UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO
DOCTORADO EN CIENCIAS BIOMÉDICAS

Genómica de poblaciones del maíz silvestre, el teocintle (*Zea mays* ssp. *parviglumis* y *Zea mays* ssp. *mexicana*)

TESIS
QUE PARA OPTAR POR EL GRADO DE
DOCTOR EN CIENCIAS

PRESENTA:
JONÁS ANDRÉS AGUIRRE LIGUORI

TUTOR: DR. LUIS ENRIQUE EGUIARTE FRUNS

INSTITUTO DE ECOLOGÍA

CIUDAD UNIVERSITARIA, CD. MX, NOVIEMBRE 2017

El jurado de examen doctoral estuvo constituído por:

Dr. Daniel Piñero	Presidente
Dr. Luis Eguiarte	Secretario
Dr. Jorge Nieto	Vocal
Dr. Rafael Lira	Vocal
Dra. Alejandra Moreno	Vocal

Agradecimientos
Antes que nada, quisiera agradecer a la Universidad Nacional Autónoma de México, por la formación académica y personal que me ha dado. Agradezco también al Instituto de Ecología y al Programa de Doctorado en Ciencias Biomédicas por haberme acobijado durante el doctorado. El mismo agradecimiento le tengo al Consejo Nacional de Ciencia y Tecnología, ya que fue el sustento financiero de mi proyecto (Investigación Científica Básica, CB2011/167826), y de mi manutención (Número de becario 255770). De igual forma, estoy muy agradecido con los programas ECOS Nord France -CONACYT-ANUIES (M12-A03, 207571) y con PAEP que me permitieron asistir a 3 estancias internacionales (dos en París y una en California). Estás estancias fueron muy importantes para mi formación y el desarrollo del proyecto. Agradecemos enormemente a Sarah Hearne y su equipo en CIMMYT, así como a MasAgro, por habernos secuenciado con la tecnología DarTseqTM toda nuestra colección de teocintles.
Estoy eternamente agradecido a los Drs. Maud Tenaillon y Brandon Gaut (y su equipos de trabajo en le Moulon, y UCI, respectivamente) por todo el apoyo y enseñanzas que me dieron durante mis estancias y con las discusiones y colaboraciones que hemos mantenido desde entonces. Sin duda, parte de mi madurez científica se debe a todo lo que me han enseñado y por su exigencia de que todo quede claro. En el mismo sentido, estoy muy agradecido a los miembros de mi comité tutoral (Drs. Juan Pablo Jaramillo, Luis Eguiarte Fruns y David Romero Camarena) por todas las discusiones que tuve con ellos para enfocar mejor mi proyecto y por todo el apoyo que siempre recibí de su parte. De igual forma agradezco individualmente al Dr. Juan Pablo Jaramillo, porque desde antes de que entrara al Doctorado, siempre estuvo abierto a hablar conmigo de todas mis ideas y fomentó gran parte de mi curiosidad y creatividad. Agradezco infinitamente a mi tutor, el Dr. Luis Enrique Eguiarte Fruns por haberme aceptado en participar en este proyecto, por haberme apoyado con mis ideas y durante todo el Doctorado. ¡Espero que continuemos colaborando!
Agradezco a los Drs. Daniel Piñero, Rafael Lira, Jorge Nieto, Alejandra Moreno-Letelier y Luis Eguiarte por haber leído y aportado ideas para mejorar esta tesis.
Agradezco a las Drs. Erika Aguirre-Planter y Laura Espinoza Asuar, así como a Doña Silvia por todo el apoyo técnico que me dieron en el laboratorio. También agradezco a Valeria Souza (junto con Luis) por haberme aceptado en su laboratorio hace ya tantos años.
Este trabajo no habría sido posible sin la ayuda de Salvador Montes, Enrique Scheinvar, Ricardo Colin, Jaime Gasca, Brandon Gaut, Maud Tenaillon, Luis Eguiarte y Valeria Souza, que hicieron la mayor parte de las colectas de campo. ¡Gracias! De igual forma agradezco a mis compañeros de proyecto, Ale Vázquez, Ale Gutiérrez, Felicitas Lagunas, el primo Beto y Gabriel Merino, por ayudarme con las extracciones y por las discusiones.
Agradezco a todos los miembros actuales y pasados del Laboratorio de Evolución Molecular y Experimental. Me siento muy contento de haber pasado este tiempo con ustedes y solo tengo agradecimiento y cariño.
Agradezco a todos mis amigos de la vida por todo lo que hemos convivido, reído, discutido y aprendido. Son parte importante de mi vida. ¡Los quiero!
Con mucho cariño, agradezco a toda mi familia extendida por todo su apoyo y cariño. Aunque se diga que uno no escoge a su familia, yo me siento muy afortunado de la familia que me tocó por nacimiento y por emparejamiento.
No tengo palabras para agradecer a mis padres, Ana Luisa y Carlos, y para decirles lo afortunado que me siento de que sean parte de mi vida. ¡Los quiero entrañablemente!
Finalmente, agradezco a mi compañera de vida, Paula Sosenski. Lo dije en la tesis pasada y lo repito una vez más, gracias por sacar lo mejor de mí, y ayudarme a ser mejor persona. Gracias por quererme, aguantarme, divertirme, y apoyarme. Te amo.

INDICE

RESUMEN

La especiación ecológica se define como el origen de barreras reproductivas como resultado de la adaptación divergente. Con el desarrollo reciente de la secuenciación de próxima generación ha habido un desarrollo importante de esta teoría. Sin embargo existen aún preguntas que se tienen que resolver: ¿Qué mecanismos geográficos, ecológicos y genéticos promueven la especiación ecológica? En este trabajo analicé la genómica de poblaciones de los teocintles para determinar las bases geográficas y climáticas donde ocurre la adaptación local y posteriormente estudié los procesos que están promoviendo su especiación ecológica. Encontré que los teocintles se están adaptando al límite del nicho, donde las condiciones son divergentes. Así mismo, encontré que la divergencia del nicho (posiblemente en su límite) ha generado la divergencia adaptativa de los teocintles. Específicamente encontré que los teocintles han divergido a lo largo de dos ejes del nicho, la temperatura y la disponibilidad de fósforo en el suelo. Argumento que la divergencia a lo largo de dos ejes climáticos está generando divergencia en varias regiones del genoma, reduciendo así el flujo genético entre poblaciones por efecto de desequilibrio de ligamiento. Existen cuatro regiones de alta diferenciación (posiblemente inversiones cromosómicas) altamente enriquecidas en genes candidatos a estar bajo selección, las cuales podrían contribuir al aislamiento reproductivo de las poblaciones de las subespecies de teocintle. La identificación de varios genes asociados a cambios en temperatura y disponibilidad de suelo podría ocuparse para diseñar estrategias de mejoramiento del maíz.

ABSTRACT

Ecological speciation is defined as the mechanism that generates barriers to gene flow as a consequence of ecologic divergent adaptation. The recent development of next generation sequencing has resulted in an important development of this theory. However important questions still need answering: What geographic, ecologic and genetic mechanisms promote ecological speciation? In this work, I analyzed the genomics of populations of teosintes to determine the geographic and climatic basis where local adaptation occurs and then test which effects have been responsible for their ecological divergence. I found that teosintes are adapting to the niche limit where conditions are divergent. Also I found that ecological divergence (at the limit) has generated adaptive divergence of teosintes along two ecological niche axis, temperature and phosphorus availability in the soil. I argue that divergence along two-niche axis is generating divergences along many regions of the genome, which in turn reduces gene flow by linkage disequilibrium. In accordance, I identifed four chromosomic regions of high differentiation (possibly chromosomal inversions) enriched in candidate SNPs. These regions could be further contributing to reproductive isolation between the subspecies of teosintes. The identification of genes associated to changes in temperature and phosphorus availability in the soil could be used to design improvement strategies in maize.

INTRODUCCION

El registro paleontológico muestra que a lo largo de las eras geológicas ha habido un recambio de organismos. Sin embargo, la constante es que siempre ha existido diversidad biológica, que se origina, mantiene y, en muchos casos, desaparece. Un objetivo central en la biología evolutiva ha sido entender los mecanismos que generan y conservan esta diversidad (Darwin 1859; Mayr 1942; Coyne y Orr 2004). Cuantitativamente, la diversidad biológica resulta de la relación entre la tasa de origen de nuevas especies (especiación) y de aquellas que se extinguen (extinción). Ambas tasas son heterogéneas a lo largo del tiempo y del espacio (Kozak y Wiens 2010; Donoghue y Edwards 2014; Schluter 2016) y ha sido un objetivo importante de la biología evolutiva entender el origen de esta heterogeneidad.

La especiación es el proceso que genera nuevas especies (Coyne y Orr 2004; Futuyma 2005; Nosil *et al.* 2012). Su concepción más sencilla considera que el origen de nuevas especies ocurre cuando se consolidan las barreras al flujo genético (Coyne y Orr 2004), permitiendo que las especies incipientes diverjan por selección, mutación y/o deriva. Los mecanismos que generan barreras pueden ser diversos y generalmente se dividen en función del flujo genético potencial que existe entre las poblaciones, generalmente explorando la presencia de factores naturales o geográficos que impidan la conectividad entre distintas poblaciones o no. Históricamente se ha considerado a la especiación alopátrica como el principal proceso que genera nuevas especies (Coyne y Orr 2004; Futuyma 2005); esta se define como el proceso que genera nuevas especies en respuesta a una barrera geográfica que impide el flujo genético entre poblaciones. En estos casos, la divergencia puede ocurrir tanto por procesos adaptativos como por procesos no adaptativos.

Sin embargo durante la última década se ha reconsiderado la importancia de estas barreras geográficas y se le ha dado una relevancia creciente a la selección divergente y a la adaptación local como mecanismos promotores de especiación (Schluter 2000, 2001). La especiación ecológica se define como el proceso que genera barreras reproductivas como respuesta a la selección divergente (Rundle y Nosil 2005; Nosil 2012). Este proceso es continuo, iniciando cuando las poblaciones se adaptan a ambientes contrastantes (Schluter 2000, 2001; Rundle y Nosil 2005; Nosil 2012) y finalizando cuando se generan barreras reproductivas (Rundle y Nosil 2005; Feder *et al.* 2013). Lo más interesante de este

mecanismo es que puede ocurrir en presencia de flujo genético. Por lo tanto, el proceso de especiación ecológica puede iniciarse regularmente, pero no siempre se completa (Nosil *et al.* 2009a; Soria-Carrasco *et al.* 2014; Riesch *et al.* 2017). Por esto, y dado que el proceso es continuo, es posible estudiar distintas poblaciones de una especie en distintas etapas del proceso y de esta forma comparar estados y entender los distintos procesos que promueven la especiación ecológica (Soria-Carrasco *et al.* 2014; Riesch *et al.* 2017).

Aunque los primeros estudios de especiación ecológica datan de la década antepasada (Nosil 2012), el desarrollo reciente de la secuenciación paralela masiva (también llamada secuenciación de "próxima generación") ha tenido un efecto mayor sobre el desarrollo de esta teoría (Nosil *et al.* 2009b; Rice *et al.* 2011; Shafer y Wolf 2013). La secuenciación de próxima generación permite obtener marcadores moleculares a lo largo de todo el genoma para muchos individuos y en muchas poblaciones (Ekblom y Galindo 2010; Metzker 2010; Eguiarte *et al.* 2013). En particular, el paso de contar en un estudio con una docena o menos de marcadores moleculares a tener cientos a millones de ellos a lo largo de todo el genoma ha permitido a los biólogos entender mejor la estructura genética de las poblaciones. Estos marcadores además permiten diferenciar regiones que presentan una diferenciación inusual localizada en el genoma, sugiriendo que podrían estar bajo selección (Lewontin y Krakauer 1973; Beaumont 2005; Allendorf *et al.* 2010; Stapley *et al.* 2010; Schoville *et al.* 2012; De Mita *et al.* 2013). De esta forma, es posible determinar las regiones del genoma que están bajo selección, avanzar en entender las bases genéticas y no genéticas, geográficas y ecológicas de la adaptación y especiación.

El desarrollo de la secuenciación masiva también ha permitido estudiar la evolución del genoma durante el proceso de especiación. Las predicciones indican que durante la especiación alopátrica todo el genoma debería divergir homogéneamente, dado que no hay flujo genético entre poblaciones y a que todos los genes en el genoma experimentan los mismo niveles de deriva génica y flujo genético (Nosil 2012). En contraste, en el caso de la especiación ecológica, dado que la fuerza evolutiva principal es la selección divergente, no se espera que el genoma diverja homogéneamente como en la especiación alopátrica, aunque exista flujo génico (Nosil 2012; Riesch *et al.* 2017). En este caso, primero divergirán los genes bajo selección y con el tiempo lo hará el resto del genoma a través de *hitchhiking* y deriva génica.

Dado que la especiación ecológica puede ocurrir en poco tiempo y en presencia de flujo genético (Hendry *et al.* 2007; Nosil *et al.* 2009a; Nosil 2012; Surget-Groba *et al.* 2012), esta permite a las especies ocupar nichos vacíos, incluso generando radiaciones adaptativas (Schluter 2016). Recientemente se ha incrementado el interés por determinar los mecanismos genéticos, genómicos, geográficos y ecológicos que promueven este tipo de especiación. Aunque la especiación ecológica parece ser un proceso evolutivo importante, y cada vez existen más evidencias de que es frecuente (Funk *et al.* 2006; Shafer y Wolf 2013), seguimos desconociendo cuáles son los factores que promueven, mantienen y concluyen la especiación ecológica (Nosil *et al.* 2009b; Via 2009; Riesch *et al.* 2017). En particular, son poco los estudios que han logrado determinan los mecanismos que generan la especiación ecológica y que son por lo tanto importantes para poder generar predicciones sobre este proceso. Los estudios disponibles están enfocados a especies modelo, como algunos fásmidos (insectos palo), peces espinosos y cíclidos, mariposas del género *Heliconius* y plantas del género *Mimulus*. Muchos de estos estudios están limitados a pequeñas regiones geográficas (lagos de África, zonas templadas, dunas de California, etc.) y es evidente que se necesitan nuevos modelos y en países megadiversos como México.

A continuación explicaré con mayor detalle qué es, y cómo se estudia la especiación ecológica. Después describiré la teoría detrás de los estudios a nivel genómico que se usan para encontrar genes bajo selección y analizar la diferenciación genómica. Finalmente, presentaré las dos subespecies que se usaron como modelo en esta tesis. Estas subespecies son adecuadas para hacer estudios de especiación ecológica, ya que muestran gradientes diferenciales de señales de adaptación local, divergencia ecológica y de mecanismos que ligan ambos elementos.

ESPECIACIÓN ECOLÓGICA

Los modelos de especiación se pueden clasificar en función del potencial de flujo genético entre poblaciones. Así, la especiación alopátrica es aquella en la que existe alguna barrera al flujo genético y la especiación simpátrica es aquella en la que las poblaciones mantienen flujo genético (Futuyma 2005). La especiación ecológica puede ocurrir en presencia de flujo genético, pero no es un requerimiento. De hecho, el único requerimiento es que el aislamiento reproductivo se origine como consecuencia de la adaptación a nichos

divergentes (Schluter 2001, Nosil 2012). Por lo tanto, la especiación puede ser ecológica y alopátrica si las especies divergen en aislamiento geográfico pero por adaptación a condiciones divergentes. Asimismo, la especiación asociada a la selección sexual solo será ecológica si la elección de pareja tiene una base ambiental.

La especiación ecológica es un proceso continuo. Inicia cuando distintas poblaciones se adaptan localmente a ambientes divergentes (Schluter 2000, 2001). Posteriormente, la selección en contra de los migrantes o de los híbridos genera aislamiento por adaptación (IBA en inglés) entre las poblaciones (Nosil *et al.* 2008; Shafer y Wolf 2013). El aislamiento por adaptación es un concepto análogo al aislamiento por distancia, en el cual la diferenciación genética en genes neutrales y adaptativos es una función de la divergencia ecológica entre poblaciones (Nosil, Egan, and Funk 2008; Funk, Egan, and Nosil 2011). Dado que el IBA reduce el flujo genético entre poblaciones, promueve de manera indirecta, ya que no hay base genética, el aislamiento reproductivo entre las poblaciones. La especiación ecológica finaliza cuando se originan mecanismos genéticos que ligan la selección divergente y el aislamiento reproductivo (Bolnick 2011; Servedio *et al.* 2011; Nosil 2012). Considerando la importancia de la adaptación a nichos divergentes, la especiación ecológica dependerá de la capacidad de una especie a adaptarse localmente (tamaño efectivo, diversidad), a que la selección natural sea mayor que la migración, y al tiempo.

