology, 40(7), 657–675.

Chen, M.-S. (2008). Inducible direct plant defense against insect herbivores: A review. Insect Science, 15(2), 101–114.

Creelman, R. A., and Mullet, J. E. (1997). Biosynthesis and action of jasmonates in plants. Annual Review of Plant Physiology and Plant Molecular Biology, 48(1), 355–381.

Davila Olivas, N. H., Kruijer, W., Gort, G., Wijnen, C. L., van Loon, J. J. A., and Dicke, M. (2017). Genome-wide association analysis reveals distinct genetic architectures for single and combined stress responses in *Arabidopsis thaliana*. The New Phytologist, 213(2), 838–851.

Espinas, N. A., Saze, H., and Saijo, Y. (2016). Epigenetic control of defense signaling and priming in plants. Frontiers in Plant Science, 7: 1201

Eulgem, T., Tsuchiya, T., Wang, X.-J., Beasley, B., Cuzick, A., Tör, M., Zhu, T., McDowell, J. M., Holub, E., and Dangl, J. L. (2007). EDM2 is required for RPP7-dependent disease resistance in *Arabidopsis* and affects RPP7 transcript levels: EDM2- mediated disease resistance. The Plant Journal, 49(5), 829–839.

Friedman, A. R., and Baker, B. J. (2007). The evolution of resistance genes in multi-protein plant resistance systems. Current Opinion in Genetics & Development, 17(6), 493–499.

Futuyma, D. J., and Agrawal, A. A. (2009). Macroevolution and the biological diversity of plants and herbivores. Proceedings of the National Academy of Sciences, 106(43), 18054–18061.

Halitschke, R., Schittko, U., Pohnert, G., Boland, W., and Baldwin, I. T. (2001). Molecular Interactions between the specialist herbivore *Manduca sexta* (Lepidoptera, Sphingidae) and Its natural host *Nicotiana attenuata*. Fatty acid-amino acid conjugates in herbivore oral secretions are necessary and sufficient for herbivore-specific plant responses. Plant Physiology, 125(2), 711–717.

Hancock, A. M., Brachi, B., Faure, N., Horton, M. W., Jarymowycz, L. B., Sperone, F. G., Toomajian, C., Roux, F., and Bergelson, J. (2011). Adaptation to climate across the *Arabidopsis thaliana* genome. Science, 334(6052), 83–86.

Hare, J. D. (2010). Ontogeny and season constrain the production of herbivore-inducible plant volatiles in the field. Journal of Chemical Ecology, 36(12), 1363–1374.

Heidel, A. J., and Baldwin, I. T. (2004). Microarray analysis of salicylic acid- and jasmonic acid-signalling in responses of *Nicotiana attenuata* to attack by insects from multiple feeding guilds. Plant, Cell & Environment, 27(11), 1362–1373.

Howe, G. A., and Jander, G. (2008). Plant immunity to insect herbivores. Annual Review of Plant Biology, 59(1), 41–66.

Hwang, C.-F., and Williamson, V. M. (2003). Leucine-Rich Repeat-mediated intramolecular interactions in nematode recognition and cell death signaling by the tomato resistance protein Mi. The Plant Journal, 34(5), 585–593.

Jan S, Mathys J., Maclintoc, M.A., De Coninck, Bruno, P.A, Cammue., De Bolle, M. (2008). Plant pathogenesis-related (PR) proteins: A focus on PR peptides. Plant Physiolpgy and Bioquemestry 46(11) 941-950.

Jensen, M. K., Lindemose, S., Masi, F. de, Reimer, J. J., Nielsen, M., Perera, V., Workman, C. T., Turck, F., Grant, M. R., Mundy, J., Petersen, M., and Skriver, K. (2013). ATAF1 transcription factor directly regulates abscisic acid biosynthetic gene NCED3 in *Arabidopsis thaliana*. FEBS Open Bio, 3(1), 321–327.

Johnson, M. T. J., Prashad, C. M., Lavoignat, M., and Saini, H. S. (2018). Contrasting the effects of natural selection, genetic drift and gene flow on urban evolution in white clover (Trifolium repens). Proceedings of the Royal Society B: Biological Sciences, 285(1883).

Jones, J. D. G., and Dangl, J. L. (2006). The plant immune system. Nature, 444(7117), 323–329.

Kahl, J., Siemens, D. H., Aerts, R. J., Gäbler, R., Kühnemann, F., Preston, C. A., and Baldwin, I. T. (2000). Herbivore-induced ethylene suppresses a direct defense but not a putative indirect defense against an adapted herbivore. Planta, 210(2), 336–342.

Kaminaka, H., Näke, C., Epple, P., Dittgen, J., Schütze, K., Chaban, C., Holt III, B. F., Merkle, T., Schäfer, E., Harter, K., and Dangl, J. L. (2006). BZIP10-LSD1 antagonism modulates basal defense and cell death in *Arabidopsis* following infection. The EMBO Journal, 25(18), 4400–4411.

Kessler, A., and Baldwin, I. T. (2001). Defensive function of herbivore-induced plant volatile emissions in nature. Science, 291(5511), 2141–2144.

Kessler, A., and Baldwin, I. T. (2002). Plant responses to insect herbivory: the emerging molecular analysis. Annual Review of Plant Biology, 53(1), 299–328.

Kourelis, J., and Hoorn, R. A. L. van der. (2018). Defended to the nines: 25 years of resistance gene cloning identifies nine mechanisms for R protein function. The Plant Cell, 30(2), 285–299.

Krishnaswamy, S., Verma, S., Rahman, M. H., and Kav, N. N. V. (2011). Functional characterization of four APETALA2-family genes (RAP2.6, RAP2.6L, DREB19 and DREB26) in Arabidopsis. Plant Molecular Biology, 75(1), 107–127.

Kroymann, J., Donnerhacke, S., Schnabelrauch, D., and Mitchell-Olds, T. (2003). Evolutionary dynamics of an *Arabidopsis* insect resistance quantitative trait locus. Proceedings of the National Academy of Sciences, 100(2), 14587–14592.

Lämke, J., and Bäurle, I. (2017). Epigenetic and chromatin-based mechanisms in environmental stress adaptation and stress memory in plants. Genome Biology, 18(1), 124.

Law, J. A., and Jacobsen, S. E. (2010). Establishing, maintaining and modifying DNA methylation patterns in plants and animals. Nature Reviews Genetics, 11(3), 204–220.

Lawrence, S. D., Novak, N. G., Ju, C. J.-T., and Cooke, J. E. K. (2008). Potato, *Solanum Tuberosum*, defense against colorado potato beetle, *Leptinotarsa decemlineata* (Say): Microarray gene expression profiling of potato by colorado potato beetle regurgitant treatment of wounded leaves. Journal of Chemical Ecology, 34(8), 1013–1025.

Linhart, Y. B., and Grant, M. C. (1996). Evolutionary significance of local genetic differentiation in plants. Annual Review of Ecology and Systematics, 27(1), 237–277.

Lu, S., Sun, Y.-H., and Chiang, V. L. (2008). Stress-responsive microRNAs in *Populus*. The Plant Journal, 55(1), 131–151.

Mao, Y.-B., Tao, X.-Y., Xue, X.-Y., Wang, L.-J., and Chen, X.-Y. (2011). Cotton plants expressing CYP6AE14 double-stranded RNA show enhanced resistance to bollworms. Transgenic Research, 20(3), 665–673.

McConn, M., Creelman, R. A., Bell, E., Mullet, J. E., and Browse, J. (1997). Jasmonate is essential for insect defense in *Arabidopsis*. Proceedings of the National Academy of Sciences, 94(10), 5473–5477.

Memelink, J. (2009). Regulation of gene expression by jasmonate hormones. Phytochemistry, 70(13), 1560–1570.

Meyers, B. C., Kozik, A., Griego, A., Kuang, H., and Michelmore, R. W. (2003). Genome-wide analysis of NBS-LRR–Encoding genes in *Arabidopsis*. The Plant Cell, 15(4), 809–834.

Michelmore, R. W., and Meyers, B. C. (1998). Clusters of resistance genes in plants evolve by divergent selection and a birth-and- death process. Genome Research, 8(11), 1113–1130.

Mithöfer, A., and Boland, W. (2012). Plant defense against herbivores: chemical aspects. Annual Review of Plant Biology, 63(1), 431–450.

Nallu, S., Hill, J. A., Don, K., Sahagun, C., Zhang, W., Meslin, C., Snell-Rood, E., Clark, N. L., Morehouse, N. I., Bergelson, J., Wheat, C. W., and Kronforst, M. R. (2018). The molecular genetic basis of herbivory between butterflies and their host plants. Nature Ecology & Evolution, 2(9), 1418–1427.

Nombela, G., Williamson, V. M., and Muñiz, M. (2003). The root-knot nematode resistance gene Mi-1.2 of tomato is responsible for resistance against the whitefly *Bemisia tabaci*. Molecular Plant-Microbe Interactions, 16(7), 645–649.

Pearce, G., Strydom, D., Johnson, S., and Ryan, C. A. (1991). A polypeptide from tomato leaves induces wound-inducible proteinase inhibitor proteins. Science, 253(5022), 895–897.

Peng, J., Li, Z., and Huang, Y. (2007). Plant species specific defense signal communication differentially regulates glutathione S- transferase activity and gene expression in the *Helicoverpa armigera* (Hubner). Journal of Plant Interactions, 2(2), 93–99.

Petrillo, E., Godoy Herz, M. A., Barta, A., Kalyna, M., and Kornblihtt, A. R. (2015). Let there be light: regulation of gene expression in plants. RNA Biology, 11(10), 1215–1220.

Pieterse, C. M. J., Van der Does, D., Zamioudis, C., Leon-Reyes, A., and Van Wees, S. C. M. (2012). Hormonal modulation of plant immunity. Annual Review of Cell and Developmental Biology, 28(1), 489–521.

Ramirez-Prado, J. S., Abulfaraj, A. A., Rayapuram, N., Benhamed, M., and Hirt, H. (2018). Plant immunity: from signaling to epigenetic control of defense. Trends in Plant Science, 23(9), 833–844.

Rausher, M. D., and Huang, J. (2016). Prolonged adaptive evolution of a defensive gene in the solanaceae. Molecular Biology and Evolution, 33(1), 143–151.

Rivas-San Vicente, M., and Plasencia, J. (2011). Salicylic acid beyond defence: Its role in plant growth and development. Journal of Experimental Botany, 62(10), 3321–3338.

Ryan, C. A., and Pearce, G. (1998). SYSTEMIN: a polypeptide signal for plant defensive genes. Annual Review of Cell and Developmental Biology, 14(1), 1–17.

Sahebi, M., Hanafi, M. M., van Wijnen, A. J., Rice, D., Rafii, M. Y., Azizi, P., Osman, M., Taheri, S., Bakar, M. F. A., Isa, M. N. M., and Noor, Y. M. (2018). Contribution of transposable elements in the plant's genome. Gene, 665, 155–166.

Savolainen, O., Lascoux, M., and Merilä, J. (2013). Ecological genomics of local adaptation. Nature Reviews Genetics, 14(11), 807–820.

Schmelz, E. A., Huffaker, A., Carroll, M. J., Alborn, H. T., Ali, J. G., and Teal, P. E. A. (2012). An amino acid substitution inhibits specialist herbivore production of an antagonist effector and recovers insect-induced plant defenses. Plant Physiology, 160(3), 1468–1478.

Schulze-Lefert, P., and Panstruga, R. (2011). A molecular evolutionary concept connecting nonhost resistance, pathogen host range, and pathogen speciation. Trends in Plant Science, 16(3), 117–125.

Seo, E., Choi, D., and Choi. (2015). Functional studies of transcription factors involved in plant defenses in the genomics era. Briefings in Functional Genomics, 14(4), 260–267.

Signor, S. A., & Nuzhdin, S. V. (2018). The evolution of gene expression in cis and trans. Trends in Genetics : TIG, 34(7), 532–544.

Slotkin, R. K., and Martienssen, R. (2007). Transposable elements and the epigenetic regulation of the genome. Nature Reviews Genetics, 8(4), 272–285.

Smith, C. M., and Boyko, E. V. (2007). The molecular bases of plant resistance and defense responses to aphid feeding: Current status. Entomologia Experimentalis et Applicata, 122(1), 1–16.

Smith, C. M., and Clement, S. L. (2012). Molecular bases of plant resistance to arthropods. Annual Review of Entomology, 57(1), 309–328.

Stam, J. M., Kroes, A., Li, Y., Gols, R., van Loon, J. J. A., Poelman, E. H., and Dicke, M. (2014). Plant interactions with multiple insect herbivores: from community to genes. Annual Review of Plant Biology, 65(1), 689–713.

Staswick, P. E., Su, W., and Howell, S. H. (1992). Methyl jasmonate inhibition of root growth and induction of a leaf protein are decreased in an *Arabidopsis thaliana* mutant. Proceedings of the National Academy of Sciences, 89(15), 6837–6840.

Tateda, C., Zhang, Z., Shrestha, J., Jelenska, J., Chinchilla, D., and Greenberg, J. T. (2014). Salicylic acid regulates *Arabidopsis* microbial pattern receptor kinase levels and signaling. The Plant Cell, 26(10), 4171–4187.

Thomma, B. P. H. J., Eggermont, K., Tierens, K. F. M.-J., and Broekaert, W. F. (1999). requirement of functional ethylene- insensitive 2 gene for efficient resistance of *Arabidopsis* to infection by *Botrytis cinerea*. Plant Physiology, 121(4), 1093–1101.

Thoen, M. P. M., Olivas, N. H. D., Kloth, K. J., Coolen, S., Huang, P.-P., Aarts, M. G. M., Bac-Molenaar, J. A., Bakker, J., Bouwmeester, H. J., Broekgaarden, C., Bucher, J., Busscher-Lange, J., Cheng, X., Fradin, E. F., Jongsma, M. A., Julkowska, M. M., Keurentjes, J. J. B., Ligterink, W., Pieterse, C. M. J., … Dicke, M. (2017). Genetic architecture of plant stress resistance: Multi-trait genome-wide association mapping. New Phytologist, 213(3), 1346–1362.

Torii, K. U. (2004). Leucine-Rich Repeat Receptor Kinases in plants: structure, function, and signal transduction pathways. International Review of Cytology (Vol. 234, pp. 1–46). Academic Press.

van Loon, L. C., Rep, M., and Pieterse, C. M. J. (2006). significance of inducible defense-related proteins in infected plants. Annual Review of Phytopathology, 44(1), 135–162.

Verhage, A., van Wees, S. C. M., and Pieterse, C. M. J. (2010). Plant immunity: it's the hormones talking, but what do they say? Plant Physiology, 154(2), 536–540.

Wang, L., Einig, E., Almeida-Trapp, M., Albert, M., Fliegmann, J., Mithöfer, A., Kalbacher, H., and Felix, G. (2018). The systemin receptor SYR1 enhances resistance of tomato against herbivorous insects. Nature Plants, 4(3), 152–156.

War, A. R., Paulraj, M. G., Ahmad, T., Buhroo, A. A., Hussain, B., Ignacimuthu, S., and Sharma, H. C. (2012). Mechanisms of plant defense against insect herbivores. Plant Signaling & Behavior, 7(10), 1306–1320.

Wasternack, C. (2007). Jasmonates: an update on biosynthesis, signal transduction and action in plant stress response, growth and development. Annals of Botany, 100(4), 681–697.

Wu, J., and Baldwin, I. T. (2010). New insights into plant responses to the attack from insect herbivores. Annual Review of Genetics, 44(1), 1–24.

Xu, S., Zhou, W., Pottinger, S., and Baldwin, I. T. (2015). Herbivore associated elicitor-induced defences are highly specific among closely related *Nicotiana* species. BMC Plant Biology, 15(1), 2.

Zarate, S. I., Kempema, L. A., and Walling, L. L. (2007). Silverleaf whitefly induces salicylic acid defenses and suppresses effectual jasmonic acid defenses. Plant Physiology, 143(2), 866–875.

Zhang, Y., and Turner, J. G. (2008). Wound-induced endogenous jasmonates stunt plant growth by inhibiting mitosis. PLOS ONE, 3(11), e3699.

Züst, T., Heichinger, C., Grossniklaus, U., Harrington, R., Kliebenstein, D. J., and Turnbull, L. A. (2012). Natural enemies drive geographic variation in plant defenses. Science, 338(6103), 116–119.

Table 1 Transcription factors families involved in plant defence.

Stimulus	Transcription factor	Function	Reference
Ethylene responsive factor, RF family member that is induced by salt stress and drougth.	AP2/ERF	The roles of AP2/ERF TFs in biotic and abiotic stress responses as well as in developmental processes have been reported.	Krishnaswamy *et al.* 2011
The bHLH TFs acting in plant defense against pathogens.	bHLH/	In animals, bHLH TFs mainly play roles in cell differentiation and neurogenic and myogenic processes. bHLH TFs in plants are involved in diverse biological processes as well	Seo *et al.* 2015
Plays important roles in plant adaptation to environmental stress and development	ATAF1/2	Cup-shaped cotyledon	Jensen *et al.* 2013
Responses to abiotic and biotic stress.	bZIP	bZIP TFs regulate diverse biological processes such as seed formation, floral development.	Kaminaka *et al*. 2006
MYB TFs functioning in response to biotic and abiotic stresses as well as primary and secondary metabolism.	MYB	MYB TF family is large and involved in controlling various processes like responses to biotic and abiotic stresses, development, differentiation, metabolism, defense etc.	War *et al.* 2012
Defense responses, especially in sensing PAMPs or pathogen effectors and in downstream signalin.	NAC	NAC TFs, response to biotic and abiotic stresses and in growth and development. For example, cold signals enhanced the proteolytic activation of a plasma membrane-bound NAC TF, NTL6, in *Arabidopsis thaliana*	War *et al.* 2012
The play roles as transcriptome analyses revealed that many *WRKY* genes were induced following infection by pathogens.	WRKY	WRKY TFs are involved in PAMP signaling downstream of mitogen-activated protein kinase (MAPK) cascades.	Seo *et al.* 2015
Pathogen attack and insect herbivory	MYC (MYC-12)	Leads to its interaction with JAZ and subsequent proteasomal degradation JAZ.	Seo *et al.* 2015
Cold stress, NTL6 (NTM (NAC with transmembrane motif1)-like 6) is induced and processed to relocate to the nucleus, activating *pathogenesis-related* (*PR*) gene expression.	NTL6	Play crucial roles in diverse processes such as shoot apical meristem maintenance, lateral root formation.	Tateda *et al*. 2014
Family coordinates stress signaling with wound healing. Different abiotic stresses: salt, drought, cold, ultraviolet B, heat, osmotic stress, as well as hormones such as ABA and JA.	ERF108, ERF109, ERF110, ERF111	Heterodimerization turns the ERFs into highly potent cell division activators.	Jan *et al.* 2008
Pathogenesis-related (PR) proteins.	TGA	Regulate defence gene expression for the generation of reactive oxygen species (ROS) and regulate specific plant responses to reactive oxylipins.	Jan *et al.* 2008
RDR proteins present in plants, role of RDR1, RDR2 and RDR6 for providing resistance against various biotic stresses endogenous small (sm) RNAs (primarily si- and miRNAs).	RDR 1, 2, 3, 4,5,6	Are important *trans/cis*-acting regulators involved in diverse cellular functions	Seo *et al.* 2015

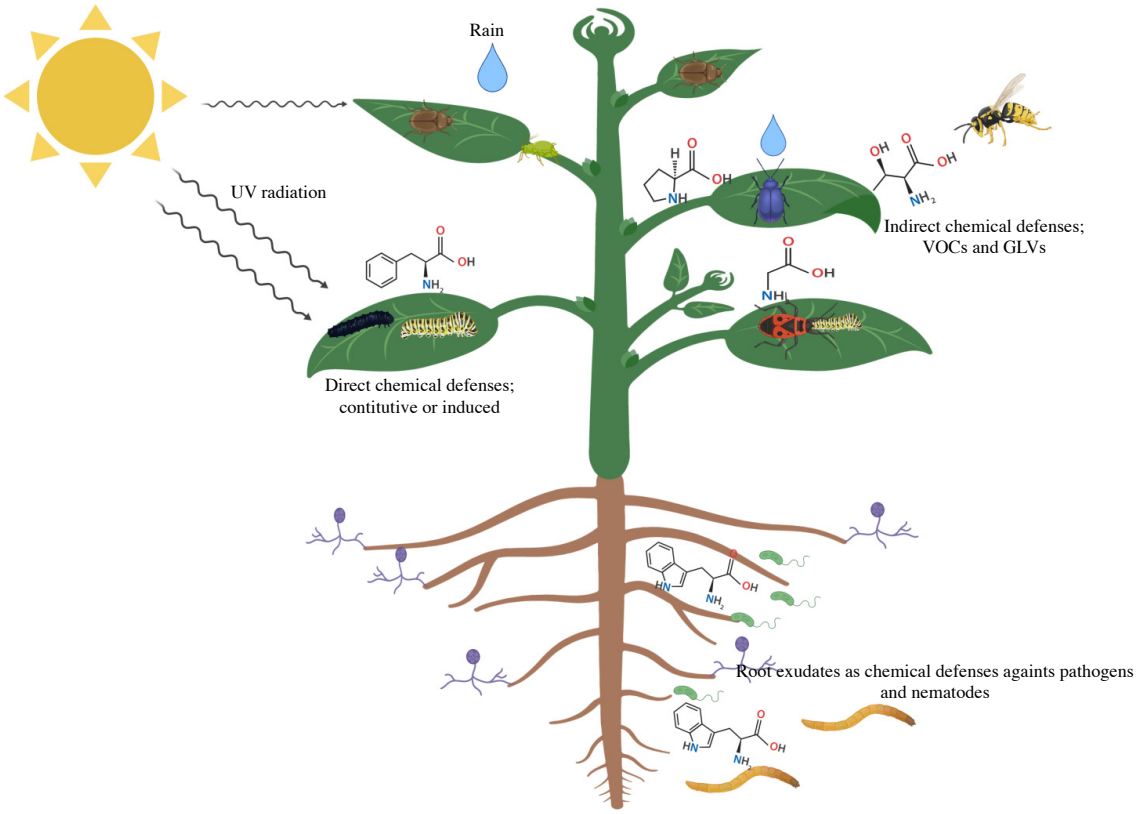


Figure 1. Plants are subject to different environmental pressures from biotic or abiotic factors. In particular, plants interact with different organisms below and above ground. Below ground such interactions include associations with mycorrhizal fungi, bacteria and worms that feed upon the roots. Above ground, insects feed on different plant tissues. Therefore, plants must deploy different defence strategies. These can be direct or indirect. Direct defences involve the production of different chemical weapons that can be constitutive or induced. Likewise, indirect defences include the release of VOCs and GLVs that attract predators of insects that feed on plants. Likewise, root exudates have been reported as defence against nematodes and pathogens. These exudates may have a role as an "attractant" to beneficial organism such as mycorrhizal fungi and bacteria. VOCs; volatile organic compounds, GLVs; green leaf volatiles.

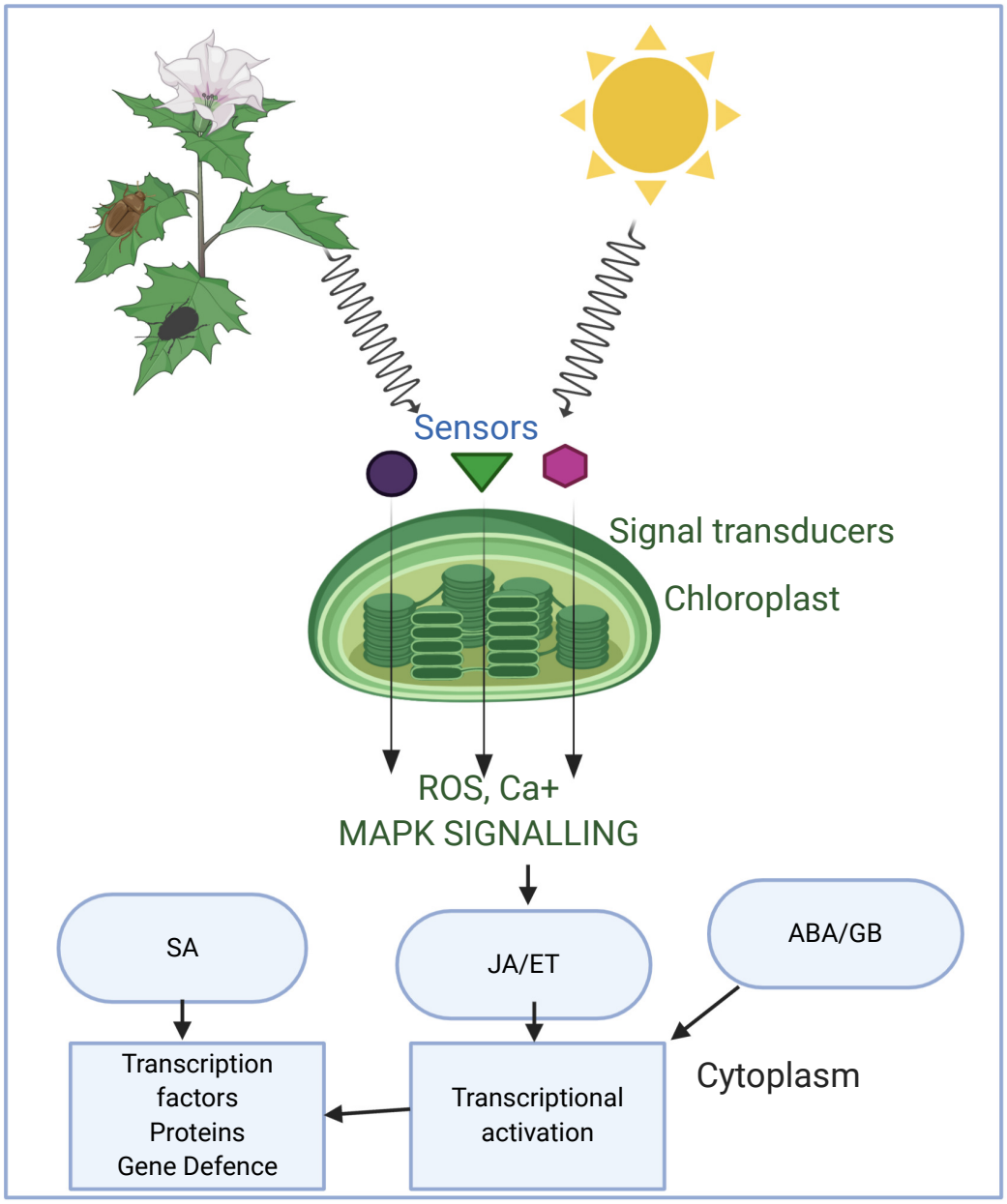


Figure 2. This figure summarizes the regulation of jasmonate-based defences in response to herbivory. Biotic factors (herbivory) and abiotic factors such as red-light assimilation activate MAPK kinases, calcium signalling or reactive oxygen species. These signals activate the production of jasmonates, ethylene and other plant phytohormones. Transcriptional regulons control direct and indirect defensive traits. The hormone signalling pathways also regulates plant responses to developmental cues and other stress conditions.

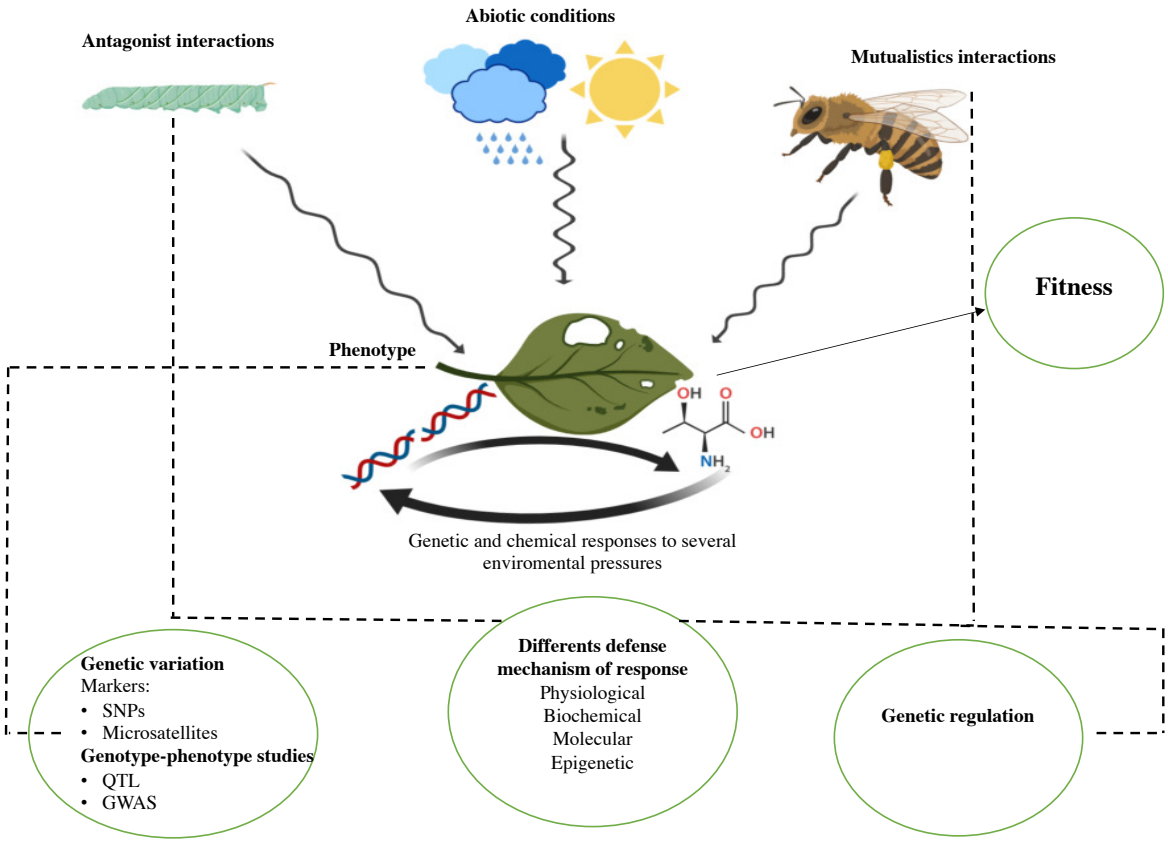


Figure 3. Figure based on Alonso *et al.* 2019. Plants are able to defend themselves against its natural enemies by producing chemical secondary compounds when they are being attacked. However, plants are part of complex interactions including pollinators, parasitoids and they are also subject to changes in the abiotic conditions. Therefore, they have to "decide" how to face with its selective pressures. Different responses may be used by the plants as defense which include physiological, biochemical, molecular and epigenetic mechanisms. Advancement in technology and DNA sequencing provides new directions to study plant-herbivore interactions. For instance, association between phenotypes and genotypes can be reached by quantitative trait loci analysis (QTL) or Genome-wide association analysis (GWAS).

CHAPTER 2

Research article

2020

EVOLUTIONARY RESPONSE TO HERBIVORY: POPULATION DIFFERENTIATION IN MICROSATELLITE LOCI, TROPANE ALKALOIDS AND LEAF TRICHOME DENSITY IN *Datura stramonium*

This manuscript has been published in ***Arthropod-Plant Interactions***

https://doi.org/10.1007/s11829-019-09735-7

Supplemental materials are provided at the end of the manuscript.

Arthropod-Plant Interactions
https://doi.org/10.1007/s11829-019-09735-7

ORIGINAL PAPER

Evolutionary response to herbivory: population differentiation in microsatellite loci, tropane alkaloids and leaf trichome density in *Datura stramonium*

I. M. De-la-Cruz[1] · L. L. Cruz[1] · L. Martínez-García[1] · P. L. Valverde[2] · C. M. Flores-Ortiz[3] · L. B. Hernández-Portilla[3] · J. Núñez-Farfán[1]

Received: 2 August 2018 / Accepted: 15 November 2019

Abstract

It is thought that natural selection exerted by herbivores on plants has promoted the evolution of plant traits that function as defence. However, such selective pressures may vary spatially differentiating populations in plant defence phenotypes. Yet, to ascertain the role of natural selection on phenotypic differentiation between populations, it is necessary to discard other evolutionary processes like genetic drift. Evolutionary biologists have designed approaches to determine whether population differentiation has been produced by natural selection in contrast to random processes as a null hypothesis. To accomplish this, we compare the magnitude of differentiation among populations in plant defence against herbivores (selection) and in neutral loci (genetic drift). Our study system is the plant *Datura stramonium*, whose anti-herbivore defence includes tropane alkaloids and foliar trichomes, and its specialized herbivorous insects. We selected two geographically close natural populations of *D. stramonium* in Central Mexico and estimated, under controlled conditions, population differentiation at neutral loci (microsatellites) and defence traits (concentration of tropane alkaloids and leaf trichome density). Results indicate very low genetic differentiation at neutral loci between populations but strong and significant phenotypic differentiation in putative defence traits. The average values of tropane alkaloids and leaf trichome density were higher in Ticumán than in Teotihuacán. Twelve out of 21 individual tropane alkaloids were significantly more abundant in plants from Ticumán, and the relative proportion of three of them contrasted markedly. Thus, results point that differentiation between populations of *D. stramonium* results from natural selection on defence traits.

Keywords Chemical defence · *Datura stramonium* · Liquid chromatography–TOF–MS · Leaf trichomes · Natural selection · P_{ST} vs. F_{ST} · Tropane alkaloids

Handling Editor: Rupesh Kariyat and Heikki Hokkanen.

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s11829-019-09735-7) contains supplementary material, which is available to authorized users.

✉ J. Núñez-Farfán
farfan@unam.mx

I. M. De-la-Cruz
imda@xanum.uam.mx

[1] Departamento de Ecología Evolutiva, Instituto de Ecología, Universidad Nacional Autónoma de México, Ciudad de México, Mexico

[2] Departamento de Biología, Universidad Autónoma Metropolitana-Iztapalapa, Ciudad de México, Mexico

[3] Facultad de Estudios Superiores Iztacala, Universidad Nacional Autónoma de México, Tlanepantla, Estado de México, Mexico

Introduction

Genetic differentiation among populations of a species is a ubiquitous phenomenon in nature driven by an interplay between natural selection, genetic drift and gene flow (Schluter 2000). In the case of the interaction between plants and herbivores, a large amount of evidence suggests that phenotypic differentiation in traits that mediate the interaction is a common outcome (e.g. Toju and Sota 2006; Laukkanen et al. 2012). Since plant populations occur along wide geographic areas and are attacked by different herbivore species, adaptive differentiation can result from selection on defence traits across populations (Thompson 2005). However, although natural selection is the ultimate cause of adaptive phenotypic differentiation among populations (Schluter 2000), trait differentiation may be adaptively

Published online: 05 December 2019

neutral, promoted by drift/migration interaction (Merilä and Cnokrak 2001).

Different approaches commonly used to test whether populations attain different phenotypic peaks in the adaptive landscape include direct measurement of natural selection, reciprocal transplant experiments and comparison with the neutral expectation (reviewed by Schluter 2000). The latter approach compares the observed differentiation between populations in quantitative characters (Q_{ST}) against the estimate of differentiation of adaptively neutral loci (F_{ST}; Spitze 1993; Schluter 2000; Merilä and Cnokrak 2001). Of three possible outcomes (i.e. $Q_{ST}=F_{ST}$, $Q_{ST}<F_{ST}$, $Q_{ST}>F_{ST}$), the finding of higher differentiation in quantitative traits than in neutral loci ($Q_{ST}>F_{ST}$) implies that directional selection is favouring different phenotypes in different populations (Merilä and Cnokrak 2001).

Tropane alkaloids are distinctive of the Solanaceae plant family and are especially abundant in the tribe Datura (see Wink 2003). Recently, lines of evidence indicate that tropane alkaloids are implicated in resistance against herbivores in the annual weed *Datura stramonium* (Shonle and Bergelson 2000; Castillo et al. 2013, 2014; Bustos-Segura et al. 2014; Miranda-Pérez et al. 2016), and that the selective value of tropane alkaloids and leaf trichomes preventing or reducing herbivory varies among populations of this plant species, depending on the type of herbivore (specialists or generalists) (Castillo et al. 2014). Prior work examining 13 natural populations of *D. stramonium* found that differences in Atropine and Scopolamine are better explained by natural selection than by genetic drift (Castillo et al. 2015). Here, we assess whether the magnitude of population differentiation in the concentration of tropane alkaloids, and leaf trichomes, exceeds genetic differentiation at neutral loci in two selected populations of *D. stramonium.* Although many tropane alkaloids have been described for this species (El Bazaoui et al. 2011), their role as plant defence has been assessed for few of them (Shonle and Bergelson 2000; Castillo et al. 2013, 2014; Bustos-Segura et al. 2014; Miranda-Pérez et al. 2016). We analyse in depth, through liquid chromatography/time-of-flight/mass spectra (HPLC–TOF–MS), the concentration tropane alkaloids in two Mexican populations of *D. stramonium* under controlled environmental conditions, to rule out the potential effects of environmental factors. For instance, the amount of alkaloid varies in relation to factors such as nitrogen level in the soil, temperature, water and light incidence (Nowacki et al. 1975; Waller and Nowacki 1978).

The two selected Mexican populations of *D. stramonium*, Teotihuacán (State of México) and Ticumán (State of Morelos), constitute an ideal system to contrast neutral and phenotypic differentiation in defence traits for several reasons. First, the two populations occur in different habitats, xerophytic shrub and tropical dry forest, respectively, and geographically close enough to maintain gene flow (ca. 110 km) (Valverde et al. 2001). Second, herbivore community differs between populations (De-la-Cruz, pers, obs 2018); plants of *D. stramonium* are attacked mainly by the specialist flea beetles *Epitrix* sp. in Ticumán (Valverde et al. 2001; Fornoni et al. 2004), while in Teotihuacán, plants are consumed mainly by the specialists *Epitrix parvula* and *Lema trilineata daturaphila* and the specialist seed predator *Trichobaris soror* (Bello-Bedoy and Núñez-Farfán 2010; Miranda-Pérez et al. 2016). Third, Teotihuacán population receives higher foliar damage by herbivorous insects than Ticumán (Valverde et al. 2001; Castillo et al. 2013). Fourth, in both populations, at least Scopolamine and leaf trichomes are implicated in plant defence (Valverde et al. 2001; Castillo et al. 2013; Miranda-Pérez et al. 2016). Finally, variation in resistance to herbivory is positively selected in both populations (Valverde et al. 2001, 2003).

In this study, we aim to measure the extent of (1) genetic differentiation at neutral loci and (2) in the concentration of each tropane alkaloid produced by plants of two populations of *D. stramonium,* in order to determine whether the differentiation is in agreement with an adaptive scenario.

Materials and methods

Experimental setting

We selected two natural populations of *D. stramonium* on which prior studies have detected differences in herbivory (Valverde et al. 2001; Castillo et al. 2013), to analyse chemical and physical defence to herbivores as well as genetic differentiation at neutral loci (Ticumán in the State of Morelos, 18° 45′ 39.90″ N, 99° 7′ 13.86″ W, and Teotihuacán, in the State of Mexico, 19° 41′ 6.96″ N, 98° 52′ 19.63″ W). The vegetation at Ticumán is a tropical seasonal dry forest, while at Teotihuacán, it is a temperate xerophytic shrubland (Valverde et al. 2001). The distance between populations is ca. 110 km but at contrasting altitudes (Valverde et al. 2001). In each population, we randomly collected the fruits of each of 47 (Ticumán) and 45 (Teotihuacán) plants. Fruits were labelled and individually bagged since they constitute natural progenies (families). The experiment was carried out between January and June 2016. Seeds of each plant ($n=10$) were imbibed in Petri dishes and maintained in a controlled-environment chamber at 12:12 L:D of photoperiod, and temperatures of 30°C during the day and 25°C at night, and at constant humidity at 85%. Seeds were scarified to promote germination (Fornoni and Núñez-Farfán 2000). Emerged seedlings were individually transferred to plastic pots (237 mL) and randomly allocated to positions in benches in the glasshouse. After 2 weeks, when the first true leaves appeared, each plant was transplanted into plastic pots (10 L) filled with a mix of 50% sand and 50%

vermiculite; again, pots were allocated randomly to positions on benches. One plant per each field-collected mother was allowed to grow until reproduction. We only sampled plants at the same developmental stage. When plants reached the second bifurcation and produced two flowers, 6–8 leaves per plant were harvested to quantify the richness and concentration of tropane alkaloids. At the same time, a sample of leaves ($n = 6$–8) was collected to quantify foliar trichome density per plant.