La especiación ecológica se puede estudiar al nivel genómico y, dependiendo de la fase en la que se encuentra el proceso, se puede traducir en las siguientes señales (Feder *et al.* 2013, 2014, Flaxman, Feder, and Nosil 2012, 2013; Flaxman *et al.* 2014). Inicialmente, se observarán regiones localizadas con alelos altamente diferenciados (F_{ST}) que promuevan la adaptación local a ambientes divergentes. Dado que la selección natural genera desequilibrio de ligamiento (DL) alrededor del gen bajo selección (Nosil 2012; Feder *et al.* 2013; Beaumont 2005; Schoville, Bonin, François, *et al.* 2012), se promoverá una reducción del flujo genético en estas regiones adyacentes (*hitchhiking*). La probabilidad de que un gen bajo selección aumente en frecuencia depende de la interacción entre el flujo genético y la selección natural que homogeniza o hace que diverjan las frecuencias alélicas, respectivamente (Hedrick 2011). Por lo tanto, genes adyacentes a las regiones de alto DL, y que tengan ventajas selectivas muy bajas (nuevos mutantes o *standing genetic variation*),

podrán incrementar su frecuencia, ya que el flujo genético no será suficientemente fuerte para contrarrestar la selección. Conforme estos nuevos genes se vayan seleccionando, se incrementará la extensión del DL y el grado de diferenciación genómica y se generarán "islas de diferenciación", regiones del genoma que contienen los genes que confieren adaptación local y los genes neutrales ligados a estos. Finalmente, conforme las islas de diferenciación se extienden, el IBA entre poblaciones irá aumentando a lo largo de una mayor extensión del genoma. De acuerdo a simulaciones informáticas (Feder *et al.* 2014; Flaxman, Feder, and Nosil 2012, 2013; Flaxman *et al.* 2014), la reducción de flujo genético entre poblaciones eventualmente será tan importante que la selección será lo suficientemente fuerte para generar una barrera al flujo genético a lo largo de todo el genoma. En inglés este fenómeno se ha llamado *genomic congealing* y se refiere al momento de divergencia del genoma en el cual el flujo genético ya no puede homogenizarlo, equiparándose así a un proceso de especiación alopátrica (Feder *et al.* 2014).

Durante el origen de estas islas de diferenciación puede haber re-arreglos cromosómicos que refuercen el aislamiento reproductivo. Un ejemplo interesante son las inversiones cromosómicas, las cuales pueden generar y mantener el aislamiento genético entre poblaciones y especies (Kirkpatrick and Barton 2006, Feder *et al.* 2013, Nosil 2012). Por un lado, las inversiones cromosómicas pueden generar inviabilidad o esterilidad en los heterócigos, al impedirse una correcta alineación de los cromosomas durante la meiosis (Stebbins 1958). Asimismo, la inviabilidad puede deberse a combinaciones alélicas incompatibles en distintas inversiones que se encontrarían en la forma heteróciga (incompatibilidades de Dobzhansky-Muller). Finalmente, las inversiones cromosómicas pueden suprimir la recombinación entre cromosomas provenientes de distintas poblaciones, ya que al no ser regiones homólogas no se pueden alinear durante la meiosis y no pueden recombinar los genes asociados a la inversión (Figura 1), evitando así que alelos mal-adaptados localmente ingresen a la población por medio de flujo genético (Noor *et al.* 2001; Kirkpatrick y Barton 2006, Feder *et al.* 2013). Esta reducción en la viabilidad de los "híbridos" puede jugar un papel importante en mantener el aislamiento reproductivo entre poblaciones adaptadas a condiciones locales y sobre todo en incrementar el aislamiento por ambiente en las regiones asociadas a la inversión cromosómica. De hecho, se ha visto que

muchos de estos genes adaptativos se tienden a agregar en inversiones cromosómicas, contribuyendo al aislamiento reproductivo entre especies adaptadas localmente (Noor *et al.* 2001, Nosil 2012, Feder *et al.* 2013). Aunque hipotéticamente las inversiones cromosómicas son mecanismos intuitivos para promover la especiación ecológica, estos son difíciles de identificar. Recientemente, el desarrollo de la secuenciación de próxima generación ha facilitado la identificación de inversiones cromosómicas y ha dado un mayor entendimiento de su función en la reducción de flujo genético entre poblaciones adaptadas a condiciones contrastantes (Kirkpatrick y Barton 2006; Lowry y Willis 2010; Andrew y Rieseberg 2013; Fishman *et al.* 2013; Twyford y Friedman 2015; He y Knowles 2016). Sin embargo faltan estudios que muestren si estas inversiones son importantes para mantener la adaptación local entre poblaciones y promover la especiación ecológica (Kirkpatrick y Barton 2006).

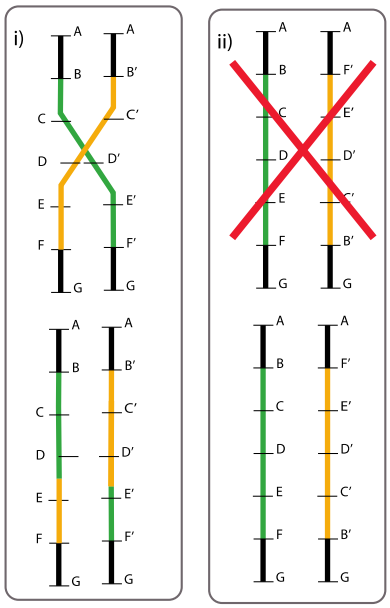


Figura 1. Esquema de inversión cromosómica (IC). i) no hay IC, por lo que las regiones de color pueden alinearse y recombinar. ii) Al estar invertidas, las regiones no homólogas no pueden recombinar

Aunque por sí sola, la especiación ecológica podría producirse por una fuerte selección divergente, existen algunos ejemplos que demuestran que otros factores son necesarios para completar la divergencia y especiación (Nosil y Sandoval 2008; Nosil *et al.* 2009a; Soria-Carrasco *et al.* 2014; Riesch *et al.* 2017). Un ejemplo interesante es el de los fásmidos *Timema cristinae*. Estos insectos palo crecen en dos tipos de plantas a los cuales se han adaptado a través del camuflaje. El aislamiento reproductivo de estos ecotipos está entonces correlacionado con la adaptación al tipo de hospedero, y esto ha ocurrido paralelamente en múltiples ocasiones (Soria-Carrasco *et al.* 2014; Riesch *et al.* 2017). Sin embargo, esta divergencia no se ha traducido en la generación de especies nuevas, mostrando que la selección divergente no es suficiente para concluir un proceso de especiación.

Diversas evidencias empíricas y teóricas sugieren que ciertos escenarios favorecen a que la especiación ecológica se complete (Chevin, *et al.* 2014; Flaxman *et al.* 2012, 2013;

Flaxman *et al.* 2014). Primero, es más probable que la especiación ecológica ocurra en poblaciones que presenten cierto aislamiento geográfico durante la divergencia (Aguilée *et al.* 2011; Nosil y Feder 2012; Surget-Groba *et al.* 2012; Riesch *et al.* 2017), ya que esto reduce el flujo genético entre poblaciones y favorece la formación y el crecimiento de islas de diferenciación. De esta forma, si las poblaciones entran en contacto otra vez, el proceso de especiación ecológica estará más avanzado y el IBA, y/o los mecanismos que generan aislamiento reproductivo, serán suficientes para el aislamiento reproductivo de las poblaciones. Otra predicción es que la especiación ecológica se puede facilitar si "un gen" que confiere adaptación local también genera apareamiento selectivo entre poblaciones, ya que habrá una reducción importante de flujo genético (Bolnick 2011; Servedio *et al.* 2011; Nosil 2012). Este mecanismo genético puede ser complejo, y puede en realidad incluir muchos genes ligados. Finalmente, una tercera predicción es que la especiación ecológica puede facilitarse si la selección divergente ocurre a lo largo de muchos ejes del nicho ecológico (Nosil y Sandoval 2008; Nosil *et al.* 2009a; b; MacColl 2011; Lenormand 2012; Chevin *et al.* 2014). A esto se le ha llamado la hipótesis de selección multifaria (Nosil, Harmon, and Seehausen 2009; Nosil, Funk, and Ortiz-Barrientos 2009) y postula que si el genoma responde a múltiples presiones de selección simultáneamente, entonces las islas de divergencia ocurrirán a lo largo de varias secciones del mismo y por lo tanto incrementará el aislamiento por adaptación entre poblaciones. Aunque se ha propuesto que la selección multifaria puede ser importante, pocos estudios la han estudiado explícitamente (pero ver Nosil y Sandoval 2008; Michel *et al.* 2010; Scholl *et al.* 2012; Liu *et al.* 2013; Arnegard *et al.* 2014; Malinsky *et al.* 2015). Sin embargo, con el uso de métodos de secuenciación masiva y estadísticos e información ecológica detallada, es posible avanzar en estudios que aborden la adaptación local y la divergencia ecológica. En la próxima sección detallaré cómo funcionan los métodos para detectar selección natural en el genoma.

MÉTODOS DE DETECCION DE GENES BAJO SELECCIÓN

Existen dos grandes aproximaciones para detectar genes que están bajo selección (Ross-Ibarra, Morrell, and Gaut 2007; Via 2009). La primera se llama en inglés *top-down* y se basa en determinar un rasgo que consideramos "ventajoso" y posteriormente identificar qué regiones genéticas se asocian a esa característica. Este método tiene dos principales

limitaciones. Una es que para utilizar este método se necesitan muchos individuos y cruzas para así determinar las asociaciones genéticas con el carácter o ambiente. Otra limitación importante es que esta aproximación únicamente permite identificar regiones que se asocian con el rasgo estudiado, dejando de lado otros genes que podrían estar bajo selección.

Caja 1. Secuenciación masiva

La identificación de genes candidatos a estar bajo selección depende de la correcta estimación de la historia demográfica de las poblaciones (por ejemplo, mediante F_{ST}). Al permitir la obtención de muchas secuencias a lo largo del genoma, tanto adaptativas como neutrales, la secuenciación masiva a potenciado los estudios de selección. Muy resumidamente (para más detalles ver el Apéndice I: Eguiarte *et al.* 2013), la secuenciación masiva consiste en fragmentar el ADN en millones de regiones, amplificarlas y secuenciarlas. Posteriormente, estos fragmentos son alineados entre ellos o respecto a un genoma de referencia. Finalmente, con estas secuencias es posible ensamblar genomas y/o obtener polimorfismos a lo largo del mismo.

Aunque la teoría es intuitiva, ensamblar genomas y/o obtener polimorfismos a lo largo de este es complejo. Por ejemplo, en genomas que presentan muchas duplicaciones, pseudogenes y elementos móviles, resulta difícil alinear los fragmentos entre ellos y distinguir las duplicaciones de los polimorfismos. Métodos como los Radtags (Baird *et al.* 2008), el GBS (Elshire *et al.* 2011) y el DARTseq (Sansaloni *et al.* 2011; Ren *et al.* 2015) han sido utilizados para simplificar la complejidad de genoma. Estos métodos permiten reducir el número de fragmentos que se secuencían. Así mismo, si estos se asocian a regiones no metiladas, en principio solo se amplificarían regiones codificantes. En este caso, la ventaja es que se obtienen datos sin ningún tipo de sesgos, aunque son muy complicados de analizar informáticamente (especialmente si no hay genoma de referencia), son costosos y generan altos porcentajes de datos faltantes.

Otra aproximación interesante ha sido el diseño de arreglos que genotipan marcadores polimórficos a lo largo del genoma. Los datos generados con estos métodos son considerablemente más sencillos de analizar, tienen pocos datos faltantes y son menos costosos. Sin embargo, se necesita conocer muy bien el genoma para poder desarrollarlos, ya que por su diseño presentan sesgos en exceso de polimorfismos (*ascertainment bias*), y por lo tanto pueden generar inferencias demográficas incorrectas (CHIP), lo que a su vez puede afectar los estudios de selección.

La segunda aproximación, llamada *bottom-up* se apoya en el desarrollo de la secuenciación masiva (ver Caja 1) y se basa en la detección de señales de selección en el genoma y posteriormente la identificación de los rasgos que están determinados por esos genes. En teoría, la recombinación genera un mosaico de polimorfismos a lo largo del genoma, de tal forma que los genes puedan responder más o menos individualmente a las distintas fuerzas evolutivas dependiendo de la tasa de recombinación. De esta forma, la gran mayoría del genoma, que es neutral, evolucionará de acuerdo a un equilibrio migración/deriva. Por el contrario, las frecuencias alélicas de los genes que estén bajo selección aumentarán o disminuirán en mayor magnitud que lo predicho bajo un escenario

de neutralidad. Basándose en eso, Lewontin y Krakauer (1973) propusieron utilizar la medida de diferenciación genética (F_{ST}) para comparar los valores globales esperados por neutralidad y posteriormente distinguir aquellos genes que tiene valores atípicos de F_{ST}. Así, genes que presenten valores más elevados podrían indicar la presencia de selección diferencial, mientras que aquellos que presenten una menor diferenciación podrían estar bajo selección balanceadora o direccional.

Los métodos *bottom up* (ver Tabla 1) más utilizados en el contexto de genómica de poblaciones son aquellos que detectan polimorfismos cuyas frecuencias alélicas se correlacionan fuertemente con el ambiente (selección correlativa) y aquellos que detectan marcadores con diferenciación genética muy marcada (selección diferencial). En el caso de la selección correlativa, los métodos más populares (por ej. Bayenv, Baypass) se basan en primero estimar matrices de auto-correlación entre la estructura genética y la distancia geográfica de las poblaciones, para posteriormente detectar marcadores que presenten una correlación estrecha con el ambiente (utilizando la matriz de auto-correlación como co-variable). Estos métodos son robustos, y tienen la ventaja (desventaja) de que permiten detectar genes asociados a un ambiente de interés (biótico y abiótico).

En el caso de la detección de marcadores asociados a selección diferencial, se han desarrollado dos aproximaciones. En la primera, se simula la distribución de la heterocigocis (H_S) y diferenciación genética (F_{ST}) de miles a millones de marcadores. Posteriormente se estima el intervalo de confianza de la F_{ST} dada la H_S y se definen como candidatos aquellos marcadores que se encuentren fuera del intervalo (Fdist, Arlequin). La ventaja más importante de estos métodos es que permiten simular escenarios demográficos y de estructura genético-poblacional complejos (Excoffier *et al.* 2009a). En la segunda aproximación se usan métodos bayesianos para estimar la distribución posterior de la estructura genética (F_{ST}) de los marcadores en función de un componente poblacional (compartido por todos los marcadores) y un componente específico del locus. Aquellos marcadores que tengan un componente local significativo podrían estar bajo selección diferencial (si la diferenciación es elevada) o balanceadora (si la diferenciación es baja). La ventaja de los métodos diferenciales es que permiten identificar loci candidatos sin tener una hipótesis ambiental. Sin embargo, estos métodos pueden ser propensos a falsos

positivos, por lo que es conveniente tener muchas poblaciones y estimar mejor los parámetros de diferenciación (de Mita *et al.* 2013).

Recientemente, de Villemeureuil y Gaggiotti (2015) desarrollaron *bayescenv*, un método que combina la información de los métodos correlativos y diferenciales. Este método tiene la ventaja de que permite incorporar un tercer componente que considera la asociación entre el ambiente y las frecuencias alélicas del locus. Este método ha mostrado ser muy robusto y generar pocos falsos positivos (de Villemereuil y Gaggiotti 2015; Aguirre-Liguori *et al.* 2017; Aguirre-Liguori *et al.* en preparación).

Tabla 1. Comparación de métodos para detectar genes candidatos a estar bajo selección. Para una descripción sencilla de los métodos y sesgos de acuerdo a parámetros poblacionales ver de Mita *et al.* (2013).

TIPO DE SELECCIÓN	EJEMPLO	VENTAJAS	DESVENTAJAS
Correlativo Identifica marcadores que tengan una correlación muy estrecha con el ambiente	Regresión Logística (Joost *et al.* 2007)	- Corre muy rápido - Sencillo	- Falsos positivos - No estima estructura genética - No controla por autocorrelación espacial
	Bayenv (Coop *et al.* 2010)	- Bayesiano - Considera estructura genética y autocorrelación	- Muy tardado - Complicado de correr
Diferencial Identifica marcadores que presentan diferenciación genética por encima de lo esperado por neutralidad	Fdist (Beaumont and Nichols 1996)	- Permite simular modelos demográficos complejos	- Susceptible a falsos positivos - Difícil de generar formato de entrada
	Bayescan (Foll y Gaggiotti 2008)	- Robusto - eficiente para distinguir *outliers* - permite distinguir marcadores bajo selección balanceadora	- Susceptible a falsos positivos - Tardado
	Bayescenv (de Villemeureuil y Gaggiotti 2015)	- Robusto - Permite identificar marcadores relacionados a ambientes específicos	- Tardado - Muy estricto - No permite identificar marcadores bajo selección balanceadora

Aunque en teoría la detección de genes candidatos es intuitiva, un problema serio ha sido que resulta muy complicado distinguir la distribución de F_{ST} real de genes neutrales (De Mita *et al.* 2013). Esto genera que muchos de los genes que se detectan como atípicos puedan ser falsos positivos, y que su alta diferenciación genética se deba a procesos no

adaptativos, como el desequilibrio de ligamiento, los cambios en las tasas de mutación, o los mismos procesos demográficos. Se han desarrollado muchas aproximaciones para tratar de corregir estos problemas. Por ejemplo, se han desarrollado diseños experimentales elegantes que buscan paralelismos entre las poblaciones (Hohenlohe *et al.* 2010; Fustier *et al.* 2017) o que observan cambios alélicos en las poblaciones después de una generación de selección (Soria-Carrasco *et al.* 2014; Egan *et al.* 2015; Riesch *et al.* 2017). Aparte del desarrollo y mejoramiento de los métodos estadísticos mencionados anteriormente, las simulaciones y resultados empíricos muestran que para reducir el número de falsos positivos en los estudios *bottom-up* es importante tener muestreos extensos y que ocupen toda la distribución geográfica y ecológica de la especie de interés (De Mita *et al.* 2013; Schoville, Bonin, Francois, *et al.* 2012; Jones *et al.* 2013). Otra propuesta ha sido utilizar distintos métodos para detectar genes candidatos y retener aquellos que sean identificados por más de dos de ellos. Sin embargo, esto podría reducir la información adaptativa, ya que los métodos se basan en distintos supuestos y por lo tanto podrían estar identificando marcadores asociados a distintos tipos de selección (por ejemplo correlativa y diferencial).

MAICES SILVESTRES, LOS TEOCINTLES

Zea mays ssp. *parviglumis* (de ahora en adelante *parviglumis*) y *Zea mays* ssp. *mexicana* (de ahora en adelante *mexicana*) son dos subespecies anuales de pastos endémicos de México (Figura 2, para más detalles ver Capítulo 1: Aguirre-Liguori *et al.* 2016). Son principalmente estudiados por tener una relación muy estrecha con el maíz domesticado *Zea mays* ssp. *mays* (de ahora en adelante maíz), del que son su ancestro silvestre más cercano. El maíz se domesticó en las tierras bajas de México hace alrededor de 9,000 años a partir de *parviglumis* (Matsuoka *et al.* 2002) y posteriormente, por introgresión adaptativa con *mexicana*, adquirió genes que le ayudaron a adaptarse a las tierras altas de México (Hufford *et al.* 2012a) donde posteriormente diversificó en nuevas razas (Matsuoka *et al.* 2002; van Heerwaarden *et al.* 2011).

Los maíces silvestres y cultivados se caracterizan por presentar tamaños efectivos grandes y una elevada diversidad genética (Ross-Ibarra, Tenaillon, and Gaut 2009). Asimismo tienen genomas muy dinámicos, que cambian de tamaño en función de la proliferación y pérdida de elementos móviles (Chia *et al.* 2012; Tenaillon *et al.* 2011; Zerjal

et al. 2012). Esta combinación de características genéticas y genómicas, así como sus ciclos de vida cortos, han permitido que los teocintles y maíces se adapten localmente a una gran variedad de ambientes, suelos y otros tipos de ambientes bióticos y abióticos (Pyhäjärvi *et al.* 2013; Aguirre-Liguori *et al.* 2017; Fustier *et al.* 2017). Por todas estas características, resultan modelos muy interesantes para estudiar procesos adaptivos y evolutivos (Ross-Ibarra *et al.* 2007; Hufford *et al.* 2012b; Aguirre-Liguori *et al.* 2016).

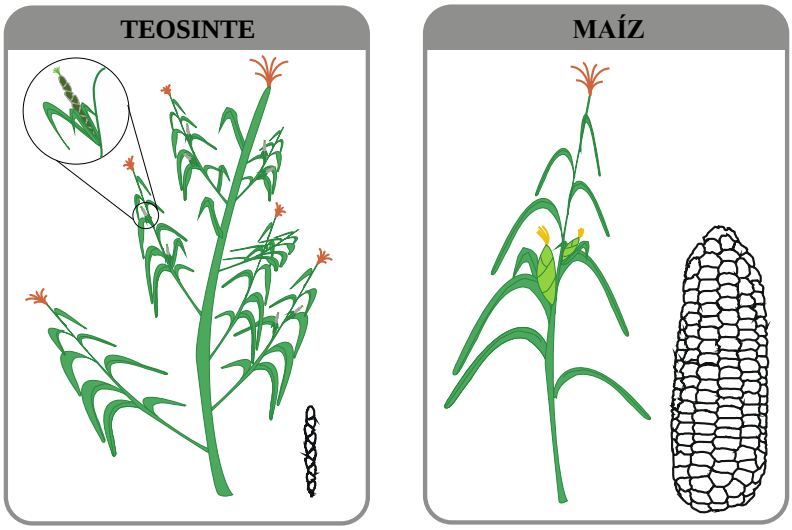


Figura 2. Comparación morfológica y del fruto entre teocintles y maíz. La figura se obtuvo del artículo de divulgación de Aguirre-Liguori 2017 (Oikos=).

Por otro lado, las dos subespecies de teocintle están adaptadas a dos ambientes divergentes (Fukunaga *et al.* 2005; Hufford *et al.* 2012a; Aguirre-Liguori *et al.* 2017). *Mexicana* crece en tierras altas, húmedas, frías y con poca disponibilidad de nutrientes en el suelo. *Parviglumis* crece en tierras bajas, húmedas y calientes con mayor disponibilidad de nutrientes en el suelo. Estas dos subespecies crecen en nichos bien definidos, y tienen pocas zonas de sobrelape (Hufford *et al.* 2012a) donde ocasionalmente hibridizan (van Heerwaarden *et al.* 2011; Aguirre-Liguori *et al.* 2016). *Mexicana* se originó posiblemente a partir de *parviglumis* y es probable que su divergencia haya ocurrido por adaptación local a nichos divergentes (Aguirre-Liguori *et al.* 2017). Finalmente estudios citológicos y genómicos han evidenciado la existencia de posibles inversiones cromosómicas entre ambas subespecies (Fang *et al.* 2012; Pyhäjärvi *et al.* 2013; Aguirre-Liguori *et al.* 2017; Aguirre-Liguori *et al.* en preparación Fustier *et al.* 2017), que podrían estar altamente

enriquecidas en genes candidatos. Por lo tanto, es posible que existan mecanismos genéticos que reduzcan el flujo genético entre subespecies. Debido a que las dos subespecies están relativamente bien aisladas geográficamente (Hufford *et al.* 2012a) y a que presentan diferencias en los tiempos de floración (Rodríguez *et al.* 2006), es altamente probable que estas estén divergiendo por especiación ecológica (Aguirre-Liguori *et al.* en preparación).

Finalmente, por la relación tan estrecha entre el maíz y los teocintles, la gran mayoría de los recursos genéticos que se han desarrollado para el primero (Schnable *et al.* 2009; Chia *et al.* 2012), se pueden utilizar en los segundos (Hufford *et al.* 2012b). Así mismo, todos los resultados genéticos obtenidos para teocintles, pueden ser relevantes para la conservación y mejoramiento del maíz. Por ejemplo, hay evidencia creciente que muestra que la alta diversidad genética de las especies silvestres puede ser importante en el mejoramiento de las especies domesticadas (Warschefsky *et al.* 2014). En ese sentido, la detección de genes de importancia agronómica en teocintle puede ser importante para el mejoramiento del maíz (Aguirre-Liguori *et al.* 2017; Aguirre-Liguori *et al.* en preparación).

OBJETIVOS

El objetivo principal de este trabajo es analizar los mecanismos que generan la adaptación local y promueven la especiación ecológica entre dos subespecies de teocintle, analizando este proceso a lo largo de un continuo de especiación.

Para responder esto, tenemos tres objetivos principales, que respondemos en cada uno de los capítulos.

- Mi primer objetivo consistió en hacer una revisión sobre la historia natural y la evolución genética de los teocintles, con el fin de discutir y mostrar que estas dos subespecies son modelos interesantes para estudiar las bases genéticas y ecológicas de la adaptación local, así como para estudiar la especiación ecológica.
- El segundo objetivo fue analizar las bases geográficas y ecológicas de la adaptación local en teocintles, para identificar en qué regiones y condiciones ecológicas está ocurriendo la adaptación local.
- Finalmente, el tercer objetivo fue analizar si la adaptación local en el límite del nicho ha promovido la divergencia genómica y ecológica de los teocintles.

CAPÍTULO 1: HISTORIA NATURAL DEL TEOCINTLE

Capítulo de libro: Genetics and Ecology of Wild and Cultivated Maize: Domestication and Introgression

Chapter 16
Genetics and Ecology of Wild and Cultivated Maize: Domestication and Introgression

Jonás Andrés Aguirre-Liguori, Erika Aguirre-Planter, and Luis E. Eguiarte

Abstract Maize (*Zea mays* subspecies *mays*) has been culturally and economically a very important crop since it was domesticated from its wild relatives, the teosintes (both the lowlands teosinte, *Zea mays* subspecies *parviglumis* and the highlands teosisnte, *Zea mays* subspecies *mexicana*) in Mexico. In this chapter, we review molecular studies analyzing different aspects of the genetic resources, domestication, phylogeography, and other aspects of the evolution of maize and teosintes, including niche modeling. The genetic studies range from isoenzymes to single nucleotide polymorphisms and other genomic and transcriptomic studies. Both cultivated maize and wild teosintes have high levels of genetic variation and signals of strong local adaptation. Currently, the most accepted hypothesis on maize origin indicates that domestication occurred 9000 years ago in a single event in southern Mexico from the lowland subspecies, *Z. m. parviglumis*. According to these ideas, later maize spread into higher elevations through adaptive introgression with highland teosintes, *Z. m. mexicana*. But these ideas are still open to discussion, as better data are needed. Since the origin of maize, there has been strong ongoing artificial selection that has allowed maize to diversify and spread globally and to highly new environments. This intensive selection in maize has left strong molecular signals of selection on a variety of genes that go from domesticated genes to improvement genes. To help respond to climate and global changes, it will be important to determine genes of agronomic importance for tolerance (weather, plagues) and improvement (increase in productivity) to cope with these changes.

Keywords Genetic resources • Domestication • Phylogeography • Maize • Mesoamerica

J.A. Aguirre-Liguori (✉) • E. Aguirre-Planter, Ph.D.
Instituto de Ecología, Universidad Nacional Autónoma de México, México, Distrito Federal, Mexico
e-mail: jonas_aguirre@hotmail.com

L.E. Eguiarte, Ph.D.
Departamento de Ecología Evolutiva, Instituto de Ecología, Universidad Nacional Autónoma de México, Ciudad Universitaria, México, Distrito Federal, Mexico
e-mail: jonas_aguirre@hotmail.com

R. Lira et al. (eds.), *Ethnobotany of Mexico*, Ethnobiology,
DOI 10.1007/978-1-4614-6669-7_16

Introduction

Maize, *Zea mays* subspecies *mays*, was the central crop species of Mesoamerican pre-Hispanic cultures, and it is still a fundamental part of Mexican and Central American cultures. Maize is also an important crop for human and cattle consumption in many countries (see [1] for a recent review about uses in maize). In addition, the wild relatives are also used in different regions of Mexico, for instance, for cattle consumption during the dry season, when other plants are scarce [2]. Since its domestication from teosinte, different maize races and varieties have been developed for different uses and growth conditions [1, 3], resulting in an impressive range of morphological, physiological, and genetic variation. There have been important, traditional genetics, genomic and genetic engineering developments in maize, all of them having the potential for further improving this crop production. We consider it relevant to understand the interactions that humans have had with maize, which have allowed their continuous improvement since its domestication, ca. 11,000–9000 years ago.

Zea mays is a monoecious monocotyledon annual grass; it is wind pollinated and it is widely planted in Mexico and several other countries, being adapted to grow in different soil and climatic conditions [4, 5]. Given its economic and cultural importance, its genome is relatively well known, specially compared with other Mexican plants. The maize genome, which was first published in 2009 [6, 7], consists of ten chromosomes, and has an extension of ca. 2.3 gigabases, comprising 32,590 genes clustered in 11,892 families. In addition, almost 85 % of the maize genome is composed of different families of transposable elements, which were initially described in this crop.

As we mentioned above, maize was domesticated in Central Mexico, 11,000 to 9000 years ago from their wild relatives, commonly called teosintes in scientific literature [1, 8, 9]. Although some experts are not convinced by the subspecies nomenclature (see [10]), currently three subspecies of teosinte closely related to maize are usually recognized. The Balsas teosinte, *Zea mays* subspecies *parviglumis* (from now on *parviglumis*) Iltis et Doebley, mainly growing in the Balsas river basin and in the state of Jalisco including also a population in southern Oaxaca. *Parviglumis* is adapted to growth at low elevations in tropical seasonal (with marked dry season in winter) regions, between 340 and 1929 m.a.s.l. at an average of 1058 m.a.s.l. [5]. The ethnobotanical information about this subspecies is scarce. However, besides being used for cattle consumption, there is a report of medicinal use [2], in which seeds are imbibed in water and consumed for stomach pain. Also, Mondragon-Pichardo and Vibrans [2] report people who have heard that crossing *parviglumis* with maize during many generations hardens the maize kernel. The second subspecies is the highland teosinte *Z. m.* spp. *mexicana* (Schrader) Iltis. (from now on *mexicana*) that grows in the volcanic region of central Mexico, at higher altitudes, at elevations between ca. 2000 and 2600 m.a.s.l., with an average of 2105 m.a.s.l., in colder, drier regions, and generally more variable temperatures [5]. Although both subspecies are mainly allopatric, both overlap in sites of northern

Balsas River Basin, where they occasionally hybridize [11]. Finally another subspecies, *Z. m. huehuetenangensis* is found only in populations in western Guatemala, near its border with Mexico, and it has been shown that this taxa is a more distant relative to the two other teosintes subspecies and to the cultivated corn [11].

In this Chapter, we review the genetic resources and phylogeography of maize and its wild relatives, *Z. m*. ssp. *parviglumis* and *Z. m*. spp. *mexicana*, discussing traditional and current ideas on the origin, spread and improvement of maize, using recent genomic and molecular evolution information.

The Evolutionary Biology and Ecology of Teosinte and Maize

Zea is divided into two sections, *Luxuriantes* and *Zea* [1]. The *Luxuriantes* section is characterized by ruderal species adapted to disturbed environments that includes *Z. luxurians*, an annual diploid species that grows in Guatemala and Nicaragua (although the Nicaraguan populations were reclassified as *Z. nicaraguensis* by Iltis and Benz [12]); *Z. diploperennis*, which is a perennial diploid species that is principally found in the Jalisco state and used for cattle consumption during the dry season [13]; and *Z. perennis*, which is a perennial tetraploid species also found in Jalisco. Sánchez et al. [14] described some teosintes populations within the *Luxuriantes* section as having morphological, ecographic, cytological, and molecular traits that suggest they may represent new species. The first one is a perennial diploid population from Nayarit, another one, a perennial tetraploid population from Michoacán, and the third one a diploid annual plant from Oaxaca.