DNA isolation and microsatellite amplification

DNA was extracted from fresh leaves with a modified CTAB mini-prep protocol (Doyle and Doyle 1987). Five microsatellite loci were amplified (Andraca-Gómez 2009). PCR reactions were elaborated with labelled primers (Applied Biosystems). Each PCR mixture (10 µL) contained 2 µL of buffer, 1.2 µL of the $MgCl_2$ solution, 1 µL of dNTP's, 1 µL of DNA (20 ng), 1 µL of each fluorescent-labelled forward primer and 1 µL of each reverse primer (10 µM) and 2.75 µL of nuclease-free water. PCR reactions were performed in a Veriti 96-Well Thermal Cycler (Applied Biosystems). All samples were genotyped at the Institute of Biology (UNAM, Mexico City, Mexico). The genotypes were analysed by GeneMarker V.2.6.3 software (SoftGenetics, State College, Pennsylvania, USA).

Tropane alkaloids extraction

Extraction of tropane alkaloids was made for each plant from each population (Ticumán $n = 47$ and Teotihuacán $n = 45$). Leaves were oven-dried at 50 °C for 3 days and then pulverized (Berkov et al. 2005; Kariñho-Betancourt et al. 2015). We recorded ground dry leaf weight per plant and maintained it for 12 h in 20 mL of methanol (MeOH). After this, we applied several modifications to the alkaloid extraction protocol for *D. stramonium* of Castillo et al. (2013) as follows: the supernatant was filtered using nylon membrane filters, and MeOH was entirely evaporated in a Multivapor (P-6/P-12, Büchi) at 60 °C running at 240 mbar. Thenceforward, 10 mL of hydrochloric acid (HCl) 0.1 N was added and washed twice with 10 mL of chloroform ($CHCL_3$) in a decanting funnel, recovering the dense phase. Then, to solubilize the alkaloids, the mixture was neutralized with 0.082 g of sodium bicarbonate ($NaHCO_3$); subsequently, 20 mL of $CHCL_3$ was set, and the dense phase was recovered again. $CHCL_3$ was evaporated in the Multivapor at 50 °C running at 300 mbar. Finally, all samples were re-suspended into glass vials (1.5 mL) with MeOH. All vials were kept at room conditions to allow MeOH evaporation.

Liquid chromatography/time-of-flight/mass spectra (HPLC–TOF–MS)

All samples were re-suspended in 1 mL of MeOH and injected into an Agilent 1260 Infinity, coupled to an Accurate-Mass Time-of-Flight (TOF) LC/MS-6230, with an auto-sampler Agilent technology 1200 infinity. The chromatographic separations were performed in a column HPLC Agilent ZORBAX. For this, a gradient of mobile phase A (1% v/v formic acid in water) and mobile phase B (1% v/v formic acid in methanol) was used. The gradient profile was set to 0.00 min 90% A eluent, 10.00 min 10% A eluent, 17.00 min 90% A eluent, 17.10 min 90% A eluent. Conditions of this last step were maintained for 5 min to balance the column. The flow rate was 0.200 µL min^{-1} each 5 min, so each sample was analysed for 20 min, and the column temperature was 50 °C. The injection volume was 5 µL for all samples. The electrospray source (ESI) was operated in the positive mode, and the interface conditions were as follows: the fragmentor of 200 V; skimmer 65 V; OCT 1 RF Vpp 750 V; gas temperature of 350 °C; drying gas flow rate of 6 L min^{-1}; and the nebulizer worked at 50 psig. The ions of the compounds and their retention times are referred to in Table 1.

To standardize the method and optimize the detection of tropane alkaloids in the HPLC–TOF–MS system, we prepared standard solutions of Atropine and Scopolamine

Table 1 *R*- and *F*-statistics of five microsatellite loci of *D. stramonium*. Inbreeding coefficient (R_{IS}) was relatively high in the loci A09100150, B08100250, G08100350, while in overall, the negative value R_{IS} suggests an excess of heterozygotes (i.e. loci E01250400 and F10100400)

Locus	R_{IS}	F_{IS}	R_{ST}	F_{ST}	R_{IT}	F_{IT}
A09100150	0.999	0.167	−0.008	0.005	0.999	0.171
B08100250	0.993	−0.211	0.007	0.203	0.993	0.035
E01250400	−0.312	−0.157	0.106	0.086	−0.174	−0.057
G08100350	0.981	0.51	0.039	0.02	0.982	0.52
F10100400	−0.034	−0.245	0.292	0.084	0.268	−0.14
Total	0.487	0.093	0.061	0.066	0.518	0.153

The major value of R_{ST} (subpopulation structure) is shown in the locus F10100400. F_{IS} was high in the locus G08100350. Total R_{ST} and F_{ST} values are very similar

(Sigma-Aldrich, St. Louis, MO, USA) with 1 mg/mL of MeOH and injected these at volumes of 1, 2 and 3 µL. Since Atropine showed better calibration curves, the average concentration of the three injection volumes (259831.32 ng/g) was used to calculate the concentration for each identified alkaloid, per plant.

Identification and quantification of tropane alkaloids in plants leaves

First, we surveyed in the literature all alkaloids found in the leaves of *D. stramonium*. Thus, we obtained a list of 21 tropane alkaloids. Each alkaloid was searched individually in each chromatogram of each plant. The MassHunter Workstation software was used to identify the alkaloids using data of mass spectra, retention time and molecular formula obtained in the chromatograms. The total concentration for each alkaloid per plant was obtained with the equation:

$$\left[\left(\left(x_i/Xs\right)\times 1000\right)/Iv_i\right]/W_i$$

where x_i is the concentration obtained from each of 21 alkaloids per plant, Xs is the mean concentration of the standard solution of Atropine, Iv_i is the injection volume for the plant i, and W_i is the dry weight of the leaf sample of plant i. Alkaloid concentration was obtained in ng/g units of leaf weight. Total alkaloid is the sum of the 21 alkaloids per plant (Kariñho-Betancourt et al. 2015). Since alkaloids 15 and 21 possess identical molecular formula (i.e. identical mass fragmentation), identification was accomplished through differences in their retention time reported in the literature. Similarly, we applied this methodology for the following pairs of alkaloids: 16 and 20; 3 and 5; 13 and 9; 17 and 19; 2 and 12 and 8 and 7 (see Table 3 for details). The rest of the alkaloids were identified according to their fragmentation pattern reported in the literature as also indicated in Table 3.

Density of leaf trichomes

We took a random sample between six and eight leaves per plant and used a plastic stencil overlay under a dissection microscope to calculate the foliar trichome density as the total number of trichomes within an area of 2.986 mm^2 (plastic stencil observation field) (Valverde et al. 2001). Trichomes were counted in four observation fields set in the adaxial side of the leaf in order to account for spatial variation within the leaf: (1) bottom right edge, (2) lower left edge, (3) top right edge and (4) upper left edge (Kariñho-Betancourt et al. 2015). Trichome density was measured in these four-leaf zones because they provide a good estimate of the whole-leaf average trichome density (Valverde et al. 2001).

Data analysis

Five microsatellite loci were amplified for each plant of each population. In each population, we calculated genetic diversity statistics, including the mean number of different alleles (N_a), observed and expected heterozygosity (H_O and H_E, respectively), as well as fixation index (F). We also calculated the same parameters for every locus in each population. The estimate of population differentiation F_{ST} and its analogue for microsatellite loci R_{ST} (Slatkin 1995) were calculated for each locus. An AMOVA was carried out to evaluate the genetic variance partition. We estimated the indirect estimate of gene flow (Nm). All the analyses were carried out in GenAlEx 6.5 (Peakall and Smouse 2012).

To explore and detect differences in chemical concentration profiles between populations, we performed a multivariate discriminant analysis (DA). Two-tailed t tests were performed for each alkaloid as well as for the total alkaloid concentration.

Because we did not have a design necessary to estimate genetic variance of defence traits and calculate Q_{ST} (i.e. $Q_{ST}=V_{AB}/(V_{AB}+V_{AW})$ where V_{AB} and V_{AW} are the additive genetic variance between and within populations; Spitze 1993; Schluter 2000), we used the non-dimensional measure of phenotypic differentiation (P_{ST}) between populations for each alkaloid, and then, we obtained a chemical differentiation average (i.e. $P_{ST}=1-[\mathrm{Teo}_{\mathrm{Avg.}}/\mathrm{Tic}_{\mathrm{Avg.}}]$). Likewise, we performed a two-tailed t test model to compare the density of foliar trichomes between populations. Discriminant analysis and t tests were performed with log-transformed data of the density of foliar trichomes and the concentration of alkaloids (Sokal and Rohlf 1995).

Results

Genetic diversity and differentiation

R_{ST} and F_{ST} indicated low genetic differentiation between populations at microsatellite loci (Total $R_{ST}=0.061$, $p=0.001$; Total $F_{ST}=0.066$, $p=0.001$) (Table 1). Two out of five loci had high values of R_{ST}. The locus F10100400 displayed the higher R_{ST} value (Table 1). Mean observed heterozygosity was higher in Ticumán ($H_O=0.591\pm0.10$; $H_E=0.499\pm0.07$) than in Teotihuacán ($H_O=0.309\pm0.09$; $H_E=0.465\pm0.13$) (Table 2), indicating excess and deficiency of heterozygous individuals, respectively. Likewise, the Teotihuacán population has higher inbreeding than Ticumán (0.275 ± 0.07 vs. -0.274 ± 0.22). The mean number of different alleles was higher in Teotihuacán than in Ticumán (6 ± 2.38 and 3.80 ± 1.11, respectively) (Table 2). The AMOVA indicated that 6% of the molecular variance is accounted for differences between populations, 46% among individuals within populations, and 48% within

Table 2 Genetic diversity per locus in two populations of *D. stramonium*

Populations	Locus	*N*	N_a	H_o	H_e	*F*
Ticumán	A09100150	37	2.000	0.432	0.339	−0.276
	B08100250	37	2.000	0.541	0.394	−0.370
	E01250400	38	4.000	0.684	0.467	−0.465
	G08100350	37	8.000	0.351	0.771	0.544
	F10100400	38	3.000	0.947	0.525	−0.805
	Mean (SE)	37.40 (0.24)	3.80 (1.11)	0.591 (0.10)	0.499 (0.07)	−0.274 (0.22)
Teotihuacán	A09100150	30	2.000	0.233	0.433	0.461
	B08100250	30	2.000	0.033	0.033	−0.017
	E01250400	30	6.000	0.267	0.376	0.291
	G08100350	26	15.000	0.577	0.852	0.323
	F10100400	30	5.000	0.433	0.633	0.315
	Mean (SE)	29.20 (0.800)	6.00 (2.38)	0.309 (0.09)	0.465 (0.13)	0.275 (0.07)

N sample size, N_a number of different alleles per locus, *Ho* and *He* observed and expected heterozygosity, respectively, *F* fixation index. Mean value and standard error (SE) are given for each estimate

Table 3 Tropane alkaloids identified in leaves of *D. stramonium*

	Alkaloid	Formula	RT (min)	*m/z*	MS Ref.
1	3-(3′-Methoxytropoyloxy) tropane	$C_{18}H_{25}NO_3$	7.2	304.1907	El Bazoui et al. (2011)
2	3-Phenylacetoxy-6,7-Epoxytropane	$C_{17}H_{23}NO_3$	1.4	290.1751	Vitale et al. (1995)
3	3-Tigloyloxy-6-hydroxytropane	$C_{13}H_{21}NO_3$	10.1	240.1594	Witte et al. (1987)
4	3α-Tigloyloxytropane (4)	$C_{13}H_{21}N0_2$	9.8	224.1645	Witte et al. (1987)
5	3-Hydroxy-6-Tigloyloxytropane	$C_{13}H_{21}NO_3$	11.2	240.1594	Witte et al. (1987)
6	3-Phenylacetoxytropane	$C_{16}H_{21}NO_2$	12.9	260.1645	Philipov and Berkov (2002)
7	6-Hydroxyapoatropine	$C_{17}H_{21}NO_3$	13.7	288.1594	Witte et al. (1987)
8	6,7-Dehydrohyoscyamine	$C_{17}H_{21}NO_3$	14	288.1594	El Bazoui et al. (2011)
9	6-Hydroxyhyoscyamine (Anisodamine)	$C_{17}H_{23}NO_4$	13.2	306.1699	Ionkova et al. (1994)
10	Apoatropine	$C_{17}H_{21}NO_2$	10.4	272.1641	Witte et al. (1987)
11	Aposcopolamine	$C_{17}H_{19}NO_3$	11.7	286.1438	Witte et al. (1987)
12	Atropine (Hyoscyamine)	$C_{17}H_{23}NO_3$	8.7	290.1751	Witte et al. (1987)
13	Atropine impurity E (7-hydroxyhyoscyamine)	$C_{17}H_{23}NO_4$	12.6	306.17	Ionkova et al. (1994)
14	Cuscohygrine	$C_{13}H_{24}N_{20}$	4.4	225.1961	Witte et al. (1987)
15	Cyclotropine	$C_8H_{13}NO$	6.2	140.1069	El Bazoui et al. (2011)
16	Hygrine	$C_8H_{15}NO$	12.3	142.1226	Witte et al. (1987)
17	Scopine	$C_8H_{13}NO_2$	5.4	156.1019	Ionkova et al. (1994)
18	Scopolamine	$C_{17}H_{21}NO_4$	15	304.1543	Witte et al. (1987)
19	Scopoline	$C_8H_{13}NO_2$	5.5	156.1019	Ionkova et al. (1994)
20	Tropine	$C_8H_{15}NO$	9.4	142.1226	Witte et al. (1987)
21	Tropinone	$C_8H_{13}NO$	15.6	140.1069	El Bazoui et al. (2011)

RT the retention time of each alkaloid, *m/z* mass/charge, *MS* mass spectrometry reference

individuals (Table S1). The indirect estimate of gene flow between populations was high (*Nm* = 3.84 migrants).

Variation in tropane alkaloids and leaf trichome density between populations

The use of HPLC–TOF–MS provided a rapid separation of the complex mixture of the 21 tropane alkaloids contained in the leaves of *D. stramonium* (Table 3; Fig. 1). Discriminant analysis indicated that populations differed in the chemical profile of alkaloids (Wilks' λ = 0.60, Exact $F_{21,60} = 1.92$, $p = 0.025$) (Fig. 2). Aposcopolamine (11), Tropinone (21), Tropine (20), Scopoline (19), 7-Hydroxyhyoscyamine (13), Anisodamine (9), 3-(3′-Methoxytropoyloxy)tropane (1), 3-Tigloyloxy-6-hydroxytropane (3) alkaloids had the higher scoring coefficients for the canonical functions 1 and

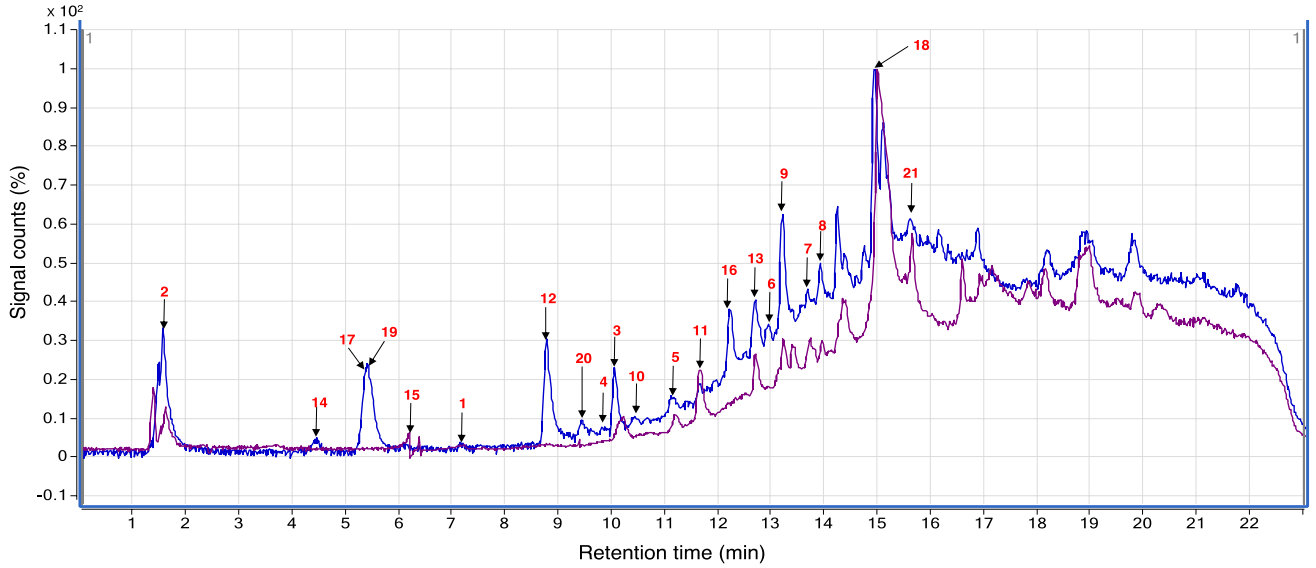


Fig. 1 Example of a chromatogram of the 21 tropane alkaloids identified in the leaves of *D. stramonium*. The blue line corresponds to the chromatogram of a Ticumán plant, while the purple line corresponds to the chromatogram of a Teotihuacán plant. There is a clear differentiation in the diversity and concentration of the 21 alkaloids between the two plants. Retention time, in minutes (m), is shown in the *x*-axis, while the signal intensity (%) is in the *y*-axis. Each alkaloid is highlighted in red colour. Names of alkaloids are given in Table 4

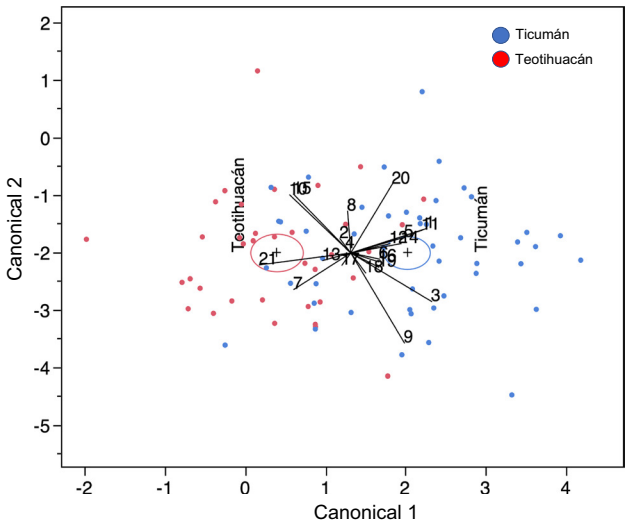


Fig. 2 Discriminant analysis for the 21 tropane alkaloids of Teotihuacán and Ticumán populations of *D. stramonium*. The normal ellipse region (50%) contours are shown in red for Teotihuacán and in blue for Ticumán. Centroids are indicated by "+"

2 (Table S2). The alkaloid Cuscohygrine was not detected in Teotihuacán, whereas it was detected in only two plants in Ticumán (Table 4).

Ticumán plants produced, on average, more total alkaloid than Teotihuacán ($t=3.01$, df = 81, $p=0.003$) (Table 4). The total alkaloid concentration was 12364.59 ng/g (SE = 2285.35) vs. 27697.76 ng/g (SE = 3537.87) for Teotihuacán and Ticumán, respectively (Table 4). Nevertheless, individual alkaloids showed more contrasting differences (Table 4; Fig. 3). For instance, Scopolamine and Anisodamine, two of the main alkaloids of *D. stramonium*, are on average 5.11- and 6.04-fold, respectively, higher in plants of Ticumán than Teotihuacán (cf. Table 4; Fig. 3). We found that 12 out of 21 alkaloids differ significantly between populations (Table 4). The relative proportion of some alkaloids also varied between populations. For instance, the alkaloid 3-phenylacetoxy-6,7-epoxytropane constitutes 0.46 of total alkaloids in Teotihuacán but only 0.20 in Ticumán. Likewise, while Scopolamine proportion is 0.25 in Ticumán, it is 0.11 in Teotihuacán (Table 4). The proportion of Anisodamine was 0.04 and 0.11 in Teotihuacán and Ticumán, respectively.

Foliar trichome density slightly differs between populations ($t=1.96$, df = 81, $p=0.050$). The mean values were 4.76 (SE = 0.207) and 3.77 trichomes × mm^{-2} (SE = 0.221) for Ticumán and Teotihuacán, respectively.

P_{ST} vs. F_{ST}

The mean estimate of phenotypic differentiation between populations ($P_{ST}=0.58$, SE = 0.059) was almost tenfold higher than mean genetic differentiation in microsatellite loci ($F_{ST}=0.06$, SE = 0.034) (Fig. 4).

Table 4 Average (SE) of total concentration of each of the 21 alkaloids in plants of two populations of *D. stramonium*

Alkaloid	Mean total alkaloid concentration (ng/g)							*t*	*p*
	Teotihuacán	SE	p_i	Ticumán	SE	p_i	Proportion Tic/Teo		
(1) 3-(3′-Methoxytropoyloxy)tropane	4.9224	2.1493	0.0004	75.2014	25.7759	0.0027	15.28	2.70	**0.0082**
(2) 3-Phenylacetoxy-6,7-epoxytropane	5737.0470	1534.4619	0.4640	5535.0539	1132.8674	0.1998	0.96	0.11	0.9092
(3) 3-Tigloyloxy-6-hydroxytropane	648.6590	148.3448	0.0525	711.1760	96.3901	0.0257	1.10	2.00	**0.0478**
(4) 3α-Tigloyloxytropane	0.7497	0.7497	0.0001	1.1101	0.4908	0.0000	1.48	1.09	0.2780
(5) 3-Hydroxy-6-tigloyloxytropane	220.3967	37.7854	0.0178	303.1317	78.2061	0.0109	1.38	1.21	0.2260
(6) 3-Phenylacetoxytropane	28.7530	5.2728	0.0023	68.0428	12.1256	0.0025	2.37	2.38	**0.0195**
(7) 6-Hydroxyapoatropine	19.1974	4.7345	0.0016	44.6846	10.3871	0.0016	2.33	1.39	0.1674
(8) 6,7-Dehydrohyoscyamine	22.3884	5.4728	0.0018	65.9704	16.6380	0.0024	2.95	1.76	0.0819
(9) 6-Hydroxyhyoscyamine (Anisodamine)	518.7318	104.6044	0.0420	3132.0905	760.5406	0.1131	6.04	3.75	**0.0003**
(10) Apoatropine	28.3447	8.3359	0.0023	116.5794	41.6242	0.0042	4.11	1.75	0.0828
(11) Aposcopolamine	13.8167	3.8216	0.0011	135.5438	41.8606	0.0049	9.81	3.76	**0.0003**
(12) Atropine (Hyoscyamine)	1940.3485	398.5295	0.1569	5474.8265	905.6186	0.1977	2.82	2.45	**0.0162**
(13) Atropine impurity E (7-hydroxyhyoscyamine)	242.1488	59.4495	0.0196	1116.9462	336.1092	0.0403	4.61	3.21	**0.0019**
(14) Cuscohygrine	0.0000	0.0000	0.0000	19.9715	19.4745	0.0007	>19.00	1.17	0.2449
(15) Cyclotropine	225.6546	49.4015	0.0183	298.1744	43.0693	0.0108	1.32	1.74	0.0853
(16) Hygrine	77.1408	37.7853	0.0062	334.2082	66.8511	0.0121	4.33	3.36	**0.0012**
(17) Scopine	504.8607	123.3744	0.0408	805.0952	202.4704	0.0291	1.59	2.21	**0.0293**
(18) Scopolamine	1385.2839	309.9245	0.1120	7077.1466	2441.0330	0.2555	5.11	2.39	**0.0187**
(19) Scopoline	572.6980	138.3338	0.0463	1893.7175	367.9377	0.0684	3.31	2.73	**0.0076**
(20) Tropine	94.8547	27.5414	0.0077	371.1753	100.6746	0.0134	3.91	2.93	**0.0043**
(21) Tropinone	78.5949	8.1374	0.0064	117.9141	12.8235	0.0043	1.50	1.80	0.0747
Total alkaloid concentration	12364.5918	2285.35		27697.7601	3537.87		2.24	3.01	**0.0035**

p_i is the relative proportion of each alkaloid in relation to the total concentration. Significant differences between means are indicated with bold-type face (*t* tests with 81 df)

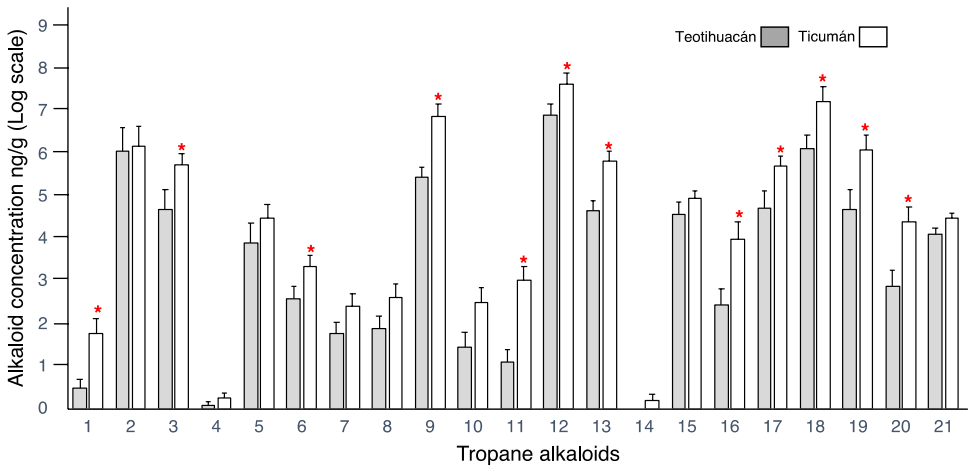


Fig. 3 Average (± SE) concentration of each of 21 tropane alkaloids in two populations of *D. stramonium*. Each alkaloid number in the *x*-axis corresponds to the name listed in Table 4. Asterisks indicate statistically significant differences. Teotihuacán, $N=45$, Ticumán, $N=47$

Discussion

When the magnitude of phenotypic differentiation among populations surpasses genetic differentiation at adaptively neutral loci, it is likely that divergent natural selection has took place ($Q_{ST} > F_{ST}$; Spitze 1993; Schluter 2000; Pujol et al. 2008). Here, we estimated the genetic differentiation between two populations of *D. stramonium* at microsatellite loci and phenotypic differentiation in the concentration

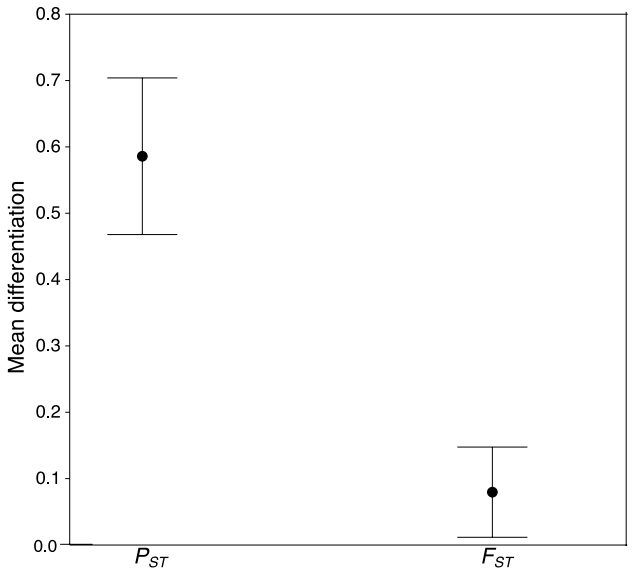


Fig. 4 Average and 95% confidence interval of phenotypic (P_{ST}) and genetic (F_{ST}) differentiation between two populations of *D. stramonium*

of 21 tropane alkaloids and leaf trichome density of plants grown in a common environment. Populations are genetically and phenotypically differentiated; differentiation in the concentration of tropane alkaloids exceeds the differentiation at neutral loci. These results indicate that phenotypic differentiation is not explained by genetic drift but likely by natural selection exerted by herbivores.

Does the evidence derived from this study support a scenario of adaptive differentiation in tropane alkaloids? We think it does; however, different sources of variation can produce phenotypic differences in defence traits. First, phenotypic differences between populations can be entirely due to environmental factors. For instance, it is well known that alkaloid concentration and trichome density in *D. stramonium* can be affected by physical factors such as nitrogen level in the soil, temperature, water and light incidence (Waller and Nowacki 1978; Castillo et al. 2013). Alkaloid content in leaves increases with an increased availability of nitrogen in the soil (Nowacki et al. 1975). Although these environmental factors affect alkaloid concentration, they cannot explain the differences found between populations studied here, since plants were grown under very uniform conditions of soil, light, temperature and watering. Leaf trichomes also show plasticity, but differences between populations are maintained in a common environment (Valverde et al. 2001). Second, it is also known that plant age, developmental stage and tissue can affect alkaloid concentration (cf. Weaver and Warwick 1984; Kariñho-Betancourt et al. 2015). Ontogenetic effects, likewise, can be ruled out since alkaloids were extracted from the same tissue, developmental stage and age of plants (see "Materials and methods" section). Third, differentiation in alkaloid production may result from different selective pressures exerted by herbivores between populations. The primary function of alkaloids in plants is considered to mediate the interaction of plants with their natural enemies (Wink 2003; Mithöfer and Boland 2012). There is evidence that the Ticumán and Teotihuacán populations differ in the community of herbivores. In the former population, plants are defoliated by its specialist herbivore, the flea beetles *Epitrix* spp. and other generalist herbivores (Valverde et al. 2001, 2003; Fornoni et al. 2004). In contrast, plants in Teotihuacán are consumed by the specialist herbivores of *Datura*, the leaf beetles *Epitrix parvula* and *Lema trilineata daturaphila,* and the specialist seed predator *Trichobaris soror* (Bello-Bedoy and Núñez-Farfán 2010; Miranda-Pérez et al. 2016). Further evidence indicates that positive natural selection on resistance to herbivores (i.e. 1-relative leaf damage) has been found in both populations (Valverde et al. 2001); in general, the Ticumán population shows higher resistance to herbivores than other populations in natural and common garden conditions (Valverde et al. 2001, 2003; Castillo et al. 2013). Finally, Scopolamine (Miranda-Pérez et al. 2016) and leaf trichomes (Valverde et al. 2001; Kariñho-Betancourt and Núñez-Farfán 2015) are implicated in plant resistance to herbivores. Therefore, we think that the between-population differentiation in leaf tropane alkaloids and trichome density found in the common environment do reflect genetic differentiation between populations brought about by differences in local herbivore pressure.

How evolutionary processes affect population differentiation? Gene flow can erode differentiation between populations (Slatkin 1987). Nevertheless, our estimate of indirect gene flow ($Nm = 3.84$) is large enough to overcome differentiation, and thus, differences in tropane alkaloids or leaf trichome density found between populations are not expected under high gene flow (i.e. $Q_{ST} = F_{ST}$). On the other hand, balancing selection may also reduce differentiation by favouring the same or similar phenotypic values across populations (i.e. $Q_{ST} < F_{ST}$; Merilä and Cnokrak 2001; Castillo et al. 2015). This scenario is not in agreement with our results. In contrast, genetic drift and divergent natural selection can produce differentiation between populations. Genetic drift can be ruled out since both populations and genetically diverse, and the indirect estimator of gene flow (Nm) is higher than the theoretical balance between drift and migration (i.e. $Nm = 1$). Here, we found that P_{ST} is higher than F_{ST}. Finally, we believe that divergent selection is responsible of differentiation in tropane alkaloids chemodiversity between Ticumán and Teotihuacán populations of *D. stramonium.* (P_{ST} is ca. tenfold higher than F_{ST}.) In fact, several lines of evidence point out that the differences in the levels of chemical and physical defence (trichomes) between the two studied populations possess a genetic variance (Valverde

et al. 2001, 2003; Fornoni et al. 2003; Carmona and Fornoni 2013; Kariñho-Betancourt and Núñez-Farfán 2015), suggesting different histories of interaction with herbivores. For instance, the Ticumán population of *D. stramonium* is consistently more resistant to herbivorous insects than Teotihuacán. The proportion of leaf damage per plant in different years is consistently higher in Teotihuacán (Valverde et al. 2001; Castillo et al. 2013).

This is the first study describing the variation in twenty-one tropane alkaloids in natural populations of *D. stramonium* and, together with a direct estimate of genetic divergence at neutral loci between two populations, allowed us to compare phenotypic differentiation (P_{ST}) against the neutral expectation (F_{ST}) (Schluter 2000). Results show that *D. stramonium* plants from Ticumán population produce higher concentration of total tropane alkaloids than those from Teotihuacán. Ticumán plants are more defended against herbivores. However, it must be recognized that we do not know, as yet, how the differences in alkaloid profiles between populations translate in plants defence. For instance, Ticumán produces on average higher proportions of Scopolamine and Atropine than Teotihuacán, but the latter produces a higher proportion of 3-phenylacetoxy-6,7-epoxytropane. Thus, we need to determine which specific alkaloids are involved in resistance to herbivores.

Population differentiation in leaf trichome density is less marked, and the role of trichomes as defence to herbivory seems to be very variable among populations of *D. stramonium* (Castillo et al. 2014). It has been suggested that trichomes in *D. stramonium* may serve other functions such heat dissipation on leaf surface thus reducing water loss (Kariñho-Betancourt and Núñez-Farfán 2015). Variation in the role of trichomes as defence character preventing herbivory has been documented in 11 species of Solanaceae. Kariyat et al. (2018) found that non-glandular trichomes were far more effective than glandular trichomes in deterring the initiation of feeding by first- and second-instar caterpillars of *Manduca sexta*. However, neither glandular nor non-glandular trichomes significantly affected the ability of third-instar caterpillars to initiate feeding.

Finally, although different factors may produce phenotypic differentiation in plant defence against herbivory between populations of *D. stramonium*, it seems that the most parsimonious explanation involves the historical interaction with different herbivorous insect communities. These findings open the opportunity to test adaptive evolution driven by divergent natural selection acting on quantitative trait loci (QTLs) of defence to herbivores.

Acknowledgements We are very grateful to Dr. R. Tapia-López for laboratory work and logistic support. We thank two anonymous reviewers for the criticisms that greatly improved the manuscript. This study is part of the PhD Dissertation of Ivan Mijail De-la-Cruz—Arguello—at the Programa de Doctorado en Ciencias Biomédicas, UNAM. IDLCA acknowledges the scholarship (# 283799) granted by CONACyT. Financial support was provided by a PAPIIT UNAM Grant #IG200717 to JNF and CMFO, and CONACyT FC 2016-2/1527 to JNF.

References

Andraca-Gómez G (2009) Genética de poblaciones comparada entre *Datura stramonium* y su herbívoro especialista *Lema trilineata*. B.Sc. Thesis, Biology, Faculty of Sciences, National Autonomous University of Mexico, Mexico City, pp 48. http://oreon.dgbiblio.unam.mx/

Bello-Bedoy R, Núñez-Farfán J (2010) Cost of inbreeding in resistance to herbivores in *Datura stramonium*. Ann Bot 105(5):747–753

Berkov S, Doncheva T, Philipov S, Alexandrov K (2005) Ontogenetic variation of the tropane alkaloids *in Datura stramonium*. Biochem Syst Ecol 33(10):1017–1029

Bustos-Segura C, Fornoni J, Núñez-Farfán J (2014) Evolutionary changes in plant tolerance against herbivory through a resurrection experiment. J Evol Biol 27:488–496

Carmona D, Fornoni J (2013) Herbivores can select for mixed defensive strategies in plants. New Phytol 197:576–585

Castillo G, Cruz LL, Hernandez-Cumplido J, Oyama K, Flores-Ortiz CM, Fornoni J, Valverde PL, Núñez-Farfán J (2013) Geographic association and temporal variation of chemical and physical defense and leaf damage in *Datura stramonium*. Ecol Res 28(4):663–672

Castillo G, Cruz LL, Tapia-López R, Olmedo-Vicente E, Carmona D, Anaya-Lang AL, Fornoni J, Andraca-Gómez G, Valverde PL, Núñez-Farfán J (2014) Selection mosaic exerted by specialist and generalist herbivores on chemical and physical defense of *Datura stramonium*. PLoS ONE 9(7):e102478

Castillo G, Valverde PL, Cruz LL, Hernández-Cumplido J, Andraca-Gómez G, Fornoni J, Sandoval-Castellanos E, Olmedo-Vicente E, Flores-Ortiz CM, Núñez-Farfán J (2015) Adaptive divergence in resistance to herbivores in *Datura stramonium*. PeerJ 3:e1411

Doyle J, Doyle JL (1987) Genomic plant DNA preparation from fresh tissue-CTAB method. Phytochem Bull 19(11):11–15

El Bazaoui A, Bellimam MA, Soulaymani A (2011) Nine new tropane alkaloids from *Datura stramonium* L. identified by GC/MS. Fitoterapia 82(2):193–197

Fornoni J, Núñez-Farfán J (2000) Evolutionary ecology of *Datura stramonium*: genetic variation and costs for tolerance to defoliation. Evolution 54(3):789–797

Fornoni J, Valverde PL, Núñez-Farfán J (2003) Quantitative genetics of plant tolerance and resistance against natural enemies of two natural populations of *Datura stramonium*. Evol Ecol Res 5:1048–1065

Fornoni J, Valverde PL, Núñez-Farfán J (2004) Population variation in the cost and benefit of tolerance and resistance against herbivory in *Datura stramonium*. Evolution 58(8):1696–1704

Ionkova I, Witte L, Alfermann H-A (1994) Spectrum of tropane alkaloids in transformed roots of *Datura innoxia* and *Hyoscyamus* x *gyorffyi* cultivated in vitro. Planta Med 60:382–384

Kariñho-Betancourt E, Núñez-Farfán J (2015) Evolution of resistance and tolerance to herbivores: testing the trade-off hypothesis. PeerJ 3:e789. https://doi.org/10.7717/peerj.789

Kariñho-Betancourt E, Agrawal AA, Halitschke R, Núñez-Farfán J (2015) Phylogenetic correlations among chemical and physical plant defenses change with ontogeny. New Phytol 206(2):796–806

Kariyat R, Hardison SB, Ryan AB, Stephenson AG, De Moraes CM, Mescher MC (2018) Leaf trichomes affect caterpillar feeding in an instar-specific manner. Commun Integr Biol 11(3):1–6

Laukkanen L, Leimu R, Muola A, Lilley M, Salminen JP, Mutikainen P (2012) Plant chemistry and local adaptation of a specialized folivore. PLoS ONE 7:e38225

Merilä J, Cnokrak P (2001) Comparison of genetic differentiation at marker loci and quantitative traits. J Evol Biol 14:892–903

Miranda-Pérez A, Castillo G, Hernández-Cumplido J, Valverde PL, Borbolla M, Cruz LL, Tapia-López R, Fornoni J, Flores-Ortiz CM, Núñez-Farfán J (2016) Natural selection drives chemical resistance of *Datura stramonium*. PeerJ 4:e1898

Mithöfer A, Boland W (2012) Plant defense against herbivores: chemical aspects. Annu Rev Plant Biol 63:431–450

Nowacki E, Jurzyesta M, Gorski P (1975) Effect of availability of nitrogen on alkaloid synthesis in Solanaceae. Bull Acad Pol Sci Ser Sci Biol 23:219–225

Peakall R, Smouse PE (2012) GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research-an update. Bioinformatics 28:2537–2539

Philipov S, Berkov S (2002) GC-MS investigation of tropane alkaloids in *Datura stramonium*. Zeitschrift für Naturforschung C 57(5–6):559–561

Pujol B, Wilson AJ, Ross RIC, Pannell R (2008) Are Q_{ST}-F_{ST} comparisons for natural populations meaningful? Mol Ecol 17:4782–4785

Schluter D (2000) The ecology of adaptive radiation. Oxford University Press, Oxford

Shonle I, Bergelson J (2000) Evolutionary ecology of the tropane alkaloids of *Datura stramonium* L. (Solanaceae). Evolution 54(3):778–788

Slatkin M (1987) Gene flow and the geographic structure of natural populations. Science 236:787–792

Slatkin M (1995) A measure of population subdivision based on microsatellite allele frequencies. Genetics 139:457–462

Sokal RR, Rohlf FJ (1995) Biometry: the principles and practice of statistics in biological research. Freeman, New York

Spitze K (1993) Population structure in *Daphnia obtusa*: quantitative genetic and allozymic variation. Genetics 135(2):367–374

Thompson JN (2005) The geographic mosaic of coevolution. University of Chicago Press, Chicago, p 442

Toju H, Sota T (2006) Adaptive divergence of scaling relationships mediates the arms race between a weevil and its host plant. Biol Lett 2(4):539–542

Valverde PL, Fornoni J, Núñez-Farfán J (2001) Defensive role of leaf trichomes in resistance to herbivorous insects in *Datura stramonium*. J Evol Biol 14(3):424–432

Valverde PL, Fornoni J, Núñez-Farfán J (2003) Evolutionary ecology of *Datura stramonium*: equal plant fitness benefits of growth and resistance against herbivory. J Evol Biol 16(1):127–137

Vitale AA, Acher A, Pomilio AB (1995) Alkaloids of *Datura ferox* from Argentina. J Ethnopharmacol 49(2):81–89

Waller GR, Nowacki EK (1978) Alkaloid biology and metabolism in plants. Plenum Press, New York

Weaver S, Warwick S (1984) The biology of Canadian weeds. *Datura stramonium* L. Can J Plant Sci 64:979–991

Wink M (2003) Evolution of secondary metabolites from an ecological and molecular phylogenetic perspective. Phytochemistry 64:3–19

Witte L, Müller K, Arfmann HA (1987) Investigation of the alkaloid pattern of *Datura innoxia* plants by capillary gas-liquid-chromatography-mass spectrometry. Planta Med 53:192–197

Supplemental materials chapter 2

Table S1. Analysis of Molecular Variance (AMOVA) of microsatellite loci of populations of *Datura stramonium*.