On the other hand, *Zea* section is composed exclusively by diploid taxa, and includes the cultivated maize (*Z. mays* ssp. *mays*), the two teosintes mentioned above (*Z. mays* ssp. *mexicana*, and *Z. mays* ssp. *parviglumis*), and the subspecies *Z. mays* ssp. *huehuetenangensis*, found in a few populations in eastern Guatemala. Using molecular data (26 nuclear loci), the divergence between *luxurians* and *parviglumis* was dated 140,000 years ago, and the divergence between *Z. m. mexicana* and *Z. m. parviglumis* at 60,000 years ago [15]. Afterwards, teosintes and maize have experienced subsequent demographic expansions that resulted in their actual high levels of genetic variation [15]. *Zea. mays* ssp. *mexicana*, and *Z. mays* ssp. *parviglumis* are genetically and evolutionary very close, as demonstrated by the number of intermediate and admixed populations and genotypes [8, 9, 11].

Recent genetic studies suggest that *Z. m. parviglumis* is the ancestor of the cultivated maize [8, 9, 15], as we will explain below. However, *mexicana* can be very similar to cultivated maize, but this has been interpreted as a result of subsequent introgression with maize after domestication [9, 16].

Obviously, since teosintes are the wild ancestors of maize, understanding their evolutionary biology is important. Besides, given their abundance, their high genetic diversity [15], and the diversity of environmental, ecological, and edaphic conditions in which they grow [5], teosintes are ideal for the study of population genetics and for understanding the basis of adaptation. Finally, given the synteny

(the conserved order of the genes in the chromosome) between maize and teosintes and its evolutionary closeness, genomic data, and genetic tools developed for maize can be used most of the times in teosintes.

The recent studies in teosintes have undergone a gradual improvement in the nature of the genetic markers used, in the number of loci, populations and individuals analyzed, as well as a shift to wild population-based studies with larger sampling numbers, in contrast to the original studies, where few individuals per accession from a seed collection were used (i.e., [8, 11]). For instance, using 93 microsatellite nuclear loci, Fukunaga et al. [11] conducted a genetic analysis of 237 plants obtained from 172 accessions (collection sites) of the wild subspecies *mexicana*, *parviglumis*, and *huehuetenanguensis*, as well as Z. *diploperennis*, *Z. luxurians y Z. perennis*. In general, *Z. m. parviglumis* has a higher genetic diversity than *Z. m. mexicana*. A phylogenetic analysis suggests that *mexicana* was originated from *parviglumis*, and that together they form a monophyletic group. This study, as is the case of others that analyze cultivated maize (i.e., [17, 18]), was conducted using accessions, which has the limitation that it is not possible to determine accurate genetic frequencies and therefore it is difficult to conduct detailed population genetic and phylogeographic analyses.

Using 61 populations (45 populations of maize described in [19] and 16 teosinte populations from [20]), Buckler et al. [21] analyzed with 21 isoenzymes 9 to 50 plants per population. According to this study, *Z. m. parviglumis* is basal and paraphyletic, including *Z. m. mexicana* as a monophiletic clade (supporting the results of 11). In addition, they reported that *Z. m. huhuetenanguensis* is basal to the other taxa, and sister taxon to the other two. Finally, these authors analyzed the phylogeography of teosintes using chloroplast RFPLs, finding five haplotypes (four in *parviglumis*, three in *mexicana* and two shared haplotypes). Although their results do not show a high resolution, they found isolation by distance, and some isolation generated by altitude, perhaps related to the climate.

Merino-Díaz [22] analyzed in average 27 individuals in 10 populations, which covered most of the distribution of both subspecies. He obtained 139 ISSRs polymorphic loci, which are nuclear dominant markers related to microsatellites [23]. With these markers, he found a mean diversity of $H_S = 0.261$ and a mean polymorphic variation of 77.74 %. This diversity is high compared, for instance, to the genetically diverse species of the *Agave* genus found in Mexico, that in average have a lower mean genetic diversity (33 studies $H_S = 0.19$) and a lower mean polymorphic diversity ($P = 56$ %; [23]). In addition, [22] found a high genetic differentiation among populations ($\theta_{WC} = 0.1837$, $\theta_H = 0.23$) using, respectively, the Weir and Cockerham theta [24] and the Hickory bayesian estimation [25, 26]. Given that teosintes are wind and cross-pollinated, their genetic differentiation would be expected to be low [27, 28], but it was higher than, for instance, the average found in the animal pollinated *Agave* genus (23 studies, $F_{ST} = 0.15$; [23]). Of the 139 loci, Aguirre-Liguori et al. (in prep) identified with Bayescan V.2 [29] three loci that appear to have been under directional selection and one additional locus that shows evidence of balancing selection. We analyzed the correlation between genotype frequencies of the three loci that appear to be under directional selection and their

population environmental data, and found that two are associated with altitude, suggesting local adaptation. Although ISSRs are anonymous markers, we suggest that this approximation will help us to detect complex patterns of genetic adaptation and genetic structure.

Villasante-Barahona [30] amplified nine nuclear microsatellites in 26–37 individuals from five populations of both subspecies, detecting, not surprisingly, higher levels of genetic variation than those estimated with ISSRs—as it generally occurs with microsatellites, given their high mutation rate and high number of alleles. Mean population diversity ranged from $H_S = 0.727$ to 0.807 and a number of alleles that ranged from 5 to 23. Comparing this data with an *Agave* (*A. parryi* average $H_S = 0.621$, 4 loci [31]) indicates again that teosinte is genetically very diverse, consistent with [15]. This study reported low F_{ST} values for pairs of populations (from 0.0389 to 0.139). However, it was interesting that he found significant and positive F_{IS} values (range 0.103–0.219), given the species is wind pollinated and has monoecious flowers. This suggests there is inbreeding; either originated by selfing and/or crosses with relatives, or generated by genetic structure within the sampled populations (i.e., Wahlund effect). In order to answer this question, it is essential to study the mating systems of wild populations along the patterns of gene flow within populations with detailed paternity analyses.

Moeller et al. [32] used five nuclear genes and two chloroplast sequences to analyze *Z. m. parviglumis* using 84 individuals in seven populations (four in Jalisco and three in the Balsas region). The Balsas region had more genetic variation than the Jalisco region, and an AMOVA showed that the majority of the differentiation was found within regions. Analyzing the chloroplast, they found a strong phylogeographical structure, but not clear patterns, with many populations presenting a unique haplotype confined to a single region. However, they found evidence of gene flow through seeds between regions.

In detailed analyses of two Balsas basin *parviglumis* populations [33] used 468 SNPs in 389 and 575 individuals in two populations finding similar levels of genetic diversity in both sites, and a low genetic differentiation between them. However, the genomic resolution allowed them to detect low, but significant genetic structure within each site that could be correlated to the sites environmental and topographic heterogeneities. This study is interesting since it shows a complex and fine genetic structure in teosintes, despite being wind pollinated, as well as the power of resolution achieved using numerous genetic markers.

Pyhäjärvi et al. [34] increased considerably the number of loci to more than 36,000 SNPs, and studied 250 individuals that belonged to 21 teosinte populations (11 from *parviglumis* and 10 from *mexicana*). These populations covered most of the distribution of the subspecies. These authors found that teosintes present hierarchical genetic structure [35], which means that populations within a neighborhood present more gene flow than population between neighborhoods. Using different methods [34] found numerous SNPs associated to environmental variables. Interestingly, several SNPs that had a signal of selection were in intergenic spacers or were synonymous. In addition, they identified four large regions with high linkage disequilibrium (more than 10 million base pairs) that might correspond to

inverted regions in the genome that inhibit recombination. These regions are rich in SNPs that are statistically associated with temperature and altitude. Among the inversions found, one (*Inv1n*) was at mid frequencies in low altitudes and at low frequencies at higher altitudes. Another one (*Inv4n*) was exclusively found at high altitudes in both subspecies, suggesting that it is relevant for high altitude adaptation. A third inversion was exclusive to the highest populations of *mexicana*.

Fang et al. [36] analyzed the *Inv1n* inversion, which corresponds to 50 Mb found in Chromosome 1, with 941 SNPs from 542 mapped genes. This inversion has a high linkage disequilibrium (i.e., some allelic combinations are more frequent that what is expected under total random recombination) compared to the rest of the chromosome. However, when they analyzed the linkage disequilibrium within the inversion, they found similar values compared to the rest of the chromosome, suggesting that recombination occurs within the inverted region. The inversion divides *parviglumis* into two distinct groups that are not detected in the rest of the chromosome. They also sequenced four loci within the inversion region that support their results. The inversion diverged between *parviglumis* and *mexicana* at the same time, indicating that it is not from an introgressed origin and estimated the divergence time of the inversion at ca. 300,000 generations. This time predates the divergence between *Z. luxurians* and the ancestor of *parviglumis*, suggesting that it was lost by genetic drift in the other taxa, perhaps because of their smaller effective population size. Finally, as this inversion is not found in maize (even if it was domesticated from *parviglumis*), the authors suggest it was lost due to selection against the inversion. To test whether the inversion is adaptive, these authors correlated their frequency with environmental variables, and found negative but significant associations with altitude and some associated climatic variables.

As is the case of inversions, other types of genome architecture changes can influence local adaptation. These are non-codifying but functional elements, such as transposable elements (TE), heterochromatic knobs, or copy-number variants and presence/absence polymorphisms [37]. Transposable elements were first described in maize [38], and there have been recent improvements in understanding their role in the evolution of genome size, and their effects on fitness.

Transposable elements are interesting because there are many families that behave differently in the genomes, in terms of where they insert themselves, either doing it in genic or non-genic regions. Given their nature, transposable elements dynamics in the genomes are normally regulated by purifying selection, explaining for instance why animals that have high effective population sizes tend to have fewer transposable elements [39]. Using Next Generation Sequencing (NGS), which are approaches that allow genomic-wide sequencing, Tenaillon et al. [40] analyzed the TE components of *Z. mays* (B73) and *Z. luxurians*, and compared their genome sizes (GS). These authors found that maize has a 1.5 fold shorter genome compared to *Z. luxurians*, which is in part explained by differences in abundance of TE. According to their results, TE explains 70 % of GS differences between *Z. luxurians* and maize, with the former presenting more TE families and abundance than

the latter. These differences could be associated with changes in physiology, phenology, and life history traits [41]. Although there could be purifying selection against transposable elements, which could explain their reduction in maize, there are families that discriminate where they insert themselves, such as the *Class 2 miniature inverted repeat elements* (MITEs). These transposable elements are able to insert themselves in genic regions, which could affect the functioning of genes or change their regulation mechanism, and may help in some cases to adapt.

Chia et al. [37] analyzed the genomic diversity of 103 inbreed lines including elite inbreed lines, landraces, and teosintes, for a total of 55 million SNPs. Twenty-one percent of the SNPs were associated with genic regions (825,000 synonymous and 571,000 nonsynonymous, and 10,000 were non-sense). In the case of the *Zea* section, these authors found that heterochromatic knobs correlated positively with GS, while in an apparent paradox, transposable elements abundance correlated negatively with GS. This means that while there has been a reduction in GS associated with loss of heterochromatic knobs (probably through purifying selection), there has been an increase in the number of TE. Overall, the data [37, 40] suggest that transposable elements are responsible of GS variation in grasses, but in maize there has been a shift to a major variation associated with heterochromatic knobs.

Diez et al. [42] found that GS in teosinte and maize varies among populations and within populations, and correlates with environmental variables, suggesting it could be under selection. Diez et al. [42] analyzed and compared the GS of 5 individuals from 21 populations of both subspecies of teosintes and 22 Mexican traditional landraces, which were distributed at diverse environments, at altitudinal clines and at two parallel transects. If the sequenced maize B73 is used as a reference (i.e., GS = 1), a significant variation in GS among individuals (ranging from 0.948 to 1.299) and a difference in average GS between maize (1.095) and teosinte (1.129) is found, although both groups had a similar coefficient of variation. Most variation occurred among populations for both groups, but it was higher in maize. In particular, we found a stronger reduction in GS for inbred elite lines, suggesting that there has been a reduction in GS, associated with domestication [40]. Two gradients were studied for each group. For the teosintes, we sampled gradients in the Balsas and Jalisco region and for maize in the Balsas and Oaxaca gradients. When we considered these gradients, we found a significant variance caused by the gradients for both groups, but it was higher for the cultivated maize. When we compared the GS variation, an association between bioclimatic (temperature and precipitation) and geographic (latitude and longitude) variables and GS in the Balsas gradient of maize samples (while it was not significant for the Oaxaca gradient) was detected. In the case of teosintes, only some variables, which were associated with seasonality variables (precipitation in the warmest and coldest quarters), were significantly correlated with GS. For maize, we also found an association with geographic coordinates, which could reflect complex environmental or cultural scenarios. Overall, these results suggest that GS evolves in complex ways with different selection pressures and/or random process changing GS in alternative ways.

The Domestication of Maize and the Problem of Introgression Teosinte-Maize

Given the importance of maize and the high diversity of landraces, there have been many efforts to identify the origin of maize. Many hypotheses have been developed to answer this question. For a recent review on traditional hypothesis, see [1]. However, with the development of molecular analyses it has been possible to advance in answering this question, although the results are not as straightforward. It is important to consider that given the strength of selection that occurred during maize domestication, comparing wild and domesticated maize has allowed to start understanding the genetic basis of adaptation, as well as the processes that have been involved [43].

As it has been pointed out earlier in this chapter, maize and teosinte are characterized by their enormous genetic and phenotypic diversity, which has led to hypothesize that it was domesticated multiple times, as it has occurred in other crops. In an effort to determine the origin of maize, Matsuoka et al. [8] used 93 microsatellite markers and 264 individuals of teosinte and maize that cover a broad distribution. According to their phylogenetic analyses, they found strong support (930 out of 1000 bootstrap samples) that maize was domesticated only once from *parviglumis*, thus making maize monophyletic. In addition, they dated the origin of maize around 9188 years BP, with a 95 % confidence limits ranging from 5689 to 13,093 years BP. Using a principal component analysis, they clustered different groups of maize and teosintes, finding that *mexicana* is separated from maize, supporting the evidence that *parviglumis* was the only ancestor of maize. When trying to identify the closest ancestor to maize, they found that *parviglumis* of the central region of the Balsas River is the closest candidate, and placed the domestication in central Oaxaca (although they suggested that a finer sampling would help defining better the site of domestication). Also, they detected admixture between subspecies, and notably introgression from *mexicana* into maize, explaining (at least in part) the high genetic and morphologic diversity encountered and the adaptation of the highland maize. Finally, the genetic clusters identified by Matsuoka et al. [8] suggest that the initial diversification of maize occurred in the highland landraces, and originated two main lineages, one that dispersed to the North of Mexico and North America, and the other one that dispersed to the western and southern lowlands of Mexico and subsequently to the Caribbean, Central and South America. There are opposite archeological and genetic evidences that place the diversification of maize at different altitudes, suggesting that either maize diversified in the highlands and subsequently spread to the lowlands, or that maize from the lowlands was first domesticated and then rapidly diversified into the different landraces, particularly in the highlands; see review in [5].

In particular, most genetic data suggest that primitive maize landraces that grow in the highlands are more similar to *parviglumis* than maize that grow in lowlands, and that they are also the ancestors to the rest of the cultivated maize. In an attempt to unravel this paradox, van Heerwaarden et al. [9] used 964 SNP from 547 genes

in 1127 accessions of maize landraces in addition to more than 100 accessions of *Z. m. parviglumis* and 96 of *Z. m. mexicana*. Using a principal components analysis (PCA), they concluded that *parviglumis* is closer to maize landraces. However, they found similar genetic patterns among highland races and *mexicana*. In addition, they found admixture among the three subspecies, and particularly strong between *mexicana* and the highland races. Given that admixture between ancestral maize and teosintes could affect the genetic signals, these authors used a method that attempts to estimate the ancestral allele frequencies of maize. When they compared the estimated ancestral frequencies of maize with those of the maize landraces, they found that the closest frequencies were those found in the west of Mexico lowland landraces. van Heerwaarden et al. [9] concluded that maize was thus domesticated in the lowlands, and subsequently diversified into other landraces. To explain the apparent paradox mentioned above, these authors proposed that the maize highland races had strong admixture with *mexicana*, which introduced many teosintes alleles into the maize gene pool, making them genetically more similar to teosintes. For example, the palomero toluqueño, a highland race, seems to have an important proportion of *mexicana* genome [8]. Overall, this study shows that, either introgression can mislead our inferences, or that we still know little about the true origin of maize.