Source	df	SS	MS	Estimated Variance	%
Among Pops	1	25468.947	25468.947	283.290	6%
Among Individuals	66	427143.170	6471.866	2119.977	46%
Within Individuals	68	151770.000	2231.912	2231.912	48%
Total	135	604382.118		4635.179	100%

Table S2. Canonical functions of the discriminant analysis carried out of 21 alkaloids of Teotihuacán and Ticumán populations of *Datura stramonium*. Values highlighted in bold type show the highest scoring coefficient for each function.

Tropane alkaloids	Function 1	Function 2
(1) 3-(3´-Methoxytropoyloxy) tropane	**0.54167894**	0.21404204
(2) 3-Phenylacetoty-6,7-Epoxytropane	0.0305629	-0.1607834
(3) 3-Tigloyloxy-6-hydroxytropane	**0.55354261**	0.21352469
(4) 3α-Tigloyloxytropane	-0.0632268	0.25553774
(5) 3-Hydroxy-6-Tigloyloxytropane	0.34740251	0.01273465
(6) 3-Phenylacetoxytropane	0.08176889	0.2395041
(7) 6-Hydroxyapoatropine	**-0.4811487**	-0.2102165
(8) 6,7-Dehydrohyoscyamine	0.05401038	-0.1785733
(9) 6-Hydroxyhyoscyamine (Anisodamine)	**0.46240479**	**0.51873185**
(10) Apoatropine	-0.3974083	0.29262842
(11) Aposcopolamine	**0.77012896**	**-1.0938858**
(12) Atropine (Hyosciamine)	0.09916512	0.05865772
(13) Atropine impurity E (7-Hydroxyhyoscyamine)	-0.3213752	**0.54326926**
(14) Cuscohygrine	0.34461045	0.07710876
(15) Cyclotropine	-0.3664321	-0.1227268
(16) Hygrine	0.09773301	0.14201001
(17) Scopine	0.00042894	-0.0525863
(18) Scopolamine	-0.0107994	-0.1731047
(19) Scopoline	0.16829841	**0.56045293**
(20) Tropine	**0.61081331**	**-0.7146543**
(21) Tropinone	**-0.6364386**	-0.1270673

CHAPTER 3

Research article

2020

GENOMIC AND CHEMICAL EVIDENCE FOR LOCAL ADAPTATION IN RESISTANCE TO DIFFERENT HERBIVORES IN *Datura stramonium*

This chapter has been peer-reviewed and resubmitted in ***Evolution***

Supplemental material files are provided at the end of the manuscript. Scripts and commands to obtain identity by descent (IBD) used in this *ms* can be found in https://github.com/icruz1989/IBDcalculation

Genomic and chemical evidence for local adaptation in resistance to different herbivores in *Datura stramonium*

Ivan M. De-la-Cruz[1,4], Juha Merilä[4], Pedro L. Valverde[2], César M. Flores-Ortiz[3], Juan Núñez-Farfán[1*]

[1]*Laboratory of Ecological Genetics and Evolution, Department of Evolutionary Ecology, Institute of Ecology, Universidad Nacional Autónoma de México, Mexico City, Mexico.*

[2]*Department of Biology, Universidad Autónoma Metropolitana Campus Iztapalapa, Iztapalapa, Mexico City, Mexico.*

[3]*Facultad de Estudios Superiores Iztacala, Universidad Nacional Autónoma de México, Los Reyes Iztacala, State of Mexico, Mexico.*

[4]*Ecological Genetics Research Unit, Organismal and Evolutionary Biology Research Programme, Faculty of Biological and Environmental Sciences, University of Helsinki, Helsinki, Finland.*

Abstract

Since most species are collections of genetically variable populations distributed to habitats differing in their abiotic/biotic environmental factors and community composition, the pattern and strength of natural selection imposed by species on each others' traits are also expected to be highly spatially variable. Here, we used genomic and quantitative genetic approaches to understand how spatially variable selection operates on the genetic basis of plant defenses to herbivores. To this end, an F_2 progeny was generated by crossing *Datura stramonium* (Solanaceae) parents from two populations differing in their level of chemical defense. This F_2 progeny was reciprocally transplanted into the parental plants' habitats and by measuring the Identity by Descent (IBD) relationship of each F_2 plant to each parent, we were able to elucidate how spatially variable selection imposed by herbivores operated on the genetic background (IBD) of resistance to herbivory, promoting local adaptation. The results highlight that plants possessing the highest total alkaloid concentrations (sum of all alkaloid classes) were not the most well-defended or fit. Instead, specific alkaloids and their linked loci/alleles were favored by selection imposed by different herbivores. This has led to population differentiation in plant defenses and thus, to local adaptation driven by plant-herbivore interactions.

Key words. *Datura stramonium*, identity by descent, local adaptation, plant-herbivore interactions, phenotypic selection, resistance.

Introduction

Coevolution between plants and insects that feed on them is thought to be fueled by reciprocal selection imposed by traits (or trait states) that mediate the interaction, potentially given rise to arms races (Ehrlich and Raven 1964; Dawkins and Krebs 1979; Thompson 2005; Janz 2011). At the microevolutionary scale, spatial environmental variation may result in a selection mosaic that favors different traits (or trait states), hence promoting phenotypic and genetic/genomic divergence among populations, and thereby local adaptation (Gomulkiewicz *et al*. 2002; Thompson and Cunningham 2002; Thompson 2005; Briscoe Runquist *et al*. 2020). For instance, plant populations are likely to encounter different communities of herbivores both in space and time (Stam *et al*. 2014), making it highly improbable that selection pressures on plant defense traits (*e. g*., chemical secondary compounds) would be homogenous across populations (Berenbaum *et al*. 1986; Charlesworth 1998; Züst *et al*. 2012). Thus, it is expected that natural selection on plant-herbivore interactions between environments can lead to population differentiation of plant defense traits and ultimately to local adaptation. However, evidence of how varying herbivore communities impose selection of phenotypic defense variation, and their role in shaping the genomic constitution of populations is still scarce (Briscoe Runquist *et al*. 2020).

Local adaptation of plant defense against insect herbivores has been primarily studied using traditional quantitative genetic approaches such as common garden and reciprocal transplant experiments (Kawecki and Ebert 2004; de Villemereuil *et al*. 2016). These traditional approaches along with recent advances in genomics and mass spectrometry have made it possible to conduct detailed analyses of the genetic basis of chemical-based plant defense (Savolainen *et al*. 2013). For example, Identity by Descent analyses (IBD), genome-wide

association analyses (GWAS), quantitative trait loci mapping (QTL), or F_{ST} *vs*. Q_{ST} comparisons, provide methodologies to conduct in-depth studies on how plant chemical defense have evolved in response to spatial variation in plant-insect interactions (Browning and Browning 2012; Savolainen *et al*. 2013; Anderson *et al*. 2014; Flood and Hancock 2017).

In particular, IBD analysis estimates to what extent two or more individuals inherit a similar nucleotide sequence from a common ancestor (Thompson 2013) and describes the degree of genetic/familial similarity among a group of individuals (*e*. *g*., parents-offspring; Albrechtsen *et al*. 2010; Thompson 2013). Thus, IBD can be used to evaluate whether the genetic background of a plant is associated with its ability to face its herbivores. Furthermore, it also allows to detect patterns of very recent or ongoing selection in the genome (Albrechtsen *et al*. 2010). For instance, if insect herbivores are reducing the fitness of individual plants, one might suppose that more resistant plants to herbivory will produce more progeny than less resistant plants (Núñez-Farfán *et al*. 2007). If so, then ongoing natural selection will increase, across generations, the amount of IBD sharing in a population in the region surrounding the allele(s) that confer(s) resistance to herbivory (Browning and Browning 2012). The reasoning behind this is that as a positively selected allele increases in frequency, the region containing the resistance allele will increase in homozygosity and experience less intra-allelic recombination at the population level (Albrechtsen *et al*. 2010). While IBD analysis has been used to identify how recent or ongoing selection operates on human diseases (Albrechtsen *et al*. 2009; 2010), to best of our knowledge, no studies have used this approach to evaluate how the genetic background of plant resistance to herbivory is driven by natural selection.

The main aim of this study was to assess the extent to which the evolution of plant defenses to insect herbivores has been driven by natural selection. To this end, we generated an

F_2 progeny derived from the cross between two populations of the annual herb *Datura stramonium* (Asteridae; Solanaceae), known to differ in their level of chemical defense and herbivore community (De-la-Cruz *et al*. 2020). The F_2 plants were reciprocally transplanted to the natural environments (populations) of the grandparents. In this way, we were able to 1) determine the level of infestation and damage exerted by different herbivores on plants sowed in each locality, 2) to determine whether the seven most abundant constitutive alkaloids of *D. stramonium* are linked to the level of herbivore infestation. 3) by estimating the Identity by Descent (IBD) relationship of each F_2 plant to each grandparent, we were able to evaluate whether genomic similarity to either of the grandparents predicts survival/fitness and resistance to herbivores in each experimental site. Finally, 4) by quantifying the strength of natural selection on plant defense traits in the two experimental sites, we assessed whether natural selection favors an increase in plant resistance against herbivores in each of the two study sites.

Materials and Methods

The study sites

The two study sites, Teotihuacán (State of Mexico, 19°41'6.96"N, 98°52'19.63"W) and Ticumán (State of Morelos, 18°45'39.90"N, 99° 7'13.86"W), were selected for four main reasons. First, the two populations occur in different habitats with distinct climatic characteristics (xerophytic shrub and tropical dry forest, respectively; Valverde *et al*. 2001). Second, species of herbivores that infest upon *D. stramonium* differ between the sites (see also Results); in Ticuman, *D. stramonium* is attacked mainly by the specialist flea beetle *Epitrix* sp. (Valverde *et al*. 2001; Fornoni *et al*. 2004), whereas in Teotihuacán it is consumed mainly by the specialists beetles *Lema daturaphila, Epitrix parvula* and the specialist seed predator *Trichobaris soror* (Bello-

Bedoy and Núñez-Farfán 2010; Miranda-Pérez *et al*. 2016). In fact, *L. daturaphila* is absent in the Ticumán population. Third, low degree of genetic differentiation at neutral loci has been detected between the two populations (F_{ST} = 0.06; De-la-Cruz *et al*. 2020). Finally, the two populations are highly differentiated in their level of tropane alkaloid concentrations (De-la-Cruz *et al*. 2020). Previous evidence indicates that tropane alkaloids are central for resistance against herbivores in this species (Shonle and Bergelson 2000; Miranda-Pérez *et al*. 2016). Indeed, at least the tropane alkaloid scopolamine has been implicated as a defense against herbivores in one population (Castillo *et al*. 2014).

Experimental design

To produce the F_2 generation progeny for this study, we randomly collected fruits from 45 and 47 different plants from the Teotihuacán and Ticumán sites, respectively. Ten seeds from each of the 92 plants were soaked in water containers and maintained in an environmental chamber at a photoperiod of 12:12 L:D, and at a temperature of 30ºC during the day and 25ºC at night, at constant humidity of 85%. Seeds were scarified to facilitate germination (Fornoni *et al*. 2000). Germinated seeds were transferred to plastic pots (237 ml) and randomly allocated to positions on benches in the greenhouse. When the first true leaves developed, each plant was transplanted into 10 L plastic pots filled with a 1:1 mix of sand and vermiculite, and again, the pots were placed randomly on the benches. Each plant received the same daily quantity of water (500 ml) during the entire experiment. When the plants reached the flowering stage, flowers were hand pollinated. Plants from Teotihuacán were used as pollen receptors and plants from Ticumán were used as pollen donators. Prior to manual pollination, flowers from Teotihuacán were emasculated before dehiscence and covered with bags to avoid pollen contamination from other plants.

Cross-pollination was achieved by rubbing anthers of pollen donors onto the stigma of a flower. Mating pairs were set at random. After pollination, flowers were tagged and bagged. Because a plant can produce several flowers, each flower could be pollinated by different pollen donors. Thus, we produced *ca*. 200 crosses. Fruits of each cross (F_1 generation progeny) were tagged and collected in paper bags and stored at room temperature. When plants reached the flowering stage and second bifurcation (~30 days after planting in pots), 6-8 leaves from each plant were harvested to quantify the diversity and concentration of tropane alkaloids. There is evidence that the highest concentration of tropane alkaloids in *D. stramonium* occurs at the flowering stage, which is related to the timing of infestation by the main herbivores of *D. stramonium* (Kariñho-Betancourt *et al*. 2015). A total of 21 tropane alkaloids were identified and analyzed for all parental plants using methods described in De-la-Cruz *et al*. (2020).

Once the total tropane alkaloid concentration of the parental plants was completed, we selected the individual plant with the lowest (Teotihuacan) and the highest (Ticumán) concentration of tropane alkaloids (grandparents Teotihuacán 1 and Ticumán 23) (S1). These plants differed 58-fold in their total alkaloid concentration (1,013 *vs*. 59,000 ng/g of leaf, respectively) (S1). F_1 seeds derived from the cross between these two parental plants were sowed, following the procedure described above, to produce the F_2 progeny (single family: S1). To this end, we used seeds from three fruits of the same crossing. From the germinated F_1 plants (n = 8), we randomly chose one plant whose flowers were bagged to avoid pollen contamination from other plants (although plants were grown in a glasshouse; S1). We allowed this F_1 individual to self-pollinate to produce the F_2 generation progeny (S1).

Transplant experiment in the two sites

Experiment. F_2 seeds, taken randomly, were germinated and grown in the greenhouse as described above. When the two true leaves appeared, F_2 seedlings (n = 430) were transplanted to experimental plots in Teotihuacán (n = 230) and Ticumán (n = 200) in order to expose the F_2 plants to the local herbivores and natural environmental conditions of their grandparents (S1). During the first days after transplanting, high seedling mortality occurred at the tropical site (Ticumán), reducing sample size to 103 plants. In each site, seedlings were planted in the experimental plot according to a complete randomized design. Plants were spaced 1 m apart in a regular grid. Experimental plots were regularly weeded to prevent interference and competition by other species.

Damage by herbivores. Leaf damage to plants by herbivores was measured with the mobile application BioLeaf (Machado *et al*. 2016) on four sampling periods (15, 30, 45, 60 days after planting). On each sampling date, we took photographs of eight randomly chosen full expanded leaves per plant using a mobile phone (Samsung Galaxy S6 edge). The app automatically calculates the injured leaf regions caused by insect herbivory and then estimates the damage (in percentage) relative to the total leaf area (Machado *et al*. 2016). Thus, we were able to quantify the damage inflicted by herbivores to the plants during the experiment. Likewise, the average proportion of leaf damage by herbivores per plant was obtained. However, it is important to highlight that in the Teotihuacán site, most leaf tissue was completely eaten by herbivores in many plants. In these cases, we assigned 100% of the damage to these plants.

Herbivore infestation. At the Teotihuacán site, three species of herbivores were recorded during three sampling periods (15, 30, 45 days after planting). In each plant, we counted the number of 1) adults *Epitrix parvula*, 2) adults *Lema daturaphila,* 3) larvae of *Lema daturaphila*, and 4)

adults *Trichobaris soror*. Since larval development and pupation of *E. parvula* occur in the soil, we were unable to record these stages. Therefore, only the number of adults on plants was obtained for this insect species, as well as for *T. soror*. To minimize bias in insect counting, only one person counted the herbivores on each plant in all the sampling periods. In the Ticumán site, we recorded the infestation accounted only by *Epitrix* sp., since *L. daturaphila* is absent and *T. soror* is very rare (only 3 individuals registered at this site). At the end of the experiments, we had a measurement of total infestation that each plant experienced by each herbivore in both sites.

Leaf tissue sampling. In order to determine alkaloid concentration, we collected one leaf (10 cm in length) per plant when plants reached their second bifurcation and were flowering (~25 days after sowing). The leaf sampled was packed in aluminum foil, labeled and immediately frozen in liquid nitrogen. All samples were transported and stored in a freezer at -80ºC. In order to obtain DNA from each F_2 plant for genetic analyses, one additional leaf was collected, frozen and stored as described above.

Plant survival and reproduction. In the Teotihuacán site, plant mortality was caused by heavy damage exerted by insects (n = 66). We recorded plant survival as a nominal variable (dead/alive). At the Ticumán site, however, there was no record of single plant mortality due to damage exerted by herbivores.

At the end of the experiment (two months after sowing), we collected all fruits produced by each plant in each experimental site. Fruits were bagged individually and labelled. In the lab, seed set per fruit was counted and total number of seeds per plant was used as an estimator of maternal plant fitness (*see* statistical analyses section; Motten and Antonovics 1992; Núñez-Farfán *et al*. 1996; Mauricio and Rausher 1997).

Alkaloid extraction of F_2 plants

In order to extract tropane alkaloids from each plant, frozen leaf tissue was first transferred to 2 mL Eppendorf tubes, grinding it with a plastic pestle while keeping it frozen by adding liquid nitrogen. Second, we weighted the pulverized frozen leaf tissue in Eppendorf tubes. Third, we added two steel balls to each Eppendorf tube along with 1.5 mL of extraction buffer (80% methanol; MeOH and 1% formic acid); the tubes were then shaken for 60 s at 30 Hz in a TissueLyser II (QIAGEN). Finally, the samples were centrifuged for 20 min at 14,000 rpm; 700 µL of supernatant was collected and stored in glass vials (1.5 mL) and maintained at -4ºC until quantified in the Liquid Chromatography/Time-of-Flight/Mass Spectra (HPLC-TOF-MS).

Liquid Chromatography/Time-of-Flight/Mass Spectra

Before analysis, 300 µL of MeOH was added to each sample (stored in a glass vial; *see* above) and then injected into an Agilent 1260 Infinity, coupled to an Accurate-Mass Time-of-Flight (TOF) LC/MS-6230, with an auto-sampler Agilent Technology 1200 Infinity. The chromatographic separations were performed in a HPLC Agilent ZORBAX column. Before samples were injected into the column, it was cleaned with 15 mL of MeOH. For this, a gradient of mobile phase A (1% (v/v formic acid in water) and mobile phase B (1% (v/v formic acid in methanol) were used. The gradient profile was set to 0.00 min 90% A eluent, 10 min 10% A eluent, 17 min 90% A eluent, 17.10 min 90% A eluent. Conditions of this last step were maintained for 5 min to balance the column. The flow rate was 0.200 µL l min^{-1} each 5 min, so each sample was analyzed for 23 min, and the column temperature was 50°C. The injection volume was 1 μl for all samples. The electrospray source (ESI) was operated in the positive

mode, and the interface conditions were as follows: the fragmentor of 200 V; Skimmer 65 V; oct 1 RF Vpp 750 V; gas temperature of 350ºC; drying gas flow rate of 6 L min^{-1}; the nebulizer worked at 50 psig. The ions of the compounds and their retention times are given in S2.

To standardize the method and optimize the detection of alkaloids in the HPLC-TOF-MS system, we prepared standard solutions (1:1000; mg/ml) of Atropine and Scopolamine (Sigma-Aldrich, St. Louis, MO, USA) of MeOH and injected these at volumes of 2, 4 and 8 µl. Since atropine showed a better calibration curve, we used this curve to calculate the concentration for each identified alkaloid per plant.

Identification and quantification of alkaloids in *D. stramonium* leaves

First, we identified the seven most abundant constitutive alkaloids in *D. stramonium*: four tropane alkaloids (atropine, scopolamine, 3-hydroxy-6-tigloyloxytropane and anisodamine; De-la-Cruz *et al*. 2020), one alkaloid derived from the phenylalanine biosynthesis (phenylacetaldehyde), one pyrrolizidine alkaloid (pyrroline), and one triterpenoid of unknown name but of similar structure and molecular weight to azadirone triterpenoid (Álvarez-Caballero and Coy-Barrera 2019) (S2). Each alkaloid was searched and integrated (peak integration) individually in each chromatogram of each plant. The MassHunter Workstation software (v. B. 06.00; Agilent Technologies) was used to identify the alkaloids using data of mass spectra, retention time, and molecular formula obtained in the chromatograms (S2). The total concentration for each alkaloid per plant was obtained using the slope and the intersect from the regression equation of the calibration curve (curve from atropine standard):

$$\left(\left(\frac{(a + bX)\, x\, 1000}{d}\right) x\, 1000\right)$$

where *a,* is the intercept obtained from the regression of the calibration curve; *b,* is the slope obtained from the regression of the calibration curve, *X* is the concentration of given alkaloid in each plant and *d* is the dry weight of the sample. Alkaloid concentration was expressed in μg/g units of leaf weight. Total alkaloid concentration was obtained as the sum of the seven alkaloids per plant (Kariñho-Betancourt *et al*. 2015).

DNA extraction, library preparation for ddRad- sequencing

Genomic DNA (gDNA) was extracted from 163 individuals planted in Teotihuacán and 51 individuals planted in Ticumán. Since we had high mortality of seedlings at the beginning of the experiment in Ticumán, we extracted DNA from more individuals sowed in Teotihuacán. gDNA was isolated from fresh leaves with a modified CTAB mini-prep protocol for ddRad-seq (Doyle and Doyle 1987). The total amount of gDNA was measured using Qubit dsDNA HS Assay Kit (Invitrogen, Thermo Fisher Scientific, Waltham, USA). A total of 200 ng of gDNA was used for library preparation. The qualified DNA samples were digested with EcoRI and Hin1II (NlaIII) restriction enzymes (Takara, Osaka, Japan) and subjected to adapter ligation. The digestion and ligation were performed at 37°C for 16 hrs. The ligation products barcoded with unique P1 adapter were pooled and purified by size selection using E-Gel SizeSelect 2% agarose (Life Technologies, Carlsbad, CA, USA). Approximately 400-600 bp fragments were retrieved. The selected size and adaptor-ligated DNA was subsequently amplified by PCR. The PCR products were purified using AMpure XP beads (Beckman Coulter, Brea, CA, USA). The purified library was sequenced using Illumina Hiseq X Ten platform (Illumina, San Diego, CA, USA). Library preparation and sequencing were carried out by CD Genomics company (Shirley, NY, USA). For the two grandparents, gDNA was isolated and measured as above. However, whole genome

sequencing was carried out for both, rather than ddRad-seq. Libraries were sheared on the Covaris and then prepped for 150PE (paired-end) Illumina HiSeq 4000 sequencing using the Kapa Hyper prep Illumina library prep kits. Final libraries were visualized on the Agilent Fragment Analyzer, then quantified and pooled at equimolar amounts with Kapa qPCR Illumina library quant Universal Kits. The sequencing and library preparations for the grandparents were carried out in the QB3 Functional Genomics and Vincent J. Coates Sequencing Laboratories at the University of California, Berkeley.

Identity by Descent (IBD)

Two haplotypes are identical by descent (IBD) if they share the same alleles inherited from a common ancestor (Thompson 2013). Thus, closely related individuals have a high proportion of IBD (Thompson 2013). We estimated IBD between each individual F_2 plant and each of the two grandparents (214 F_2 plants *vs*. grandparent from Teotihuacán/grandparent from Ticumán). This information was used to evaluate whether F_2 plants more related to a given grandparent (*i. e.*, grandparent from Ticumán or grandparent from Teotihuacán) were more or less resistant to herbivory or had higher or lower fitness/survival in the experimental sites (*see* below).

For IBD estimation, demultiplexing was performed with the Illumina bcl2fastq v2.19 software, which returned sequence data in fastq format for each individual. Barcodes and indexes had been removed previously by CD Genomics and QB3 services. Illumina reads were trimmed using a Phred quality score > 20 in TRIMMOMATIC v0.32 (Bolger *et al*. 2014). We visually verified the quality of the grandparents and some individuals (~80) before and after trimming with FastQC (Andrews 2010). This allowed us to keep only high-quality reads for IBD analyses. Sequences of each individual were aligned to the *Datura stramonium* reference genome (De-la-

Cruz *et al*. under review; available in NCBI, BioProject PRJNA622882, biosample SAMN14531593, accession JAAWWY000000000) using the BWA v0.7.17 software (Li and Durbin 2009) with default parameters. SAM files from BWA were converted to BAM format and these BAM files were sorted using SAMtools v1.10 (Li *et al*. 2009).

The calculation of IBD values between each individual in relation to the grandparents was based on the genotype posterior probabilities (GPP) of each individual (Li *et al*. 2009; Rastas 2017). These GPPs were calculated using SAMtools mpileup (Li *et al*. 2009) and custom scripts provided in the tutorial of the Lep-MAP3 program (Rastas 2017; https://sourceforge.net/p/lep-map3/wiki/LM3%20Home/). These scripts also account for the alignment quality and filtering. The Lep-MAP3 program (Rastas 2017) was then used to calculate the IBD values between each individual and their grandparents.

Statistical analysis

All statistical analyses were performed using the JMP statistical package (v. 15.0; SAS Institute). Plotting was made using ggplot2 (Wickham 2016) in RStudio version 1.1.463 (R Core Team 2020).

Damage, herbivore infestation and concentration of alkaloids

Estimates of herbivore damage, herbivore infestation and alkaloid concentration were natural log transformed (log $n + 1$) to meet normality assumptions. An ANOVA was performed to test for differences in the levels of infestation by different herbivore species. A repeated measures ANOVA was used to test for differences in the level of damage between sampling dates in each experimental site. To assess the severity of damage as a function of the total infestation rate by

each species of herbivore in each of the two populations, a Pearson correlation analysis (Zar 1999) was performed between the average damage and the total infestation by each herbivore in each population.

IBD and survival in Teotihuacán

To analyze if genomic similarity among the F_2 individuals to their Ticumán and Teotihuacán grandparent was associated with their survival probability (alive/dead), two-tailed *t*-tests were used to compare mean IBD of individuals that survived or died. This analysis was only carried out for plants grown at the Teotihuacan site, because there was not plant mortality due to damage exerted by herbivores at the Ticumán site (*see* above).

Relationships between resistance, IBD and herbivory

Prior to analyses, all variables were standardized to a mean of zero and a standard deviation of one ($\bar{x} = 0$, SD = 1). Generalized linear models (GLMs) were employed to evaluate the relationships between resistance and IBD and herbivory. The GLMs described hereafter, were selected based on the statistical significance of the model and on the lowest corrected AIC values, *i. e.*, models that best explained the relationship between the variables (Akaike 1974).

First, general plant resistance (R_i) of the plant *i* was defined as $R_i = 100\text{-}def$, where *def* is the average proportion of leaf damage experienced by each plant (Núñez-Farfán and Dirzo 1994). To evaluate the relationship between resistance and IBD, two GLMs (link = identity, distribution = normal) were constructed; one using the IBD values between F_2 plants and the Teotihuacán grandparent, and the other using the IBD values between F_2 plants and the Ticumán grandparent. In these models, the response variable was resistance, whereas IBD, the

experimental site, and their interaction were used as predictors. Adding the interaction between experimental site and the covariate in the models allowed us to assess whether the effect of the IBD to each one of the grandparents (Teotihuacán or Ticumán) differed depending on the site of testing (*cf*. Zar 1999).

Since herbivore species differed between the sites, we independently assessed resistance as a function of herbivore species by GLMs (link = identity, distribution = normal). In Teotihuacán, we used the abundance of adults and/or larvae of *L. daturaphila*, *T. soror* and *E. parvula* as covariates, whereas in Ticumán, only the abundance of *Epitrix* sp. (the only herbivore detected in this site) was used as a predictor.

Relationship between herbivore infestation and alkaloid concentration

To assess the effect of the alkaloids on herbivore infestation, we also carried out GLMs (link = identity, distribution = normal) in which the response variables were *E. parvula*, *Epitrix* sp. adults or larvae of *L. daturaphila* or *T. soror* abundances on plants. The predictors in these models were the concentrations of the seven alkaloids. In addition, we performed stepwise GLMs (link = identity, distribution = normal) following a backward selection, which starts with all predictors in the model (seven alkaloids), and iteratively removes the least contributive predictors (Sokal and Rohlf 1994; Zar 1999). This allowed us to detect which alkaloid configuration had a greater positive or negative effect (or both) on the infestation of each herbivore species. The best GLMs were selected based on the statistical significance of the model and on the lowest corrected AIC values (Akaike 1974). An additional GLM (link = identity, distribution = normal) with the total alkaloid concentration as a predictor was carried out to see the impact of total alkaloid concentration on herbivores. Tests for the interaction

between experimental site and one particular herbivore were not possible because different species were present in the two sites.

Natural selection on alkaloids, resistance, herbivore infestation and IBD

To quantify the magnitude and direction of natural selection acting on the seven alkaloids, we used the number of seeds produced by each plant as a fitness proxy to perform phenotypic-selection analyses (Lande 1979; Lande and Arnold 1983). For this purpose, standardized individual fitness (relative fitness) was calculated as $w_i = x_i/\bar{x}$, where x_i is the total number of seeds produced per plant, and $\bar{x}$ is the average number of seeds per plant in the population in each site. In all analyses, w_i was used as a response variable. Thus, one GLM (link = identity, distribution = normal) was constructed using the concentrations of seven alkaloids, the experimental site, and their interactions. An additional two separate GLMs were constructed using resistance and total alkaloid concentration as predictors, as well as experimental site as a factor. Interactions between site and predictors allowed us to test if the effects of predictors on fitness differed between the two sites.

Two separate models, one for each experimental site, were carried out to assess selection on the infestation by each herbivore (independent variables). As pointed out earlier, we could not evaluate the effect of the experimental site and its interaction with predictors, since different species of herbivores were present in the two populations.

Finally, to evaluate the effects of identity by descent (IBD) on fitness (seed production), two GLMs were constructed (one using the IBD values between F_2 plants and the Teotihuacán grandparent, and the other using the IBD values between F_2 plants and the Ticumán

grandparent). In these models, the response variable was relative fitness, whereas IBD, experimental site and its interaction were predictors.

The generalized linear coefficients (*i. e.*, the selection gradients; β_i, Lande and Arnold 1983) obtained from the selection analyses represent the strength and direction of selection acting directly on each alkaloid, resistance, infestation by each herbivore and IBD in comparable units (standard deviations; Wise and Rausher 2013).

Results

Damage, herbivore infestation and alkaloid concentrations in the two experimental sites

Damage by herbivores varied between sampling dates in each site (Teotihuacán: $F_{710} = 110.98$, $R^2 = 0.41$, $p = 0.0001$; Ticumán: $F_{262} = 27.16$, $R^2 = 0.27$, $p = 0.0001$; S3 a, b, S4). There were clear differences in level of infestation by the different species of herbivores ($F_{545} = 215.32$, $R^2 = 0.61$, $p = 0.0001$; S3 c, S4). Correlation analyses indicated that plant damage in Teotihuacán site was mainly imposed by larvae *of L. daturaphila*, whereas that in the Ticumán site mainly by *Epitrix* sp. (S5).

Effect of the Identity by Descent (IBD) on survival/fitness and on resistance

The GLM between fitness and the IBD to the Ticumán grandparent as measured by genome wide IBD was significant, revealing a positive effect of increasing IBD on fitness (*L-R* chi-square$_3$ = 12.91, *AICc* = 273.68, $p = 0.0048$, Table 1, Fig. 1 a, b). The model between fitness and the IBD with the Teotihuacán grandparent was not significant (L-R chi-square$_3$ = 2.68, AICc = 283.90, $p = 0.4424$, Table 1, Fig. 1 c, d). Our results also showed that plants more related to the Ticumán grandparent had higher survival than F_2 plants less related to the Ticumán grandparent in the

Teotihuacán site ($F_{129} = 17.52$, $R^2 = 0.12$, $p = 0.0001$; Fig. 1 e). In contrast, plant survival was not significantly associated with IBD to the Teotihuacán grandparent in Teotihuacán ($F_{129} = 1.92$, $R^2 = 0.014$, $p = 0.1682$; Fig. 1 f). The mean F_2 full-sibs relatedness (IBD) was 0.47 (range 0.006-0.803, standard error = 0.0007). Identity by descent between the F_2 plants to each grandparent range between 0.006-0.5 (Teotihuacán grandparent) and 0.031-0.5 (Ticumán grandparent). Relatedness between the two grandparents was zero (Fig. 1 g).

The GLM between resistance and the IBD with the Ticumán grandparent was significant (*L-R* chi-square$_3$ = 54.21, *AIC*c = 441.47, $p = 0.0001$). Significant effects included population, and the interaction between population and IBD to the Ticumán grandparent (positive relationship in Teotihuacán site, while an opposite effect was observed in the Ticumán site; Table 2, Fig. 2 a, b). The GLM between resistance and the IBD with the Teotihuacán grandparent was significant (*L-R* chi-square$_3$ = 32.50, AICc = 463.18, $p = 0.0001$). However, only the population effect was significant (Table 2, Fig. 2 c, d).

Resistance against herbivore infestation levels in the two experimental sites

Resistance to herbivory was significantly related to herbivore infestation levels in Teotihuacán (*L-R* chi-square$_4$ = 111.09, *AIC*c = 426.37, $p = 0.0001$). Resistance was only positively related to levels of *E. parvula* infestation and negatively related to infestation by larvae of *L. daturaphila* in Teotihuacán (Table 2, Fig. 2 e, f). However, resistance and *Epitrix* sp. infestation levels were negatively related in Ticumán (*L-R* chi-square$_1$ = 14.96, *AIC*c = 223.89, $p = 0.0001$; Table 2, Fig. 2 g).

Effect of the alkaloid concentration on herbivore infestation levels

The effect of the seven alkaloid concentrations on *E. parvula* infestation level was significant (*L-R* chi-square$_7$ = 15.67, *AIC*c = 444.78; $p = 0.0282$). However, the effects of individual alkaloids were different: triterpenoid had a positive significant effect, whereas negative significant effects were detected in case of phenylacetaldehyde and pyrroline (S6). The effect of the total alkaloid concentration on *E. parvula* infestation was negative and significant (*L-R* chi-square$_1$ = 3.91, *AIC*c = 462.23; $p = 0.0479$; S6).

The effect of alkaloid concentrations on infestation levels by larvae of *L. daturaphila* was significant (*L-R* chi-square$_5$ = 12.53, *AIC*c = 460.49, $p = 0.0281$). Significant effects included phenylacetaldehyde (positively) and the triterpenoid (negative effect; S6). The GLM of infestation levels by larvae of *L. daturaphila* against the total alkaloid concentration was not significant (*L-R* chi-square$_1$ = 1.45, *AIC*c = 473.76; $p = 0.2272$; S6).

The effect of alkaloid concentration on infestation levels by adults of *L. daturaphila* was significant (*L-R* chi-square$_3$ = 9.49, *AIC*c = 453.93; $p = 0.023$). In this model we observed that 3-hydroxy-6-tigloyloxytropane was significantly and negatively related with *L. daturaphila* adults, whereas atropine showed a significant positive relationship (S6). The GLM of infestation levels by adults of *L. daturaphila* against the total alkaloid concentration was also not significant (*L-R* chi-square$_1$ = 0.49, *AIC*c = 469.65; $p = 0.4797$; S6).

The GLM testing for the effects of alkaloid concentration on *T. soror* infestation rate was significant (*L-R* chi-square$_4$ = 10.95, *AIC*c = 467.18, $p = 0.027$). In this model, *T. soror* infestation level was negatively related to scopolamine concentration, but positively related to the concentration level of the triterpenoid (S6). The effect of the total alkaloid concentration on *T. soror* infestation was also not significant (*L-R* chi-square$_1$ = 0.11, *AIC*c = 478.92; $p = 0.7367$; S6).

The GLM between *Epitrix* sp. infestation levels and the alkaloid concentrations in Ticumán was significant (*L-R* chi-square$_7$ = 17.89, *AIC*c = 134.49, $p = 0.0125$; Table 2). Significant effects included 3-hydroxy-6-tigloyloxytropane (negative effect) and pyrroline (positive effect; Table 2). The effect of the total alkaloid concentration on *Epitrix* sp. infestation was not significant (*L-R* chi-square$_1$ = 0.70, *AIC*c = 141.09; $p = 0.4000$; S6).

Natural selection on resistance, alkaloids and herbivore infestation in the two experimental sites

The GLM of relative fitness against resistance to herbivory was significant (L-R chi-square$_3$ = 20.77, AICc = 487.04; $p = 0.0001$). While there was no main effect of the population, the population × resistance interaction was significant, revealing that the relative fitness was positively related to resistance in Teotihuacán, but negatively in Ticumán (Table 1, S7 a, b).

The GLM of relative fitness as a function of the concentration of seven alkaloids was significant (L-R chi-square$_{15}$ = 26.14, AICc = 400.98; $p = 0.036$). There was a significant positive main effect of the pyrroline concentration on fitness (Table 1), but the experimental site × pyrroline interaction was also significant, revealing that the fitness decreased with increasing pyrroline concentration in Teotihuacán, whereas the opposite was true in Ticumán (Table 1, S8). Similarly, the significant experimental site × triterpenoid interaction revealed that fitness was positively related to triterpenoid concentration in Teotihuacán, but with a negative effect in Ticumán (Table 1, S8). The GLM of relative fitness against total alkaloid concentration was not significant in either of the populations (*L-R* chi-square$_3$ = 2.86, *AIC*c = 418.28; $p = 0.41$; Table 1).

The GLM of relative fitness against level of herbivore infestation (larvae and adults of *Lema*, *E. parvula* and *T. soror*) was significant in Teotihuacán (L-R chi-square$_4$ = 17.29, AICc = 342.00; p = 0.002) (Table 1). Nevertheless, only the negative effect of *L. daturaphila* larvae on fitness was significant (Table 1). Likewise, a significant positive effect of *Epitrix* sp. infestation level on fitness was detected in Ticumán (*L-R* chi-square$_1$ = 13.06, AICc = 130.18, p = 0.001; Table 1).

Discussion

Our results revealed differentiation in plant-herbivore interactions among the study sites. First, different herbivore species are present in each population, and the infestation levels and the amount of foliar damage exerted by each herbivore on plants differed within and between populations. Second, different chemical compounds were related to infestation by each specific herbivore. Third, variable spatial selection was detected on identity by descent (IBD), resistance, chemical defensive traits and herbivore infestation levels.

A number of studies have also documented geographic variation in the level of herbivory and chemical defenses (Castells *et al.* 2005; Muola *et al*. 2010; Agrawal *et al*. 2012; Züst *et al*. 2012; Castillo *et al.* 2014; Verçosa *et al.* 2019; Hanh *et al.* 2019). However, there has been no previous attempts to determine how the plants' genetic background (IBD) is driven by ongoing natural selection-imposed by herbivores. Furthermore, the results provide strong evidence of local adaptation in plant-herbivore interactions in both populations of *D. stramonium*.

In the locality of Teotihuacán, F_2 plants more related to the local grandparent (selected as a parent due to its low alkaloid concentration) were less resistant and had higher mortality due to herbivory than F_2 plants more related to the Ticumán grandparent. Furthermore, we did not

detect any relationship between fitness and IBD to the Teotihuacán grandparent in this site. This result was anticipated since it is not expected that natural selection would favor poorly defended plants in a habitat where damage by herbivores can be lethal (*e. g.*, plant deaths due to herbivory caused by larvae of *Lema daturaphila*) (Valverde *et al.* 2001; 2003; Fornoni *et al.* 2004). Hence, it is plausible that F_2 plants more related to the Teotihuacán grandparent inherited the loci/alleles that do not confer resistance (Albrechtsen *et al.* 2010; Browning and Browning 2012). Likewise, alkaloid concentration in F_2 plants more related to the Teotihuacán grandparent remained at low levels after damage by different herbivores in Teotihuacán. This result indicates that the chemical defenses studied here are not induced and have a genetic basis, since a positive significant relationship between plant resistance and IBD to the Teotihuacán grandparent would be expected if plant defenses were induced after herbivore damage (Baldwin 1998; Karban and Baldwin 2007).

In marked contrast, we detected strong positive selection on IBD to the Ticumán grandparent in the locality of Teotihuacán. Also, plant resistance to herbivores and IBD to the Ticumán grandparent were positively related in this site. We suggest that F_2 plants more related to the Ticumán grandparent - selected as a parent on the basis of its high alkaloid concentration - had higher survival in Teotihuacán as they inherited the loci/alleles that confer resistance to herbivores; positive selection of these loci/alleles would be associated with different defensive chemical compounds which are produced in high concentration (Albrechtsen *et al.* 2010; Lowry *et al.* 2019). In fact, our findings indicate that the higher resistance of F_2 plants more related to the Ticumán grandparent in the Teotihuacán site was provided by specific alkaloids that are produced in very high concentration to face different herbivore species. Total alkaloid concentration (sum of the concentration of all classes of alkaloids; Moore *et al.* 2014) only seems

to affect negatively the infestation levels of *E. parvula*. Since alkaloid concentrations vary in wild Teotihuacán plants (Castillo *et al*. 2014; Miranda-Pérez *et al*. 2016; De-la-Cruz *et al*. 2020), we think that wild plants from Teotihuacán that produce specific alkaloids in very high concentrations (*i. e*., plants more related to the Ticumán grandparent) have strong chemical defense against the herbivores in this site. For instance, we observed strong positive selection to increase the concentration of the triterpenoid compound in Teotihuacán, which seems to affect negatively the infestation levels of the most harmful herbivore of *D. stramonium*, the larvae of *L. daturaphila*.

The defensive role of specific alkaloids in the Teotihuacán site revealed unexpected results, namely, changing the sign of their relationship with the infestation by different herbivores. For instance, while the triterpenoid compound appears to reduce the infestation of *Lema* larvae (the most dangerous herbivore of *D. stramonium*), it was also positively associated with infestation levels by *E. parvula* and *T. soror*. Triterpenoids are structurally similar to insect hormones known as ecdysones (Oliveira *et al*. 2019) known to control metamorphosis as insects pass from larva to pupa to adult (Yamanaka *et al*. 2013). It has been reported that many triterpenoids function as ecdysone blockers (*e. g*., azadirone; Ujváry 2010; Oliveira *et al*. 2019). Therefore, the most parsimonious explanation for our observations is that this triterpenoid of *D. stramonium* is acting mainly on larvae of *Lema* (Miller *et al*. 1989; Ujváry 2010), and since this compound is structurally similar to insect hormones (Ujváry 2010), it may be used by *E. parvula* and *T. soror* adults to trace *D. stramonium* plants (and potential mates on them). Complex interactions where one compound is toxic to insects at one developmental stage (*e. g*., larvae) or to a particular herbivore species, but functioning as an attractant at other stage (adults) or to other herbivore species have been reported, for instance, in *Nicotiana attenuata* (Zhou *et al*. 2017).

Local adaptation of plant defenses to herbivores depends on (1) the strength of selection as a result of the interaction, (2) the level of specificity on the interaction (*e. g.*, folivores, seed predators, stem-borers) (Thompson *et al*. 2005; Cogni and Futuyma 2009; Agrawal *et al*. 2012). In the Teotihuacán site, our results suggest that the strong selection pressure exerted by one herbivore (the folivore *L. daturaphila*) on *D. stramonium* plants may affect the interaction between plants and other insects, leading to local adaptation of plant defenses to different herbivore species (Wise 2009, 2010).

On the other hand, in the Ticumán site, we detected strong positive selection on pyrroline alkaloid. It has been reported that pyrroline is a defensive compound against many insect species and pathogens (bacteria, virus, fungi) (Qamar *et al*. 2015; Martins *et al*. 2015; Tamariz *et al*. 2018). Pyrroline has also been related to different physiological processes such as plant growth (Chen *et al*. 2018; Tamariz *et al*. 2018). It is worth mentioning that polyamine oxidase, an enzyme involved in the biosynthesis of the pyrroline, is a growth-regulating enzyme (Chen *et al*. 2018). Nevertheless, an unexpected finding is that we observed a positive association between pyrroline concentration and infestation level by *Epitrix* in Ticumán. It has reported that some herbivore insects can tolerate pyrrolizidine alkaloids and use them for defense against their predators or as precursors of insect hormones (Martins *et al*. 2015). Thus, our most parsimonious explanation is that *Epitrix* sp. is surpassing the defensive role of the pyrroline alkaloid in Ticumán. This could explain why the F_2 plants more genetically related to the Ticumán grandparent (with higher concentration of pyrroline) had lower resistance towards *Epitrix* sp. infestation. Furthermore, since pyrroline could be positively related to plant growth (Chen *et al*. 2018; Tamariz *et al*. 2018), it is also possible that *Epitrix* sp. searches for more vigorous plants, which have more biomass to feed (Agrawal 2005; Wise and Rausher 2013). On the other hand,

we observed that 3-hydroxy-6-tigloyloxytropane negatively affected the infestation levels of *Epitrix* sp. in Ticumán. Then, it seems that the latter alkaloid is providing resistance against this herbivore in this site.

Interestingly, pyrroline affected negatively the infestation levels of *E. parvula* in Teotihuacán. However, negative selection on this compound was detected also in Teotihuacán. Thus, while *Epitrix* sp. appears to be adapted to this compound in Ticumán, the production of this compound in high concentrations in Teotihuacán may involve physiological costs, as plants also have to allocate resources for production of other compounds (*e. g.,* triterpenoid) to tackle their most harmful herbivore (*Lema* larvae). Indeed, as we mentioned above, it seems that total alkaloid concentration should be the option to face with *E. parvula* infestation in Teotihuacán.

De-la-Cruz *et al*. (2020) found that plants from Ticumán have on average higher alkaloid concentration than those in Teotihuacán. Why we did not observe strong selection to increase the IBD to the Ticumán grandparent (higher alkaloid concentration) in Ticumán? Our most parsimonious explanations are, first, as mentioned above, that *Epitrix* sp. (the main herbivore in this site) seems locally adapted to plant chemical defenses (pyrroline) in Ticumán, and that other alkaloids could now be providing defense against this herbivore. Second, since these compounds are expressed constitutively, it is possible that all these powerful chemical weapons are being used to face other natural enemies (virus, bacteria, nematodes, fungi, oomycete, other herbivore species) that we did not detect or that were not present during our experiment. Third, it is also possible that these compounds have other functions in this habitat (*e. g*., growth, plant-plant communication). For instance, recent genomic evidence from *D. stramonium* indicates that tropane alkaloids such as atropine and scopolamine also act as defenses against pathogens and viruses (De-la-Cruz *et al*. under review).

Finally, the lack of association between the IBD to the Teotihuacán grandparent with resistance or fitness in Ticumán suggests that the chemical defenses studied here are not induced (see above; Karban and Baldwin 2007).

Conclusions

The methodology used in this study allowed us to get insights on how natural selection imposed by herbivores drives the genetic underpinnings of plant resistance traits. The lack of association between plant fitness and IBD to the Teotihuacán grandparent (low resistance) in both populations, as well as different magnitude and direction of selection on the IBD to the Ticumán grandparent (high resistance) across populations, provides evidence of how ongoing natural selection operates on plant resistance and promotes local adaptation. Likewise, the results of this study shed some new light on how plants defend themselves against the attack from different herbivores. It seems that in populations where plants are suffering frequent or heavy damage by different herbivores, plants are able to produce and "use" different chemical defensive compounds to face each insect species that feed on them (Wittstock and Gershenzon 2002). The same alkaloids were produced by plants in both populations, but plants possessing the highest total alkaloid concentrations were not the most well-defended or fit in either of the populations. Instead, different specific alkaloids appear to be favored by natural selection imposed by herbivores in the two study populations.

Our results also revealed that the strong negative selection imposed by one herbivore species (*e. g.*, larvae of *L. daturaphila*) on plants likely affects interactions with other insects. Most importantly, the results provide evidence for local adaptation by showing that selection favors different loci/alleles related to plant resistance to herbivores in the two populations

(Briscoe Runquist *et al.* 2020). Hence, intraspecific diversity in secondary metabolites of *D. stramonium* seems to be maintained and selected to cope with varying local conditions among populations (Moore *et al*. 2014), giving rise to a geographic coevolutionary mosaic (Thompson 2005).

Declarations

Author contributions

Conceived and designed the experiments: JNF, IMDC. Performed the experiments: IMDC, JNF, PLV. Analyzed the data: IMDC, JM, JNF. Contributed reagents/materials/analysis tools: IMDC, JNF, CMFO, PLV, JM. Wrote the paper: IMDC, JNF, JM.

Acknowledgments

We are very grateful to M.Sc. Laura Cruz Rodríguez, Dr. Rosalinda Tapia-López, M.Sc. Luis Barbo Hernández-Portilla, Dr. Pasi Rastas, Dr. Antoine Fraimout for field and laboratory, logistic, and chemical and genetic analyses support. We are deeply grateful with two anonymous reviewers, Dr. Erik Sotka, Dr. David Hall and Dr. Luis E. Eguiarte which provided highly valuable corrections, comments and suggestions to improve this manuscript. Thanks to the members of the *Ecological genetics and evolution* lab for their help in the field. Thanks to Jacquelin DeFaveri at University of Helsinki language services for help in correcting the language. Ivan Mijail De la Cruz Arguello is graduate student in the "Programa de Doctorado en Ciencias Biomédicas, National Autonomous University of Mexico (UNAM)". IMDA acknowledges the scholarships granted by CONACyT (# 283799) and by the Finnish National Agency for Education (application KM-19-11065). We are most grateful by the strong financial support, to perform genomic research in *Datura*, provided by the grants "Fronteras de la

Ciencia" CONACyT Mexico (# 2016-2/1527) to JNF and the PAPIIT UNAM #IG200717 to JNF and CMFO.

Data availability

Scripts, commands and workflow to obtain the identity by descent (IBD) can be consulted in https://github.com/icruz1989/IBDcalculation. Raw data used in this manuscript will be stored in Dryad upon acceptance. Supplemental materials are provided in a separate file of this manuscript.

Ethics approval and consent to participate

All authors approved the manuscript

Competing interests

The authors declare that they have no competing interests

References

Agrawal, A. A. (2005). Natural selection on common milkweed (*Asclepias syriaca*) by a community of specialized insect herbivores. Evolutionary Ecology Research, 7(5), 651-667.

Agrawal, A. A., Hastings, A. P., Johnson, M. T. J., Maron, J. L., and Salminen, J.-P. (2012). Insect herbivores drive real-time ecological and evolutionary change in plant populations. Science, 338(6103), 113-116.

Albrechtsen, A., Moltke, I., and Nielsen, R. (2010). Natural selection and the distribution of Identity by Descent in the human genome. Genetics, 186(1), 295-308.

Albrechtsen, A., Sand Korneliussen, T., Moltke, I., van Overseem Hansen, T., Nielsen, F. C., and Nielsen, R. (2009). Relatedness mapping and tracts of relatedness for genome-wide data in the presence of linkage disequilibrium. Genetic Epidemiology, 33(3), 266-274.

Álvarez-Caballero, J. M., and Coy-Barrera, E. (2019). Chemical and antifungal variability of several accessions of *Azadirachta indica* A. Juss. from six locations across the Colombian caribbean coast: identification of antifungal azadirone Limonoids. Plants, 8(12), 555.

Anderson, J. T., Wagner, M. R., Rushworth, C. A., Prasad, K. V. S. K., and Mitchell-Olds, T. (2014). The evolution of quantitative traits in complex environments. Heredity, 112(1), 4-12.

Andrews, S., 2010 FastQC: a quality control tool for high throughput sequence data. http://www.bioinformatics.babraham.ac.uk/projects/fastqc

Akaike, H. (1974). A new look at the statistical model identification. IEEE Transactions on Automatic Control 19(6), 716-723.

Baldwin, I. T. (1998). Jasmonate-induced responses are costly but benefit plants under attack in native populations. Proceedings of the National Academy of Sciences, 95(14), 8113-8118.

Bello-Bedoy, R., and Núñez-Farfán, J. (2010). Cost of inbreeding in resistance to herbivores in *Datura stramonium*. Annals of Botany, 105(5), 747-753.

Berenbaum, M. R., Zangerl, A. R., and Nitao, J. K. (1986). Constraints on chemical coevolution: wild parsnips and the parsnip webworm. Evolution, 40(6), 1215-1228.

Briscoe Runquist, R. D., Gorton, A. J., Yoder, J. B., Deacon, N. J., Grossman, J. J., Kothari, S., Lyons, M. P., Sheth, S. N., Tiffin, P., and Moeller, D. A. (2020). Context dependence of local adaptation to abiotic and biotic environments: a quantitative and qualitative synthesis. The American Naturalist, 195(3), 412-431.

Bolger, A. M., Lohse, M., and Usadel, B. (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 30:2114-20.

Browning, S. R., and Browning, B. L. (2012). Identity by Descent between distant relatives: detection and applications. Annual Review of Genetics, 46(1), 617-633.

Cantelo, W. W., and Jacobson, M. (1979). Phenylacetaldehyde attracts moths to bladder flower and to blacklight traps. Environmental Entomology, 8(3), 444-447.

Castillo, G., Cruz, L. L., Tapia-López, R., Olmedo-Vicente, E., Carmona, D., Anaya-Lang, A. L., Fornoni, J., Andraca-Gómez, G., Valverde, P. L., and Núñez-Farfán, J. (2014). Selection mosaic exerted by specialist and generalist herbivores on chemical and physical defense of *Datura stramonium*. PLoS ONE, 9(7), e102478.

Castells, E., Berhow, M. A., Vaughn, S. F., and Berenbaum, M. R. (2005). Geographic variation in alkaloid production in *Conium maculatum* populations experiencing differential herbivory by *Agonopterix alstroemeriana*. Journal of Chemical Ecology, 31(8), 1693-1709.

Charlesworth, B. (1998). Measures of divergence between populations and the effect of forces that reduce variability. Molecular Biology and Evolution, 15(5), 538-543.

Chen, D., Shao, Q., Yin, L., Younis, A., & Zheng, B. (2019). Polyamine Function in Plants: Metabolism, Regulation on Development, and Roles in Abiotic Stress Responses. *Frontiers in Plant Science*, *9*.

Cogni, R., and Futuyma, D. J. (2009). Local adaptation in a plant herbivore interaction depends on the spatial scale. Biological Journal of the Linnean Society, 97(3), 494-502.

Dawkins, R., Krebs, J. R., Maynard Smith, J., and Holliday, R. (1979). Arms races between and within species. Proceedings of the Royal Society of London. Series B. Biological Sciences, 205(1161), 489-511.

De-la-Cruz, I. M., Cruz, L. L., Martínez-García, L., Valverde, P. L., Flores-Ortiz, C. M., Hernández-Portilla, L. B., and Núñez-Farfán, J. (2020). Evolutionary response to herbivory: population differentiation in microsatellite loci, tropane alkaloids and leaf trichome density in *Datura stramonium*. Arthropod-Plant Interactions, 14(1), 21-30.

de Villemereuil, P., Gaggiotti, O. E., Mouterde, M., and Till-Bottraud, I. (2016). Common garden experiments in the genomic era: New perspectives and opportunities. Heredity, 116(3), 249-254.

Doyle, J. J., and Doyle, J. L. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochemical Bulletin, 19, 11-15.

Ehrlich, P. R., and Raven, P. H. (1964). Butterflies and plants: a study in coevolution. Evolution, 18(4), 586-608.

Falconer, D. S., and Mackay, T. (1996). Introduction to quantitative genetics. Harlow: Longmans Green. 386 pp.

Flood, P. J., and Hancock, A. M. (2017). The genomic basis of adaptation in plants. Current Opinion in Plant Biology, 36, 88-94.

Fornoni, J., and Núñez-Farfán, J. (2000). Evolutionary ecology of *Datura stramonium*: genetic variation and costs for tolerance to defoliation. Evolution, 54(3), 789-797.

Fornoni, Juan, Valverde, P. L., and Núnez-Fárfan, J. (2004). Population variation in the cost and benefit of tolerance and resistance against herbivory in *Datura stramonium*. Evolution, 58(8), 1696-1704.

Gomulkiewicz, R., Thompson, J. N., Holt, R. D., Nuismer, S. L., and Hochberg, M. E. (2000). Hot Spots, Cold Spots, and the Geographic Mosaic Theory of Coevolution. The American Naturalist, 156(2), 156–174.

González-Coloma, López-Balboa, A., Santana, C., Reina, O., and Fraga, B. M. (2011). Triterpene-based plant defenses. Phytochemistry Reviews, 10(2), 245-260.

Hahn, P. G., Agrawal, A. A., Sussman, K. I., and Maron, J. L. (2019). Population variation, environmental gradients, and the evolutionary ecology of plant defense against herbivory. The American Naturalist, 193(1), 20-34.

Ionkova, I., Witte L., and Alfermann H.-A. (1994). Spectrum of tropane alkaloids in transformed roots of *Datura innoxia* and *Hyoscyamus x gyorffyi* cultivated in vitro. Planta Medica, 60, 382-384.

Janz, N. (2011). Ehrlich and Raven revisited: Mechanisms underlying codiversification of plants and enemies. Annual Review of Ecology, Evolution, and Systematics, 42(1), 71-89.

Karban, R., and Baldwin, I. T. (2007). Induced responses to herbivory. University of Chicago Press.

Kariñho-Betancourt, E., Agrawal, A. A., Halitschke, R., and Núñez-Farfán, J. (2015). Phylogenetic correlations among chemical and physical plant defenses change with ontogeny. New Phytologist, 206(2), 796-806.

Kawecki, T. J., and Ebert, D. (2004). Conceptual issues in local adaptation. Ecology Letters, 7(12), 1225-1241.

Lande, R. (1979). Quantitative genetic analysis of multivariate evolution applied to brain: body size allometry. Evolution, 33(1Part2), 402-416.

Lande, R., and Arnold, S. J. (1983). The measurement of selection on correlated characters. Evolution, 37(6), 1210-1226.

Li, H., and Durbin, R. (2009) Fast and accurate short read alignment with Burrows-Wheeler Transform. Bioinformatics, 25:1754-60.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and 1000 Genome Project Data Processing Subgroup. The sequence alignment/map (SAM) format and SAMtools, Bioinformatics (2009) 25(16) 2078-9.

Lowry, D. B., Lovell, J. T., Zhang, L., Bonnette, J., Fay, P. A., Mitchell, R. B., Lloyd-Reilley, J., Boe, A. R., Wu, Y., Rouquette, F. M., Wynia, R. L., Weng, X., Behrman, K. D., Healey, A., Barry, K., Lipzen, A., Bauer, D., Sharma, A., Jenkins, J., … Juenger, T. E. (2019). QTL × environment interactions underlie adaptive divergence in switchgrass across a large latitudinal gradient. Proceedings of the National Academy of Sciences, 116(26), 12933-12941.

Martins, C. H. Z., Cunha, B. P., Solferini, V. N., and Trigo, J. R. (2015). Feeding on host plants with different concentrations and structures of pyrrolizidine alkaloids impacts the chemical-defense effectiveness of a specialist herbivore. PLoS ONE, 10(10): e0141480.

Machado, B. B., Orue, J. P. M., Arruda, M. S., Santos, C. V., Sarath, D. S., Goncalves, W. N., Silva, G. G., Pistori, H., Roel, A. R., and Rodrigues-Jr, J. F. (2016). BioLeaf: a professional mobile application to measure foliar damage caused by insect herbivory. Computers and Electronics in Agriculture, 129, 44-55.

Mauricio, R., and Rausher, M. D. (1997). Experimental manipulation of putative selective agents provides evidence for the role of natural enemies in the evolution of plant defense. Evolution, 51(5), 1435-1444.

Miller, A. J., and Chamberlain, W. F. (1989). Azadirachtin as a larvicide against the horn fly, stable fly, and house fly (Diptera: Muscidae). Journal of Economic Entomology, 82(5), 1375-1378.

Miranda-Pérez, A., Castillo, G., Hernández-Cumplido, J., Valverde, P. L., Borbolla, M., Cruz, L. L., Tapia-López, R., Fornoni, J., Flores-Ortiz, C. M., and Núñez-Farfán, J. (2016). Natural selection drives chemical resistance of *Datura stramonium*. PeerJ, 4.

Moore, B. D., Andrew, R. L., Külheim, C., and Foley, W. J. (2014). Explaining intraspecific diversity in plant secondary metabolites in an ecological context. New Phytologist, 201(3), 733-750.

Motten, A. F., and Antonovics, J. (1992). Determinants of outcrossing rate in a predominantly self-fertilizing weed, *Datura stramonium* (Solanaceae). American Journal of Botany, 79(4), 419-427.

Muola, A., Mutikainen, P., Lilley, M., Laukkanen, L., Salminen, J.-P., and Leimu, R. (2010). Associations of plant fitness, leaf chemistry, and damage suggest selection mosaic in plant–herbivore interactions. Ecology, 91(9), 2650-2659.

Núnez-Farfán, J., Cabrales-Vargas, R. A., and Dirzo, R. (1996). Mating system consequences on resistance to herb ivory and life history traits in *Datura stramonium*. American Journal of Botany, 83(8), 1041-1049.

Núñez-Farfán, J., and Dirzo, R. (1994). Evolutionary ecology of *Datura Stramonium* L. in central Mexico: natural selection for resistance to herbivorous insects. Evolution, 48(2), 423-436.

Núñez-Farfán, J., Fornoni, J., and Valverde, P. L. (2007). The Evolution of resistance and tolerance to herbivores. Annual Review of Ecology, Evolution, and Systematics, 38(1), 541-566.

Rastas, P. (2017). Lep-MAP3: Robust linkage mapping even for low-coverage whole genome sequencing data, Bioinformatics, 33(23), 3726-3732.

Oliveira, D. A. B. de, and Niculau, A. V. da S. and E. dos S. (2019). Molecular docking of Azadirachtin in nuclear Ecdysone receptor. Current Physical Chemistry, 6(1), 50-57.

Pedrol, N., and Tiburcio, A. F. (2001). Polyamines determination by TLC and HPLC. In M. J. Reigosa Roger (Ed.), Handbook of Plant Ecophysiology Techniques (pp. 335-363). Springer Netherlands.

Qamar, A., Mysore, K. S., & Senthil-Kumar, M. (2015). Role of proline and pyrroline-5-carboxylate metabolism in plant defense against invading pathogens. *Frontiers in Plant Science*, *6*.

Savolainen, O., Lascoux, M., and Merilä, J. (2013). Ecological genomics of local adaptation. Nature Reviews Genetics, 14(11), 807-820.

Shonle, I., and Bergelson, J. (2000). Evolutionary Ecology of the tropane alkaloids of *Datura stramonium* L. (Solanaceae). Evolution, 54(3), 778-788.

Sokal, R. R., and Rohlf, F. J. (1994). Biometry: The principles and practice of statistics in biological research. W H Freeman & Co.

Stam, J. M., Kroes, A., Li, Y., Gols, R., van Loon, J. J. A., Poelman, E. H., and Dicke, M. (2014). Plant interactions with multiple insect herbivores: from community to genes. Annual Review of Plant Biology, 65(1), 689-713.

Tamariz, J., Burgueño-Tapia, E., Vázquez, M. A., and Delgado, F. (2018). Pyrrolizidine alkaloids. In H.-J. Knölker (Ed.), The alkaloids: chemistry and biology (Vol. 80, pp. 1-314). Academic Press.

Thompson, J. N., and Cunningham, B. M. (2002). Geographic structure and dynamics of coevolutionary selection. Nature, 417(6890), 735-738.

Thompson, E. A. (2013). Identity by descent: variation in meiosis, across genomes, and in populations. Genetics, 194(2), 301-326.

Thompson, J. N. (2005). Coevolution: the geographic mosaic of coevolutionary arms races. Current Biology, 15(24), R992-R994.

Ujváry, I. (2010). Chapter 3-pest control agents from natural products. In R. Krieger (Ed.), Hayes' Handbook of Pesticide Toxicology (Third Edition) (pp. 119-229). Academic Press.

Valverde, P. L., Fornoni, J., and Núñez-Farfán, J. (2001). Defensive role of leaf trichomes in resistance to herbivorous insects in *Datura stramonium*. Journal of Evolutionary Biology, 14(3), 424-432.

Valverde, P. L., Fornoni, J., and Núñez-Farfán, J. (2003). Evolutionary ecology of *Datura stramonium*: Equal plant fitness benefits of growth and resistance against herbivory. Journal of Evolutionary Biology, 16(1), 127-137.

Verçosa, D., Cogni, R., Alves, M. N., and Trigo, J. R. (2019). The geographical and seasonal mosaic in a plant-herbivore interaction: Patterns of defences and herbivory by a specialist and a non-specialist. Scientific Reports, 9(1), 15206.

Wickham, H. (2016). ggplot2: Elegant graphics for data analysis (2nd ed.). Springer International Publishing.

Wise, Michael Joseph. (2009). Competition among herbivores of *Solanum carolinense* as a constraint on the evolution of host-plant resistance. Evolutionary Ecology, 23(3), 347–361.

Wise, Michael J. (2010). Diffuse interactions between two herbivores and constraints on the evolution of resistance in horsenettle (*Solanum carolinense*). Arthropod-Plant Interactions, 4(3), 159-164.

Wise, M. J., and Rausher, M. D. (2013). Evolution of resistance to a multiple-herbivore community: genetic correlations, diffuse coevolution, and constraints on the plant's response to selection. Evolution, 67(6), 1767-1779.

Wittstock, U., and Gershenzon, J. (2002). Constitutive plant toxins and their role in defense against herbivores and pathogens. Current Opinion in Plant Biology, 5(4), 300-307.

Witte, L., Müller, K., and Arfmann, H. A. (1987). Investigation of the alkaloid pattern of *Datura innoxia* plants by capillary gas-liquid-chromatography-mass spectrometry. Planta Medica, 53:192-197.

Yamanaka, N., Rewitz, K. F., and O'Connor, M. B. (2013). Ecdysone control of developmental transitions: lessons from drosophila research. Annual Review of Entomology, 58(1), 497-516.

Zar, J. H. (1999). Biostatistical Analysis. Prentice Hall.

Zhou, W., Kügler, A., McGale, E., Haverkamp, A., Knaden, M., Guo, H., Beran, F., Yon, F., Li, R., Lackus, N., Köllner, T. G., Bing, J., Schuman, M. C., Hansson, B. S., Kessler, D., Baldwin, I. T., and Xu, S. (2017). Tissue-specific emission of (E)-α-Bergamotene helps resolve the dilemma when pollinators are also herbivores. Current Biology, 27(9), 1336-1341.

Züst, T., Heichinger, C., Grossniklaus, U., Harrington, R., Kliebenstein, D. J., and Turnbull, L. A. (2012). Natural enemies drive geographic variation in plant defenses. Science, 338(6103), 116-119.

Table 1. Analyses of natural selection testing the effects of (a) the seven alkaloids, (b) total alkaloid concentration, (c) Identity by Descent to Ticumán grandparent, (d) Identity by Descent to Teotihuacán grandparent, (e) resistance, (f) herbivores at Teotihuacán and (g) *Epitrix* sp. infestation at Ticumán. N = number of individuals, d.f. = degrees of freedom, βi = selection gradients (generalized linear coefficients), se = standard error, t = t-ratio. Significant p-values (p) are in bold. Pop = effect of the experimental site; TEO = Teotihuacán.

Response variable		Effects	N	d.f.	βi	se	t	p
(a)	w_i fitness	3-hydroxy-6-tigloyloxytropane	136	1	-0.36	0.21	-1.71	0.0899
		Anisodamine	136	1	-0.07	0.28	-0.25	0.7961
		Atropine	136	1	-0.07	0.12	-0.56	0.5719
		Triterpenoid	136	1	0.26	0.18	1.46	0.1459
		Scopolamine	136	1	-0.24	0.19	-1.24	0.2151
		Phenylacetaldehyde	136	1	-0.21	0.21	-0.98	0.3254
		Pyrroline	136	1	**0.32**	0.14	2.19	**0.0304**
		Pop (TEO)	136	1	-0.12	0.09	-1.29	0.1997
		Pop (TEO) × 3-hydroxy-6-tigloyloxytropane	136	1	0.13	0.21	0.62	0.5362
		Pop (TEO) × Anisodamine	136	1	-0.06	0.28	-0.24	0.8104
		Pop (TEO) × Atropine	136	1	0.19	0.12	1.55	0.1224
		Pop (TEO) × Triterpenoid	136	1	**0.38**	0.18	2.10	**0.0378**
		Pop (TEO) × Scopolamine	136	1	-0.04	0.19	-0.20	0.8375
		Pop (TEO) × Phenylacetaldehyde	136	1	-0.14	0.21	-0.67	0.5015
		Pop (TEO) × Pyrroline	136	1	**-0.43**	0.14	-2.93	**0.0040**
(b)	w_i fitness	Total alkaloid concentration	144	1	-0.05	0.08	-0.58	0.5657
		Pop (TEO)	144	1	-0.12	0.09	-1.34	0.1838
		Pop (TEO) × Total alkaloid concentration	144	1	0.09	0.08	1.04	0.2988
(c)	w_i fitness	IBD-Ticumán grandparent	135	3	**2.19**	0.99	2.20	**0.0302**
		Pop (TEO)	135	3	-0.13	0.10	-1.22	0.2262
		Pop (TEO) × IBD	135	3	1.39	0.99	1.40	0.1646
(d)	w_i fitness	IBD-Teotihuacán grandparent	135	3	0.41	0.89	0.46	0.6484
		Pop (TEO)	135	3	-0.17	0.11	-1.49	0.1390
		Pop (TEO) × IBD	135	3	-0.71	0.89	-0.80	0.4252
(e)	w_i fitness	Resistance	177	1	0.11	0.27	0.43	0.6624
		Pop (TEO)	177	1	-0.15	0.31	-0.50	0.6143
		Pop (TEO) × Resistance	177	1	**0.64**	0.27	2.36	**0.0179**
(f)	w_i fitness	Adults of *Lema daturaphila*	113	1	0.08	0.09	0.90	0.3692
		Adults of *Epitrix parvula*	113	1	0.09	0.09	0.97	0.3324
		Adults of *Trichobaris soror*	113	1	0.09	0.08	1.03	0.3070
		Larvae of *Lema daturaphila*	113	1	**-0.27**	0.08	-3.22	**0.0017**
(g)	w_i fitness	Adults of *Epitrix* sp.	63	1	**0.30**	0.081	3.75	**0.0004**

Table 2. General linear models testing the effect of (a) Identity by Descent to the Ticumán grandparent, (b) Identity by Descent to the Teotihuacán grandparent, (c) herbivore infestation at Teotihuacán and (d) *Epitrix* sp. infestation at Ticumán, on whole plant resistance. N = number of individuals, d.f. = degrees of freedom, Estimate = generalized linear coefficients, se = standard error, *t* = t-ratio. Significant *p*-values (*p*) are in bold. Pop = effect of the experimental site; TEO = Teotihuacán.

Response variable		Effects	N	d.f.	Estimate	se	*t* ratio	*p*
(a)	Resistance	IBD-Ticumán grandparent	182	3	0.98	0.64	1.52	0.1256
		Pop (TEO)	182	3	**-0.40**	0.06	-5.95	**0.0001**
		Pop (TEO) × IBD	182	3	**2.06**	0.64	3.18	**0.0015**
(b)	Resistance	IBD-Teotihuacán grandparent	182	3	0.31	0.48	0.64	0.5182
		Pop (TEO)	182	3	**-0.40**	0.07	-5.69	**0.0001**
		Pop (TEO) × IBD	182	3	0.51	0.48	1.05	0.2896
(c)	Resistance	Adults of *Lema daturaphila*	185	1	0.02	0.05	0.38	0.7009
		Epitrix parvula	185	1	**0.27**	0.06	4.51	**0.0001**
		Trichobaris soror	185	1	0.03	0.05	0.53	0.5944
		Larvae of *Lema daturaphila*	185	1	**-0.52**	0.05	-9.00	**0.0001**
(d)	Resistance	*Epitrix* sp.	82	1	**-0.40**	0.10	-4.00	**0.0001**

Figure legends

Fig. 1. Relationships between plant fitness (log scale) and identity by descent (IBD) relationship to the Ticumán grandparent for (a) plants grown in the Teotihuacán population and (b) plants grown in the Ticumán population. Relationships between plant fitness and IBD relationship to the Teotihuacán grandparent for (c) plants grown in the Teotihuacán population and (d) plants grown in the Ticumán population. (e) Median IBD to the Ticumán grandparent for plants that survived and died. (f) Median IBD to the Teotihuacán grandparent for plants that survived and died. (g) Distribution of the relatedness between all F_2 full-sibs, and relatedness of F_2 plants to each grandparent. A relatedness of ~0.5 is the mean expected value between all F_2 full-sibs. A relatedness of ~0.5 is the maximum expected value between the F_2 progeny and each one of the grandparents (Falconer and Mackay 1996). *p*-values of full GLMs are shown in each plot (figures a-f). Each dot depicts observation for an individual. TEO = Teotihuacán, TIC = Ticumán. See also Table 1.

Fig. 2. Relationships between resistance to herbivory and identity by descent (IBD) relationship to the Ticumán grandparent for (a) plants grown in the Teotihuacán population and (b) plants grown in the Ticumán population. Relationships between resistance to herbivory and IBD relationship to the Teotihuacán grandparent for (c) plants grown in the Teotihuacán population and (d) plants grown in the Ticumán population. Relationship between resistance and (e) adults of *Epitrix parvula* in Teotihuacán, (b) larvae of *Lema daturaphila* in Teotihuacán, (c) adults of *Epitrix* sp. in Ticumán. *p*-values of GLMs are shown in each plot (a-g). Each dot depicts observation for an individual. TEO = Teotihuacán, TIC = Ticumán. See also Table 2.

Fig. 1

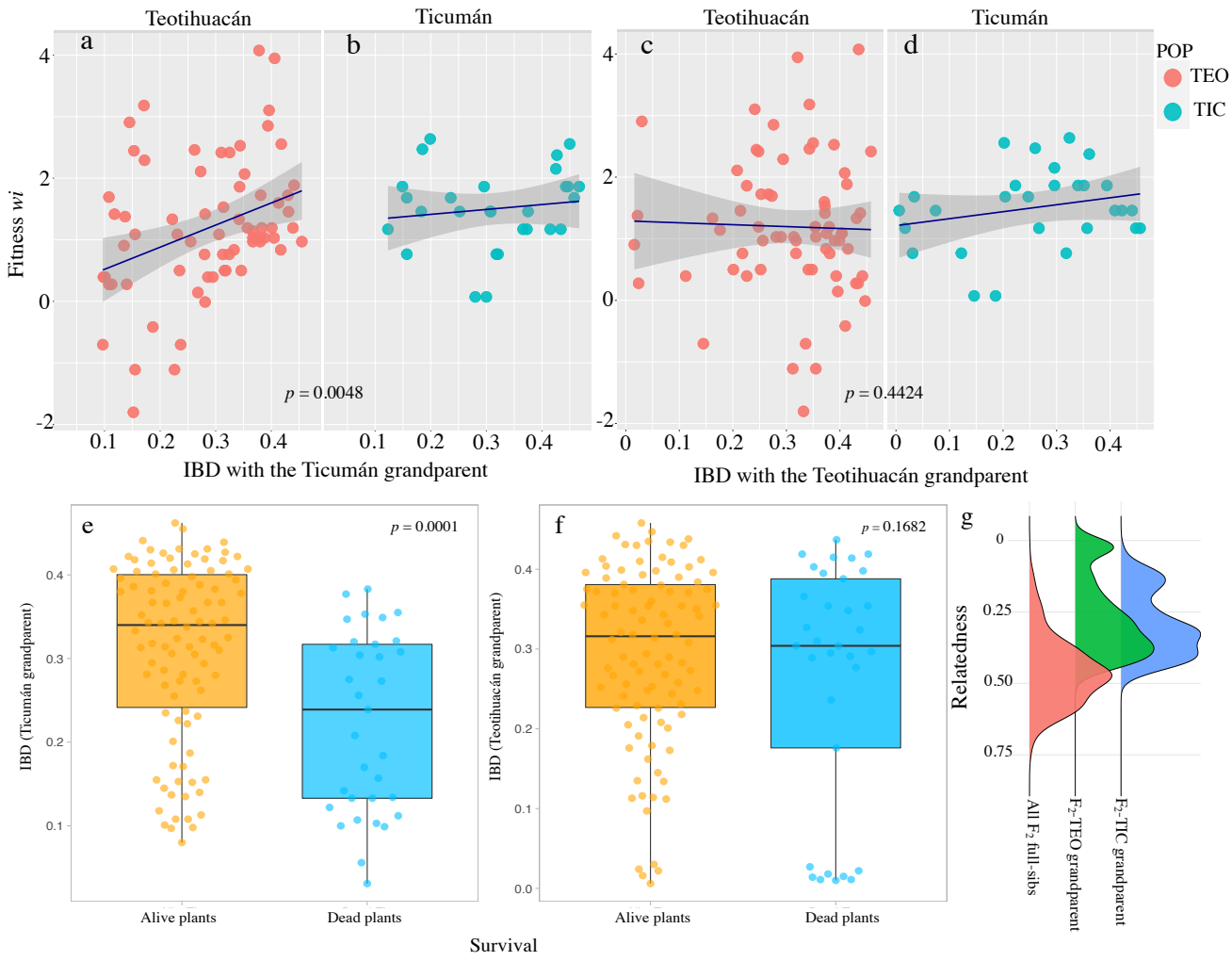
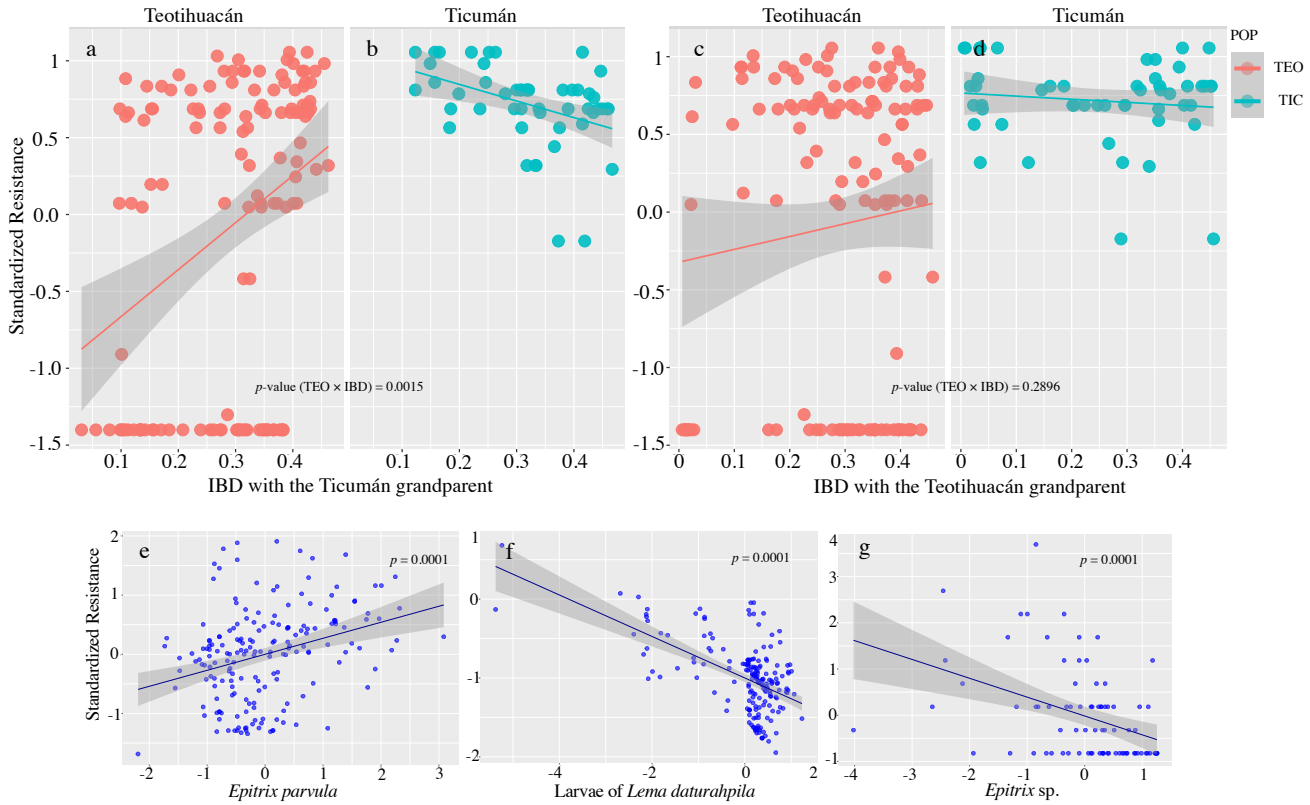
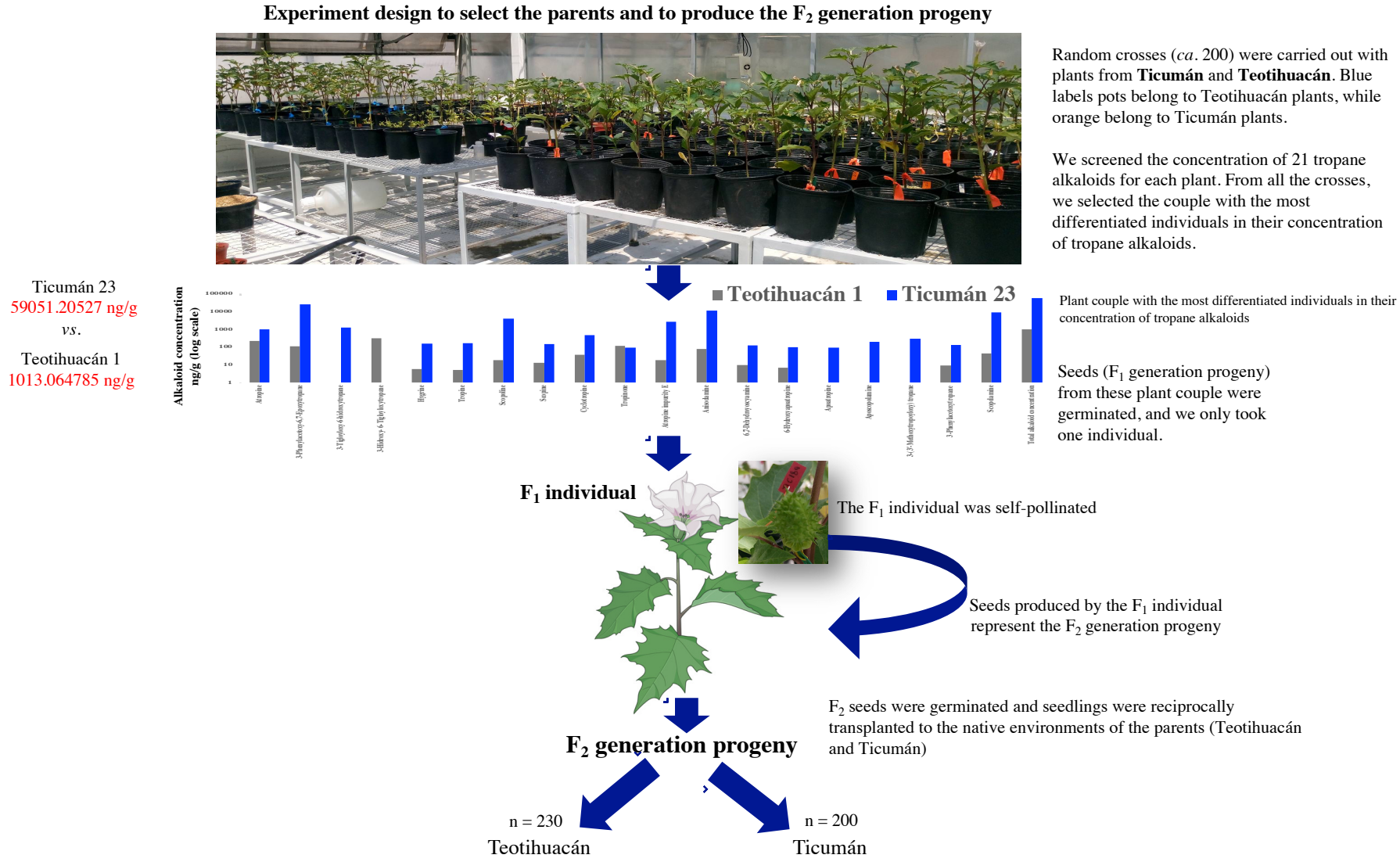

Teotihuacán
Ticumán
Teotihuacán
Ticumán
a
b
c
d
POP
TEO
TIC
Fitness wi
p = 0.0048
p = 0.4424
IBD with the Ticumán grandparent
IBD with the Teotihuacán grandparent
e
f
g
p = 0.0001
p = 0.1682
IBD (Ticumán grandparent)
IBD (Teotihuacán grandparent)
Alive plants
Dead plants
Survival
Relatedness
All F2 full-sibs
F2-TEO grandparent
F2-TIC grandparent

Fig. 2


Teotihuacán
Ticumán
Teotihuacán
Ticumán
a
b
c
d
POP
TEO
TIC
Standardized Resistance
p-value (TEO × IBD) = 0.0015
p-value (TEO × IBD) = 0.2896
IBD with the Ticumán grandparent
IBD with the Teotihuacán grandparent
e
f
g
p = 0.0001
p = 0.0001
p = 0.0001
Standardized Resistance
Epitrix parvula
Larvae of Lema daturahpila
Epitrix sp.