Recently, Hufford et al. [16] analyzed the putative introgression between *mexicana* and maize. These authors analyzed nine sympatric *mexicana* and maize populations and one isolated (allopatric) *mexicana* population, using 189 individuals and 39,029 SNPs. Although maize and teosintes have well-defined genetic "membership," there is important admixture in the sympatric populations. However, they found that gene flow is asymmetric, with *mexicana* contributing more to the gene pool of maize, and that apparently this process is ancient (ca. 174 generations in maize according to a likelihood of introgression analysis). In addition, Hufford et al. [16] studied the genomic regions that were introgressed and found that within these regions there are many shared SNPs, and were more similar to the non-introgressed regions of the taxa of origin, suggesting similar evolutionary histories, i.e., that introgressed regions in maize had similar diversity than *mexicana* genome. When comparing introgressed and non-introgressed regions, they found that introgressed regions were not rich in domestication genes, but were associated with local adaption to highlands, such as genes associated with pigmentation and macro-hairs, which are important in adaptation to cold and higher lands [34]. When Hufford et al. [16] analyzed the introgressed parts of *mexicana*, they did not find genes associated with domestication, suggesting that *mexicana* has resisted the gene flow from these genes and indicating that *Z. m. mexicana* has always been adapted to disturbed environments, as it happens for species of the section *Luxuriantes*. This resistance to gene flow from domesticated genes could happen either because gene flow from maize to *mexicana* is rare, and probably not advantageous; or that humans could select against *mexicana* hybrids that present intermediate phenotypes. However, the fact that *Z. m. mexicana* is adapted to disturbed environments suggests it might be the true ancestor of maize, and thus the similarity between highland races and *mexicana* could be explained by this alternative hypothesis. In order to have a definitive test if there is ongoing introgression from *mexicana* into maize, and its magnitude,

it would be necessary to analyze with paternity tests the contribution of each subspecies to new seeds. Furthermore, it would be interesting to introduce in experimental fields, lowland maize and highland teosintes and evaluate if adaptive introgression to highlands occurs.

The current distribution of teosinte *Z. m. mexicana* and *parviglumis* is mainly allopatric, and mostly determined by altitude, precipitation, and temperature [5]. There are a few geographic regions were they overlap and where there seems to occur gene flow [11]. In contrast, as mentioned above, cultivated maize as a total has a wider niche, giving its current and extended distribution, varying in elevation, temperature, and seasonality; but each race and variety has its own adapted environmental conditions [4, 5].

We have already reviewed the current ideas and data on the origin and phylogeography of teosintes and maize; however, little is known about the environmental context in which they occur. Using niche modeling, we analyzed the change in potential distribution of these subspecies, as well as four ancestral landraces, to determine how its ecological history has changed [5], in general supporting the genetic analyses we previously described, while helping to determine the climatic environments where domestication took place. The study of [5] used the MaxEnt program and a set of 19 bioclimatic variables to analyze the current potential climatic niche distribution, the potential niche during the Last Glacial Maximum (LGM), 21,000 years ago—when temperature was 4 to 6 °C cooler and 10–30 % drier that today—and during the Last Interglacial (LI), 135,000 years ago. Paleoclimatic evidences suggest that there were important climatic shifts, and particularly at 10,300 and 8200 years ago, resulting in important vegetation shifts during the time of domestication. This has made it difficult to determine the climatic context during domestication and the past distribution of the wild taxa. The analyses indicated in general terms that while *mexicana* is able to grow in a wider diversity of environments, *parviglumis* is confined to a more tropical and seasonal temperature and precipitation. According to our niche modeling, there was an important increase in the distribution from the LI to the LGM, and a minor increase to the present time, suggesting a continuous population increase of teosintes populations, as inferred previously [15]. In addition, we found that *parviglumis* has apparently expanded into higher areas, a shift from 524 m.a.s.l. to the current mean of 1058 m.a.s.l., while *mexicana* changed from a mean altitude of 1836 m.a.s.l. to the current average of 2015 m.a.s.l. 21,000 years ago (at the LGM). Given these results, we proposed that *parviglumis* colonized the Central Balsas region, while *mexicana* expanded through the Transverse Volcanic Axis and into the state of Oaxaca. Since the LGM, *parviglumis* expanded to Nayarit, Northern Jalisco, and Eastern Guerrero, and *mexicana* increased its geographic range to the State of Mexico, Tlaxcala, Puebla, and Oaxaca. We found areas of overlap in the three models, which could correspond to potential Pleistocenic refugees and in consequence these areas may be richer in genetic variation, and relevant sites of field study for the understanding of maize domestication. For *parviglumis*, the proposed refugia are in Michoacan and Colima, in the border with Jalisco, while in *mexicana* we identified a similar area in Jalisco and the north of Michoacán.

Hufford et al. [5] also analyzed the potential climatic niche of several traditional and putatively old landraces including Arrocillo Amarillo and Palomero Toluqueño from the highlands and Nal Tel and Chapolote from the lowlands, finding that their distributions have expanded beyond the distribution of their wild relatives, showing that maize adapted to novel environments since its origin and that the diversification process was very fast. This may explain why there are so many old archeological vestiges found outside the distribution of *parviglumis*, which makes it difficult to determine where they were domesticated. An alternative hypothesis is that it was domesticated from an ancestor of the recent *parviglumis* and *mexicana*, but perhaps more closely related, in climatic adaptations, to the current *mexicana* genomes.

From the above sections, it is clear that the vast diversity of maize was shaped by strong natural and artificial selection, coupled by huge effective population sizes allowed by its out crossing, open pollinated system. There was first a domestication process followed by an improvement process that occurred according to the necessities of the environments in which maize was selected. With the recent development of genomic analysis, it has been possible to determine the genes associated with domestication and improvement. For instance, Hufford et al. [44] analyzed over 21 million SNPs in the genomes of 35 improved lines, 23 land races, and 17 individuals of both subspecies of teosinte. Comparing teosintes, traditional landraces of maize and elite recently derived inbreed lines, 484 genetic regions associated with domestication and 695 genetic regions associated with improvement were suggested. In addition, the intensity of selection was estimated to have been, in average, stronger during the domestication process (a selection coefficient $s=0.015$), than during the improvement process ($s=0.003$). It has been an important objective to determine which genes were involved first in domestication and in later improvement processes. In particular, several genes associated with domestication have been identified, some previously well known (*tb1)*, and some other worth of carefully analyzing their population genetics [44]. After domestication and initial development of traditional landraces by Mesoamerican people, there was a subsequent weaker selection in the improved lines used in the USA, on many genes already associated with domestication, highlighting the importance of this old landraces in the actual and future improvement of maize [9]. Besides, Hufford et al. [44] determined the genomic basis of features associated with both domestication and improvement. They found that domestication features contained in average 3.4 genes that covered 322 kilo-base pairs and 7.6 % of the maize genome. On the other hand, traits associated with improvement were in average smaller, and involved fewer genes.

Although domestication occurred through selection on genes, it can also occur through changes in transcription. In the case of maize, phenotypic change from the ancestor has been substantial, which suggests that there should be differences in transcription or in transcription networks. In order to answer this, Swanson-Wagner et al. [45] used an array that covered over 18,000 expressed genes to see differences in gene expression and gene co-expression between 24 accessions of wild relatives and 28 of cultivated maize. The eight days plants profile (i.e., basal transcripts) showed that there is not a whole genome change in transcripts between wild and

domesticated lines, indicating that domestication and improvement did not shift in general the expression. However, they found 612 genes that had a significant differential expression between *parviglumis* and maize and from these, 288 had a two-fold difference, with a slightly general higher expression in maize. Interestingly, these differences were not fixed for subspecies, but instead different lines shared teosinte like-expression and vice versa. Nevertheless, adaptation can occur not only because of differences in expression, but also because of changes in co-expression, which means that two or more genes are expressed simultaneously, resulting in a genetic network interaction. Comparing the topologies of co-expression, they found a significant change in gene network during domestication. In total, they identified 1115 genes with altered co-expression, from which only 276 of these also had differences in expression. When they compared the expression of genes with the traits of domestication and improvement identified by Hufford et al. [44], for many there were increases in expression and changes in co-expression, although only the former were significant. In general, it was found that genes that had a change in expression were higher in maize, although they had a better connectivity (co-expression) in teosintes. The limitation in this study is that they used plants at the initial grow, and not older tissues that could be more easily associated with domestication traits (i.e., ears or flowering time). Nevertheless, it is interesting to see that they found important differences in plants that had similar morphologies.

According to Hufford et al. [5], different races and populations of maize and teosinte are sensitive to patterns of seasonality and temperature. Taking this as guidance, it will be relevant to determine genes of agronomic importance for tolerance (weather, plagues) and improvement (increase in size) in both cultivated and wild relatives to define strategies that will help respond to climate and global change.

Acknowledgements This research was supported by CONACYT Grant CB2011/167826; SEP-CONACYT-ANUIES-ECOS France, Grant M12-A03; and DGAPA, PAPIIT grant IN202712 to LEE. We deeply acknowledge the support and discussions on maize and teosinte of Brandon S. Gaut, Maud Tenaillon, and Daniel Piñero.

References

1. Kato TA, Mapes C, Mera LM, Serratos JA, Bye RA. Origen y diversificación del maíz: una revisión analítica. Mexico, D.F.: Universidad Nacional Autónoma de México, Comisión Nacional para el Conocimiento y Uso de la Biodiversidad; 2009.
2. Mondragon-Pichardo J, Vibrans H. Ethnobotany of the Balsas teosinte (*Zea mays* ssp. *parviglumis*). Maydica. 2005;50(2):123–8.
3. Vielle-Calzada JP, Padilla J. The Mexican landraces: description, classification and diversity. In: Bennetzen JL, Hake SC, editors. Handbook of maize: its biology. New York: Springer; 2009. p. 543–61.
4. Ruiz-Corral JA, Sanchez JDJ, Aguilar M. Potential geographical distribution of teosinte in Mexico: a GIS approach. Maydica. 2001;46(2):105–10.
5. Hufford MB, Martinez-Meyer E, Gaut BS, Eguiarte LE, Tenaillon M. Inferences from the historical distribution of wild and domesticated maize provide ecological evolutionary insight. PLoS One. 2012;7(11), e47659.

6. Schnable PS, Ware D, Fulton RS, Stein JC, Wei F, Pasternak S, et al. The B73 maize genome: complexity, diversity and dynamics. Science. 2009;326(5956):1112–5.
7. Vielle-Calzada JP, De La Vega OM, Hernandez-Guzman G, Ibarra-Laclette E, Alvarez-Mejia C, Vega-Arreguin JC, Jimenez-Moraila B, Fernandez-Cortes A, Corona-Armenta G, Herrera-Estrella L, Herrera-Estrella A. The Palomero genome suggests metal effects on domestication. Science. 2009;326(5956):1078.
8. Matsuoka Y, Vigouroux Y, Goodman MM, Sanchez MM, Buckler E, Doebley J. A single domestication for maize shown by multilocusmicrosatellite genotyping. Proc Natl Acad Sci. 2002;99(9):6080–4.
9. van Heerwaarden J, Doebley J, Briggs WH, Glaubitz JC, Goodman MM, Sanchez GJ, Ross-Ibarra J. Genetic signals of origin, spread and introgression in a large sample of maize landraces. Proc Natl Acad Sci. 2011;108(3):1088–92.
10. Wilkes HG. Teosintes distribution in Mexico. In: Serratos JA, Wilcox MC, Castillo F, editors. Proceedings of a forum. Mexico: CIMMyT; 1995. p. 10–39.
11. Fukunaga K, Hill J, Vigouroux Y, Matsuoka Y, Sanchez GJ, Liu K, Buckler E, Doebley J. Genetic diversity and population structure of teosintes. Genetics. 2005;169(4):2241–54.
12. Iltis HH, Benz BF. *Zea nicaraguensis* (Poaceae): a new teosinte from Pacific Coastal Nicaragua. Novon. 2000;10(4):382–90.
13. Benz BF, Sanchez-Velasquez LR, Santana MFJ. Ecology and ethnobotany of *Zea diploperennis*: preliminary investigations. Maydica. 1990;35(2):85–98.
14. Sanchez GJJ, De la Cruz L, Vidal MVA, Ron PJ, Taba S, Santacruz-Ruvalcaba F, Sood S, Holland JB, Ruiz CJA, Carvajal S, Aragon CF, Chavez TVH, Morales RMM, Barba-Gonzalez R. Three new teosintes (*Zea* spp., Poaceae) from Mexico. Am J Bot. 2011;98(9):1548–73.
15. Ross-Ibarra J, Tenaillon M, Gaut B. Historical divergence and gene flow in the genus *Zea*. Genetics. 2009;181(4):1399–413.
16. Hufford MB, Lubinsky P, Pyhäjärvi T, Devegenzo MT, Ellstrand NC, Ross-Ibarra J. The genomic signature of crop-wild introgression in maize. PLoS Genet. 2013;9(5):e1003477.
17. Vigouroux Y, Mitchell S, Matsuoka Y, Hamblin M, Kresovich S, Smith JS, Jaqueth J, Smith OS, Doebley J. An analysis of genetic diversity across the maize genome using microsatellites. Genetics. 2005;169(3):1617–30.
18. Loáisiga CH, Rocha O, Brantestam AK, Salomon B, Merker A. Genetic diversity and gene flow in six accessions of Meso-America teosintes. Gen Resour Crop Evol. 2012;59(1):95–111.
19. Doebley JF, Goodman MM, Stuber CW. Isoenzymatic variation in *Zea* (Gramineae). Syst Bot. 1984;9(2):203–18.
20. Sanchez G JJ, Kato TA, Aguilar M, Hernandez JM, Lopez A, Ruiz JA. Distribución y caracterización del teocintle. Guadalajara: Instituto Nacional de Investigaciones Forestales, Agrícolas y Pecuarias; 1998.
21. Buckler ES, Goodman MM, Holtsford TP, Doebley JF, Sanchez GJ. Phylogeography of the wild subspecies of Zea mays. Maydica. 2006;51(1):123–34.
22. Merino-Díaz G. Genética de poblaciones de dos subespecies de maíz (*Zea mays* ssp. *mexicana* y *Zea mays* ssp. *parviglumis*). Master thesis, Departamento de Ecología Evolutiva, Instituto de Ecología, UNAM, Mexico D.F., México; in preparation.
23. Eguiarte LE, Aguirre-Planter E, Aguirre X, Colin R, Gonzalez A, Rocha M, Scheinvar E, Trejo L, Souza V. From isozymes to genomics: population genetics and conservation of Agave in Mexico. Bot Rev. 2013;79(4):483–506.
24. Weir BS, Cockerham CC. Estimating F-statistics for the analysis of population structure. Evolution. 1984;38(6):1358–70.
25. Holsinger KE, Lewis PO. Hickory: a package for analysis of population genetic data, version 1.0. Storrs; University of Connecticut; 2003.
26. Holsinger KE, Lewis PO. Computer program and documentation. Storrs: University of Connecticut; 2007.