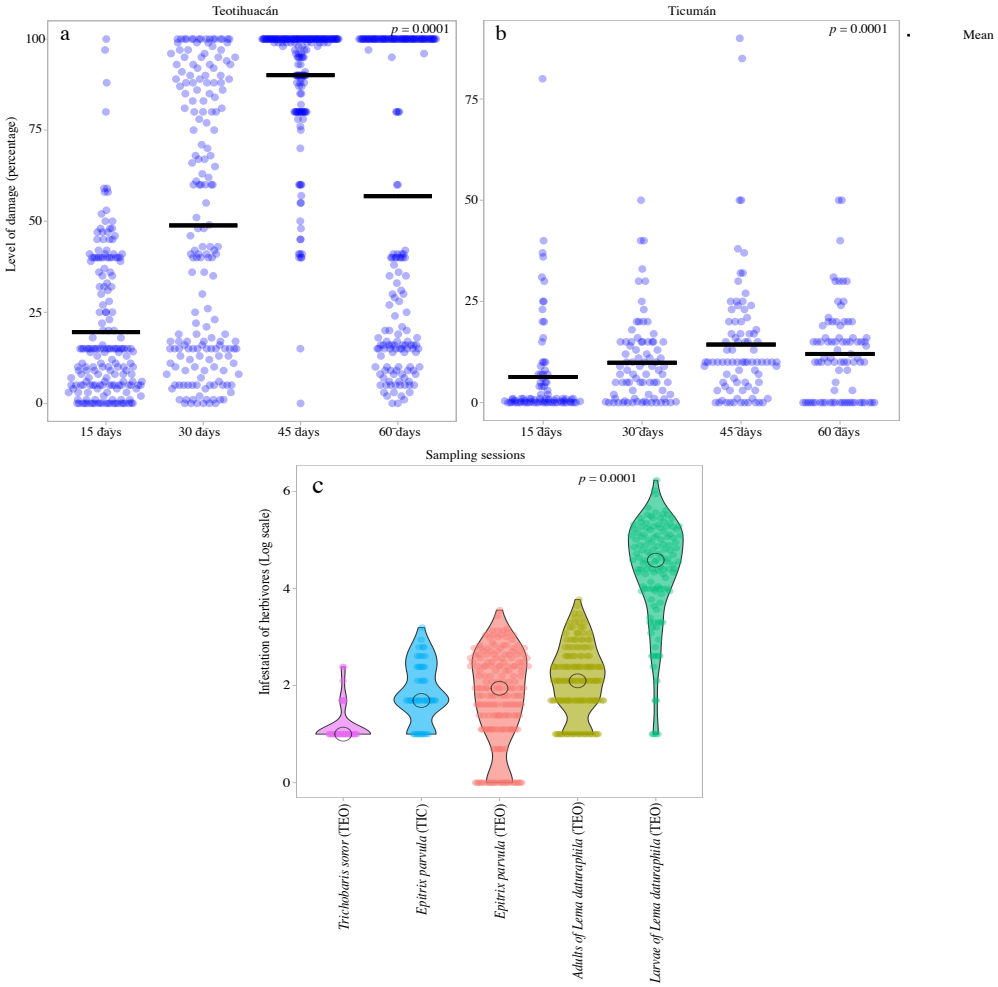
S1. Depiction of the experimental design used to produce the F_2 generation progeny used in the study (*see* Methods for details).



S2. Alkaloids identified in leaves of *Datura stramonium*. RT, the retention time of each alkaloid; *m/z* = mass/charge, MS = mass spectrometry reference.

Alkaloid	Formula	RT (min)	*m/z*	MS Ref.
3-Tigloyloxy-6-Hydroxytropane	$C_{13}H_{21}NO_3$	10.1	240.1594	Witte 1987
6-Hydroxyhyoscyamine (Anisodamine)	$C_{17}H_{23}NO_4$	13.2	306.1699	Ionkova *et al*. 1994
Atropine (Hyosciamine)	$C_{17}H_{23}NO_3$	8.7	290.1751	Witte 1987
Scopolamine	$C_{17}H_{21}NO_4$	15	304.1543	Witte 1987
Pyrroline	C_4H_7N	5.5	70.0655	Pedrol and Tiburcio 2001
Phenylacetaldehyde	C_8H_8O	9.4	121.0641	Cantelo and Jacobson 1979
Triterpenoid	$C_{28}H_{36}O_4$	15.6	437.2721	González-Coloma *et al*. 2011

S3. Plant damage (in percentage) experienced by each F_2 plant during four sampling dates at (a) Teotihuacán and (b) Ticumán. Plants experienced the most severe damage at 45 days after transplanting in the two localities, and damage levels were higher in Teotihuacán than in Ticumán. (c) Violin plot showing total infestation accounted by different herbivore (log scale). The circle inside each violin depicts the mean value. Overall, *Lema daturaphila* larvae was the most abundant insect herbivore. A quasi-random jittering was used to reduce datapoint overlap. TEO = Teotihuacán, TIC = Ticumán. *p*-values of ANOVAs are showed in each plot. For figures (a) and (b): black line = mean.



S4. Mean differentiation in the level of damage (expressed as percentage) between sampling sessions in (a) Teotihuacán and (b) Ticumán. (c) Mean differentiation in the level of herbivore infestation (Log transformed data). N = number of individuals, se = standard error, F = Fisher-statistic, p = p-values of ANOVAs.

ANOVA	N	Mean	se	F	p
(a) Damage at Teotihuacán					
15 days	164	19.55	0.07	110.98	0.0001
30 days	179	48.84	0.07		
45 days	184	90.08	0.07		
60 days	183	56.86	0.07		
(b) Damage at Ticumán					
15 days	83	6.40	0.15	27.16	0.0001
30 days	83	9.96	0.14		
45 days	83	14.51	0.13		
60 days	82	12.15	0.15		
(c) Herbivore					
E. parvula (TEO)	185	1.74	0.06	215.32	0.0001
Epitrix sp. (TIC)	47	1.87	0.12		
Adults of *L. daturaphila* (TEO)	143	2.15	0.07		
Larvae of *L. daturaphila* (TEO)	131	4.36	0.07		
T. soror	40	1.16	0.13		

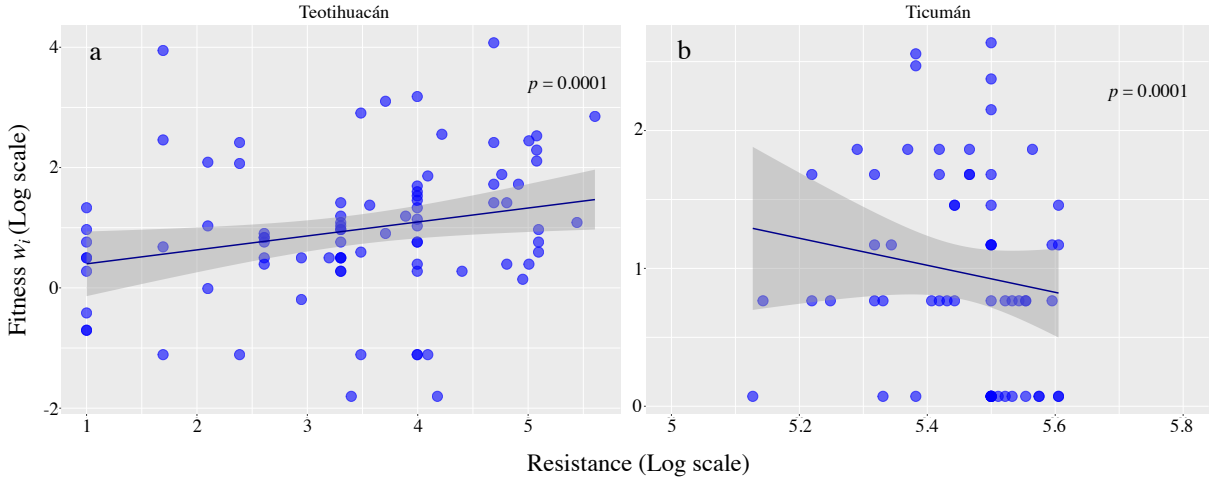
S5. Correlations between infestation by each herbivore and plant leaf average damage. a = Teotihuacán and b = Ticumán. ALd = adults of *Lema daturaphila*, Ep = *Epitrix parvula*, Ts = *Trichobaris soror*, LLd = larvae of *Lema daturaphila*.

Variable	by Variable	Correlation	Lower 95%	Upper 95%	*p*-value
		(a) Teotihuacán			
Damage	LLd	0.615779	0.517482	0.698035	**1.079E-20**
Ts	Ep	0.241138	0.100361	0.37245	**9.446E-04**
Ep	ALd	0.168386	0.024718	0.305239	**2.195E-02**
Ts	ALd	0.137232	-0.00718	0.276036	**6.250E-02**
LLd	ALd	-0.03421	-0.1776	0.110605	**6.439E-01**
Damage	Ts	-0.123	-0.26261	0.021649	**9.531E-02**
Damage	ALd	-0.13037	-0.26956	0.014169	**7.694E-02**
LLd	Ts	-0.15396	-0.29175	-0.00991	**3.640E-02**
LLd	Ep	-0.31356	-0.43802	-0.17731	**1.386E-05**
Damage	Ep	-0.32261	-0.44611	-0.18704	**7.516E-06**
		(b) Ticumán			
Damage	*Epitrix* sp.	0.387549	0.187536	0.556704	**2.949E-04**

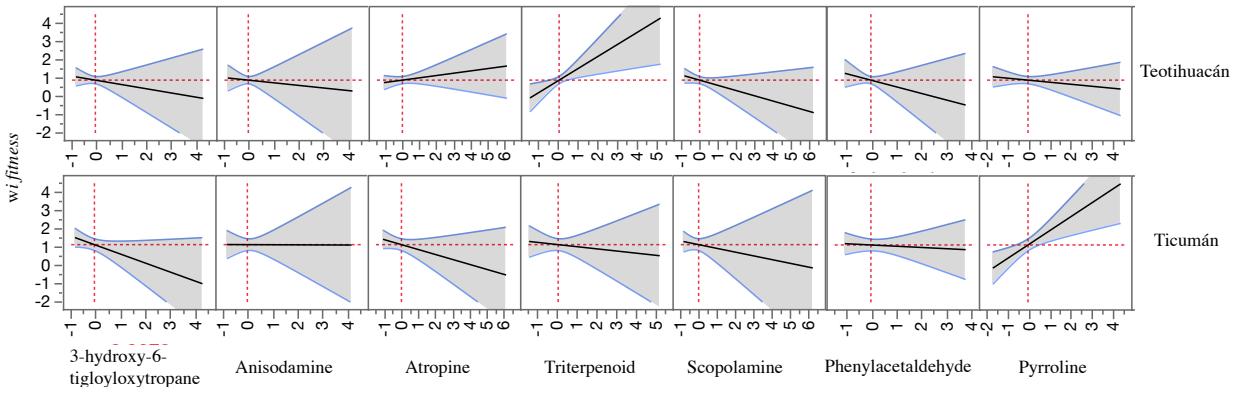
S6. Generalized linear models (GLMs) between herbivore infestation levels and the seven alkaloids and total alkaloid concentration: (a) infestation by adults of *E. pavula*, (b) stepwise GLM between larvae of *Lema* and alkaloids. (c) stepwise GLM between adults of *Lema* and alkaloids, (d) stepwise GLM between adults of *T. soror* and alkaloids. (e) GLM between *Epitrix* sp. and alkaloids in Ticumán. Significant *p*-values are in bold. N = number of individuals, se = standard error. Notice that for the total alkaloid concentration was performed a different model for each herbivore (*i. e.*, the effect of this variable was not included in the GLMs between each herbivore and the seven alkaloids (see methods).

Teotihuacán								
Response variable		**Effects**	**N**	**d.f.**	**Estimate**	**se**	***t***	*p*
(a)	# of adult of *Epitrix parvula*	3-hydroxy-6-tigloyloxytropane	153	1	-0.29	0.17	3.09	0.0808
		Phenylacetaldehyde	153	1	**-0.32**	0.14	5.24	**0.0234**
		Pyrroline	153	1	**-0.26**	0.12	4.41	**0.0375**
		Triterpenoid	153	1	**0.46**	0.18	6.12	**0.0145**
		Scopolamine	153	1	-0.22	0.16	1.78	0.1831
		Anisodamine	153	1	0.27	0.21	1.65	0.1997
		Atropine	153	1	0.01	0.11	0.00	0.9272
		Total alkaloid concentration	161	1	**-0.15**	0.079	-1.98	**0.0496**
(b)	# of larvae of *Lema daturaphila*	Phenylacetaldehyde	156	1	**0.31**	0.13	2.38	**0.0188**
		Pyrroline	156	1	0.10	0.12	0.88	0.3820
		Triterpenoid	156	1	**-0.56**	0.17	-3.24	**0.0015**
		Scopolamine	156	1	0.08	0.16	0.48	0.6326
		Anisodamine	156	1	0.06	0.12	0.55	0.5837
		Total alkaloid concentration	161	1	-0.09	0.08	-1.20	0.2308
(c)	# adult of *Lema daturaphila*	3-hydroxy-6-tigloyloxytropane	156	1	**-0.24**	0.09	-2.65	**0.0089**
		Phenylacetaldehyde	156	1	-0.08	0.09	-0.84	0.4031
		Atropine	156	1	**0.25**	0.10	2.45	**0.0152**
		Total alkaloid concentration	161	1	0.05	0.08	0.70	0.4832
(d)	# adult of *Trichobaris soror*	Phenylacetaldehyde	158	1	-0.17	0.13	-1.29	0.1978
		Triterpenoid	158	1	**0.33**	0.16	2.05	**0.0419**
		Scopolamine	158	1	**-0.37**	0.14	-2.65	**0.0088**
		Atropine	158	1	0.13	0.11	1.16	0.1978
		Total alkaloid concentration	161	1	-0.02	0.08	-0.33	0.7387
Ticumán								
(e)	# of adult of *Epitrix* sp.	3-hydroxy-6-tigloyloxytropane	48	1	**-0.43**	0.20	4.56	**0.0388**
		Phenylacetaldehyde	48	1	0.044	0.17	0.06	0.7982
		Pyrroline	48	1	**0.54**	0.18	8.52	**0.0057**
		Triterpenoid	48	1	-0.19	0.20	0.90	0.3463
		Scopolamine	48	1	-0.21	0.18	1.32	0.2566
		Anisodamine	48	1	0.14	0.24	0.34	0.5596
		Atropine	48	1	-0.00	0.16	0.00	0.9807
		Total alkaloid concentration	50	1	0.11	0.15	0.83	0.4412

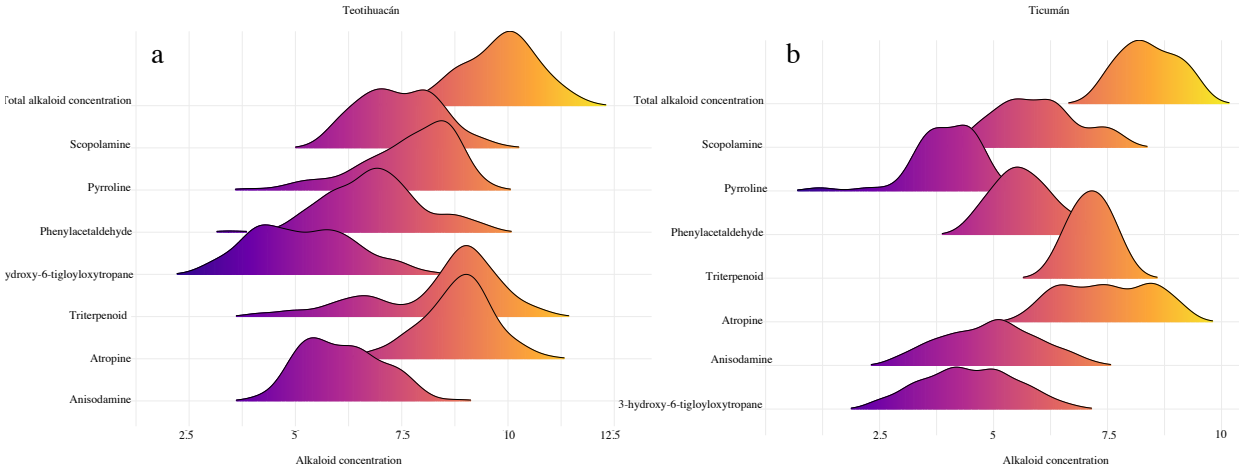
S7. Relationships between fitness and resistance in (a) F_2 plants grown in Teotihuacán and (b) F_2 plants grown in Ticumán. Positive selection to resistance was observed in Teotihuacán while a negative trend was detected in Ticumán. Each dot depicts observation for an individual. *p*-value of full model is shown in the plots. See also Table 1.



S8. Prediction profilers from the GLM between fitness (w_i) and concentrations of the seven alkaloids. Response effects are shown separately for the two experimental sites. Envelops = confidence intervals (95% CI).



S9. Ridgeline plot showing the distribution of the concentration of the seven alkaloids and total alkaloid concentration in (a) F_2 plants grown in Teotihuacán and (b) F_2 plants grown in Ticumán. Data was log-transformed.



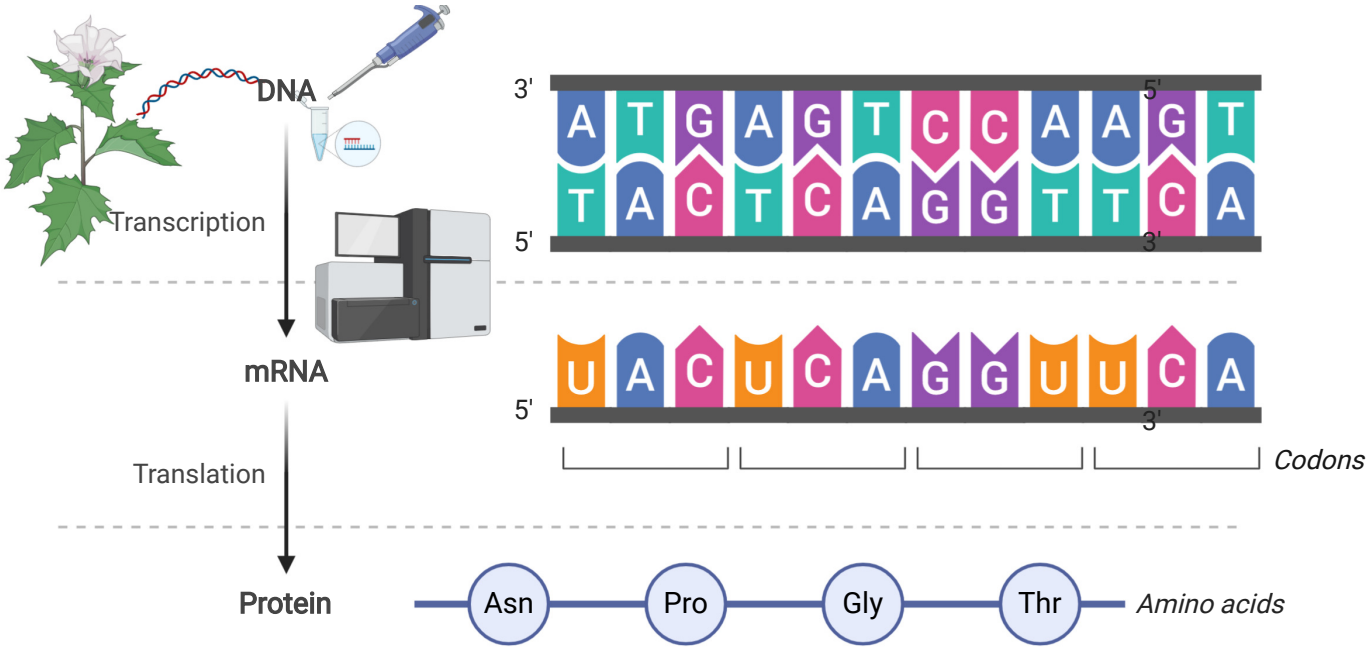
CHAPTER 4

Research article

2020

GENOMIC SIGNATURES OF THE EVOLUTION OF DEFENCE AGAINST ITS NATURAL ENEMIES IN THE POISONOUS AND MEDICINAL PLANT *Datura stramonium* (SOLANACEAE)

This manuscript has been peer-reviewed in Peerage of Science and submitted in ***Scientific Reports***



This manuscript has several additional files, these were deposited in https://github.com/icruz1989/Datura-stramonium-genome-project/tree/master/Supplementalmaterials. The complete workflow; commands and scripts that support this manuscript have been also deposited in https://github.com/icruz1989/Datura-stramonium-genome-project. Please follow these links

Genomic signatures of the evolution of defence against its natural enemies in the poisonous and medicinal plant *Datura stramonium* (Solanaceae)

I. M. De-la-Cruz[1], A. Hallab[4], U. Olivares[2], R. Tapia-López[1], S. Velázquez-Márquez[1], D. Piñero[1], K. Oyama[3], B. Usadel[4,5], J. Núñez-Farfán[1]*

[1]*Department of Evolutionary Ecology, Institute of Ecology, National Autonomous University of Mexico (UNAM), Mexico City, Mexico.*

[2]*National School of Higher Education, National Autonomous University of Mexico (UNAM), Campus Juriquilla, Queretaro, Mexico.*

[3]*Institute of Research in Ecosystems and Sustainability, National School of Higher Education, National Autonomous University of Mexico (UNAM), Campus Morelia, Michoacan, Mexico.*

[4]*Research group of Bioinformatics, Cell Walls and Secondary Metabolites, Jülich Research Centre, Cologne, Germany.*

[5]*Research group of Bioinformatics, Cell Walls and Secondary Metabolites, Institute for Biology I, RWTH Aachen University, Aachen, Germany.*

*E-mail of the corresponding authors: farfan@unam.mx

Abstract

Tropane alkaloids and terpenoids are widely used in the medicine and pharmaceutic industry and have been related as chemical defenses against herbivores and pathogens in the annual herb *Datura stramonium* (Solanaceae). In this study, we present the first draft genomes of two plants from contrasting environments of *Datura stramonium*. Using these *de novo* assemblies, along with other previously published genomes from 11 Solanaceae species, we carried out comparative genomic analyses to provide insights into the genome evolution of *D. stramonium* within the Solanaceae family and to elucidate adaptive genomic signatures to biotic and abiotic stresses in this plant. We also studied in detail the evolution of eight genes of *Datura stramonium*; Putrescine N-methyltransferase, Primary-amine oxidase, Tropinone reductase I, Tropinone reductase II, Aspartate aminotransferase, Tyrosine aminotransferase, Hydroxyphenylpyruvate reductase, and Hyoscyamine-6S-dioxygenase, which are involved in the production of the tropane alkaloid biosynthesis. Our analyses revealed that the genomes of *D. stramonium* show signatures of expansion, physicochemical divergence and/or positive selection on proteins related to the production of tropane alkaloids, terpenoids, and glycoalkaloids as well as on *R* defensive genes and other important proteins related with biotic and abiotic pressures such as defense against natural enemies and drought.

Keywords. Comparative genomics, Genome assembly, Solanaceae, Secondary compounds, Plant natural enemies.

Introduction

Plant species from the Solanaceae family, which contains numerous economically and ecologically important species (Tomato, Potato, Eggplant, Bell peppers, Tobacco, Jimsonweed) produce various secondary metabolites (tropane alkaloids, terpenoids and glycoalkaloids), that affect herbivore insect pest and pathogens (bacteria, fungi, virus) (Chowanski *et al*. 2016). In particular, tropane alkaloids belong to the world's oldest plant medicines and compounds are abundantly present in the Solanaceae family but also in Erythroxylaceae, Convolvulaceae, Brassicaceae, and Euphorbiaceae families (Jirschitzkaa *et al*. 2012; Chowanski *et al*. 2016; Kohnen-Johannsen and Oliver Kayser 2019). Within the Solanaceae family, the annual herb *Datura stramonium* produces the highest concentration of tropane alkaloids (Kohnen-Johannsen and Kayser 2019).

Scopolamine, atropine (hyoscyamine), and anisodamine are the main tropane alkaloids of *D. stramonium* (Castillo *et al*. 2013; De-la-Cruz *et al*. 2019 a). Scopolamine and atropine (hyoscyamine) historically have been used for asthma, rheumatism as anesthetic and spasmolytic drugs (Hightower 1979; Kohnen-Johannsen and Kayser 2019). The cultivation and production of plants to produce scopolamine and atropine are still of economic interest, due to its pharmaceutical applications and increasing global demand (Patocka and Jelinkova 2017; Kohnen-Johannsen and Kayser 2019). In fact, scopolamine is one of the essential active medical compounds according to the World Health Organization (WHO) (WHO 2015).

In Mexico, plants of the genus *Datura*, known as *Toloache*, have been used by native cultures since pre-Columbian times (De la Cruz 1552; De Sahagún 1577; Hightower 1979; see Castillo *et al*. 2019). The species of the genus *Datura* (Solanaceae) are native to dry, temperate, tropical and subtropical regions of North America, and occur mostly in Mexico, considered its

center of origin (Barclay 1959; Symon and Haegi 1991). *Datura stramonium*, although native to North America, has expanded its distribution, owing to humans, worldwide except to polar and subpolar climate zones (Weaver and Warwick 1984). This species occurs, distinctively, in human-disturbed habitats (Weaver and Warwick 1984; Núñez-Farfán and Dirzo 1994).

Recent advances in DNA sequencing have allowed the genome assembly of several important model species from the Solanaceae family (*e. g.*, tomato, potato; Xu *et al*. 2011; Sato *et al*. 2012; Sierro *et al*. 2013; Kim *et al*. 2014; Qin *et al*. 2014; Bolger *et al*. 2014b; Bombarely *et al*. 2016; Edwards *et al*. 2017; Razali *et al*. 2018; Barchi *et al*. 2019). However, genome sequences of non-model species from this plant family are scarce (Xu *et al*. 2017). Non-model species such as *D. stramonium* could be of wide interest because they offer new modes of investigating the ecological and evolutionary processes that plants face in their natural environments and how they respond to pollution, human disturbance and climate change (Savolainen *et al*. 2013). Furthermore, the availability of non-model Solanaceae species genomes would be of great value to better understand the evolution of the Solanaceae family (Savolainen *et al*. 2013).

Here, we present the first draft genomes of two plants of *D. stramonium* that were selected from two contrasting populations of Mexico (Teotihuacán, State of Mexico and Ticumán, State of Morelos) (Valverde e*t al*. 2003; Fornoni 2004). Plants from Ticumán produce a higher concentration of tropane alkaloids than those from Teotihuacán (De-la-Cruz *et al*. 2020 a). Evidence points out that this differentiation is adaptive likely due to different herbivores pressures between populations (Castillo *et al*. 2014; Miranda-Pérez *et al*. 2016; De-la-Cruz *et al*. 2020 a). The two selected individuals for this study are also the parents of an F_2 generation

progeny used to map quantitative trait loci (QTLs) of complex phenotypes such as chemical defensive traits (De-la-Cruz *et al*. unpublished).

In this study, we also carried out extensive comparative genomic analyses with a total of 13 Solanaceae species (including both genomes of *D. stramonium*) to explore the adaptation of these plants to biotic and abiotic stresses. Furthermore, we studied eight genes (Putrescine *N*-methyltransferase, Primary-amine oxidase, Tropinone reductase I, Tropinone reductase II, Aspartate aminotransferase, Tyrosine aminotransferase, Hydroxyphenylpyruvate reductase and Hyoscyamine-6S-dioxygenase) of *D. stramonium* that are involved in the biosynthesis of the tropane alkaloids and we relate this genetic information with the concentration of 19 tropane alkaloids that were quantified for the two genomes using Liquid chromatography-time-of-flight-mass spectrometry (HPCL-TOF-MS).

Results and discussion

Genome sequencing and assembly

DNA was isolated and assembled from two diploid plants collected from two populations of *D. stramonium*; Ticumán State of Morelos, Mexico and Teotihuacán, State of Mexico, Mexico (Additional file 1). 323M PE (paired-end) raw sequences (2 X 150b) were obtained from Illumina HiSeq 4000 sequencing; corresponding to 112 Gb and an average 30.85-fold genome coverage for the Ticumán individual, while 318M PE sequences corresponding to 110 Gb and 30.29-fold genome coverage were generated for the Teotihuacán individual (Table 1). After trimming the PE sequences, we obtained 305M and 303M reads for Ticumán and Teotihuacán, respectively (Table 1). For PacBio Sequel II sequencing, we obtained 9,995,713 subreads corresponding to 37 Gb for the Ticumán individual, while for Teotihuacán individual 9,505,413

subreads were generated, corresponding to 30 Gb (Table 1). The frequency of K-mers estimated a genome size of 1.38 Gb for Teotihuacán and 1.57 Gb for Ticumán, using a k-mer size of 95 and 93 (best k- mer sizes estimated by Kmergenie), respectively (Fig. 1 a, b). Cell flow cytometry analysis did not reveal differences in genome sizes between both individuals, estimating the genome size between 1.7 and 2 Gb (Additional file 2 a). Overall, the Ticumán genome was better assembled than the Teotihuacán genome (Table 2, Fig. 1 c, d, e). This differentiation is because we obtained lower quality sequences from PacBio platform for the Teotihuacán plant. The total length of the assembly was 1.47 Gb and 1.28 Gb for Ticumán and Teotihuacán, respectively (Table 2). Both assemblies showed a normal pattern for relative GC content (Fig. 1 d). These assessments of genome size are similar to the KmerGenie estimates and the results of the cell flow cytometry analysis. The total number of scaffolds was 27,915 and 30,392 for Ticumán and Teotihuacán genomes, respectively. Approximately, 1.05 Gb and 730 Mb of the Ticumán and Teotihuacán *Datura* genomes, respectively, showed contigs ≥ 50,000 bp (Table 2). The largest scaffold for Ticumán was 3.13 Mb, while for Teotihuacán was 2.11 Mb. N50 scaffold length of 84,687- and 58,197-bp for Ticumán and Teotihuacán respectively (Table 2). The number of unknown bases (N's) was 725.89 and 110.68 per 100 kbp for Ticumán and Teotihuacán, correspondingly (Table 2).

The alignment between the two genome assemblies revealed a total of 6,673,981 SNPs (Additional file 2 b). The average identity of 1-to-1 alignment blocks (number of alignment blocks comprising the 1-to-1 mapping of Ticumán to Teotihuacán) was 97.92 % (Additional file 2 b). The Ticumán genome showed 935 more relocations than the Teotihuacán genome (Additional file 2 b). In addition, Ticumán genome showed 6,202 more translocations than

Teotihuacán. While only four more inversions were found in Ticumán than the Teotihuacán genome (Additional file 2 b).

We mapped the raw PE sequences from each individual to its corresponding assembly for genome validation quality, the overall mapping rates were 96.14 % and 89.47 % for Ticumán and Teotihuacán, respectively (Additional file 3). Our *Datura* genomes covered 91% and 81.7 % complete single copy orthologs (BUSCOs) for Ticumán and Teotihuacán, respectively, of the 3,052 total BUSCOs searched (Table 3, Fig. 1 e). 4.8 % and 12.1 % correspond to missing BUSCOs for Ticumán and Teotihuacán, respectively (Table 3, Fig. 1 e). Completeness of single copy orthologs and mapping rates indicate a good assembly for both genomes but especially for the Ticumán individual. The MAKER annotation pipeline included 33,856 and 30,934 protein-coding genes (Table 4). The total genome covered by the genes for Teotihuacán was 14.2 % and for Ticumán was 16 % (Table 4). The total number of exons was 188,426 and 163,107 for Ticumán and Teotihuacán genomes, respectively (Table 4). The mean exons per mRNA was 5.2 for Ticumán and 5.3 for Teotihuacán. A total of 99% gene models showed high confidence matches (E-value ≤ $1e^{-5}$) in the UniProtKB/TrEMBL database. Other non-model Solanaceae species that have been sequenced with similar genome size of *D. stramonium* are *Petunia inflata* (genome size = 1.29 Gb) and *Petunia axilaris* (genome size = 1.26 Gb), and they were assembled in 83,639 and 136,283 scaffolds, respectively (Bombarely *et al*. 2016). Our workflow using iteratively short and long reads to generate contigs and scaffolds revealed an accurate and continuous assembly. PacBio sequences from Teotihuacán individual showed a higher error rate and this produced a shorter and more fragmented genome assembly than the Ticumán individual. This also affected the number of genes annotated. Nonetheless, this number in both genomes approximately is equal to the expected number in Solanaceae species. Furthermore, the

percentage of missing BUSCOs was relatively low for both genomes (Simão *et al*. 2015). Here, the gene completeness (BUSCOs) of our genome assembly is very similar to that reported for Tomato, Potato, Eggplant, Pepper, Tobacco and its wild relatives, as well as *P. inflata* and *P. axilaris* (Xu *et al*. 2011; Sato *et al*. 2012; Bolger *et al*. 2014 b; Sierro *et al*. 2014, Bolbarely *et al*. 2016; Xu *et al*. 2017, Hulse-Kemp *et al*. 2018; Barchi *et al*. 2019).

Repetitive landscape of *Datura* genomes

Datura genomes are rich in repetitive DNA (as are most other plant genomes, Kubis *et al*. 1998). The repetitive landscape of our genomes revealed that 76.04 % and 74.11 % of the genomes are composed by repetitive elements (Table 5, Fig. 2). These results reveal a higher proportion of repetitive elements than other Solanaceae genomes, such as tomato, potato and *Petunia* species, and nearly similar to the repetitive landscapes of *Nicotiana* and *Capsicum* genomes (Additional file 4) (Xu *et al*. 2011; Sato *et al*. 2012; Bolbarely *et al*. 2016; Xu *et al*. 2017, Hulse-Kemp *et al*. 2018). Long terminal repeats (LTR) elements are the most abundant in the *D. stramonium* genomes (Table 5, Fig. 2), covering 65.88 % and 63.41 % of the genomes for Ticumán an Teotihuacán, respectively (Table 5, Fig. 2). The Gypsy family is the most represented in both genomes covering 61.33 % and 58. 71 % for Ticumán and Teotihuacán genomes, respectively (Additional file 5). The Copia family represents almost the rest of the repetitive landscape for both genomes (Additional file 5). In *N. attenuata,* an analysis of the history of repetitive elements revealed that all *Nicotiana* species experienced a recent wave of Gypsy retrotransposon expansion (Xu *et al*. 2017) and this seems to have happened also in the *Datura* species. A recent study showed that *Capsicum* species also experienced a large expansion of Gypsy transposable elements (Quin *et al*. 2014), albeit earlier than in *Nicotiana*, indicating that after whole genome

triplication, the different Solanaceae lineages independently experienced the processes of Gypsy proliferation (Xu *et al*. 2017).

Comparative genomic analyses

OrthoFinder assigned 480,594 genes out of 536,483 (89.6% of total) to 35,458 orthogroups or protein families. Mean gene family size is 13.6 proteins, while fifty percent of all proteins were in proteins families with 19 or more proteins (G50 = 19) (Additional file 6). There were 10,141 protein families with all species present (Fig. 3) and 181 of these consisted entirely of single-copy genes. The two species which shared the most protein families were *S. pimpinellifolium* and *S. lycopersicum* (Fig. 3). The range of proteins in the genomes between species comprise of 30,034 (from *D. stramonium* Teotihuacán) to 69,500 (from *N. tabacum*) (Additional file 7). The species phylogeny shows four mayor clades, the group of *Nicotiana* species, the clade of *Datura*, the group of *Capsicum* species and the *Solanum* group. *P. inflata* was selected as the outgroup species. *P. inflata* diverged from all the Solanaceae species studied here approximately 35 Mya (Fig. 4) while *D. stramonium* diverged ~30.1 Mya from *Solanum*, *Capsicum* and *Nicotiana* species. The divergence dates reported here are consistent with other phylogenies reported for the Solanaceae species (Särkinen *et al*. 2013; Bombarely *et al*. 2016). The rate of gene gain and lost (λ) resulted from CAFE analysis was 0.015 for the whole tree (Fig. 4). The internal branch with the largest numbers of significant rapidly evolving gene families corresponds to the most recent common ancestor of *N. tomentosiformis* (Fig. 4). The terminal branch with the most rapidly significant evolving gene families is the one leading to *D. stramonium* clade. The internal branch with the largest numbers of significant contractions corresponds also to the most recent common ancestor of *Datura* species clade. The terminal branches with the most contractions are

the one leading to *D. stramonium* Teotihuacán. While *P. inflata* is the species with least contractions (Fig. 4). Likewise, the internal branch with the largest significant number of expansions corresponds to the most recent common ancestor of *N. tomentosiformis*. *D. stramonium* Ticumán showed the highest number of gene family expansions (Fig. 4).