27. Hamrick JL, Godt MJW. Effects of life history traits on genetic diversity in plant species. Philos Trans R Soc London. 1996;351(1345):1291–8.
28. Nybom H. Comparison of different nuclear DNA markers for estimating intraspecific genetic diversity in plants. Mol Ecol. 2004;13(5):1143–55.
29. Foll M, Gaggiotti O. A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: A Bayesian perspective. Genetics. 2008;180(2):977–93.
30. Villasante-Barahona A. Diversidad y estructura genética del teocintle anual Zea mays ssp. *parviglumis* Iltis & Doebley y *Zea mays* ssp. *mexicana* (Schrader) Iltis, en un gradiente geográfico y ambiental. Mexico D.F.: Facultad de Ciencias, UNAM; 2013.
31. Parker KC, Trapnell DW, Hamrick JL, Hodgson WC, Parker AJ. Inferring ancient Agave cultivation practices from contemporary genetic patterns. Mol Ecol. 2010;19(8):1622–37.
32. Moeller DA, Tenaillon M, Tiffin P. Population structure and its effects on patterns of nucleotide polymorphism in teosinte (Zea mays ssp. parviglumis). Genetics. 2007;176(3): 1799–809.
33. van Heerwaarden J, Ross-Ibarra J, Doebley J, Glaubitz JC, Sanchez GJJ, Gaut BS, Eguiarte LE. Fine scale structure in the wild ancestor of maize (*Zea mays* ssp. *parviglumis*). Mol Ecol. 2010;19(6):1162–73.
34. Pyhäjärvi T, Hufford M, Mezmouk S, Ross-Ibarra J. Complex patterns of local adaptation in teosinte. Genome Biol Evol. 2013;5(9):1594–609.
35. Slatkin M, Voelm L. F_{ST} in a hierarchical island model. Genetics. 1991;127(3):627–9.
36. Fang Z, Pyhäjärvi T, Weber AL, Dawe RK, Glaubitz JC, Sanchez GJJ, Ross-Ibarra C, Doebley J, Morrell PL, Ross-Ibarra J. Megabase-scale inversion polymorphism in the wild ancestor of maize. Genetics. 2012;191(3):883–94.
37. Chia JM, Song C, Bradbury PJ, Costich D, de Leon N, Doebley J, Elshire RJ, Gaut B, Geller L, Glaubitz JC, Gore M, Guill KE, Holland J, Hufford MB, Lai J, Li M, Liu X, Lu Y, McCombie R, Nelson R, Poland J, Prasanna BM, Pyhäjärvi T, Rong T, Sekhon RS. Maize HapMap2 identifies extant variation from genome in flux. Nat Genet. 2012;44(7):803–7.
38. McClintock B. The origin and behavior of mutable loci in maize. Proc Natl Acad Sci U S A. 1950;36(6):344–55.
39. Cutter A, Jovelin R, Dey A. Molecular hyperdiversity and evolution in very large populations. Mol Ecol. 2013;22(8):2074–95.
40. Tenaillon MI, Hufford MB, Gaut BS, Ross-Ibarra J. Genome size and transposable element content as determined by high-throughput sequencing in Maize and *Zea luxurians*. Genome Biol Evol. 2011;3:219–29.
41. Gaut BS, Ross-Ibarra J. Selection on major components of angiosperm genomes. Science. 2008;320(5875):484–6.
42. Diez CM, Gaut BS, Meca E, Scheinvar E, Montes-Hernandez S, Eguiarte L, Tenaillon MI. Genome size variation in wild and cultivated maize along altitudinal gradients. New Phythol. 2013;199(1):264–76.
43. Ross-Ibarra J, Morrel PL, Gaut BS. Plant domestication, a unique opportunity to identify the genetic basis of adaptation. Proc Natl Acad Sci. 2007;104 Suppl 1:8641–8.
44. Hufford MB, Xu X, van Heerwaarden J, Pyhäjärvi T, Chia JM, Cartwright RA, Elshire RE, Glaubitz JC, Guill KE, Kaeppler SM, Lai J, Morrell PL, Shannon LM, Song C, Springer NM, Swanson-Wagner RA, Tiffin P, Wang J, Zhang G, Doebley J, McMullen MD, Ware D, Buckler ES, Yang S, Ross-Ibarra J. Comparative population genomics of maize domestication and improvement. Nat Gen. 2012;44(7):808–11.
45. Swanson-Wagner R, Briskine R, Schaefer R, Hufford MB, Ross-Ibarra J, Myers CL, Tiffin P, Springer NM. Reshaping of the maize transcriptome by domestication. Proc Natl Acad Sci. 2012;109(29):11878–83.

CAPÍTULO 2: DISTRIBUCIÓN GEOGRÁFICA Y ECOLÓGICA DE LA ADAPTACIÓN LOCAL EN TEOCINTLE

Artículo aceptado en Molecular Ecology: Connecting genomic patterns of local adaptation and niche suitability in teosinte

Received: 19 September 2016 | Revised: 14 April 2017 | Accepted: 1 June 2017

DOI: 10.1111/mec.14203

ORIGINAL ARTICLE

WILEY MOLECULAR ECOLOGY

Connecting genomic patterns of local adaptation and niche suitability in teosintes

J. A. Aguirre-Liguori[1] | M. I. Tenaillon[2] | A. Vázquez-Lobo[3] | B. S. Gaut[4] | J. P. Jaramillo-Correa[1] | S. Montes-Hernandez[5] | V. Souza[1] | L. E. Eguiarte[1]

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

[2]Génétique Quantitative et Evolution - Le Moulon, INRA, Univ. Paris-Sud, CNRS, AgroParisTech, Université Paris-Saclay, Gif-sur-Yvette, France

[3]Centro de Investigación en Biodiversidad y Conservación, Universidad Autónoma del Estado de Morelos, Cuernavaca, Morelos, México

[4]Department of Ecology and Evolutionary Biology, UC Irvine, Irvine, CA, USA

[5]Campo Experimental Bajío, Instituto Nacional de Investigaciones Forestales, Agrícolas y Pecuarias, Celaya, Guanajuato, Mexico

Correspondence
Luis E. Eguiarte, Departamento de Ecología Evolutiva, Instituto de Ecología, Universidad Nacional Autónoma de México, Mexico City, Mexico.
Email: fruns@unam.mx

Funding information
Consejo Nacional de Ciencia y Tecnología, Grant/Award Number: 255770; CONACYT Investigación Científica Básica, Grant/Award Number: CB2011/167826; UC MEXUS-CONACYT, Grant/Award Number: CN-10-393; SEP-CONACYT-ANUIES-ECOS Nord France, Grant/Award Number: M12-A03, CONACYT-ANUIES 207571

Abstract

The central abundance hypothesis predicts that local adaptation is a function of the distance to the centre of a species' geographic range. To test this hypothesis, we gathered genomic diversity data from 49 populations, 646 individuals and 33,464 SNPs of two wild relatives of maize, the teosintes *Zea mays* ssp. *parviglumis* and *Zea. mays.* ssp. *mexicana*. We examined the association between the distance to their climatic and geographic centroids and the enrichment of SNPs bearing signals of adaptation. We identified candidate adaptive SNPs in each population by combining neutrality tests and cline analyses. By applying linear regression models, we found that the number of candidate SNPs is positively associated with niche suitability, while genetic diversity is reduced at the limits of the geographic distribution. Our results suggest that overall, populations located at the limit of the species' niches are adapting locally. We argue that local adaptation to this limit could initiate ecological speciation processes and facilitate adaptation to global change.

KEYWORDS
central abundance hypothesis, conditional neutrality, ecological speciation, local adaptation, niche centroid, *Zea mays*

1 | INTRODUCTION

The dynamics of spatial heterogeneity, reduced migration and differential selection contribute to establish local adaptation (Leimu & Fisher, 2008; Schluter, 2000, 2001). In recent years, studies of local adaptation have been stimulated by the availability of both new statistical tools (e.g., Beaumont & Nichols, 1996; Coop, Witonsky, Di Rienzo, & Pritchard, 2010; Foll & Gaggiotti, 2008; Joost et al., 2007) and high-throughput genotyping technologies (e.g., Hancock et al., 2011; Hohenlohe et al., 2010; Turner, Bourne, Von Wettberg, Hu, & Nushdin, 2010). The results from these approaches have highlighted the importance of conditional neutrality, whereby alleles are adaptive in some populations but neutral in others (Anderson, Inouye, McKinney, Colautti, & Mitchell-Olds, 2012; Anderson, Willis, & Mitchell-Olds, 2011; Schnee & Thompson, 1984) and of antagonistic pleiotropy, whereby different alleles have contrasting selective effects in different habitats (Anderson et al., 2012; Kawecki & Ebert, 2004; Savolainen, Lascoux, & Merilä, 2013).

The range of ecological conditions a species can experience, that is, the species niche, is a function of its geographic distribution.

Conversely, the limits of its geographic distribution are dictated by abiotic (i.e., climate, soil characteristics) and biotic conditions (i.e., mutualistic species, competitors, pathogens) where it can survive and reproduce in nature. The central abundance hypothesis (CAH) formalizes the relationships among geographic distribution, ecological conditions and the adaptation of populations (Eckert, Samis, & Lougheed, 2008; Hengeveld & Haeck, 1982; Sexton, McIntyre, Angert, & Rice, 2009). The CAH predicts that populations at geographic limits *i)* experience biotic and abiotic conditions that are divergent from the centre of the niche, *ii)* have smaller effective population sizes, and in turn less genetic variation than more central populations, and therefore *iii*) are the most likely to become extinct (Bridle & Vines, 2006). However, as edge populations often grow in marginal or distinct environments relative to the range centre, an alternative scenario is that these populations could adapt to the environmental challenges (Eckert et al., 2008; Sexton et al., 2009).

Several experimental observations contradict the CAH predictions. For example, Sagarin and Gaines (2002a) found that abundance was not associated with the centre of the distribution in 89 of 145 studies (61%). The fact that diversity and distance to the geographic centre does not correlate may relate to a decoupling between the limits of the geographic distribution and the environmental limits of a species (Sagarin & Gaines, 2002a,b; Sagarin, Gaines, & Gaylord, 2006). Hence, it has been proposed that the limits of a species' range are more accurately defined by environmental conditions, rather than geographic distance (Lira-Noriega & Manthey, 2014; Martínez-Meyer, Díaz-Porras, Peterson, & Yañez-Arenas, 2013; Yañez-Arenas, Martínez-Meyer, Mandujano, & Rojas-Soto, 2012). From this point of view, suitability can be estimated as a combination of environmental variables in which populations exhibit positive growth; its maximum being at the centre of the niche distribution—the centroid (Hutchinson, 1978; Maguire, 1973). The distance to the niche centroid (not to the geographic centre) thus becomes a measure of environmental suitability, with populations further away growing in less suitable conditions than those growing at the centre of the niche. This measure has been used to test predictions of the CAH. Consistent with the CAH, Martínez-Meyer et al. (2013) found a negative correlation between population size and distance to the niche centroid for seven species of mammals. Moreover, Lira-Noriega and Manthey (2014) observed the same trend with genetic diversity taken as a proxy for population size for 45 taxa, including insects, plants, birds and mammals. But there is still an open question: Do populations at niche limits tend to be hotspots for local adaptation or for local extinction (Eckert et al., 2008; Sexton et al., 2009)?

In this study, we assembled a large data set of 49 populations from two subspecies of teosintes, *Zea mays* ssp. *parviglumis* (hereafter *parviglumis*) and *Zea mays* ssp. *mexicana* (hereafter *mexicana*), to test hypotheses about local adaptation in relation to the niche centroid. *Parviglumis* and *mexicana* are two subspecies of annual plants found in central Mexico and are the closest wild relatives of domesticated maize, *Zea mays* ssp. *mays*. They occupy distinct ecological niches; *mexicana* grows in the central highlands of Mexico (1500–2800 m) under cooler and drier conditions than *parviglumis*, which grows in the west coast lowlands (300–1900 m) under warmer and more humid conditions (Fukunaga et al., 2005; Hufford, Martínez-Meyer, Gaut, Eguiarte, & Tenaillon, 2012). This ecological differentiation results in a reduced area of overlap (Hufford et al., 2012), and reduced gene flow between subspecies (Aguirre-Liguori et al. unpublished data). Because teosintes' geographic and environmental ranges are well defined, and local adaptation is widespread within the two subspecies, given the complex landscape of Mexico (Pyhäjärvi, Hufford, Mezmouk, & Ross-Ibarra, 2013a,b), teosintes are ideal models to study local adaptation along their geographic and ecologic ranges and verify CAH predictions.

To verify CAH predictions, we used combined genotyping data at 33,464 biallelic SNP markers and environmental information. We identified SNPs putatively involved in local adaptation along the geographic and ecological distribution of each subspecies and tested the predictions of the CAH by analysing the patterns of genetic diversity and the number of putatively adaptive SNPs as a function of the distance to the niche and geographic centroids. Our results reveal that while genetic diversity is reduced in populations further away from the geographic centroid, signatures of local adaptation seem to increase in populations located at the edge of the niche, where conditions are less suitable.

2 | MATERIALS AND METHODS

2.1 | Sampling and genotyping

We assembled seeds from 12 to 15 individuals from 16 of the populations described in Díez et al. (2013) and from 12 additional locations (Table S1 and Fig. S1). We collected seeds from random plants along the entire population, to assure obtaining a good genetic representation. These populations were selected to cover the geographic and environmental range of both *mexicana* and *parviglumis*. Seeds were sown in a greenhouse, and leaf samples were harvested 3 weeks after germination. DNA was isolated using a modified version of the CTAB extraction method (Doyle & Doyle, 1987) and RNase treated.

After verifying DNA integrity (agarose gel at 0.8% and NanoDrop Lite (Thermo Scientific)), concentrations were set to a 50 ng/μl for genotyping with the MaizeSNP50 Genotyping BeadChip on the Infinium platform available at the Genome Center of the University of California, Davis, USA. Genotype calling was conducted using the Genotyping Module v1.0 of Genome Studio (Illumina) with a threshold of 0.15 for the GC_{50} scores. A visual inspection of clusters was conducted to control for quality, filtering monomorphic data and clusters with a call rate <0.85.

We expanded the data set by including data published by Pyhäjärvi et al. (2013a,b), which includes 11 populations of *parviglumis* and 10 populations of *mexicana*. Overall, we analysed 49 populations, 646 individuals and 33,464 SNPs (Table S2, Data are available from the Dryad Digital Repository https://doi.org/10.5061/dryad.tf556).

2.2 | Genetic diversity and differentiation

We examined the genetic structure of populations through a principal component analyses (PCA) using the package Adegenet (Jombart,

2008) implemented in R 3.02 (R Development Core Team 2008). We then followed the approximation of Pyhäjärvi et al. (2013a,b) to define the number of genetic clusters in our entire sample from the first three PCs, which were plotted using the *colorplot* function in Adegenet. Briefly, we used the Ward clustering algorithm performed by the *hclust* function in R, and the *cutree* function to group populations by genetic similarity. Then, for each population and genetic group, we calculated standard genetic population parameters (H_S, F_{IS}) and calculated pairwise Nei's (1972) genetic distances using the R package Hierfstat (Goudet, 2005) and Adegenet, respectively. Genetic distances were estimated between populations within subspecies and between populations for the entire sample.

2.3 | Distance to geographic and environmental (climatic) niche centroid

To obtain the distances to the geographic and climatic niche centroids, we followed the method of Lira-Noriega and Manthey (2014). First, we modelled the potential distribution for *mexicana* and *parviglumis* using MAXENT 3.3 (Phillips, Anderson, & Schapire, 2006) as in Hufford et al. (2012). To do so, we first downloaded a database containing the geographic coordinates of 254 populations of *mexicana* and 329 localities of *parviglumis* from the Comisión Nacional de Biodiversidad (CONABIO 2015; http://www.conabio.gob.mx/). Then, for each coordinate, we retrieved 19 bioclimatic variables from the WORLDCLIM database (Hijmans, Cameron, Parra, Jones, & Jarvis, 2005) at a resolution of 30 arcsec, which were incorporated into the model of each taxa.

Models were validated with 10 bootstrap replicates and 30 per cent of occurrence records. Further validation included the use of the area under the curve (AUC) of the receiver operating characteristic (ROC) plot (Fielding & Bell, 1997). Each bootstrapped replicate was then converted into a binary model by removing all the distribution area that had probability values with omission rate of the training and testing data under 5%. We obtained the final consensus map of the distribution of each teosinte by adding up the 10 bootstrapped replicates and retaining only the areas that were predicted by at least five of them (Hufford et al., 2012).

For determining the geographic centroid, we first used the R package Raster (Hijmans & van Etten, 2015) to extract the entire coordinates of the consensus maps. Second, we determined the geographic centroid as the median value of the latitude and longitude. Third, for each of the sampled populations of *mexicana* and *parviglumis*, we calculated their Euclidian distance to their specific geographic centroid. Similarly, for the distance to the niche centroids of each taxa, we used Raster to extract 19 bioclimatic layers from each grid of the consensus maps. Then, using the *prcomp* function of R, we performed a PCA analysis to reduce the environmental information into four principle components (PC) that were used to define the environmental niche of *mexicana* and *parviglumis* (Fig. S2). The niche centroid was defined as the mean value of each PC for *mexicana* and *parviglumis*, respectively. Finally, for each population in each panel, we calculated their multivariate Euclidian distance to the niche centroid as in Lira-Noriega and Manthey (2014).

2.4 | Outlier SNP detection

Selection analyses were performed on each subspecies separately. Such partitioning was motivated because of differences in ecological requirements across teosintes and their genetic differentiation (Fukunaga et al., 2005; Hufford et al., 2012), which increases the rate of false positives in neutrality tests (De Mita et al., 2013; Lotterhos & Whitlock, 2014). Following the same rationale, we removed from the *mexicana* panel the Nabogame and Puerta Encantada populations (Fig. S1); the former is strongly isolated and was probably originated by seeds moved by Spaniards to use the plants as cattle feed (Lumholtz, 1902), while the latter grows well below its normal altitude (Fukunaga et al., 2005) and has unusual phenology (Wilkes, 1997). Wilkes (1997) actually suggested that Puerta Encantada populations might have been grown in ancient pre-Columbian botanical gardens, making its origin unclear. For detecting patterns of selection, we therefore restricted the analyses to 23 *mexicana* populations encompassing 309 individuals, and 24 *parviglumis* populations encompassing 313 individuals. For the *parviglumis* panel, we included intermediate populations from Guerrero (see Pyhäjärvi et al., 2013a, b), as they have an extended distribution and group within *parviglumis* (see Results). As all populations have a similar sample size 12–16, we do not expect to find errors associated to sampling.

To identify SNPs exhibiting deviation from neutral expectations (and thus potentially associated to adaptive processes), we applied two outlier detection tests. First, we used Bayescenv1.1 (de Villemereuil & Gaggiotti, 2015) within the *mexicana* and *parviglumis* panels separately. Bayescenv uses a Bayesian framework to determine the expected genomewide F_{ST} value given by an overall population-specific parameter (β) shared by all loci. Inclusion of a locus-specific parameter (α) to account for a SNP-specific F_{ST} value is further tested. Then, an environmental parameter (γ) is included to account for environmental variation associated to this SNP. If for a given locus the posterior probability of the γ model is higher than the posterior probability of the α model, it is then assumed that environmental and allelic frequency variations are associated. By contrasting the γ and α model, Bayescenv compares outlier loci affected by nonenvironmental and environmental factors, and therefore reduces considerably the number of false positives (de Villemereuil & Gaggiotti, 2015). We ran Bayescenv for each subspecies and used the first two PC of the environmental models (which explained the highest percentage of the variance; see Table S3 and Results) to test the γ models. To run Bayescenv, we used the following parameters: $\pi = 0.1$, $p = .5$, the upper bound γ was set at 10 and $\alpha = -1$. For each analyses, we ran 20 pilot of 5,000 iterations each, and a burn-in of 50,000 iterations. We sampled 5,000 MCMC iterations. At the end, we retained all SNPs with posterior error probability (PEP) of γ lower than the PEP α. Because Bayescenv relies on genomewide F_{ST} to define the null model, it implicitly accounts for ascertainment bias inherent to the use of the MaizeSNP50 Genotyping BeadChip

designed on a restricted number of samples (Albrechtsen, Nielsen, & Nielsen, 2010).

We also used Bayenv 2.0 (Coop et al., 2010) to identify SNPs that show a strong correlation between allelic frequencies and the first two PC of the niche model. Bayenv estimates in a first step the covariance matrix between populations, and then uses this matrix to test the correlation of each SNP with environment, while controlling for genetic structure. First, we used a set of 10,000 random SNPs to obtain the covariance matrix. We ran Bayenv 2.0 for 100,000 iterations and saved the covariance matrix every 500 iterations. To test the confidence of the covariance matrix, we obtained the correlation between random covariance matrices and the final matrix, and between the final covariance matrix and the pairwise F_{ST} matrix obtained from the R Package BEDASSLE (Bradburg, Ralph, & Coop, 2013), to assure that there were no outliers in the matrix. Finally, for each SNP in *mexicana* and *parviglumis*, we ran 100,000 iterations of Bayenv to test the correlation between their allelic frequencies and the two environmental PC. At the end, we retained all SNPs that had Bayes factors in the 95th percentile of the distribution (see Pyhäjärvi et al., 2013a,b).

Finally, for each subspecies and the two first PC of the environmental variation, we defined the final set of "outlier" SNP as all those loci that were detect as outlier by both Bayenv and Bayescenv.

2.5 | Selection analysis along environmental clines

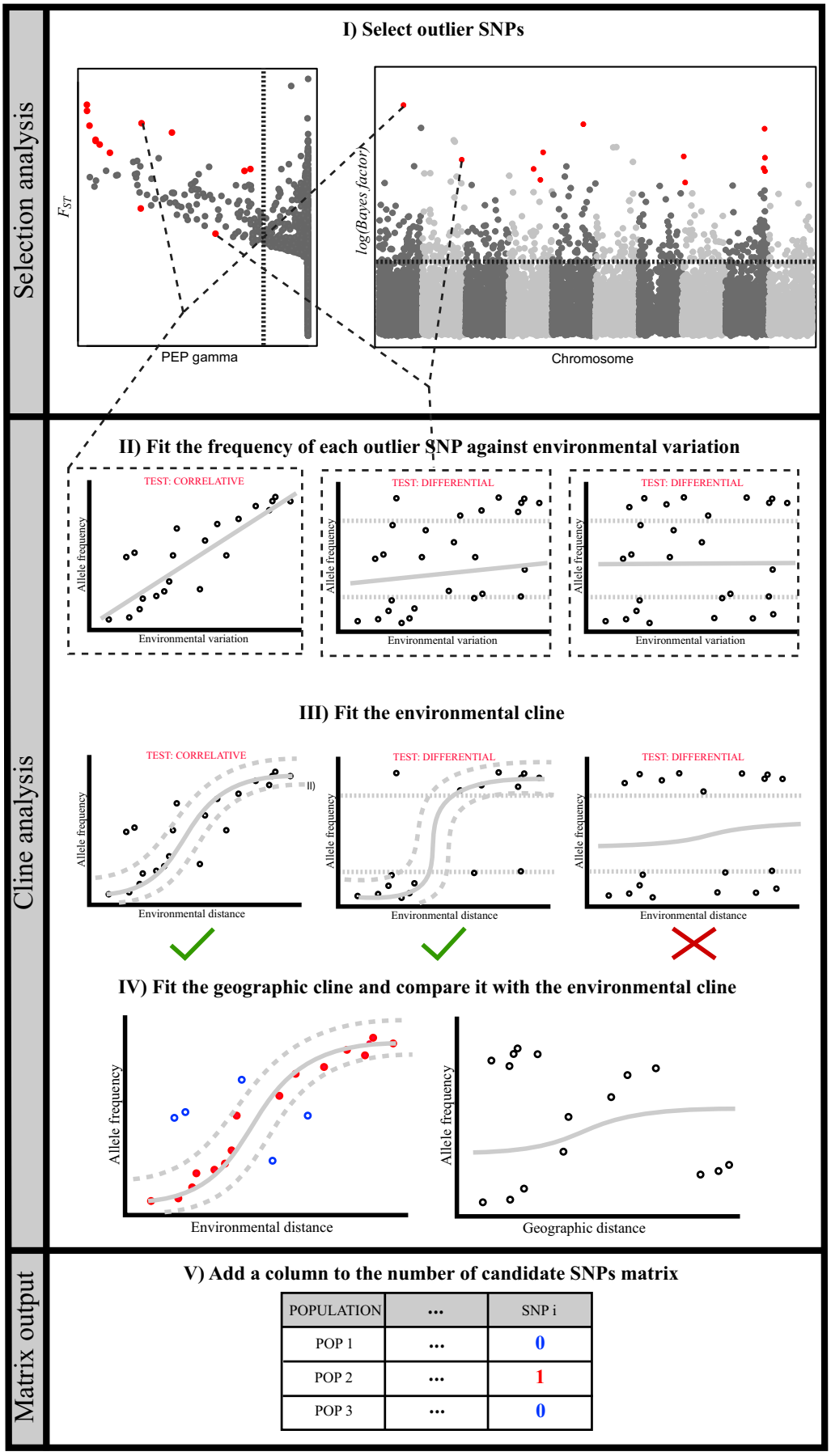
Bayescenv and Bayenv detect an overall association among populations of their allele frequency variation and environmental variables, but they are not designed to detect specific locus-by-population effects (Beaumont & Balding, 2004). Locus-by-population effects may occur for loci evolving under conditional neutrality (Anderson et al., 2011, 2012), whereby an allele may be selected for in some populations but evolve neutrally in others. Moreover, locus-by-population effects are difficult to infer due to possible effects of genetic drift driving alleles to high frequencies. In order to define locus-by-population effects, while controlling for genetic drift, we performed additional environmental and geographic cline analyses on the set of outlier SNPs detected by both Bayescenv and Bayenv to identify such situations.

Natural selection can usually be differentiated as correlative or differential. In the case of correlative selection, the frequency of an adaptive allele changes gradually with the environment (Fig. S3a; Joost et al., 2007; Coop et al., 2010), resulting in a soft environmental cline. In the case of differential selection, allele frequencies of an adaptive locus reach contrasting frequencies (towards $p = 0$ or $p = 1$) as a result of divergent selective pressures (Excoffier, Hofer, & Foll, 2009; Lewontin & Krakauer, 1973), that result in steep environmental clines (Fig. S3b; Savolainen et al., 2013). In order to differentiate between correlative and differential selection, as well as to define locus-by-population effects, we followed the approach illustrated in Figure 1. For each outlier SNP (Figure 1I), we first tested whether it was under putative correlative selection (Figure 1II). For this, we performed a linear regression model for each outlier SNP by contrasting the arcsine square root of the allele frequencies (as loci are usually not normally distributed) of each population, to the first two PC scores obtained from the niche model. When the regression was significant ($p < .01$), we considered the SNP to be putatively affected by correlative selection, and retained all populations to pursue the cline analyses. A lack of significance indicates either the absence of cline (Figure 1II-III—right plots) or a cline driven by populations with extreme allele frequencies, that is, differential selection (Figure 1II-III—middle plots) (Excoffier, Hofer, et al., 2009; Lewontin & Krakauer, 1973; Savolainen et al., 2013). To examine the latter, we discarded populations with intermediate allele frequencies—between 0.2 and 0.8 (Figure 1III, middle and right plots), and retained populations with extreme allelic frequencies to perform the cline analyses. Our rationale is that under differential selection, allelic frequencies should be extreme, and populations with intermediate frequencies should not be affected by selection (i.e., conditional neutrality), and thus could affect the estimation of the cline.

In the next step, we used the R package HZAR (Derryberry, Derryberry, Maley, & Brumfield, 2014) to fit models of clinal variation of allele frequencies with the two first PCs of the distribution model. To do so, we fitted an environmental cline over all the populations retained after the regression analysis (Figure 1III)—all populations for loci evolving under correlative selection and a restricted set of populations for loci evolving under differential selection. We first used the AIC values to compare the null model (where no cline is fitted) and the environmental cline model. When the latter was supported, we calculated the absolute distance (AD) between the observed frequency of the allele in the population and the HZAR predicted cline frequency. Then, we removed any populations that had a frequency outside the 75th percentile of the AD distribution, to retain populations that fitted best the cline.

Finally, using these populations, we fitted both a geographic and an environmental cline. For a given SNP, if the geographic cline had

FIGURE 1 Cline analysis pipeline (See Methods for more details). I) Identification of outlier SNPs using Bayescenv and Bayenv. II) Linear regression of outliers' allele frequency on environmental variation to determine whether SNPs better fit correlative or differential selection. III) Model of cline of allele frequency against environmental variation and associated confidence intervals (CI), after removal of extreme populations for SNPs evolving under differential selection. SNPs with no significant fit to a cline are discarded (right plot). IV) Only SNPs that better fit an environmental than a geographic cline are retained. For those candidate adaptive SNPs, we distinguish between populations in which the SNP is evolving neutrally—populations outside the cline CI, blue dots—versus those for which the SNP is likely selected—populations contained within the cline CI, red dots. V) Construction of the matrix of the number of candidate adaptive SNPs per population. If the SNP fits the environmental cline, we add a column to the matrix indicating for which populations the SNP was adaptive (1) or neutral (0). The sum of columns gives the number of candidate SNPs


Selection analysis
I) Select outlier SNPs
F_{ST}
PEP gamma
log(Bayes factor)
Chromosome
Cline analysis
II) Fit the frequency of each outlier SNP against environmental variation
TEST: CORRELATIVE
TEST: DIFFERENTIAL
TEST: DIFFERENTIAL
Allele frequency
Environmental variation
III) Fit the environmental cline
TEST: CORRELATIVE
TEST: DIFFERENTIAL
TEST: DIFFERENTIAL
Allele frequency
Environmental distance
IV) Fit the geographic cline and compare it with the environmental cline
Allele frequency
Environmental distance
Allele frequency
Geographic distance
Matrix output
V) Add a column to the number of candidate SNPs matrix
POPULATION
...
SNP i
POP 1
...
0
POP 2
...
1
POP 3
...
0

an AIC value lower than the environmental cline, we considered allele frequencies as being affected by isolation by distance or gene surfing, rather than by environmental conditions, and therefore discarded it. In contrast, when the environmental cline had a lower AIC value (Figure 1IV), we considered the SNP as being affected by environment and adaptive in this set of populations, but neutral in the discarded populations (red and blue dots for selected and neutral SNPs, respectively, in Figures 1IV and S4). The number of candidate adaptive SNPs per population was determined by summing the number of SNPs that fitted the environmental cline in each population (Figure 1V; Table S1). With this data set, we tested the CAH for each panel (*mexicana* and *parviglumis*) by performing linear regression analyses to evaluate the following relationships 1) between the genetic diversity (H_S, all SNPs) and the distance to the geographic centroid, 2) the genetic diversity (H_S, all SNPs) and the distance to the niche centroid, 3) the number of candidate SNPs in populations and the distance to the geographic centroid and 4) the number of candidate SNPs in populations and the distance to the niche centroid.

According to the rationale described above, the number of adaptive loci in a given population should depend on the intensity of selection occurring in the population. However, it is also possible that the number of adaptive SNPs could result from an artefact due to different probabilities of a locus being considered as adaptive in each population. As we are using cline analyses, it is possible that some populations might have a higher probability of fitting the cline (i.e., extreme populations) than others. Also, as we are using environmental differentiation to test for SNPs under selection, and then test whether those SNPs are in the most differentiated environments, there could be a circularity in the approximation. To verify these potential biases, we performed four tests (see Fig. S5, and Supporting information for more details on the methods and results of these tests). First, we determined the number of outlier SNPs fitting a geographic cline, instead of the environmental cline (Test 1). Second, we tested the number of neutral SNPs fitting the environmental clines described above (Test 2). If any of these numbers was similar to the one of the candidates retained above (i.e., those fitting the environmental clines), we could suspect a statistical bias. For the third test (Test 3), instead of fitting clines to call a SNP adaptive or neutral in a given population, we defined as adaptive SNPs those that were nearly fixed in these populations ($p < .2$ or $>.8$) and tested for the CAH. In this case, we expected to have similar results as those obtained with the cline-fitting test. However, as fixation of allelic frequencies could also occur by genetic drift, we tested the same analyses, but using 50 replicates of random neutral SNPs to compare outlier SNPs and neutral SNPs. If genetic drift is not generating the pattern, then we would expect to find a nonsignificant association with the distance to the niche centroid for neutral SNPs. Finally, for the fourth test, we identified outlier SNP and locus-by-population effects (following the same approximation as in Figure 1) but randomizing the environmental values of populations. If there is not a circularity bias, we would expect to find no association between the distance to niche randomized environmental centroid and the number of candidate SNPs.

3 | RESULTS

3.1 | Genetic diversity and population structure

After filtering for high-quality data and including genotypes from Pyhäjärvi et al. (2013a,b), we obtained a complete data set of 49 populations, 646 individuals and 33,464 SNPs (Fig. S1 and Table S2). SNPs were distributed along the 10 chromosomes of maize, with numbers ranging from 2,032 SNPs on chromosome 10 to 5,275 SNPs on chromosome 1. The median distance between two consecutive SNPs varied from 7.4 Kbp on chromosome 6 to 17.7 Kbp on chromosome 4.

After performing a PCA and a Ward clustering algorithm to determine the relatedness among populations, we found two distinct clusters separating both subspecies, *mexicana* and *parviglumis*. This last taxon was further divided into three genetic groups (thereafter GI to GIII) (Figure 2). Nei's pairwise genetic distance between individual populations varied from 0.01 to 0.25 (Tables S4 and S5). Genetic diversity (H_S) was similar between subspecies ($H_{S\ mexicana} = 0.225$, $SD = 0.02$; $H_{S\ parviglumis} = 0.226$, $SD = 0.05$), although significant variation was observed across populations of the same subspecies (Figure 3a–b, Table S1). Within *parviglumis*, GI had the highest H_S (0.25; $SD = 0.03$) and GIII the lowest (0.12; $SD = 0.03$) (Table S1). F_{IS} values, which reflects the degree of inbreeding, were low within all groups, but also exhibited variation among populations (Table S1), reaching values as high as 0.21.