Enrichment tests

We found 49 InterPro enriched domains in proteins subject to physicochemical divergence in both genomes ($p < 0.01$) (Additional file 8). 94 enriched InterPro domains were detected in proteins with signal of expansion in both genomes ($p < 0.01$) (Additional file 9). 56 enriched InterPro domains was detected in *Datura* proteins with positively selected conserved amino acids (codons) (Additional file 10). Likewise, we found with the MapMan4 annotation a total of 14 enriched proteins with signal of expansion (Additional file 20), 23 proteins with positively selected conserved amino acids (Additional file 21), and 54 enriched proteins with physicochemical divergence (Additional file 22). We found that either over-represented domains (enrichment test using InterproScan database) as well as over-represented proteins (enrichment test using MapMan4 database) with signal of expansion, positive selection or physicochemical divergence are related with immunity and defence against pathogens, virus, fungi and insect herbivores as well as related with responses to abiotic stresses such as drought and nutrients deficiency (Table 6, 7, Additional file 11).

Domains associated with defensive proteins (*R* genes)

Several domains have been associated as fundamental components of the *R* genes in *D. stramonium* genomes (Table 6) but some notable domains in expanded and positively selected proteins were found in *D. stramonium* (Table 6); the Virus X resistance protein-like (IPR038005)

and Rx, N-terminal (IPR041118) are domains that confer resistance against the potato virus X (Rairdan and Moffett 2006; Tameling and Baulcombe 2007; Hao *et al*. 2013). IPR038005 domain has been identified in a family of resistance proteins with an architecture that includes an N-terminal coiled-coil domain, a Nucleotide-binding domain, and a Leucine-rich repeat (CC-NB-LRR) (Hao *et al*. 2013). These intracellular resistance proteins recognize pathogen effector proteins and will subsequently trigger a response that may be as severe as localized cell death (Hao *et al*. 2013). The NB-ARC (IPR002182) and Leucine-rich repeat (LRR) domain superfamily (IPR032675) (Table 6), interact and release a signal to initiate an event of immunity against pathogens (Rairdan and Moffett 2006; van Ooijen *et al*. 2008).

The LRR proteins are involved in specific protein-protein interactions and have a significant role in plant defenses (Padmanabhan *et al*. 2009). Resistance to a diverse range of pathogens, including nematodes, fungi, bacteria, and viruses involves LRR proteins either as resistance proteins or as proteins required for resistance proteins to function (Padmanabhan *et al*. 2009). We found that the LRR-XII kinase and SD-1 kinase proteins had positively selected codons in *Datura* genomes (Additional file 21). Magalhães *et al*. (2016), found a large expansion of LRR-XII in *Citrus* genomes, suggesting that it might play a key role in adaptive responses in host-pathogen co-evolution, related to the perennial life cycle and domestication of the citrus crop species. It has demonstrated that SD-1 kinase protein is a plant receptor with roles in signaling and plant defense (Afzal *et al*. 2008). Moreover, we found several proteins belong to the Kinase superfamily with significant physicochemical divergence (Additional file 22). Kinase superfamily proteins have been related with different stresses including light, pathogen invasion, hormones, temperature stress, and nutrient deprivation (Stone and Walker 1995).

We also found expansion and positive selection on Bet v I/Major latex domain (IPR000916) (Table 6), a domain related with the pathogenesis-related proteins whose expression is induced by pathogen infection, wounding, or abiotic stress (Pasternak *et al*. 2005; Fernandes *et al*. 2008). The Late blight resistance domain R1 (IPR021929) (Table 6) showed a significant expansion signal. The R1 is a protein for resistance to late blight, the most destructive disease in potato cultivation worldwide (Ballvora *et al*. 2002). The R1 protein belongs to the class of plant proteins for pathogen resistance that have a leucine zipper motif, a putative Nucleotide binding domain and a Leucine-rich repeat domain (Ballvora *et al*. 2002). On the other hand, the Trichome birefringence-like 45 (IPR029981) domain had physicochemical divergence signal (Table 6). This domain is involved in non-host resistance (NHR) or plant immunity to non-adapted pathogen species (Bischoff *et al*. 2010).

Domains and proteins related with the biosynthesis of Terpenoids

We found in both enrichment analyses (InterPro and MapMan4 functional annotation) proteins with signal of expansion, physicochemical divergence and positively selected that are directly related with the biosynthesis of terpenoids (Table 7, Additional file 8-10, 22-24). For instance, Cytochrome P450 domain is related in the biosynthesis of terpenoids and has been associated as a key domain in the production of hederagenin-based saponins which mediate plant defence against herbivores and linalool metabolism (Liu *et al*. 2019). This latter compound is used by some plants to cope against floral antagonists (Boachon *et al*. 2015). We also found domains directly related in the biosynthesis of terpenoids in *D. stramonium* protein families with significant expansion and with positively selected codons. These families comprise Terpene synthases with the N-terminal domain (IPR001906), Terpenoid cyclases/protein

prenyltransferase alpha-alpha toroid (IPR008930), Terpene synthase, metal-binding domain (IPR005630), Terpene cyclase-like 1, C-terminal (IPR034741) (Table 7). Enrichment analysis with MapMan4 revealed that mono/sesquiterpene/diterpene synthase family proteins showed signal of expansion and with positively selected codons (Additional file 20). Also, protein containing the Glycoside hydrolase, family 35 domain (IPR001944) showed signal of physicochemical divergence (Table 7). This Glycoside hydrolase domain is involved in the biosynthesis of different secondary compounds including terpenoids alkaloids (Taron *et al*. 1995) (Table 7). Several studies have indicated that terpene synthases are the primary enzymes in the formation of low-molecular-weight terpene metabolites (Degenhardt *et al*. 2003). Overall, it has reported that when attacked by herbivorous insects or mites, some plant species call on other arthropods for help (Tholl 2006). Plants emit mixtures of volatile compounds, dominated by terpenes or glycosides, to attract carnivorous arthropods that prey on or parasitise herbivores and so reduce further damage (Degenhardt *et al*. 2003; Tholl 2006; Cheng *et al*. 2007; Mithöfer and Boland 2012; Aljbory and Chen 2018). Likewise, several terpenoids have been reported to act directly as defence against herbivores (Ujváry 2010).

Enriched proteins related with abiotic stresses

A notable domain, the SNF1-related protein kinase regulatory subunit beta-2 (IPR030070) (Additional file 11, 24) was detected in proteins with physicochemical divergence. This domain is implicated in the response against drought, in the efficiency of carbohydrate metabolism and response to glucose limitation (Akkasaeng *et al*. 2007; Shanker *et al*. 2014; Thangella *et al*. 2018). Another notable domain in over-represented genes with signal of expansion is the START-like domain superfamily (IPR023393), which has been related with the response to

drought, salt tolerance, wound and heat stress (Satheesh et al. 2016) (Additional file 11). Also, we found over-represented domains in proteins with signal of expansion and with positively selected codons containing Galactose oxidase/kelch, beta-propeller (IPR011043) domain. This domain is also involved in the stress responses induced under Fe deficiency in the roots and also related as defence protein (Kawahara *et al*. 2017) (Additional file 11). Kinase proteins showed signal of physicochemical divergence (Additional file 22), and these proteins have been related with different stress factors including light, pathogen invasion, hormones, temperature stress, and nutrient deprivation (Stone and Walker 1995).

Genes involved in the tropane alkaloid biosynthesis

Notable domains involved in the tropane alkaloids pathway were proteins of families expanded in the *Datura* branch and with positively selected conserved amino acids (codons) (Additional file 9, 10, 22, 23). For instance, Cytochrome P450 (IPR001128), Transferase (IPR003480), NADH:ubiquinone oxidoreductase (IPR003918) and Phosphoethanolamine *N*-methyltransferase (IPR025771) (Table 7). Cytochrome P450 is involved in the rearrangement of Littorine (a kind of tropane alkaloid) to produce atropine/hyosciamine and scopolamine (Li *et al*. 2006; Nasomjai *et al*. 2009). This step is very important in the biosynthesis of scopolamine via the Hyoscyamine (6S)-dioxygenase gene (*h6h*) (Li *et al*. 2006; Nasomjai *et al*. 2009). Indeed, the enzymes that participate in the tropane alkaloid biosynthesis belong to the classes of oxidoreductases and transferases (Kanehisa and Sato 2019) such as we detected in enriched proteins with signal of expansion, positively selected and proteins with physicochemical divergence (Table 7).

Within tropane alkaloids, the *pmt* gene family showed significant gene expansion during the evolution of the *Datura* genus; the last common ancestor of *D. stramonium* had only one

gene (Additional file 12), while *D. stramonium* Ticumán and Teotihuacán have three and two gene copies, respectively (Additional file 12). Pfam annotation of *pmt* genes showed that the gene dati7568, which belongs to the Ticumán genome, has an extra domain of spermine-synthase in comparison with its homolog from Teotihuacán (Additional file 13). It has been reported that spermine is a potent plant defense activator with broad-spectrum protective effects (Seifi *et al*. 2019). Indeed, Kasukabe *et al*. (2004), found that the overexpression of spermidine-synthase enhanced tolerance to multiple environmental stresses including herbivory and pathogenesis. Moreover, *pmt* is the key gene catalyzing the formation of *N*-methylputrescine from putrescine and S-adenosyl-L-methionine and this enzyme triggers the production of hygrine and other different tropane alkaloids (Kanehisa and Sato 2019). HPLC-TOF-MS results revealed that the plant of Ticumán showed 26.63-fold of hygrine concentration than the plant of Teotihuacán (Additional file 14; Fig. 5 a). In fact, a differentiation of 59-fold was obtained in total tropane alkaloid concentration between Ticumán and Teotihuacán (Additional file 14, Fig. 5 a). Thus, it is possible that the additional domain of spermine-synthase confers overproduction of tropane alkaloids in the Ticumán genome.

Each *Datura* plant has a single copy of the Primary-amine oxidase gene (Additional file 15). This enzyme triggers the production of tropinone. We did not observe differences in the domain's architecture of both genes (the same four domains were observed for both genes). This could be related with the fact that we also did not observe differences between Ticumán and Teotihuacán tropinone concentration (difference = 0.80 ng/g) (Additional file 14, Fig. 5 a). In fact, tropinone is the common substrate to the formation of *tpr* I and *tpr* II genes (Drager *et al*. 1988; Nakajima 1993; Keiner *et al*. 2002). Tropinone represents a branching point in the tropane alkaloid metabolic pathway (Drager *et al*. 1988; Nakajima 1993; Kanehisa and Sato 2019).

Tropine (the product of *tpr* I) is incorporated into hyoscyamine and scopolamine whereas pseudotropine (the product of *tpr* II) is the first specific metabolite on the pathway to the calystegines (Nakajima 1993). Both genes are always found together in any given tropane-alkaloid-producing species (Drager *et al*. 1988; Keiner *et al*. 2002; Kanehisa and Sato 2019). These results could be suggesting that the gene Primary-amine oxidase is a highly conserved domain to maintain the biosynthesis of the final products in the tropane alkaloid pathway such as atropine and scopolamine (Keiner *et al*. 2002).

No expansion was detected for *tpr* I gene in *D. stramonium* (Additional file 12). However, we observed four copies of this gene in both *Datura* genomes (Fig. 6). The gene date9161 (Teotihuacán) and dati33033 (Ticumán) have two domains absent in the other *tpr* I *Datura* genes; Sulfakinin (PF08257) and ECR1_N domain (PF14382) (Fig. 6). ECR1_N is an N-terminal region of the exosome complex exonuclease RRP proteins. However, in the red flour beetle *Tribolium castaneum* (Tenebrionidae) sulfakinin has been reported to be involved in food uptake functioning as a receptor (Yu *et al*. 2013). The authors found that RNAi-mediated gene silencing plants of sulfakinin stimulated food intake. In contrast, injection of a sulfakinin analog led to a significant reduction in food intake (Yu *et al*. 2013). Also, food intake by the desert locust, *Schistocerca gregaria* (Acrididae), is decreased by the injection of Lom-sulfakinin (Wei *et al*. 2000). Wei *et al*. (2000), suggest that the sulfakinins may reduce the sensitivity of taste receptor. Here, we found that sulfakinin domain is observed in *D. stramonium* but not in the other Solanaceae species (Fig. 6).

Indeed, the *Datura tpr* I genes date9170 and dati33044 had a domain architecture different from the *tpr* genes of the other studied Solanaceae (Fig. 6). However, the gene

dati33044 (Ticumán) has one additional domain adh_short_C2 (PF13561), in comparison with its homolog date9170 (Teotihuacán) (Fig. 6). Likewise, dati33027 showed an additional adh_short_C2 domain in comparison with its homolog date20542 (Fig. 6). We observed that the Ticumán plant produced ca. 32 times more tropine than the Teotihuacán plant (Additional file 14, Fig. 5 a). This chemical difference in tropine concentration could be related with this additional domain (adh_short_C2) that was observed in the Ticumán genome. It has reported that the domain adh_short_C2, also known as Enoyl-(Acyl carrier protein) reductase, plays a role as anti-microbial and anti-parasitic molecules (Massengo-Tiassé and Cronan 2009).

Other gene copies of *tpr* I, date11128 (Teotihuacán) and dati22507 (Ticumán) have the domain RHH_5 (PF07878) (Fig. 6), this domain has been described as a toxin-antitoxin system (TA), which is comprised of a set of two or more closely linked genes that together encode both a toxic protein (the "toxin") and a corresponding "antitoxin" (Shidore and Triplett 2017). Toxin-antitoxin genes are often inherited through horizontal gene transfer and are associated with pathogenic bacteria (Ramisetty and Santhosh 2016). TA systems are numerous in many plant-associated bacteria, but very little is known regarding their function and distribution in phytopathogens (Shidore and Triplett 2017). This TA system is also found in *N. attenuata*, *N. tabacum*, *N. sylvestris* and *C. annuum gabriusculum* (Fig. 6). Several studies about the production of secondary metabolites as defence against natural enemies have been reported for these species (Chowański *et al*. 2016).

Evidence of expansion for the gene family *tpr* II was detected. The analysis revealed that the last common ancestor of *D. stramonium* had one *tpr* II gene, while five and three copies of *tpr* II genes showed to be expanded from Teotihuacán and Ticumán, respectively (Additional file

12). The domain adh_short_C2 (enoyl-acyl carrier protein reductase) is present in the above mentioned four Solanaceae species (Additional file 16). An interesting domain, (PapA_C) is present in the gene copy date754 (Additional file 16). This domain is so-called Polyketide-associated protein (Pap) that belongs to the subfamily of acyltransferases and has been found to be involved in the biosynthesis of secondary metabolites (Tiassé and Cronan 2009). A Cobalamin (CobU) domain is found in the *Datura* gene dati23799 (Additional file 16). Vascular plants neither synthesize nor require vitamin B12 because they contain cobalamin-independent methionine synthase (MetE) (Smith *et al*. 2007). However, herbivores have been found to obtain their dietary quota of cobalamin from plants contaminated with cobalamin-producing soil bacteria (rhizobia) that grow in roots and nodules of plants (Antony 2018). Thus, more studies are needed to prove and assess the interaction between mycorrhizal and *D*. *stramonium*.

Secondary compounds and enzymes derived from the phenylalanine biosynthesis are involved in the tropane alkaloid biosynthesis. We found three genes (Aspartate aminotransferase, Tyrosine aminotransferase and Hydroxyphenylpyruvate reductase) from the phenylalanine biosynthesis (Kanehisa and Sato 2019). The three genes have not experienced expansion in none of the two genomes (Additional file 12). Aspartate aminotransferase protein family only comprise of the Aminotran_1_2 (PF00155) domain (Additional file 17). The same case if for the Tyrosine aminotransferase gene family except for *N*. *attenuata* (nta10614 gene contains two additional domains of Aminotran_1_2 and one zf-RVT, PF13966) and *P*. *inflata* (pin21235 gene contains one additional Aminotran_1_2 domain and one Glyco_hydro_20, PF00728) (Additional file 18). The Hydroxyphenylpyruvate reductase gene family is only constituted by three genes; two of them belong to *D*. *stramonium* and one belongs to *N*. *tabacum* (Additional file 19). The Ticumán gene has an additional 2-Hacid_dh (PF00389) and 2-Hacid_dh_C domains in contrast

with the Teotihuacán gene. Also, is notable that *N. tabacum* gene do not have the DUF3089 (PF11288) and Peptidase_M10_C domains (Additional file 19).

Hyoscyamine (6S)-dioxygenase gene (*h6h*) is the last rate-limiting enzyme directly catalyzing the formation of atropine and scopolamine in tropane alkaloids biosynthesis pathway (Kanehisa and Sato 2019). It has been the primary target gene in the genetic modification of tropane alkaloids metabolic pathway and as we have pointed out, scopolamine is the mainly secondary metabolite of *D. stramonium* with pharmaceutic and medical interest (Qiang *et al.* 2015). First, our results revealed that two copies of *h6h* are found in both genomes of *D. stramonium* (Fig. 5 b). However, these copies were distributed in two different gene families (*i. e.,* gene families "OG0028637" and "OG0043057") (Fig. 5 b). Indeed, both gene families were composed by only two genes; one from the Ticumán plant and one from the Teotihuacán plant. Therefore, we used 17 *h6h* genes (13 genes retrieved from Uniprot database and four from our *D. stramonium* genomes) to construct an artificial gene family. We carried out a multiple sequence alignment and reconstruct the phylogeny. Also, using the Pfam database the protein domain architecture of these genes was identified. The gene DaturastramoniumTic8550_OG0028637 (from our Ticumán genome) has two domains of DIOX_N (PF14226), while the above homologs have only one DIOX_N domain (Fig. 5 b). In contrast, only three genes in this *h6h* family are composed of a single domain (2OG-FeII_Oxy, PF03171); two of them belong to our *D. stramonium* genomes and one are found in the *Medicago truncatula* genome (Fig. 5 b). The DIOX_N domain is composed of a highly conserved N-terminal region of proteins with Oxoglutarate/iron-dependent dioxygenase activity (2OG-FeII_Oxy) (Hagel and Facchini 2010) and this domain could be involved in the high production of atropine, anisodamine and scopolamine in the plant of Ticumán (Additional file 14, Fig. a, b).

In summary, tropane alkaloids quantification between both genomes revealed clear differences in the production of tropane alkaloids. Overall, the Ticumán genome showed 57.31-fold of total tropane alkaloids than the Teotihuacán parent. Differences in the tropane alkaloid concentration between Ticumán and Teotihuacán seem to be related with the differences in domain architecture of almost all here studied genes involved in the tropane alkaloid biosynthesis. Recently, evidence has been published indicating that tropane alkaloids are implicated in resistance against herbivores in *D. stramonium* (Shonle and Bergelson 2000; Valverde *et al*. 2001; Castillo *et al*. 2013, 2014; Bustos-Segura *et al*. 2014; Miranda-Pérez *et al*. 2016; De-la-Cruz *et al*. 2020 a) and that the selective value of tropane alkaloids preventing or reducing herbivory varies among populations of this plant species, depending on the type of enemies (specialist or generalists herbivores, fungi, pathogens, oomycete) (Castillo *et al*. 2014).

Conclusions

The information generated in this work will provide support for future studies in *D. stramonium* and other plant species. Understanding the evolution, adaptation and the ecological role of tropane alkaloids and other secondary metabolites such terpenoids is necessary to disentangle its role in defence against natural enemies. The availability of these genomes provides a tool for future studies to better understand the genome evolution of the Solanaceae family and for other scientific fields such as medicine and pharmaceutics. Likewise, we described how the *D. stramonium* genome expanded and we detected positive selection and physicochemical divergence on terpenoids, *R* genes, and proteins related with abiotic stresses such as drought. Finally, non-model species such as *D. stramonium* could be of wide interest because they allow to investigate the ecological and evolutionary processes that plants undergone

in their natural environments and how they respond to pollution, human disturbance and climate change.

Materials and Methods

Selection of the parent genomes

The two selected genomes were extracted from two different populations, Ticumán in the State of Morelos, 18°45'39.90"N, 99°7'13.86"W, and Teotihuacán, in the State of Mexico, 19°41'6.96"N, 98°52'19.63"W (Valverde *et al*. 2003; Miranda-Perez *et al*. 2016; De-la-Cruz *et al*. 2020 a). We analyzed 21 tropane alkaloids via HPLC-TOF-MS from 47 (Ticumán) and 45 (Teotihuacán) plants under controlled conditions (green house experiment) and we selected these two individuals that were the most differentiated in their total tropane alkaloid concentration (Teotihuacán = 1,018 ng/g; Ticumán = 59,051 ng/g). Chemical conditions and details of samples extraction and mass spectrometry analyses can be consulted in De-la-Cruz *et al*. (2020 a).

DNA extraction, genomic library preparation and sequencing

To obtain a high-quality *de novo* assembly, we combined data generated from short insert paired-end libraries from Illumina sequencing, with long read sequencing by PacBio Sequel II sequencing. First, gDNA was extracted from the two individuals. gDNA was isolated from fresh leaves with a modified CTAB mini-prep protocol (Doyle and Doyle 1987). The total amount of gDNA was measured using Qubit dsDNA HS Assay Kit (Invitrogen, Thermo Fisher Scientific, Waltham, USA). A total of 200 ng of gDNA were used for library preparation. Libraries were sheared on the Covaris M220 Focused-ultrasonicator (Covaris Massachusetts, USA) then prepped for 150PE (paired-end) Illumina HiSeq 4000 sequencing, using the Kapa Hyper prep

Illumina library prep kits. Final libraries were visualized on the Agilent Fragment Analyzer, then quantified and pooled equimolar with Kapa qPCR Illumina library quant Universal Kits. Demultiplexing was then performed with the Illumina bcl2fastq v2.19 software and returned in fastq format.

The PacBio Sequel II sequencing (Pacific Biosciences) was performed by taking 20 ug of gDNA into SMRTbell library preparation for long-insert libraries with the PacBio Express Template preparation kit, followed by size selection at 15kb on the Sage BluPippin. Libraries were then run on 10-hour movies (length of time to continuously run the sequencing reaction) on a PacBio Sequel using v2.0 chemistry. The sequencing and libraries preparation for both sequencing platforms was carried out in the QB3 Functional Genomics and Vincent J. Coates Sequencing Laboratories at the University of California, Berkeley.

Preprocessing of sequenced short reads

Reads quality has a major impact on the quality of the resulting assembly, and the use of error-corrected reads increases the size of the contigs (Salzberg *et al*. 2012). Illumina paired-end reads were trimmed using a phred quality score > 30 in TRIMMOMATIC v0.32 (Bolger *et al*. 2014 a). We verified visually the quality (including contamination with Illumina paired-end adaptors) before and after trimming using the program FastQC (Andrews 2010). This allowed us to only keep high-quality reads prior to the assembly steps.

Genome size estimation

Corrected Illumina paired-end reads were used to estimate the genome size using KmerGenie v1.7016 (Chikhi and Medvedev 2014). Also, the genome size was estimated through cell flow

cytometry for both individuals carried out at National Laboratory of Flow Cytometry of the National Autonomous University of Mexico. To estimate the genome size, *Arabidopsis* col-1 ecotype and human PBMCs (male donor) were used as a reference. The nuclear DNA content of the sample was calculated with the formula:

$$Value\ 2C\ sample\ (pg)\ =\ Value\ 2C\ reference\ x\ \frac{IMF\ sample}{IMF\ reference}$$

Where *pg* is picograms and IMF is average fluorescence intensity.

***De novo* genome assembly**

Each individual of *D. stramonium* was assembled independently *de novo*. We followed the workflow of Chakraborty *et al*. (2016) for both plants, with modifications. First, PacBio raw subreads (in bam format) were transformed to fasta format and assembled with Canu v1.8 pipeline that includes three stages: correction, trimming, and assembly (Koren *et al*. 2017). PacBio only assembles of high error, long molecule sequences, depend upon redundancy between the various low-quality reads to 'vote out' errors and identify the true sequence in the sequenced individual (Chakraborty *et al*. 2016). Therefore, we used also a hybrid assembly approach suggested by Ye *et al*. (2016). For this, Illumina short reads were used to perform De Bruijn graph assembly with SparseAssembler (Ye *et al*. 2016). The generated contigs from SparseAssembler were used with PacBio raw sequences to carry out a hybrid assembly using the program DBG2OLC (Ye *et al*. 2016). An advantage of DBG2OLC program is that uses multiple sequence alignment to clean the PacBio reads and remove reads with structural errors (the so-called chimeras) (Ye *et al*. 2016).

The program MUMmer v3 (Kurtz *et al*. 2004) was used to run the NUCmer wrap and the program delta-filter to compute unique alignments between the contigs from the Hybrid assembly (DBG2OLC) and PacBio assembly (Canu). DBG2OLC assembly was used as reference and Canu assembly was used as query. This last step allowed us to merge both assemblies (DBG2OC and Canu) using the program Quickmerge (Chakraborty *et al*. 2016). As the two assemblies used for merging come from the same genome, gaps in one assembly can be bridged using the corresponding sequences from the other assembly (Chakraborty *et al*. 2016). Thus, Quickmerge program improved the contiguity of both genome assemblies.

Polishing, consensus and scaffolding

Genome assemblies were polished using the programs Pilon (Walker *et al*. 2014) and Arrow (https://github.com/PacificBiosciences/GenomicConsensus). Polishing the contigs using both programs brings the error rate down to 0.01% or lower (Chakraborty *et al*. 2016). First, raw Illumina sequences were aligned to its corresponding merged assembly (draft genome) with Bowtie2 (Langmead *et al*. 2018). We used SAMtools v1.8 (Li *et al*. 2009) to transform, sort and index the alignments outputs and then Pilon was used to polishing the draft genome with these Illumina aligned reads.

We used the program pbalign (https://github.com/PacificBiosciences/pbalign) to align the PacBio raw sequences to the new corresponding polished draft genome from Pilon. Then, the program Arrow was implemented as a second polishing step and to generate a consensus genome. After this, we used the program OPERA-LG v2.0.6 (Gao *et al*. 2016) for scaffolding and then a third step of polishing with Pilon (which implied align the raw Illumina sequences against its corresponding genome) to improve the accuracy of the final genome assembly.

To evaluate the sequence and structural similarity between the two draft genomes (*i. e.*, single nucleotide polymorphisms or SNPs, breakpoints, insertions, relocations, translocations, inversions, average sequence similarity), we used the wrapper dnadiff from MUMmer v3 (Kurtz *et al*. 2004). The Ticumán assembly was used as reference and the Teotihuacán assembly as query. Likewise, NOVOPlasty v3.8.2 (Dierckxsens *et al*. 2016) was used to extract and reconstruct the chloroplast and mitochondrial genomes from the whole genome shotgun data of the two plants of *D. stramonium*. This program is capable to assemble the incidentally sequenced chloroplast and mitochondrial DNA that is present in almost all plant sequencing projects, due to the extraction of whole cellular DNA (Dierckxsens *et al*. 2016). The complete report and results of the chloroplast and mitocondrial genomes of these plants can be consulted in De-la-Cruz and Núñez-Farfán (2020 b).

Nuclear genome validation

We evaluated the genome assemblies using the standard assembly statistics (average contig size, number of contigs, assembled genome size, N50, etc.) with the package Quast v5.0.2 (Gurevich *et al*. 2013). Also, BUSCO v.2.0.1 (Simão *et al*. 2015) was used to assess the assembly quality through the gene completeness for both genomes. BUSCO inspects *de novo* assemblies searching for single-copy orthologs (BUSCOs) and assess the completeness of the genomes according with the number of BUSCOs found (Simão *et al*. 2015). In our case, the "Solanaceae odb10*" dataset loaded in the program was used to find 3,052 orthologs. BUSCOs were classified as complete and single-copy (S), complete and duplicated (D), fragmented (F) or missing (M). As an additional evaluation of the genome assembly quality, we assessed the mapping rate of the Illumina sequences of each individual to their corresponding assembly using Bowtie2 v2.3.4.3.

Repetitive elements analysis

To characterize the repetitive elements in the genomes of *D. stramonium,* we followed the pipeline "Repeat Library Construction-Basic for MAKER v2.31.10" (http://weatherby.genetics.utah.edu/MAKER/wiki/index.php/Repeat_Library_Construction--Basic) by Campbell *et al*. (2014). With this method, we followed a *de novo* approach to identified and collect repetitive sequences from the genomes. This was achieved using RepeatModeler v2.0 (Smit and Hubley 2015). The repetitive elements derived from this pipeline were concatenated with the databases RepBase-20181026 (https://www.girinst.org/server/RepBase/index.php) and Dfam_Consensus-20181026 (https://dfam.org/help/tools). These databases contain a comprehensive number of repetitive elements from all plant species (Smit and Hubley 2015). Then, the program RepeatMasker v4.0.9 (Smit and Hubley 2015) was run to identify the final interspersed repeats and low complexity DNA sequences. The output of the program was a detailed annotation of the repeats that are present in the genome sequence, as well as a modified version of the genome sequence in which all the annotated repeats have been masked (Smit and Hubley 2015).

Gene prediction and structural annotation

The program BUSCO was used for genome assembly assessment and the annotated BUSCO gene models built during genome assessments were used to optimize the Hidden Markov search model (HMM) to train the gene predictor program Augustus v.3.2.2 (Stanke *et al*. 2006) using the --long option (BUSCO uses Augustus to search the conserved genes) and produce a trained HMM of our genes models for the program MAKER v.2.31.10 (Campbell *et al*. 2014). MAKER identifies and masks out repeat elements based on repeat annotation from RepeatMasker, aligns

RNA-seq data from the same species and/or related species to the genome; also, aligns proteins from related species and use gene predictors to synthesizes all these data into final structural annotations and produces evidence-based quality values for downstream annotation management (Campbell *et al*. 2014).

The *D. stramonium* annotation workflow consisted of a total of four MAKER runs which is the recommended number to obtain the best annotation (Campbell *et al*. 2014). For the first run, we used the gene trained models from Augustus, 443,235 proteins from UniProtKB/TrEMBL from all Solanaceae species database (searching for the word "Solanaceae" with date 30/03/2019), expression sequence tags (ESTs) of *D. stramonium* provided by an alternative experiment from our laboratory (other plants), 328,166 ESTs of five Solanaceae species (*Solanum lycopersicum*, *Solanum tuberosum*, *Nicotiana attenuata*, *Nicotiana tabacum* and *Capsicum annuum*) from EnsemblPlants and our specific repeat library to masks out the genome. This first step produced a set of draft gene models. The gene models from this first run were used to train another *ab initio* gene predictor called SNAP v.2006-07-28 (Korf 2004). Once SNAP was trained with the draft gene models, we ran MAKER again using the same parameters. This process was repeated twice to retrain SNAP for three times in total (Korf 2004). Therefore, we used the gene models from the one round to train *ab initio* SNAP program to improve the inference of gene models in the next round. Only the HMM gene models from SNAP in the MAKER configuration file was changed in each run. Retraining of SNAP was performed using gene models with an annotation edit distance (AED ≤ 0.25) and amino acids length ≥ 50 bp. AED ranges from 0 to 1 and quantifies the confidence between a gene annotation and its supporting evidence (gene models, EST, protein and mRNAseq alignments) (Ozerov *et al*. 2018). Lower AED values imply higher congruency between the intron-exon coordinates of

annotation and its aligned evidence, whereas AED = 1 indicates no evidence for support of predicted genes. Only sequences with AED < 0.5 were retained in the final set of predicted genes (Ozerov *et al*. 2018). We used the Perl scripts from the GitHub repository Genome Assembly Annotation Service (GAAS) (https://github.com/NBISweden/GAAS/tree/master/annotation) in order to retrieve summary statistics from the MAKER annotation.

Functional annotation

Blastp v.2.6.0 (Boratyn *et al*. 2013) was used to functionally annotate the genes from both *Datura* genomes against all the Solanaceae sequences from the UniProt/TrEMBL and UniProt/Swiss-Prot database. We used the program Automated Assignment of Human Readable Descriptions (AHRD) (https://github.com/asishallab/AHRD) to assign gene descriptions that were concise, informative and precise (Sato *et al*. 2012). Gene Ontology terms were annotated using MapMan4 through the Mercartor webtool (Schwacke *et al*. 2019). In addition, protein motifs and domains were annotated using Interproscan v.5.24 (Jones *et al*. 2014), by searching against publicly available databases, including TIGRFAM (Haft *et al*. 2013), SFLD (Akiva *et al*. 2014), ProDom (Bru *et al*. 2005), CDD (Marchler-Bauer *et al*. 2017), PRINTS (Attwood *et al*. 2012), PHANTER (Thomas *et al*. 2003), Gene3D (Yeats *et al*. 2006), PIRSF (Nikolskaya *et al*. 2007), Coils (Lupas *et al*. 1991), MobiDB-lite (Necci *et al*. 2017), PROSITE (Sigrist *et al*. 2013), SMART (Letunic *et al*. 2012), SUPERFAMILY (de Lima Morais *et al*. 2011), and Pfam (Finn *et al*. 2014).

Data sources for comparative genomics (gene family analysis)

Gene family analyses included 11 genomes representing almost all the Solanaceae species that have complete genomes as well as the two genomes of *D. stramonium*. Retrieval of protein coding genes and CDS from 11 genomes were sourced from the Sol Genomics Network (https://solgenomics.net/; *Nicotiana tabacum* (Edwards *et al*. 2017), *Nicotiana sylvestris* (Sierro *et al*. 2013), *Nicotiana attenuata* (Xu *et al*. 2017), *Nicotiana tomentosiformis* (Sierro *et al*. 2013), *Solanum pimpinellifolium* (Razali *et al*. 2018), *Solanum lycopersicum* (Sato *et al*. 2012), *Solanum pennellii* (Bolger *et al*. 2014 b), *Solanum tuberosum* (Xu *et al*. 2011), *Capsicum annuum*, CM334 v1.55 (Kim *et al*. 2014), *Capsicum annuum* var. *glabriusculum* (Qin *et al*. 2014) and *Petunia inflata* (Bombarely *et al*. 2016).

Orthology, reconstruction of orthogroups (protein families) and construction of species and gene family trees

To gain insight into the evolution of *D. stramonium* genome, we used the thirteen proteomes as input to OrthoFinder program (Emms and Kelly 2019). We used in OrthoFinder v2.3.3, DIAMOND blast (E-value < $1e^{-5}$) (Buchfink and Huson 2015) for orthogroup inference, and the MCL clustering algorithm for sequence similarity and clustering (Enright *et al*. 2002). For each orthogroup or gene family we used MAFFT v7 (Katoh *et al*. 2002) as multiple protein sequence aligner and FastTree2 v2.1.10 (Price *et al*. 2010) for maximum likelihood gene trees inference. The inference of species tree is constructed by OrthoFinder, using a concatenated alignment of single-copy orthogroups (those with at most one gene per species) (Emms and Kelly 2019). For some species sets which have been diverging for a very long time, there are not enough single copy orthogroups. In those cases, orthogroups that are mostly single-copy are also used for the

concatenated alignments by only using sequences for the species that are single-copy in that orthogroup and gap characters for the other species (Emms and Kelly 2019). The species tree was inferred with FastTree2 (Emms and Kelly 2019). The rooting is done via STRIDE algorithm (Specie Tree Root Inference from Duplication Events) (Emms and Kelly 2019) and according with OrthoFinder, *P. inflata*, was selected as outgroup of the whole phylogeny.

Inferring the species ultrametric phylogeny

To build an ultrametric phylogeny for the Computational Analysis of gene Family Evolution (CAFE v4.2.1) program, the rooted species tree obtained from OrthoFinder was used to search in TimeTree webtool (Kumar *et al*. 2017) the divergence times between the branches, the rooted species tree and the information of divergence times were used to create the ultrametric species tree using the chronos function of the R package ape (v.3.4 on R v.3.2.1) (Paradis and Schliep 2018). The tip to root length was adjusted to match the approximately 40 million-year evolutionary history of Solanaceae species (Bombarely *et al*. 2016, Kumar *et al*. 2017).

Identification and analysis of gene expansions/contractions

To assess the gene family expansion and contractions of the thirteen Solanaceae species, we used only the gene families with more than four genes per family (24,235) and the species ultrametric tree as inputs to CAFE (Han *et al*. 2013). The main goal of CAFE is to estimate the birth-death (λ) parameters for the provided tree and gene family counts, the λ parameter describes the probability that any gene will be gained or lost (Han *et al*. 2013). First, the python scripts provided by CAFE pipeline were used to estimate the error in our dataset and to removed gene families with large variance (Han *et al*. 2013). This last filter was carried out because gene

families that have large gene copy number variance can cause parameter estimates to be non-informative (Han *et al*. 2013). The CAFE software was then run using the mode in which the gain and loss rates are estimated together (λ) for the whole phylogeny. For the entire analysis, the CAFE overall *p*-value threshold was kept at its default value (0.01). We used a custom script (https://github.com/asishallab/SlydGeneFamsAnalyses/blob/icruz/exec/parseCafeResult.R) to parse the CAFE output for functional enrichment analysis (see below).

Physicochemical protein divergence

We used all the multiple sequence alignments of the 24,235 (protein families with more than four proteins) protein families to carried out a Multivariate Analysis of Protein Polymorphism (MAPP program) (Stone and Sidow 2005). MAPP estimates the average deviation from six physicochemical properties (hydropathy, polarity, charge, volume, free energy in alpha-helix conformation, and free energy in beta-strand conformation) at an amino acid position across a multiple sequence alignment to assess the effect of a substitution at a particular amino acid site (physicochemical divergence) (Stone and Sidow 2005). Thus, we used MAPP to estimate the physiochemical divergence in each gene family. First, we used the script readAndParseOrthogroupsTxt.R (https://github.com/asishallab/SlydGeneFamsAnalyses/blob/icruz/exec/readAndParseOrthogroupsTxt.R) to parse and create folders from each gene family and stored its corresponding protein tree and multiple sequence alignment from OrthoFinder results. Then, we used MAPP program (Stone and Sidow 2005) with default parameters in each one of the protein families. We used the script readMappResults.R (https://github.com/asishallab/SlydGeneFamsAnalyses/blob/icruz/exec/readMappResults.R) to

parse and read all the MAPP results of the gene families. This script reads the MAPP results for all families, adjust *p*-value, find *Datura* genes of families with good multiple sequence alignments (Valdar Score > 0.6) and only retains significant sites with physicochemical divergence that fell into conserved domain proteins. Valdar Score method allows to score residues in a multiple sequence alignment and assigns a score ranging from 0 for low and 1 for high conservation (Valdar 2002). This program can be found in https://github.com/asishallab/SlydGeneFamsAnalyses/blob/icruz/exec/computeValdarMsaScores.R and was used into the readMappResults.R script.

Positive selection in gene families

We performed a codon-level analysis of positive natural selection with FUBAR program (Fast, Unconstrained Bayesian AppRoximation) (Murrel *et al*. 2013) on 24,235 gene families. FUBAR is a Bayesian approach to infer non-synoymous (dN) and synonymous (dS) substitution rates on a per-site basis for a given coding alignment and corresponding gene phylogeny (Murrell *et al*. 2013). To run FUBAR, first we retrieved the coding sequences (CDS) for each of the 13 Solanaceae species mentioned above. We removed trailing stop codons from the CDS, then we applied PAL2NAL (Suyama *et al*. 2006) to produce a codon alignment for each gene family. PAL2NAL is a program that converts a multiple sequence alignment of proteins and the corresponding DNA (CDS) sequences into a codon alignment (Suyama *et al*. 2006). Thus, we used the protein tree that we already had from each protein family to run PAL2NAL. FUBAR was run for all the codon alignments of each protein family. A custom Python script was used to transform the ".json" format from FUBAR result to tabular format. Then, the R script "loadFubarResults.R" from the R package GeneFamilies

(https://github.com/asishallab/GeneFamilies/blob/master/exec/loadFubarResults.R) was used to obtain a table with the significant posterior probabilities of a codon being subject to positive selection for each gene family (significant posterior probabilities ≥ 0.98; and Bayes Factor > 100).

Enrichment analysis

For enrichment test (Fisher's exact test, Fisher 1922), we used as background all the proteins from both genomes of *D. stramonium* (64,790 proteins) to detect over-represented proteins that showed signal of expansion, physicochemical divergence (MAPP), and with positively selected conserved amino acids (codons) (FUBAR). Functional annotation of the proteins was done using MapMan4 (Schwacke *et al*. 2019) and InterproScan. MapMan4 was used to annotate the general function of the proteins in order to retrieve the function of significant proteins resulted from MAPP, FUBAR and CAFE analyses, while InterProscan was used to identify and annotate domains overlapping the proteins with significant expansion signal, proteins with physicochemical divergence as well as positively selected conserved amino acids (codons). These analyses were done using the scripts "enrichedAnnosInExpContrFams.R (CAFE)", "identifyDomainsAtSelectedSites.R (FUBAR)" and "readMappResults.R (MAPP)" of the R package SlydGeneFamsAnalyses (https://github.com/asishallab/SlydGeneFamsAnalyses).

Genes involved in the tropane alkaloids biosynthesis

We investigated eight families that contain genes involved in the pathway of tropane alkaloids that is stored in the KEEG database; https://www.genome.jp/kegg-bin/show_pathway?map=map00960&show_description=show, Kanehisa and Sato 2019). These

genes correspond to Putrescine *N*-methyltransferase (*pmt*), Primary-amine oxidase (AOC3), Tropinone reductase I (*tpr* I), Tropinone reductase II (*tpr* II), Aspartate aminotransferase (GOT 2), Tyrosine aminotransferase (TAT), Hydroxyphenylpyruvate reductase (HPPR) and Hyoscyamine (6S)- dioxygenase (*h6h*). Multiple sequence alignments and protein trees for each family were generated from the previous analyses. We analyzed into our CAFE results if these eight protein families experienced expansions. Since proteins were already functional annotated, we also investigated the differences in the protein domain architecture in each gene family.

It is important to note that the gene family storing *h6h* contained just two genes belonging to both *D. stramonium* genomes. Since special interest was pointed out to the gene *h6h* (Hashimoto and Yamada 1986), we retrieved 13 *h6h* genes from UniProt database belong to *Datura metel* (acc. Q6EZB3), *Datura stramonium* Acc A0A0M4K1P1 (acc. A0A0M4K1P1), *Brugmansia arborea* (acc. A0A0M3SG09), *Hyoscyamus niger* (acc. P24397), *Brugmansia x candida* (hybrid plant generated by *Brugmansia aurea* x *Brugmansia versicolor*, acc. B2CNC8), *Hyoscyamus senecionis* (J7HDC2), *Atropa baetica* (acc. A9Q1G4), *Atropa belladonna* (acc. *Q9XJ43*), *Capsicum chinense* (acc. A0A2G3CG79), *Medicago truncatula* (acc. I3SNT9), *Glycine soja* (acc. A0A0B2P514), *Vitis vinifera* (acc. A0A438KDU2) and *Zea mays* (acc. B6T4W5). These genes were joined as a *h6h* gene family for which we generated a multiple sequence alignment with MAFT and a gene tree using FastTree2 with default parameters. Domain architecture was annotated with Pfam 31.0 database.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and materials

The complete workflow, all supplemental materials as well as commands used in this study are available in https://github.com/icruz1989/Datura-stramonium-genome-project. Genome assemblies, Illumina and PacBio raw sequences from the two plants of *D. stramonium* have been deposited at DDBJ/ENA/GenBank under the BioProject PRJNA622882: Teotihuacan assembly, *acc*. JAAWWX000000000, Ticumán assembly *acc*. JAAWWY000000000. Illumina and PacBio sequences for the Ticumán genome: *acc*. SRR11474700, SRR11474698, respectively. Illumina and PacBio sequences for the Teotihuacán genome: SRR11474701, SRR11474699, respectively.

Competing interests

All authors read and approved the final manuscript and declare no competing interests.

Funding

This work was financed by the CONACyT "Fronteras de la Ciencia" grant (#1527) and the PAPIIT, UNAM grant (PAPIIT-DGAPA AG200717) to JNF.

Authors' contributions

Conceived and designed the experiments: IMDA, AH, JNF. Performed the experiments: IMDA, RTL. Analyzed the data: IMDA, AH. Contributed reagents/materials/analysis tools: JNF, IMDA, AH, UO, RTL, SVM, DP, KO, BU. Wrote the paper: IMDA, JNF, AH.

Acknowledgements

We are deeply grateful with two anonymous reviewers from the Peerage of Science initiative and with Luis Eguiarte Fruns for their valuable comments and suggestions to improve this manuscript. We are also grateful to Luis F. Lozano Aguirre Beltrán of the Center for Genomic Sciences of the National Autonomous University of Mexico (UNAM) for helping and advising us with the first bioinformatics analyzes. We thank to Ernesto Campos and Alicia Mastretta Yanés for logistic and for the access to the cluster of the National Commission for the Knowledge and Use of Biodiversity (CONABIO, Mexico). S. Koren and B. P. Walenz from the National Human Genome Research Institute, USA, for all the help to carry out Canu assemblies. Also, the National Laboratory for Ecological Analysis and Synthesis (LANASE, UNAM, Mexico) gave us the access to use its cluster for bioinformatics analysis. Ivan Mijail De la Cruz Arguello is a doctoral student from the "Programa de Doctorado en Ciencias Biomédicas, National Autonomous University of Mexico (UNAM)" and acknowledges the scholarship (# 283799) granted by CONACyT.

References

Afzal, A J., A. Wood and D. A 2008 Lightfoot plant receptor-like serine threonine kinases: roles in signaling and plant defense. Molecular Plant-Microbe Interactions. 21(5):507-517

Akiva, E, S. Brown, D. E. Almonacid, A. E. Barber, A. F. Custer *et al*., 2014 The structure-function linkage database. Nucleic Acids. Res. 42:D521-D530.

Akkasaeng, C, N. Tantisuwichwong, I. Chairam *et al*., 2007 Isolation and identification of peanut leaf proteins regulated by water stress. Pakistan Journal of Biological Sciences. 10(10):1611-1617.

Aljbory, Z. and M. S. Chen, 2018 Indirect plant defense against insect herbivores: a review. Insect Science. 25:2-23.

Andrews, S., 2010 FastQC: a quality control tool for high throughput sequence data. http://www.bioinformatics.babraham.ac.uk/projects/fastqc

Antony, A. C., 2018 Chapter 39-Megaloblastic Anemias. In: Hoffman R., E. J. Benz, L. E. Silberstein, H. E. Heslop, JI Weitz, J. Anastasi *et al*., editors. Elsevier. pp. 514-545.

Attwood, T. K., A. Coletta, G. Muirhead, A. Pavlopoulou, P. B. Philippou *et al*., 2012 The PRINTS database: a fine-grained protein sequence annotation and analysis resource, its status in 2012. Database. bas019.

Ballvora, A., M. R. Ercolano, J. Weiß, K. Meksem, C. A. Bormann *et al*., 2002 The *R1* gene for potato resistance to late blight (*Phytophthora infestans*) belongs to the leucine zipper/NBS/LRR class of plant resistance genes. The Plant Journal. 30:361-371.

Barclay, A. S., 1959 Studies in the genus *Datura* (Solanaceae). I. Taxonomy of subgenus *Datura*. Ph. D. Thesis. Harvard University, Cambridge, MA, USA.

Barchi, L., Pietrella, M., Venturini, L., Minio, A., Toppino, L., Acquadro, A., Andolfo, G., Aprea, G., Avanzato, C., Bassolino, L., Comino, C., Molin, A. D., Ferrarini, A., Maor, L. C., Portis, E., Reyes-Chin-Wo, S., Rinaldi, R., Sala, T., Scaglione, D., … Rotino, G. L. (2019). A chromosome-anchored eggplant genome sequence reveals key events in Solanaceae evolution. Scientific Reports, 9(1), 11769.

Bello-Bedoy, R., and J. Núñez-Farfán, 2010 Cost of inbreeding in resistance to herbivores in *Datura stramonium*. Ann. Bot.105:747-53.

Bischoff, V., S. Nita, L. Neumetzler, D. Schindelasch, A. Urbain *et al*., 2010 TRICHOME BIREFRINGENCE and its homolog AT5G01360 encode plant-specific DUF231 proteins required for cellulose biosynthesis in Arabidopsis. Plant Physiol. 153:590-602.

Boachon, B, R. R. Junker, L. Miesch, J. E. Bassard, R. Höfer *et al*., 2015 CYP76C1 (Cytochrome P450)-mediated linalool metabolism and theformation of volatile and soluble linalool oxides in *Arabidopsis* flowers: a strategy for defense against floral antagonists. Plant Cell. 27:2972LP-2990.

Bolger, A. M., M. Lohse, and B. Usadel, 2014a Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics.30:2114-20.

Bolger, A., F. Scossa, M. E. Bolger, C. Lanz, F. Maumus *et al*., 2014b The genome of the stress-tolerant wild tomato species *Solanum pennellii*. Nature Genetics. 46:1034.

Bombarely, A., M. Moser, A. Amrad, M. Bao, L. Bapaume, C. S. Barry *et al*., 2016 Insight into the evolution of the Solanaceae from the parental genomes of *Petunia hybrida*. Nature Plants. 2:16074.

Boratyn, G. M., C. Camacho, P. S. Cooper, G. Coulouris, A. Fong *et al*., 2013 BLAST: a more efficient report with usability improvements. Nucleic Acids Res. 41:W29-W33.

Bru, C., E. Courcelle, S. Carrère, Y. Beausse, S. Dalmar *et al*., 2005 The ProDom database of protein domain families: more emphasis on 3D. Nucleic Acids Res. 33:D212-D215.

Buchfink, B., C. Xie, and D. Huson, 2015 Fast and sensitive protein alignment using DIAMOND. Nature Methods. 12:59-60

Bustos-Segura, C., J. Fornoni, and J. Núñez-Farfán, 2014 Evolutionary changes in plant tolerance against herbivory through a resurrection experiment. J. Evol. Biol. 27:488-96.

Campbell, M. S., C. Holt, B. Moore, and M. Yandell, 2014 Genome Annotation and Curation Using MAKER and MAKER-P. Curr. Protoc. Bioinforma. 48:4.11.1-4.11.39.

Castillo, G., L. L. Cruz, J. Hernández-Cumplido, K. Oyama, C. M. Flores-Ortiz *et al*., 2013 Geographic association and temporal variation of chemical and physical defense and leaf damage in *Datura stramonium*. Ecol. Res. 28:663-72.

Castillo, G., L. L. Cruz, R. Tapia-López, E. Olmedo-Vicente, D. Carmona *et al*., 2014 Selection mosaic exerted by specialist and generalist herbivores on chemical and physical defense of *Datura stramonium*. PLoS One. 9(7):e102478.

Chakraborty, M., J. G. Baldwin-Brown, A. D. Long, and J. J. Emerson, 2016 Contiguous and accurate de novo assembly of metazoan genomes with modest long read coverage. Nucleic Acids Res. 44:e147-e147.

Cheng, A. X., Y. G. Lou, Y. B. Mao, S. Lu, L. J. Wang *et al*., 2007 Plant terpenoids: biosynthesis and ecological functions. Journal of Integrative Plant Biology. 49:179-186.

Chikhi, R., and P. Medvedev, 2013 Informed and automated k-mer size selection for genome assembly. Bioinformatics.30:31-7.

Chowański, S., Z. Adamski, P. Marciniak, G. Rosiński, E. Büyükgüzel *et al*., 2016 A review of bioinsecticidal activity of Solanaceae alkaloids. Toxins. 8:60

de Lima Morais, D. A., H. Fang, O. J. L. Rackham, D. Wilson, R. Pethica, C. Chothia *et al*., 2011 SUPERFAMILY 1.75 including a domain-centric gene ontology method. Nucleic Acids Res. 39:D427-34.

De la Cruz, M., 1552. *Libellus de medicinalibus indorum herbis*. Translated to Latin by Juan Badiano. Reprinted in 1964 by Instituto Mexicano del Seguro Social, Fondo de Cultura Económica Mexico City, Mexico.

De-la-Cruz, I. M., L. L. Cruz, L. Martínez-García, P. L. Valverde, C. M. Flores-Ortiz *et al*., 2020a Evolutionary response to herbivory: population differentiation in microsatellite loci, tropane alkaloids and leaf trichome density in *Datura stramonium*. Arthropod-Plant Interactions. 14:21-30.

De-la-Cruz, I. M., and Núñéz-Farfán, J. 2020b The complete chloroplast genomes of two Mexican plants of the medicinal and toxic herb *Datura Stramonium* (Solanaceae). Mitochondrial DNA Part B. 5(3):2829-2831.

De Sahagún, B., 1577 Historia general de las cosas de la Nueva España. Reprinted in 1830. Alejandro Valdés' Press, Mexico City, Mexico.

Degenhardt, J., J. Gershenzon, I. T. Baldwin, and A. Kessler, 2003 Attracting friends to feast on foes: engineering terpene emission to make crop plants more attractive to herbivore enemies. Curr. Opin. Biotechnol. 14:169-76.

Doyle, J. J., and J. L. Doyle, 1987 A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochemical Bulletin.19:11-15.

Dräger, B., A. Portsteffen, A. Schaal, P. McCabe, A. Peerless *et al*., 1992 Levels of tropinone-reductase activities influence the spectrum of tropane esters found in transformed root cultures of *Datura stramonium* (L.). Planta 188(4):81-586.

Edwards, K. D., N. Fernandez-Pozo, K. Drake-Stowe, M. Humphry, A. D. Evans *et al*., 2017A reference genome for *Nicotiana tabacum* enables map-based cloning of homologous loci implicated in nitrogen utilization efficiency. BMC Genomics. 8:448.

Emms, D. M., and S. Kelly, 2019 OrthoFinder: phylogenetic orthology inference for comparative genomics. Genome Biol. 20:238.

Enright, A., S. Dongen, and C. A. Ouzounis, 2002 An efficient algorithm for large-scale detection of protein families. Nucleic acids research. 30:1575-84.

Fernandes, H., O. Pasternak, G. Bujacz, A. Bujacz, M. M. Sikorski *et al*., 2008 *Lupinus luteus* pathogenesis-related protein as a reservoir for cytokinin. Mol. Biol. 378:1040-51.

Finn, R. D., A. Bateman, J. Clements, P. Coggill, R. Y. Eberhardt *et al*., 2014 Pfam: the protein families database. Nucleic Acids Res. 42:D222–D230.

Fisher, R. A., 1922 On the interpretation of χ2 from contingency tables, and the calculation of P. Journal of the Royal Statistical Society. 85(1):87-94.

Fornoni, J., P. L. Valverde, and J. Nunez-Farfan, 2004 Population variation in the cost and benefit of tolerance and resistance against herbivory in *Datura stramonium*. Evolution.58:1696-704.

Gao, S., D. Bertrand, B. K. H. Chia, and N. Nagarajan, 2016 OPERA-LG : efficient and exact scaffolding of large , repeat-rich eukaryotic genomes with performance guarantees. Genome Biol.1-16.

Gurevich, A., V. Saveliev, N. Vyahhi, and G. Tesler, 2013 QUAST: quality assessment tool for genome assemblies. Bioinformatics.29:1072-5.

Haft, D. H., J. D. Selengut, R. A. Richter, D. Harkins, M. K. Basu *et al*., 2013 TIGRFAMs and Genome Properties in 2013. Nucleic Acids Res. 41:D387-D395.

Hagel, J. M., and Facchini, 2010 Dioxygenases catalyze the O-demethylation steps of morphine biosynthesis in opium poppy. PJ. Nat. Chem. Biol. 6:273-5.

Han, M. V., G. W. C. Thomas, J. Lugo-Martinez, and M. W. Hahn, 2013 Estimating gene gain and loss rates in the presence of error in genome assembly and annotation using CAFE 3. Molecular Biology and Evolution 30(8):1987-1997.

Hightower, C. E., 1979 Plants that kill and cure. Vet Hum Toxicol. 21:360-2.

Hulse-Kemp, A. M., S. Maheshwari, K. Stoffel, T. A. Hill, D. Jaffe *et al*., 2018 Reference quality assembly of the 3.5-Gb genome of *Capsicum annuum* from a single linked-read library. Hortic. Res. 5:4.

Jirschitzka, J., Schmidt, G. W., Reichelt, M., Schneider, B., Gershenzon, J., & D'Auria, J. C. (2012). Plant tropane alkaloid biosynthesis evolved independently in the Solanaceae and Erythroxylaceae. *Proceedings of the National Academy of Sciences*, *109*(26), 10304–10309.

Jones, P., D. Binns, H. Y. Chang, M. Fraser, W. Li *et al*., 2014 InterProScan 5: genome-scale protein function classification. Bioinformatics. 30:1236-1240.

Kanehisa, M., and Y. Sato, 2019 KEGG Mapper for inferring cellular functions from protein sequences. Protein Science. 29:28-35.

Kasukabe, Y., L. He, K. Nada, S. Misawa, I. Ihara *et al*., 2004 Overexpression of spermidine synthase enhances tolerance to multiple environmental stresses and up-regulates the expression of various stress-regulated genes in transgenic *Arabidopsis thaliana*. Plant & cell physiology. 45:712-22.

Katoh, M., and M. Kuma, 2002 MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids. Res. 30:3059-3066.

Kawahara, Y., T. Hashimoto, H. Nakayama, and Y. Kitamura. 2017 Galactose oxidase/kelch repeat-containing protein is involved in the iron deficiency stress response in the roots of *Hyoscyamus albus*. Plant Root.11:58-63.

Keiner, R., H. Kaiser, K. Nakajima *et al*., 2002 Molecular cloning, expression and characterization of tropinone reductase II, an enzyme of the SDR family in *Solanum tuberosum* (L.). Plant Mol. Biol. 48:299-308.

Kohnen-Johannsen, K. L., and O. Kayser, 2019 Tropane Alkaloids: chemistry, pharmacology, biosynthesis and Production. Molecules. 24:796.

Koren, S., B. P. Walenz, K. Berlin, J. R. Miller, N. H. Bergman *et al*., 2017 Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. Genome Res.27:722-36.

Korf, I., 2004 Gene finding in novel genomes. BMC Bioinformatics. 5:59.

Kumar, S., G. Stecher, M. Suleski, and S. B. Hedges, 2017 TimeTree: A Resource for Timelines, Timetrees, and Divergence Times. Molecular Biology and Evolution. 34(7):1812-1819,

Kurtz, S., A. Phillippy, A. L. Delcher, M. Smoot, M. Shumway *et al*., 2004 Versatile and open software for comparing large genomes. Genome Biol. 5:R12.

Langmead, B., C. Wilks, V. Antonescu, and R. Charles, 2018 Scaling read aligners to hundreds of threads on general-purpose processors. Bioinformatics. 35:421-32.

Letunic, I., T. Doerks, and P. Bork, 2012 SMART 7: recent updates to the protein domain annotation resource. Nucleic Acids Res. 40:D302-D305.

Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer *et al*., 2009 The sequence alignment/map format and SAMtools. Bioinformatics. 25:2078-9.

Li, R., D. Reed, E. Liu, J. Nowak, L. E. Pelcher *et al*., 2006 Functional genomic analysis of alkaloid biosynthesis in *Hyoscyamus niger* reveals a cytochrome P450 involved in littorine rearrangement. Chem. Biol.13:513-20. Available from:

Liu, Q., B. Khakimov, P. D. Cárdenas, F. Cozzi, C. E. Olsen *et al*., 2019 The cytochrome P450 CYP72A552 is key to production of hederagenin-based saponins that mediate plant defense against herbivores. New Phytol. 222: 1599-1609.

Lupas, A.V., M. Van Dyke, and J. Stock, 1991 Predicting Coiled Coils from Protein Sequences. Science. 252(5009):1162-1164

Magalhães, D. M., L. L. Scholte, N. V. Silva, , G. C. Oliveira, C. Zipfel *et al*., 2016 LRR-RLK family from two *Citrus* species: genome-wide identification and evolutionary aspects. BMC genomics. 17(1):623.

Marchler-Bauer, A., Y. Bo, L. Han, J. He, C. J. Lanczycki *et al*., 2017 CDD/SPARCLE: functional classification of proteins via subfamily domain architectures. Nucleic Acids Res. 45:D200-D203.

Massengo-Tiassé, R. P., and J. E. Cronan, 2009 Diversity in enoyl-acyl carrier protein reductases. Cell. Mol. Life Sci. 66:1507-1517.

Miranda-Pérez, A., G. Castillo, J. Hernández-Cumplido, P. L. Valverde, M. Borbolla *et al*., 2016 Natural selection drives chemical resistance of *Datura stramonium*. PeerJ.4:e1898-e1898.

Mithöfer, A. and W. Boland, 2012 Plant Defense Against Herbivores: Chemical Aspects. Annual Review of Plant Biology. 63(1):431-450.

Murrell, B., S. Moola, A. Mabona, T. Weighill, D. Sheward *et al*., 2013 FUBAR: A Fast, Unconstrained Bayesian AppRoximation for Inferring Selection, Molecular Biology and Evolution. 30(5):1196-1205.

Nakajima, K., T. Hashimoto, and Y. Yamada, 1993 Two tropinone reductases with different stereospecificities are short-chain dehydrogenases evolved from a common ancestor. Proc. Natl. Acad. Sci. 90:9591LP-9595LP.

Nasomjai, P., D. Reed, D. Tozer, M. Peach, A. Slawin *et al*., 2009. Mechanistic Insights into the Cytochrome P450-mediated oxidation and rearrangement of littorine in tropane alkaloid biosynthesis. Chembiochem. 10:2382-93.

Necci, M., D. Piovesan, Z. Dosztányi, and S. C. E. Tosatto, 2017 MobiDB-lite: fast and highly specific consensus prediction of intrinsic disorder in proteins. Bioinformatics. 33:1402-1404.

Nikolskaya, A. N., C. N. Arighi, H. Huang, W. C. Barker, and C. H. Wu, 2007 PIRSF family classification system for protein functional and evolutionary analysis. Evol. Bioinform. 2:197-209.

Ozerov, M. Y., F. Ahmad, R. Gross, L. Pukk, S. Kahar, V. Kisand *et al*., 2018 Highly continuous genome assembly of Eurasian Perch (*Perca fluviatilis*) using linked-read sequencing. G3. 8:3737-3743.

Padmanabhan, M., P. Cournoyer, and S. P. Dinesh-Kumar, 2009 The leucine-rich repeat domain in plant innate immunity: a wealth of possibilities. Cellular microbiology, 11(2):191-198.

Pasternak, O., J. Biesiadka, R. Dolot, L. Handschuh, G. Bujacz *et al*., 2005 Structure of a yellow lupin pathogenesis-related *PR-10* protein belonging to a novel subclass. Acta Crystallographica Section D. 61:99-107.

Patocka, J., and R. Jelinkova, 2017 Atropine and atropine-like substances usable in warfare. Military Medical Science Letters.86:58-69.

Price, M. N., P. S Dehal, and A. P. Arkin, 2010 FastTree 2 Approximately Maximum-Likelihood Trees for Large Alignments. PLoS ONE. 5(3):e9490.

Qiang, W., Y. L. Hou, X. Li, K. Xia, and Z. H. Liao, 2015 Cloning and expression of the key enzyme hyoscyamine 6 beta-hydroxylase gene (DaH6H) in scopolamine biosynthesis of *Datura arborea*. Yao Xue Xue Bao. 50:1346-1355.

Qin, C., C. Yu, Y. Shen, X. Fang, L. Chen, *et al*., 2014 Whole-genome sequencing of cultivated and wild peppers provides insights into into *Capsicum* domestication and specialization. Proc. Natl. Acad. Sci. 111:5135-5140.

Rairdan, G. J., and P. Moffett, 2006 Distinct domains in the ARC region of the potato resistance protein *Rx* mediate LRR binding and inhibition of activation. Plant Cell. 18:2082LP-2093LP.

Ramisetty, B. C. M., and R. S. Santhosh, 2016 Horizontal gene transfer of chromosomal Type II toxin–antitoxin systems of *Escherichia coli*. FEMS Microbiology Letters. 363(3). fnv238.

Razali, R., S. Bougouffa, M. Morton, D. J. Lightfoot, I. Alam *et al*., 2018 The Genome sequence of the wild Tomato *Solanum pimpinellifolium* provides insights into salinity tolerance. Frontiers in plant science. 9:1402.

Salzberg, S. L., A. M. Phillippy, A. Zimin, D. Puiu, T. Magoc, *et al*., 2012 GAGE: A critical evaluation of genome assemblies and assembly algorithms. Genome Res.22:557-67.

Särkinen, T., L. Bohs, R. G. Olmstead *et al*., 2013 A phylogenetic framework for evolutionary study of the nightshades (Solanaceae): a dated 1000-tip tree. BMC Evol. Biol. 13:214.

Satheesh, V., P. Chidambaranathan, P. T. Jagannadham, V. Kumar, P. K. Jain *et al*., 2016 Transmembrane START domain proteins: in silico identification, characterization and expression analysis under stress conditions in chickpea (*Cicer arietinum* L.). Plant signaling & behavior. 11(2):e992698.

Sato, S., S. Tabata, H. Hirakawa. *et al*., 2012 The tomato genome sequence provides insights into fleshy fruit evolution. Nature 485:635-641.

Savolainen, O., Lascoux, M., & Merilä, J. (2013). Ecological genomics of local adaptation. *Nature Reviews Genetics*, *14*(11), 807–820.

Schwacke, R., G. Y. Ponce-Soto, K. Krause, A. M. Bolger, B. Arsova, A. Hallab *et al*., 2019 MapMan4: A refined protein classification and annotation framework applicable to multi-omics data analysis. Molecular Plant. 12:879-92.

Seifi, H. S., A. Zarei, T. Hsiang, and B. J. Shelp, 2019 Spermine is a potent plant defense activator against gray Mold disease on *Solanum lycopersicum*, *Phaseolus vulgaris*, and *Arabidopsis thaliana*. Phytopathology. 109(8):1367-1377.

Shanker, A. K., M. Maheswari, S. K. Yadav *et al*., 2014 Drought stress responses in crops. Funct. Integr. Genomics. 14:11–22.

Shidore, T., and L. R. Triplett, 2017 Toxin-antitoxin systems: implications for plant disease. Annual Review of Phytopathology. 55(1):161-179.

Shonle, I., and J. Bergelson, 2000 Evolutionary ecology of the tropane alkaloids of *Datura stramonium* (Solanaceae). Evolution. 54:778-88.

Sierro, N., J. N. D. Battey, S. Ouadi, L. Bovet, S. Goepfert *et al*., 2013 Reference genomes and transcriptomes of *Nicotiana sylvestris* and *Nicotiana tomentosiformis*. Genome Biol. 14:R60.

Sigrist, C. J. A., E. de Castro, L. Cerutti, B. A. Cuche, N. Hulo *et al*., 2013 New and continuing developments at PROSITE. Nucleic Acids Res. 41:D344-D347.

Simão, F. A., R. M. Waterhouse, P. Ioannidis, E.V. Kriventseva, and E. M. Zdobnov, 2015 BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics.31:3210-2.

Smit, A. F. A., and R. Hubley, 2008-2015 RepeatModeler Open-1.0. http://www.repeatmasker.org.

Smit, A. F. A., R. Hubley, and P. Green, 2013-2015 RepeatMasker Open-4.0. http://www.repeatmasker.org.

Stanke, M., O. Keller, I. Gunduz, A. Hayes, S. Waack *et al*., 2006 AUGUSTUS: ab initio prediction of alternative transcripts. Nucleic Acids Res. 34(2):W435-W439.

Stone, E. A., and A. Sidow, 2005 Physicochemical constraint violation by missense substitutions mediates impairment of protein function and disease severity. Genome research, 15(7):978-986.

Stone, J. M., and J. C. 1995 Walker plant protein kinase families and signal transduction. Plant Physiology. 108:451-457.

Suyama, M., D. Torrents, and P. Bork, 2006 PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. Nucleic acids research. 34:W609-W612.

Symon, D. E. and Haegi, L., 1991 *Datura* (Solanaceae) is a New World genus. In: Hawkes, J.G., Lester, R.N., Nee, M, and Estrada, R.N., editors. Solanaceae III: taxonomy, chemistry, evolution. London, UK: The Royal Botanic Gardens, Kew, 197–210.

Tameling, W. I. L., and DC. Baulcombe, 2007 Physical association of the NB-LRR resistance protein *Rx* with a ran GTPase–Activating protein is required for extreme resistance to Potato virus. Plant Cell. 19:1682LP-1694LP.

Taron, C., J. Benner, L. Hornstra, and E. A. 1995 Guthrie novel β-galactosidase gene isolated from the bacterium *Xanthomonas manihotis* exhibits strong homology to several eukaryotic β-galactosidase. Glycobiology. 5:603-10.

Thangella, P. A.V., S. N. B. S. Pasumarti, R. Pullakhandam *et al*., 2018 Differential expression of leaf proteins in four cultivars of peanut (*Arachis hypogaea* L.) under water stress. 3 Biotech. 8:157.

Tholl, D., 2006 Terpene synthases and the regulation, diversity and biological roles of terpene metabolism. Curr. Opin. Plant. Biol. 9:297–304.

Thomas, P. D., M. J. Campbell, A. Kejariwal, H. Mi, B. Karlak *et al*., 2003 PANTHER: a library of protein families and subfamilies indexed by function. Genome Res. 13:2129-2141.

Valdar, W. S., 2002 Scoring residue conservation. Proteins. 48:227-241.

Ujváry, I., 2010 Chapter 3-Pest Control Agents from Natural Products. In R. Krieger (Ed.), Hayes' Handbook of Pesticide Toxicology (Third Edition) (pp. 119–229). Academic Press.

Valverde, P. L., J. Fornoni, and J. Núñez-Farfán, 2001 Defensive role of leaf trichome in resistance to herbivorous in *Datura stramonium*. J. Evol. Biol. 14:424-432.

Valverde, P. L., J. Fornoni, and J. Núñez-Farfán, 2003 Evolutionary ecology of *Datura stramonium*: equal plant fitness benefits of growth and resistance against herbivory. J. Evol. Biol.16:127-37.

van Ooijen, G., G. Mayr, M. A. K. Mobien, M. Albrecht, B. J. C. Cornelissen *et al*., 2008 Structure–function analysis of the NB-ARC domain of plant disease resistance proteins. Journal of Experimental Botany. 59(6):1383-1397.

Walker, B. J., T. Abeel, T. Shea, M. Priest, A. Abouelliel, S. Sakthikumar *et al*., 2014 Pilon: An integrated tool for comprehensive microbial variant detection and genome assembly improvement. PLoS One. 9:e112963.

Weaver, S. E., and S. I. Warwick, 1984 The biology of Canadian weeds: 64. *Datura stramonium* L. Canadian Journal of Plant Science: 64: 979-991.

Wei, H., S. Collier, P. Moffett, and J. Chai, 2013 Structural basis for the interaction between the potato virus X Resistance Protein (*Rx*) and its cofactor ran GTPase-Activating Protein 2 (RanGAP2). The Journal of biological chemistry. 288. 10.1074/jbc.M113.517417.

Wei, Zhu., B. Geert, R. Nachman, G. Goldsworthy, P. Verhaert, A. Loof, L. Schoofs, 2000 Sulfakinins reduce food intake in the desert locust, *Schistocerca gregaria*. Journal of Insect Physiology. 46:1259-1265.

WHO, 2015 Annex 1 19th WHO Model List of Essential Medicines; WHO: Geneva, Switzerland.

Xu, S., T. Brockmöller, A. Navarro-Quezada, H. Kuhl, K. Gase *et al*., 2017 Wild tobacco genomes reveal the evolution of nicotine biosynthesis. Proc. Natl. Acad. Sci. 114:6133LP-LP6138.

Xu, X., S. Pan, S. Cheng *et al*., 2011 Genome sequence and analysis of the tuber crop potato. Nature. 475:189-195.

Ye, C., C. M. Hill, S. Wu, J. Ruan, and Z. Ma, 2016 DBG2OLC: Efficient assembly of large genomes using long erroneous reads of the third-generation sequencing technologies. Sci Rep. 6:31900.

Yeats, C., M. Maibaum, R. Marsden, M. Dibley, D. Lee *et al*., 2006 Gene3D: modelling protein structure, function and evolution. Nucleic Acids Res. 34:D281-284.

Yu, N., R. J. Nachman, and G. Smagghe, 2013 Characterization of sulfakinin and sulfakinin receptor and their roles in food intake in the red flour beetle *Tribolium castaneum*. Gen. Comp. Endocrinol. 188:196-203.

Table 1. DNA reads produced for *Datura stramonium* genome assembly

Library technology	Plant ID	Sequences	Number of total sequences	Length of the sequences (bp)	Trimmed sequences	Length of the trimmed sequences	% GC
Illumina HiSeq 4000	Teotihuacán	PE reads	159,031,241	150	151,898,993	36-151	39
	Ticumán	PE reads	161,985,268	150	152,665,899	36-151	41
PacBio Sequel	Teotihuacán	SMRTbell library Subreads	9,505,413	~ 9,000-15,000	--	--	39
	Ticumán		9,995,713	~ 9,000-18,000	--	--	39

Table 2. Assembly statistics for both genomes of *Datura stramonium*.

Assembly	**Ticumán**	**Teotihuacán**
# contigs (>= 0 bp)	27,915	30,392
# contigs (>= 1,000 bp)	27,843	30,344
# contigs (>= 5,000 bp)	26,349	29,471
# contigs (>= 10,000 bp)	24,397	27,709
# contigs (>= 25,000 bp)	17,380	17,390
# contigs (>= 50,000 bp)	9,343	7,299
# contigs	27,900	30,385
Largest contig	3,131,142	2,112,153
Total length	1,482,568,706	1,288,884,002
GC (%)	38.47	38.45
N50	84,121	58,197
N75	44,455	32,640
L50	4,557	5,713
L75	10,641	13,166
# N's per 100 kbp	721.42	110.68

Table 3. Single-copy othologs (BUSCOs) statistics for both Genomes of *Datura stramonium*.

BUSCO statistics	Ticumán	Teotihuacán
Complete BUSCOs (c)	2,779	2,493
Complete and single-copy BUSCOs (S)	2,693	2,418
Complete and duplicated BUSCOs (D)	86	75
Fragmented BUSCOs (F)	128	189
Missing BUSCOs (M)	145	370
Total BUSCO groups searched	3,052	3,052

Table 4. Annotation statistics for the *Datura stramonium* genomes.

Features	Ticumán	Teotihuacán
Number of genes	33,856	30,934
Number of exons	176,756	163,107
Number of exons in cds	170,946	157,410
Number of introns in cds	137,090	126,476
Number of introns in exon	142,900	132,173
mean mrnas per gene	1.0	1
mean exons per mrna	5.2	5.3
Total gene length	147,987,251	132,043,897
% of genome covered by gene	16	14.2
% of genome covered by cds	2.5	2.6
% of genome covered by exon	2.7	2.9
% of genome covered by intron from cds	7.1	7.2
% of genome covered by intron from exon	7.3	7.4

Table 5. Number of elements, length occupied and Percentage of sequence of the repeat elements in the genome assemblies of both *Datura stramonium* individuals identified by RepeatModeler and RepeatMasker.

Repeat class	Number of elements	Length occupied (bp)	Percentage of sequence	Number of elements	Length occupied (bp)	Percentage of sequence
	Ticumán			Teotihuacán		
SINEs	16,542	4,402,707	0.30	18,140	3,331,482	0.26
LINEs:	43,108	23,439,474	1.58	46,737	23,593,403	1.83
LTR	534,845	976,671,332	65.88	492,830	817,336,298	63.41
DNA elements	162,833	48,087,803	3.24	171,198	46,018,189	3.57
Unclassified	114,168	52,221,075	3.52	115,278	46,571,233	3.61
Small RNA	14,535	5,997,859	0.40	16,014	6,557,612	0.51
Satellites	2,934	875,492	0.06	2,956	661,284	0.05
Simple repeats	184,608	12,181,935	0.82	171,120	10,393,978	0.81
Low complexity	43,493	3,524,503	0.24	44,144	3,089,946	0.24
Total	1,117,066	1,127,402,180	76.04	1,078,417	957,553,425	74.11

Table 6. Classifications of domains related with defence against natural enemies (pathogens, viruses, fungi, oomycete, herbivores) subject to expansion, positive selection or physicochemical divergence. Some domains were detected to be expanded and positively selected. Ex = expanded, PS = positive selected, FQ = physicochemical divergence. P value is showed for each analysis. The entire list for each analysis is showed in additional files 8, 9 and 10.

Defensive domains	InterProscan ID	*p*-value	Analysis
Virus X resistance protein-like	IPR038005	2.851E-13/1.120E-17	Ex, PS
Rx, N-terminal	IPR041118	3.412E-04/4.928E-09	Ex, PS
NB-ARC	IPR002182	9.735E-10/5.184E-21	Ex, PS
Actin family	IPR004000	7.157E-09	Ex
Actin, conserved site	IPR004001	9.950E-06/6.619E-07	Ex, PS
Actin/actin-like conserved site	IPR020902	2.980E-11/4.336E-04	Ex, PS
START-like domain superfamily	IPR023393	3.878E-04	Ex
Bet v I/Major latex protein	IPR000916	4.277E-08/9.020E-04	Ex, PS
Zinc finger, PMZ-type	IPR006564	4.402E-04	Ex
Late blight resistance protein R1	IPR021929	5.064E-05	Ex
Leucine-rich repeat domain superfamily	IPR032675	1.166E-05/4.319E-18	Ex, PS
Ribonuclease H-like superfamily	IPR012337	1.000E-04	PS
DNA-binding pseudobarrel domain superfamily	IPR015300	2.277E-03	Ex
Syntaxin, N-terminal domain	IPR006011	1.009E-03	FQ
Target SNARE coiled-coil homology domain	IPR000727	7.302E-03	FQ
Syntaxin/epimorphin, conserved site	IPR006012	1.041E-03	FQ
Transmembrane protein 131-like	IPR039877	1.041E-03	FQ
Ribonuclease II/R	IPR001900	7.031E-03	FQ
BRCA1-associated	IPR031099	9.441E-03	FQ
Trichome birefringence-like 45	IPR029981	4.774E-03	FQ

Table 7. Classifications of domains related with the biosynthesis of secondary compounds and that act as defence against natural enemies (pathogens, viruses, fungi, oomycete, herbivores) subject to expansion, positive selection or physicochemical divergence. Some domains were detected to be expanded and positively selected. Ex = expanded, PS = positive selected, FQ = physicochemical divergence. P value is showed for each analysis. The entire list for each analysis is showed in additional files 8, 9 and 10.

Domains related with the biosynthesis of secondary compounds	InterproScan	*p*-value	Pathway of secondary compounds	Analysis
Cytochrome P450 superfamily	IPR036396	7.888E-03/2.134E-10	Tropane, terpenoid	Ex, PS
Cytochrome P450, E-class, group I	IPR002401	4.539E-03/4.047E-12	Tropane, terpenoid	Ex, PS
Cytochrome P450	IPR001128	5.179E-03/1.033E-10	Tropane, terpenoid	Ex, PS
Cytochrome P450, conserved site	IPR017972	1.410E-08	Tropane, terpenoid	PS
Aminotransferase-like, plant mobile	IPR019557	8.030E-11	Tropane	Ex
Transferase	IPR003480	8.572E-05/6.310E-04	Tropane	Ex, PS
Isoprenoid synthase domain superfamily	IPR008949	1.909E-04	Isoprenoid	Ex
Terpene synthase, N-terminal domain	IPR001906	1.704E-05/2.295E-07	Terpenoid	Ex, PS
Terpene synthase, N-terminal domain	IPR036965	1.364E-04/3.426E-07	Terpenoid	Ex, PS
Terpenoid cyclases/protein prenyltransferase alpha-alpha toroid	IPR008930	1.140E-03/2.163E-07	Terpenoid	Ex, PS
Terpene synthase, metal-binding domain	IPR005630	5.064E-05/2.289E-06	Terpenoid	Ex, PS
Terpene cyclase-like 1, C-terminal	IPR034741	8.806E-03/2.533E-06	Terpenoid	Ex, PS
NADH:ubiquinone oxidoreductase	IPR003918	7.056E-03	Tropane	Ex
NADH-quinone oxidoreductase, subunit D superfamily	IPR038290	1.571E-04	Tropane	Ex
NADH-quinone oxidoreductase, subunit D	IPR001135	8.570E-05	Tropane	Ex
NADH-quinone oxidoreductase chain 4	IPR022997	8.806E-03	Tropane	Ex
Glutathione S-transferase, C-terminal-like	IPR010987	3.430E-07	Glutathione	PS
Glutathione Transferase family	IPR040079	4.542E-04	Glutathione	PS
Glutathione S-transferase, N-terminal	IPR004045	9.960E-07	Glutathione	PS
Glutathione S-transferase, C-terminal	IPR004046	1.970E-05	Glutathione	PS
Phosphoethanolamine N-methyltransferase	IPR025771	4.280E-07	Tropane	FQ
Glycoside hydrolase, family 35	IPR001944	4.176E-04	Tropane, terpenoid, glutathione, isoprenoids	FQ
Glycoside hydrolase 35, catalytic domain	IPR031330	3.637E-04	Tropane, terpenoid, glutathione, isoprenoids	FQ
Glycosyltransferase family 92	IPR008166	2.379E-03	Tropane, terpenoid, glutathione, isoprenoids	FQ
Glycosyl transferase, family 31	IPR002659	7.031E-03	Tropane, Terpenoid, glutathione, isoprenoids	FQ
NADPH-cytochrome P450 reductase	IPR023208	4.774E-03	Tropane, terpenoid	FQ
SNF1-related protein kinase regulatory subunit beta-2	IPR030070	4.740E-06	Terpenoids	FQ
Association with the SNF1 complex (ASC) domain	IPR006828	2.360E-05	Terpenoids	FQ

Figure legends

Fig. 1. Genome size estimation in *Datura stramonium* by the K-mer distribution of the Illumina DNA reads (a = Ticumán, b) Teotihuacán). c) GC content plot shows the distribution of GC content in the contigs (red line = Ticumán, blue line = Teotihuacán). d) Cumulative length plot shows the growth of contig lengths. On the x-axis, contigs are ordered from the largest to smallest. The y-axis gives the size of the x largest contigs in the assembly. This is the total genome assembled (red line = Ticumán, blue line = Teotihuacán). e) BUSCO plots for the two *Datura stramonium* genomes. The plot shows quantitative measures for the assessment of the genome completeness based on evolutionarily informed expectations of gene content from near-universal single-copy orthologs selected from OrthoDB.v9. The *Datura* genomes assembly covered 91% and 81.7 % complete single-copy orthologs for Ticumán and Teotihuacán, respectively.

Fig. 2. The repeat landscapes depict the relative abundance of repeat classes in the genome of *Datura stramonium* (Ticumán example) versus the Kimura divergence from the consensus. LTR/Gypsy family is the most represented repetitive element in the genome of *D. stramonium* (61.33 %) followed by LTR Copia family. Genome of Teotihuacán also presents the same pattern.

Fig. 3. a) Venn diagram shows 7,653 InterProscan domains that are shared between *Datura stramonium* Teotihuacán (Date), *Solanum lycopersicum* (Sly), *Datura stramonium* Ticumán (Dati) and *Nicotiana attenuata* (Natt). 15 and 17 domains are exclusive for *D. stramonium* Teotihuacán and Ticumán, respectively. The UpSet plot shows the intersections of the set of orthogroups from the thirteen Solanaceae genomes. Each column corresponds to an orthogroup,

and each row corresponds to one segment in a Venn diagram. Cells are either empty (grey-black), indicating that this set is not part of that intersection, or filled, showing that the set is participating in the intersection and that the species share that orthogroup. OrthoFinder assigned 480,594 genes out of 536,483 (89.6% of total) to 35,458 orthogroups or gene families. There were 10,141 orthogroups with all species present and 181 of these consisted entirely of single-copy genes. It is also showed that *Solanum pimpinellifolium* and *Solanum lycopersicum* was the species-pair with more orthogroups shared (3140), followed by the pair between *Datura stramonium* Ticumán and Teotihuacán (2857 shared orthogroups).

Fig. 4. The species phylogeny shows four clades, the group of *Nicotiana* species, the clade of *Datura*, the group of *Capsicum* species and the *Solanum* group. *Petunia inflata* was selected as the outgroup species. *P. inflata* diverged from all the Solanaceae species studied here approximately 35 Mya. *Nicotiana* species diverged almost 32 Mya. While *Datura stramonium* diverged ~30.1 Mya from *Solanum*, *Capsicum* and *Nicotiana* species. The rate of gene gain and lost (λ) resulted from CAFE analysis was 0.0153461 for the whole tree. The internal branch with the largest numbers of significant rapidly evolving gene families corresponds to the most recent common ancestor of *Nicotiana* species clade. The terminal branch with the most rapidly significant evolving gene families is the one leading to *D. stramonium* clade. The internal branch with the largest numbers of significant contractions corresponds also to the most recent common ancestor of *Datura* species clade. The terminal branch with the most contractions is the one leading to *P. inflata*. While *D. stramonium* Ticumán is the species that lost the least gene families. Likewise, the internal branch with the largest significant number of expansions corresponds to the most recent common ancestor of *Nicotiana* species clade. *N. tabacum* showed the highest number of family expansions.

Fig. 5. a) Tropane alkaloid differentiation between Ticumán and Teotihuacán plants (Log scale) that were used to sequence their genomes. Except for the alkaloids 3-hydroxy-6-tigloyloxytropane (only found in Teotihuacan) and tropinone (similar concentration between both plants), all the alkaloids showed higher concentration in Ticumán than the Teotihuacán plant. Scopolamine alkaloid is highlighted with a red box. b) *h6h* gene phylogeny was generated using 17 genes. Results revealed that two copies of *h6h* are found in both genomes of *D. stramonium* (highlighted with a red box). These two gene copies were distributed in two different gene families (OG0028637 and OG0043057). The name of the family was added at the end of the gene name. The gene DaturastramoniumTic8550_OG0028637 (Ticumán genome) have two domains of DIOX_N (PF14226), while all its homologous of the phylogeny only have one DIOX_N domain. In contrast, only three genes show to be composed by only one domain (2OG-FeII_Oxy, PF03171); two of them belong to our *D. stramonium* genomes and one for *Medicago truncatula*. The higher production of scopolamine alkaloid in the Ticumán genome could be related with this additional domain and different architecture in the Ticumán *h6h* gene.

Fig. 6. Four copies of the *tpr* I gene for both genomes were observed (highlighted with a red box). The gene date9161 (Teotihuacán) and dati33033 (Ticumán) have two different domains compared with the rest of the *tpr* I genes of *Datura*; Sulfakinin (PF08257) and ECR1_N domain (PF14382). Different architecture is showed for the *tpr* I genes, date9170 and dati33044. However, the gene dati33044 (Ticumán) have one additional domain of adh_short_C2 (PF13561), in comparison with its homologous date9170 (Teotihuacán). Likewise, dati33027 showed and additional adh_short_C2 domain in comparison with its closer homologous date20542. We observed that the Ticumán plant produced 31.95-fold tropine concentration than Teotihuacán. This chemical differentiation in tropine concentration could be related with this

additional domain (adh_short_C2) that was observed in the Ticumán genome. Another gene copies of *tpr* I, date11128 (Teotihuacán) and dati22507 (Ticumán) have the domain, RHH_5 (PF07878), this domain has been described as a toxin-antitoxin system (TA). This TA system is also found in *Nicotiana attenuata*, *Nicotiana tabacum*, *Nicotiana sylvestris* and *Capsicum annuum gabriusculum*. Several studies relating the production of secondary metabolites as defence against natural enemies have been well reported for all these species.

Fig. 1

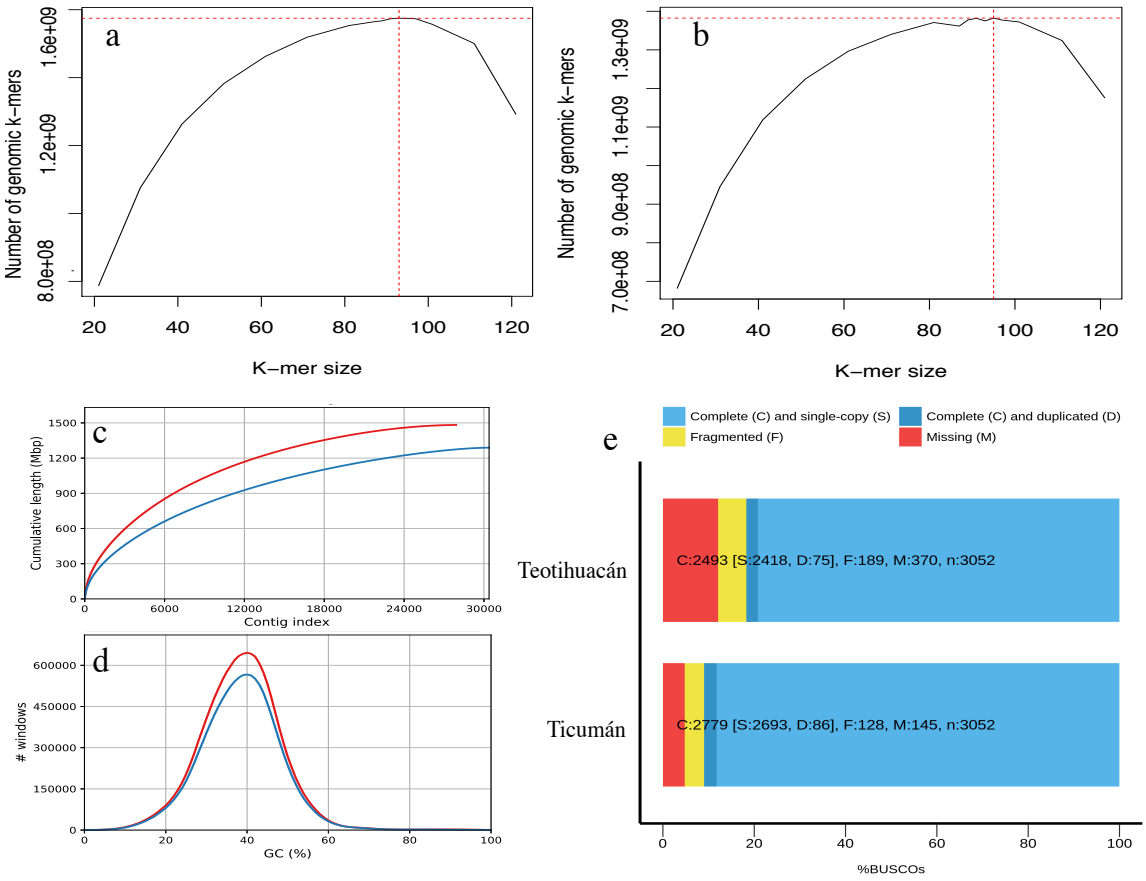

a
Number of genomic k−mers
8.0e+08
1.2e+09
1.6e+09
20
40
60
80
100
120
K−mer size
b
Number of genomic k−mers
7.0e+08
9.0e+08
1.1e+09
1.3e+09
20
40
60
80
100
120
K−mer size
c
Cumulative length (Mbp)
0
300
600
900
1200
1500
0
6000
12000
18000
24000
30000
Contig index
d
windows
0
150000
300000
450000
600000
0
20
40
60
80
100
GC (%)
e
Complete (C) and single-copy (S)
Complete (C) and duplicated (D)
Fragmented (F)
Missing (M)
Teotihuacán
C:2493 [S:2418, D:75], F:189, M:370, n:3052
Ticumán
C:2779 [S:2693, D:86], F:128, M:145, n:3052
0
20
40
60
80
100
%BUSCOs

Fig. 2

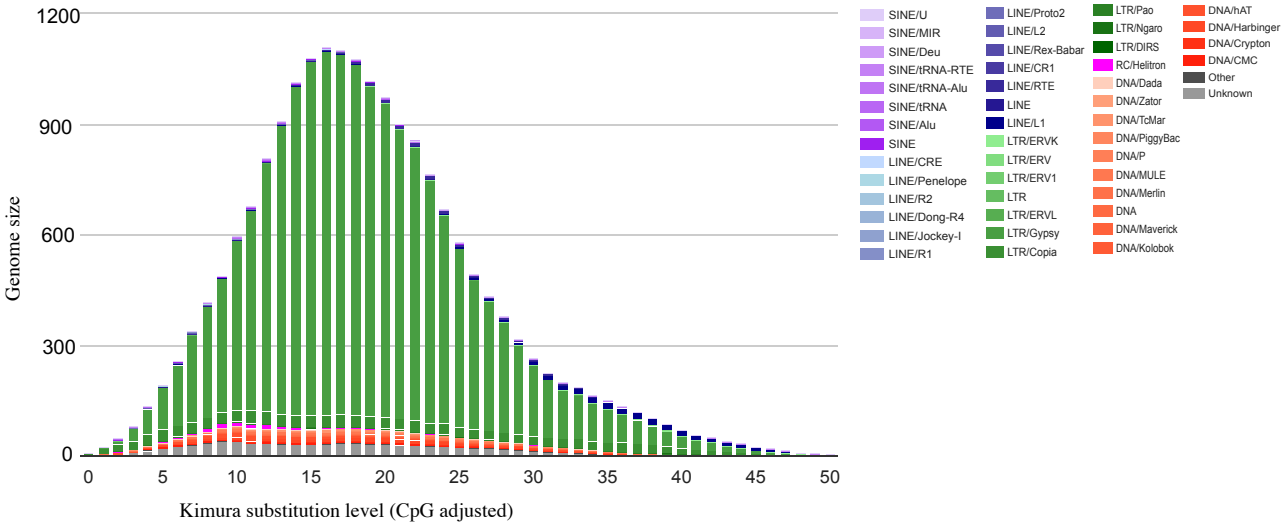

1200
900
600
300
0
Genome size
0
5
10
15
20
25
30
35
40
45
50
Kimura substitution level (CpG adjusted)
SINE/U
SINE/MIR
SINE/Deu
SINE/tRNA-RTE
SINE/tRNA-Alu
SINE/tRNA
SINE/Alu
SINE
LINE/CRE
LINE/Penelope
LINE/R2
LINE/Dong-R4
LINE/Jockey-I
LINE/R1
LINE/Proto2
LINE/L2
LINE/Rex-Babar
LINE/CR1
LINE/RTE
LINE
LINE/L1
LTR/ERVK
LTR/ERV
LTR/ERV1
LTR
LTR/ERVL
LTR/Gypsy
LTR/Copia
LTR/Pao
LTR/Ngaro
LTR/DIRS
RC/Helitron
DNA/Dada
DNA/Zator
DNA/TcMar
DNA/PiggyBac
DNA/P
DNA/MULE
DNA/Merlin
DNA
DNA/Maverick
DNA/Kolobok
DNA/hAT
DNA/Harbinger
DNA/Crypton
DNA/CMC
Other
Unknown

Fig. 3

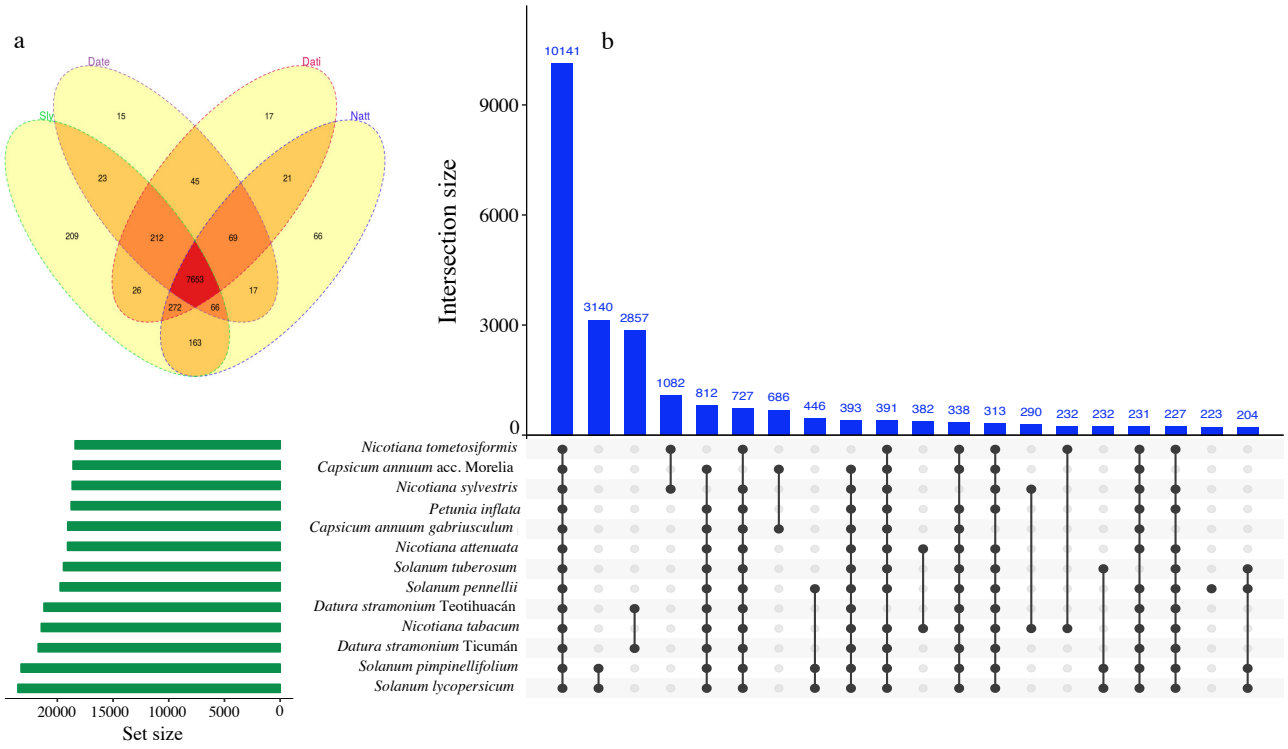

a
Date
Dati
Sly
Natt
15
17
23
45
21
209
212
69
66
7653
26
17
272
66
163
20000
15000
10000
5000
0
Set size
b
Intersection size
9000
6000
3000
0
10141
3140
2857
1082
812
727
686
446
393
391
382
338
313
290
232
232
231
227
223
204
Nicotiana tometosiformis
Capsicum annuum acc. Morelia
Nicotiana sylvestris
Petunia inflata
Capsicum annuum gabriusculum
Nicotiana attenuata
Solanum tuberosum
Solanum pennellii
Datura stramonium Teotihuacán
Nicotiana tabacum
Datura stramonium Ticumán
Solanum pimpinellifolium
Solanum lycopersicum

Fig. 4

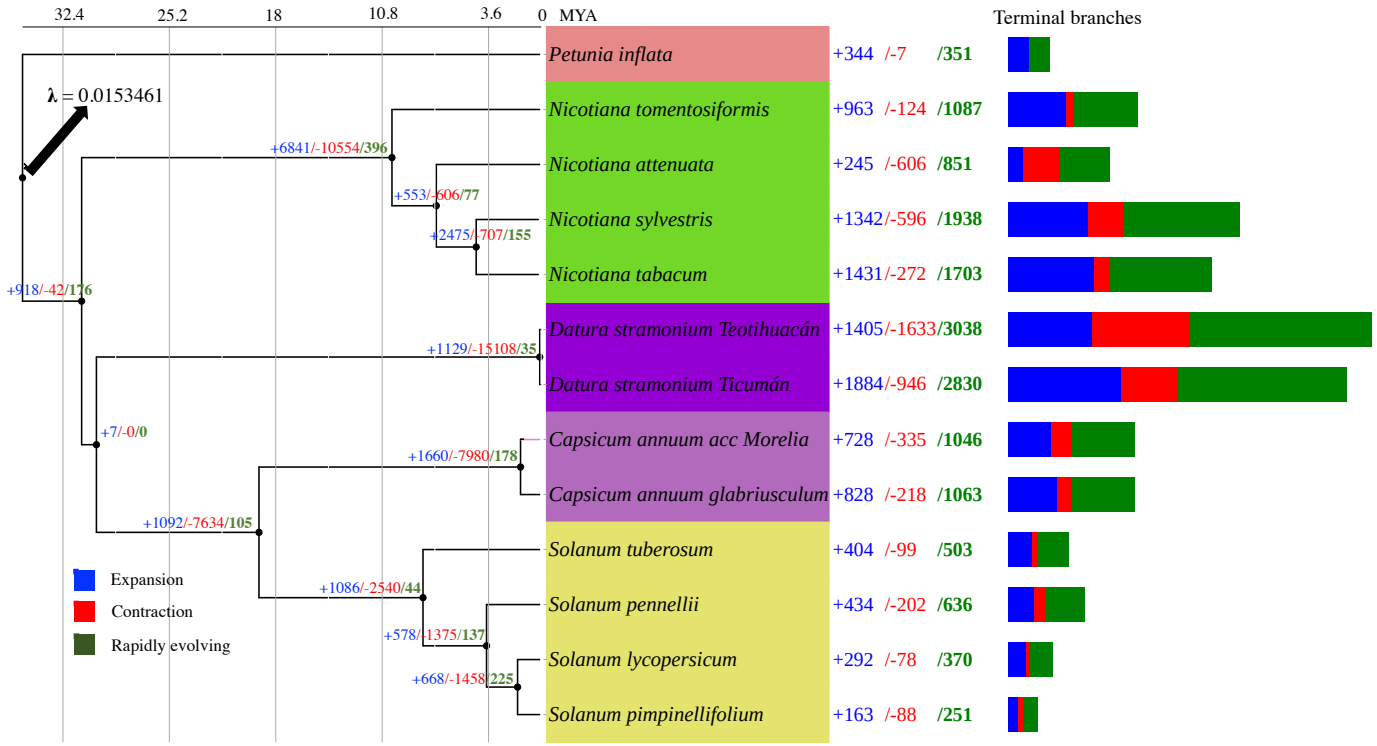

32.4
25.2
18
10.8
3.6
0 MYA
λ = 0.0153461
+6841/-10554/396
+553/-606/77
+2475/-707/155
+918/-42/176
+1129/-15108/35
+7/-0/0
+1660/-7980/178
+1092/-7634/105
+1086/-2540/44
+578/-1375/137
+668/-1458/225
Expansion
Contraction
Rapidly evolving
Petunia inflata
Nicotiana tomentosiformis
Nicotiana attenuata
Nicotiana sylvestris
Nicotiana tabacum
Datura stramonium Teotihuacán
Datura stramonium Ticumán
Capsicum annuum acc Morelia
Capsicum annuum glabriusculum
Solanum tuberosum
Solanum pennellii
Solanum lycopersicum
Solanum pimpinellifolium
+344 /-7 /351
+963 /-124 /1087
+245 /-606 /851
+1342/-596 /1938
+1431/-272 /1703
+1405/-1633/3038
+1884/-946 /2830
+728 /-335 /1046
+828 /-218 /1063
+404 /-99 /503
+434 /-202 /636
+292 /-78 /370
+163 /-88 /251
Terminal branches

Fig. 5

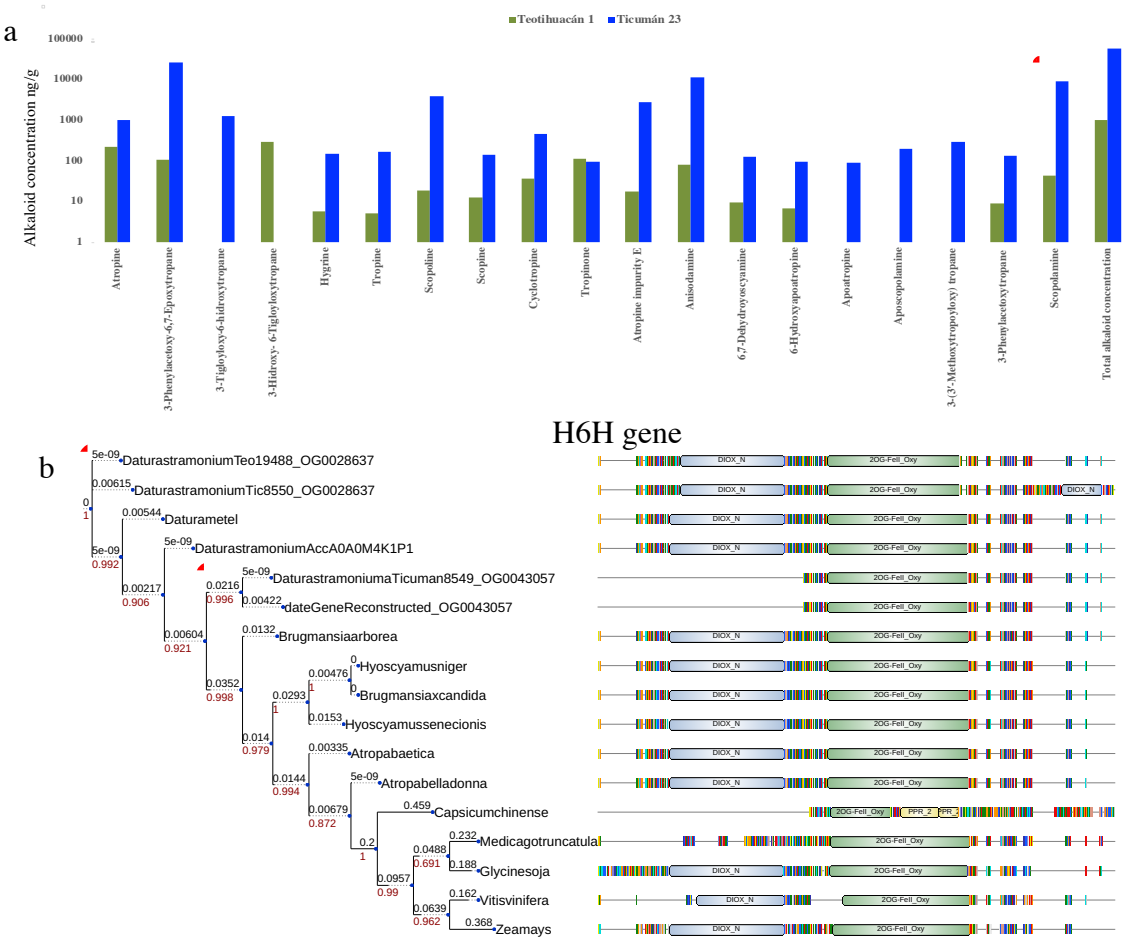
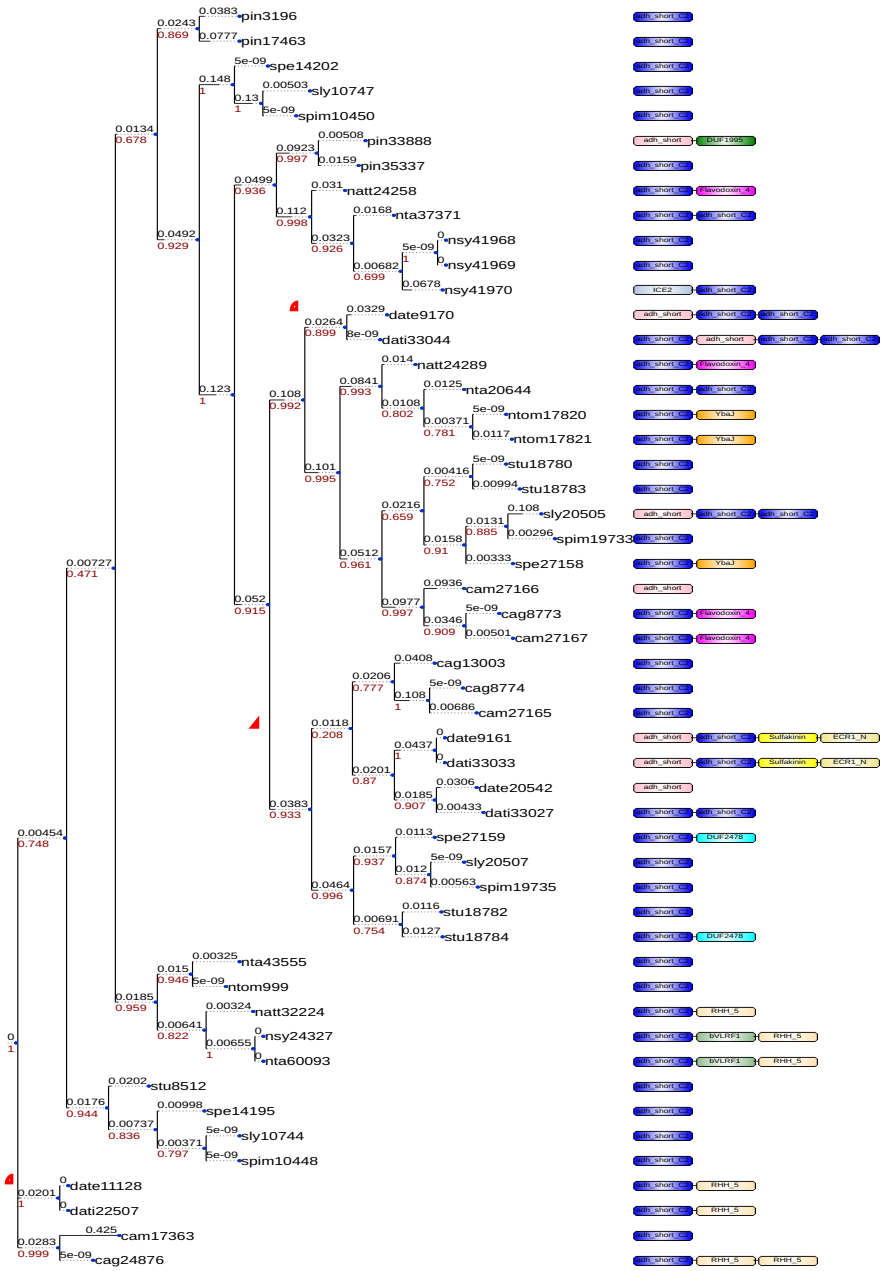

a
Teotihuacán 1
Ticumán 23
Alkaloid concentration ng/g
100000
10000
1000
100
10
1
Atropine
3-Phenylacetoxy-6,7-Epoxytropane
3-Tigloyloxy-6-hidroxytropane
3-Hidroxy- 6-Tigloyloxytropane
Hygrine
Tropine
Scopoline
Scopine
Cyclotropine
Tropinone
Atropine impurity E
Anisodamine
6,7-Dehydroxyscyamine
6-Hydroxyapoatropine
Apoatropine
Aposcopolamine
3-(3'-Methoxytropoyloxy) tropane
3-Phenylacetoxytropane
Scopolamine
Total alkaloid concentration
b
H6H gene
5e-09 DaturastramoniumTeo19488_OG0028637
0.00615 DaturastramoniumTic8550_OG0028637
0.00544 Daturametel
5e-09 DaturastramoniumAccA0A0M4K1P1
5e-09 DaturastramoniumaTicuman8549_OG0043057
0.00422 dateGeneReconstructed_OG0043057
0.0132 Brugmansiaarborea
0 Hyoscyamusniger
0.00476 Brugmansiaxcandida
0.0153 Hyoscyamussenecionis
0.00335 Atropabaetica
5e-09 Atropabelladonna
0.459 Capsicumchinense
0.232 Medicagotruncatula
0.188 Glycinesoja
0.162 Vitisvinifera
0.368 Zeamays
0.0216
0.996
0.00217
0.906
0.00604
0.921
0.0352
0.998
0.0293
1
0.014
0.979
0.0144
0.994
0.00679
0.872
0.2
1
0.0488
0.691
0.0957
0.99
0.0639
0.962
5e-09
0.992
1
0
DIOX_N
2OG-Fell_Oxy
PPR_2
PPR

Fig 6.


pin3196
pin17463
spe14202
sly10747
spim10450
pin33888
pin35337
natt24258
nta37371
nsy41968
nsy41969
nsy41970
date9170
dati33044
natt24289
nta20644
ntom17820
ntom17821
stu18780
stu18783
sly20505
spim19733
spe27158
cam27166
cag8773
cam27167
cag13003
cag8774
cam27165
date9161
dati33033
date20542
dati33027
spe27159
sly20507
spim19735
stu18782
stu18784
nta43555
ntom999
natt32224
nsy24327
nta60093
stu8512
spe14195
sly10744
spim10448
date11128
dati22507
cam17363
cag24876
adh_short
adh_short_C2
DUF1995
Flavodoxin_4
YbaJ
ICE2
Sulfakinin
ECR1_N
DUF2478
RHH_5
bVLRF1

Discusión general y conclusiones

Las plantas terrestres (briofitas) responden al ataque de los herbívoros a través de un complejo sistema de defensas variables, que involucra tanto diferentes barreras físicas y químicas o el reclutamiento de depredadores para sus enemigos naturales (Howe & Jander 2008; Rasmann & Agrawal 2009; Ali & Agrawal 2012). Por otra parte, los herbívoros son capaces de sobrepasar estás barreras defensivas, ya sea excretando, secuestrando, o modificando la estructura química de los compuestos de las plantas a otras sustancias que incluso pueden llegar ser benignas para el insecto (Després *et al.* 2007).

A pesar de los grandes avances en el estudio de las interacciones planta-herbívoros, nuestro entendimiento sobre la comunicación de las plantas con sus vecinos, simbiontes, patógenos y herbívoros todavía es limitado (Howe & Jander 2008). Más aún, la mayoría de los estudios de interacciones planta-herbívoros son hechos utilizando especies modelos (Howe & Jander 2008). Sin embargo, los estudios de plantas no-modelos, tales como *D. stramonium*, pueden ser de amplio interés debido a que estos sistemas ofrecen nuevos modos y enfoques de investigar los procesos ecológicos y evolutivos que las plantas enfrentan en sus ambientes naturales, y como ellas responden al calentamiento global, contaminación y perturbación humana (Savolainen *et al.* 2013).

Hasta hace relativamente poco, había sido difícil estudiar las bases genéticas de las defensas químicas de las plantas (Schuman and Baldwin 2018). Recientes avances en la secuenciación masiva del DNA y de la espectrometría de masas, ahora permiten llevar a cabo estudios a profundidad sobre la evolución de las armas químicas producidas por las plantas para contrarrestar el ataque de sus enemigos naturales (Schuman and Baldwin 2018). Por ejemplo, sobre-expresando o silenciando genes de interés actualmente es posible comprender el papel específico de una molécula defensiva o su modo de acción. Sin embargo, este tipo de experimentos por sí solos no son suficientes para entender holísticamente el papel fisiológico y ecológico de las defensas de las plantas (Howe & Jander 2008; Rasmann & Agrawal 2009; Ali & Agrawal 2012).

Por lo anterior, en está tesis utilizamos una combinación de metodologías que provienen de diferentes subdisciplinas de la ciencia moderna incluyendo la genética cuantitativa y ecológica, experimentos en campo y en condiciones controladas, herramientas genómicas y de la química analítica con el fin de estudiar la evolución adaptativa de la resistencia contra herbívoros en nuestro sistema modelo, *Datura stramonium*. Así, esta tesis aporta evidencia de adaptación local en plantas de *D. stramonium* para enfrentar a sus distintas especies de herbívoros que varían geográficamente. Nuestros resultados están en concordancia con la teoría del mosaico geográfico coevolutivo, la cual

postula que existe un mosaico selectivo a lo largo de la distribución de una especie (Thompson 2005). En particular, dado que la comunidad de herbívoros cambia a lo largo de la distribución de las plantas, se espera que la fuerza y dirección de la selección natural sobre la defensa de las plantas sea distintas entre las poblaciones (Thompson 1999). Tal como lo hemos evidenciado a lo largo de este estudio.

Por lo tanto, las principales conclusiones de esta tesis son las siguientes:

1. Se brinda conocimiento básico sobre los mecanismos moleculares, químicos y genéticos relacionados con la defensa en plantas contra sus herbívoros (Capitulo I).
2. Se identificaron por primera vez 21 tropano-alcaloides en plantas mexicanas de *D. stramonium* (Capitulo II).
3. Mediante el experimento de jardín común bajo condiciones ambientales controladas detectamos diferenciación genética y fenotípica en las poblaciones de estudio; Ticumán y Teotihuacán. Los niveles en la concentración de alcaloides tropanos son más altos en Ticumán (capitulo II).
4. Encontramos que la diferenciación en la concentración de los tropanos alcaloides (P_{ST}) excede la diferenciación en loci neutrales (Q_{ST}). Estos resultados indican que la diferenciación fenotípica no es explicada por la deriva génica sino probablemente por selección natural ejercida por herbívoros (capitulo II).
5. En el capitulo tres, detectamos cuales son los alcaloides que confieren resistencia a *D. stramonium* en sus ambientes naturales, específicamente que alcaloides están funcionando como defensa para cierto tipo de herbívoros, pero no para otros (Capitulo III).
6. Detectamos que la larva del escarabajo folívoro *Lema daturaphila* es el herbívoro que ejerce mayor presión selectiva sobre las plantas de *D. stramonium* (Capitulo III).
7. Encontramos que al menos el alcaloide triperpenoide (posiblemente relacionado con el alcaloide azadirone), identificado por primera vez en *Datura*, afecta negativamente los niveles de infestación del herbívoro más dañino de *D. stramonium*, las larvas de *L. daturaphila*.
8. Encontramos que la selección natural impulsada por múltiples herbívoros especialistas maneja la resistencia química en *D. stramonium*. Esto ha resultado en adaptación local en las defensas de *D. stramonium* contra sus enemigos naturales (Capitulo III).

9. Detectamos que la resistencia contra los herbívoros especialistas en *D. stramonium* tiene una base genética (capitulo III).

10. En el capitulo IV, se obtuvieron los primeros dos ensambles de genomas de alta calidad y contiguos *de D. stramonium*. El tamaño del genoma ensamblado es ~1.5 Gb. La anotación de genes concuerda con lo reportado para otras especies de Solanáceas. Asimismo, se obtuvo el paisaje de elementos repetidos del genoma. Los elementos repetidos del *toloache* cubren cerca de 75 % del genoma (capitulo IV).

11. Los análisis de genómica comparativa hechos en el capitulo IV revelaron hallazgos importantes tales como genes expandidos, positivamente seleccionados y con divergencia fisicoquímica en *D. stramonium* que revelan la evolución y adaptación de la defensa en contra de sus enemigos naturales. Muchos de estos genes pertenecen a la familia de genes *R* (genes de resistencia) (capitulo IV).

12. Otros genes positivamente seleccionados y expandidos o con divergencia fisicoquímico están involucrados en la producción de metabolitos secundarios involucrados en defensa en contra de patógenos como herbívoros, bacterias, virus, hongos (capitulo IV).

13. Esta investigación reveló diferenciación en la arquitectura de dominio en varios de los genes involucrados en la ruta de los alcaloides tropanos entre ambos parentales de *D. stramonium*, los cuales fueron seleccionados por su alta diferenciación en la producción de alcaloides (capitulo IV).

14. El flujo de trabajo presentado en el capitulo IV para ensamblar, anotar y hacer análisis extensivos de genómica comparativa (selección positiva, expansión y divergencia fisicoquímica en familias de genes) es de libre acceso. El flujo de trabajo, scripts y comandos puede consultarse libremente en https://github.com/icruz1989/Datura-stramonium-genome-project.

15. El programa para análisis genómicos comparativos puede descargarse libremente en la página de GitHub; https://github.com/asishallab/SlydGeneFamsAnalyses/tree/icruz. Una función destacada de este programa es que el enriquecimiento funcional es hecho sobre dominios de proteínas, lo cual es más detallado y exacto con respecto a los usuales flujos de trabajo los cuales únicamente hacen enriquecimiento funcional usando ontología

de genes (GO). La cual es una categoría muy amplia y que no permite conocer a detalle la evolución de las funciones de genes.

16. El flujo de trabajo, scripts, comandos para analizar la Identidad por Descendencia (IBD) a partir de ddRadseq, el cual fue ocupado en el capitulo III, puede consultarse libremente en https://github.com/icruz1989/IBDcalculation

Finalmente, considero que esta tesis aportá información clave y significativa para futuros estudios en *D. stramonium,* así como para el campo de las interacciones planta-herbívoros. Asimismo, las metodologías, flujos de trabajo, scripts, comandos y el programa construido en este estudio pueden ser de amplio interés para los biólogos evolutivos, ecólogos y bionformáticos. Los datos genómicos (secuencias de genomas, ddRadseq) y químicos (cromatogramas del HPLC-TOF-MS) desarrollados en este estudio estarán disponibles para la comunidad científica para futuros estudios en las especies de Solanáceas y particularmente en *Datura*. Los genomas reportados son potencialmente de interés para la farmacéutica y medicina, ya que esta especie es estudiada y utilizada para la producción de medicinas a partir de sus alcaloides. Asimismo, la información generada en esta tesis es la base metodológica y de conocimiento para detectar "quantitative trait loci" (QTLs) de las defensas químicas y de resistencia en *D. stramonium*. Más estudios que involucren la interacción entre varios campos de la ciencia, como los ecológicos, genómicos, químicos son necesarios para entender a profundidad como las plantas se defienden de sus plagas (Schuman and Baldwin 2018). Esto aportará información invaluable para establecer mejores estrategias en la producción de alimentos a partir de plantas y entender como las plantas pueden enfrentar la perturbación humana y el cambio climático.

Literatura citada

Després, L., David, J.-P., & Gallet, C. (2007). The evolutionary ecology of insect resistance to plant chemicals. *Trends in Ecology & Evolution*, *22*(6), 298–307.

Howe, G. A., & Jander, G. (2008). Plant immunity to insect herbivores. *Annual Review of Plant Biology*, *59*, 41–66.

Rasmann, S., & Agrawal, A. A. (2009). Plant defense against herbivory: Progress in identifying synergism, redundancy, and antagonism between resistance traits. *Current Opinion in Plant Biology*, *12*(4), 473–478.

Savolainen, O., Lascoux, M., & Merilä, J. (2013). Ecological genomics of local adaptation. *Nature Reviews Genetics*, *14*(11), 807–820.

Schuman, M. C., & Baldwin, I. T. (2018). Field studies reveal functions of chemical mediators in plant interactions. *Chemical Society Reviews*, *47*(14), 5338–5353.

Thompson, J. N. (1999). Specific Hypotheses on the Geographic Mosaic of Coevolution. *The American Naturalist*, *153*(S5), S1–S14.