3.2 | Environmental information

Maxent models of climatic niche for both subspecies had high AUC values and small standard deviations (*parviglumis*: AUC = 0.969, $SD = 0.002$; *mexicana*: AUC = 0.980, $SD = 0.002$; Figs S6–S7), indicating good performance during model construction. We used the final models to obtain the distances to the geographic and the niche centroids for each population (Figure 3). Table S3 shows the percentage of the variation explained by the first four climate PCs for both *mexicana* and *parviglumis*, as well as the climatic variables contributing to each PC. In both cases, PC1 (*mexicana*: 50.72%, *parviglumis*: 38.7%) was related to temperature and PC2 (*mexicana*: 21.95%, *parviglumis*: 25.0%) to precipitation. We observed no significant linear association between the distances to the geographic and niche centroids for any taxa (*mexicana*: $p = .28$; *parviglumis* $p = .36$; Fig. S8).

3.3 | Selection analysis along environmental clines

We performed two outlier analyses, Bayescenv (de Villemereuil & Gaggiotti, 2015) and Bayenv (Coop et al., 2010), which rely on different assumptions (De Mita et al., 2013). Bayescenv relies in the detection of SNP that present outlier F_{ST} and that have a strong

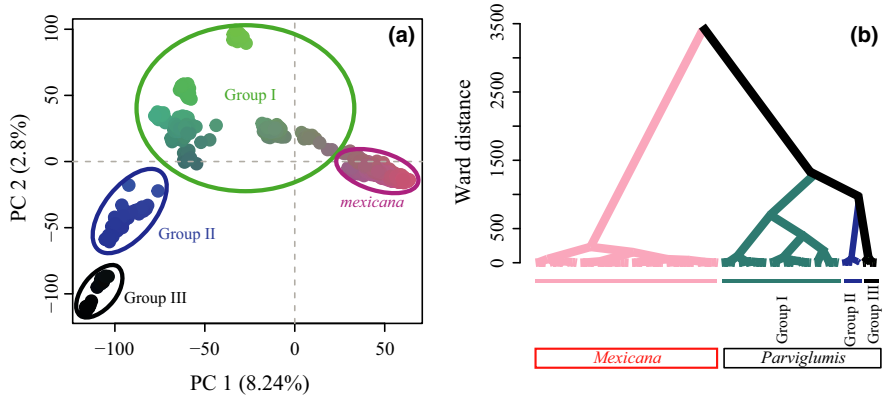


FIGURE 2 (a) PCA analysis of 33,464 SNP genotypes. The third PC is used to define the colours, which indicate genetic similarity between populations. (b) Ward Clustering of populations based on the first three PCs obtained from the PCA. According to colours and the Ward clustering analyses, we defined four groups. *mexicana* populations are in pink. Genetic group I consists of *parviglumis* populations from Guerrero, Oaxaca, Michoacan and Jalisco states. Genetic groups II and III correspond to other Jalisco populations

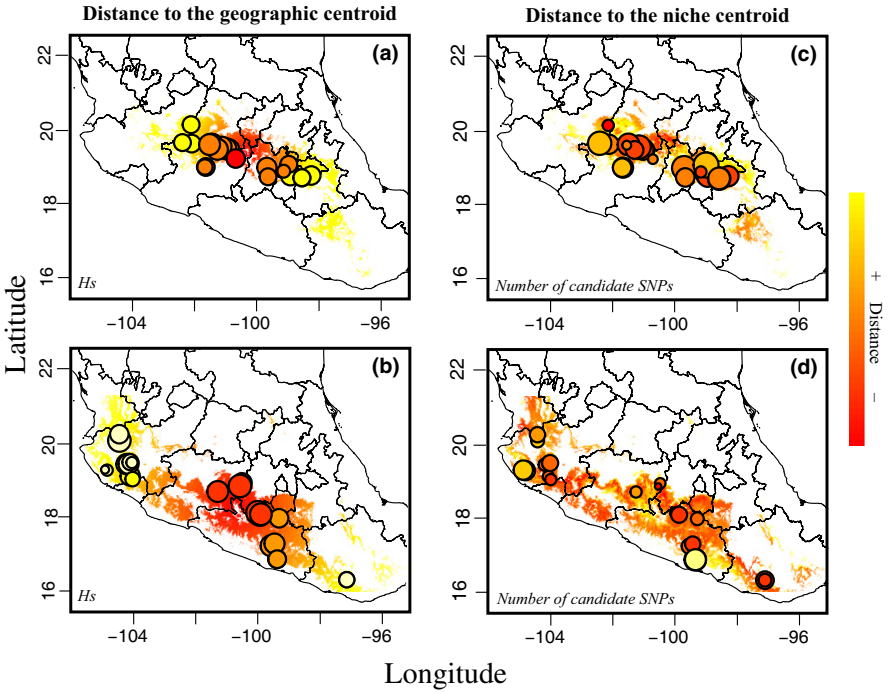


FIGURE 3 The potential distribution maps representing the distance to the geographic centroid of (a) *mexicana* and (b) *parviglumis*, and the distance to the niche centroid of (c) *mexicana* and (d) *parviglumis*. The heat-colour bar represents distance to the specific centroid, with lighter colours indicating higher distance. The size of the circle represents the amount of (a–b) genetic diversity, and (c–b) the number of candidate SNPs. The linear regressions between the distance to the geographic and the niche centroid are not significant (Fig. S8)

association with environment (de Villemereuil & Gaggiotti, 2015). Bayenv relies on the detection of strong correlations between allelic frequencies and environmental variation (Coop et al., 2010). At the end, we retained for each subspecies and both environmental PCs (1-2) all SNPs that were detected by both tests as being outliers. As these two outlier tests rely on different assumption, it is possible that we could be ignoring some adaptive SNPs (Pyhäjärvi et al., 2013a,b). However, by doing this, we are aiming at retaining the strongest candidate SNPs, and more importantly, reducing the amount of false positives. For *parviglumis*, we identified 97 SNP associated to PC1 and 80 SNPs associated to PC2. For *mexicana*, we found 81 SNP associated to PC1 and 89 SNPs associated to PC2.

For each outlier SNP, we ran the cline analysis described in Figure 1, and retained SNPs that had significant associations with environmental cline variation, and discarded those that either had null clines or fitted geographic clines (See Methods and Figure 1). At the end, for each subspecies, we identified 20 SNPs associated to PC1 and 14 SNPs associated to PC2. None of these SNPs were shared between subspecies or PCs (Table S1), which probably reflects the environmental variation between taxa (Hufford et al., 2012). We then used the Phytozome database to annotate the neutral and retained candidates SNPs (https://phytozome.jgi.doe.gov). From the 68 SNPs detected by the cline analysis, 42 SNPs were found in coding regions of annotated genes in maize (61%), many of which have previously shown associations to drought, heat and/or abiotic tolerance in grasses and other plants (Table 1 and S6), while from the 33,396 neutral SNPs, we found that 14,939 were in coding genes (44%).

TABLE 1 Annotation of strong candidate SNPs that have been found to be associated to temperature, drought or abiotic stress

Gene Name	Loci Name	Chr	Position	subspecies	PC	Annotation	Response	Species	References[a]
GRMZM2G141596	SYN32365	9	140846908	*mexicana*	PC1	CONSTANS interacting protein 4	Flowering time	Poaceae	Fjellheim, Boden, and Trevaskis (2014)
GRMZM2G162949	SYN6394	9	124028957	*mexicana*	PC1	RING/U-box superfamily protein	Heat response	*Oryza sativa* (Poaceae)	Zhang et al. (2012)
GRMZM2G164358	SYN34677	9	129186069	*mexicana*	PC1	Predicted: E3 ubiquitin-protein ligase RING1-like	Abiotic stress	*Arabidopsis thaliana* (Brassicaceae)	Mazzucotelli et al. (2006)
GRMZM2G166780	SYN27200	9	139374933	*mexicana*	PC1	RNA recognition motif in THO complex subunit 4 (THOC4)	Abiotic stress	*Oryza sativa* (Poaceae)	Sharma, Kaur, Singla-Pareek, and Sopory (2016)
GRMZM2G015159	PUT.163a.50338335.2262	5	14343518	*mexicana*	PC2	Ras-related small GTP-binding family protein/RAB homolog 1	Desiccation tolerance	*Sporobolus stapfianus* (Poaceae)	O'Mahony and Oliver (1999)
GRMZM2G069773	PZA00291.7	2	110826348	*mexicana*	PC2	Polyketide cyclase/dehydrase and lipid transport superfamily protein	Drought stress	*Arabidopsis thaliana* (Brassicaceae)	Zhou et al. (2013)
GRMZM2G033544	PZE.104001592	4	1986674	*parviglumis*	PC1	Cyclopropane-fatty-acyl-phospholipid synthase	Response to darkness	*Arabidopsis thaliana* (Brassicaceae)	Hudson et al. (2003)
GRMZM2G073750	PZA02619.1	3	123862971	*parviglumis*	PC1	Auxin response factor 6	Seed germination	*Zea mays* (Poaceae)	Xing et al. (2011)
GRMZM2G102681	SYN22745	4	154688945	*parviglumis*	PC1	F-box family protein with a domain of unknown function (DUF295)	Root growth under abiotic stress	*Oryza sativa* (Poaceae)	Yan et al. (2011)
GRMZM2G098819	PZE.104025828	4	30949186	*parviglumis*	PC2	Mob1/phocein family protein	Root development	*Arabidopsis thaliana* (Brassicaceae)	Pinosa et al. (2013)
GRMZM2G108501	PZE.106007374	6	21317002	*parviglumis*	PC2	Flavin-binding monooxygenase family protein	Auxin biosynthesis	*Arabidopsis thaliana* (Brassicaceae)	Dai et al. (2013)
GRMZM2G111529	SYN35105	5	144114260	*parviglumis*	PC2	Glucan synthase-like 4	Drought response	*Arabidopsis thaliana* (Brassicaceae)	Maeda, Song, Sage, and DellaPenna (2014)
GRMZM2G340656	PZE.107020282	7	19055024	*parviglumis*	PC2	Seed imbibition 2	Drought and heat response	*Sorghum bicolor* (Poaceae)	Johnson et al. (2014)

[a]See Table S6 for complete reference and a complete list of candidate SNPs.

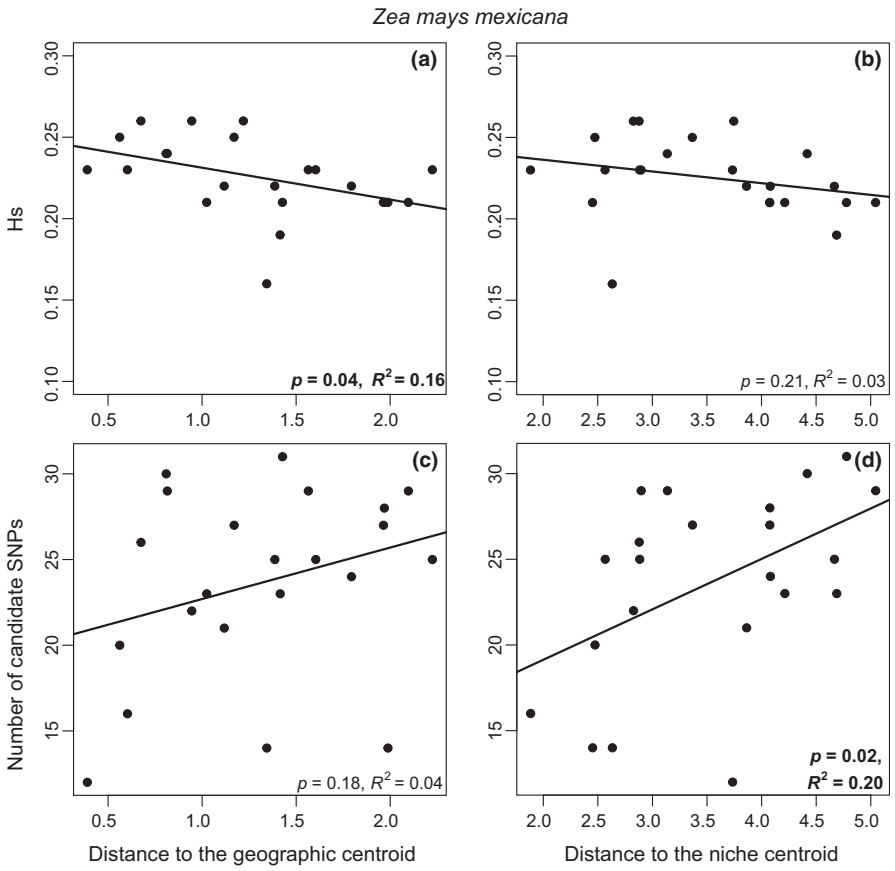


FIGURE 4 Test of the central abundance hypothesis for populations of *mexicana*. Relation between the distance to the geographic centroid and: (a) the genetic diversity (p = .04, R^2 = .16); (c) the number of candidate SNPs (p = .18, R^2 = .04). Relation between the distance to the niche centroid and: (b) the genetic diversity (p = .21, R^2 = .03); (d) the number of candidate SNPs (p = .02, R^2 = .20). The Puruandiro population has an outlier number of candidate SNPs (p < .04, see main text) and therefore was removed from plots c and d

When surveying the position of the retained candidates, we found that they were mostly distributed along the genome (Figs S9-S10). However, for the case of *mexicana*, we found eight SNPs associated to PC1 along the chromosomic inversion *Inv9e* in chromosome 9 (Pyhäjärvi et al., 2013a,b; Fig. S11). None of the other four inversions described in teosintes so far (Pyhäjärvi et al., 2013a,b) had more than one of the candidate SNPs detected on our outlier and cline analyses. As inversions often present high linkage disequilibrium (LD), which may lead to allele fixation by nonselective processes (different demographic model within the inversion or LD), these candidates could represent false positives. However, when comparing LD (r^2) differences between all surveyed loci within and outside this inversion (calculated with *plink* (Purcell, Neale, & Todd-Brown, 2007)), we obtained low median values in both cases; although they were significantly higher within (r^2 = .017) than outside (r^2 = .0036) the inversion (p < .001; See Fig. S12a). These values together with the fact that candidate SNPs were separated by long blocks of very low LD within the inversion imply that it is no likely that they are false positives (Fig. S12b).

As Bayescenv and Bayenv are not designed to determine locus-by-population interactions, the cline analysis (see Methods) further allowed us to pinpoint populations in which a given candidate could be considered adaptive or neutral, that is, those that, respectively, fitted and not fitted the cline (see Methods and Fig. S4). We found that there is variation in the number of putatively adaptive SNPs among populations (Figure 3c–d; Figures 4–5; Table S1). Overall, the *parviglumis* panel exhibited a lower average number of candidate SNPs per population (N_{cand} = 17.29; SD = 4.83), ranging from 8 to 27, than the *mexicana* panel (N_{cand} = 23.09; SD = 5.99), which ranged from 11 to 31.

3.4 | Effect of distance to geographic and climatic niche centroid

In each panel, we employed the distance to the geographic and niche centroid to test the CAH. For both subspecies, we found a significant negative association between the distance to the geographic centroid and H_S (Figure 4a, *mexicana*: p = .04, R^2 = .16; Figure 5a, *parviglumis*: p = .002, R^2 = .34; Fig. S13), but no significant association between the distance to the geographic centroid and the number of candidate SNPs (Figure 4c, *mexicana*: p = .18, R^2 = .04; Figure 5c, *parviglumis*: p = .16, R^2 = .05).

When comparing the distance to the niche centroid with genetic diversity and enrichment of candidate SNPs, we found nonsignificant associations between the distance to the niche centroid and genetic diversity (Figure 4b, *mexicana*: p = .21; R^2 = .03; Figure 5b, *parviglumis*: p = .23, R^2 = .02), and a positive and significant association between the distance to the niche centroid and the of number candidate SNPs in both taxa (Figure 4d, *mexicana*: p = .049, R^2 = .13; Figure 5d, *parviglumis*: p = .0004, R^2 = .41).

In the case of *mexicana*, we found that two populations (El Porvenir and Puruandiro) had a very low number of candidate SNPs

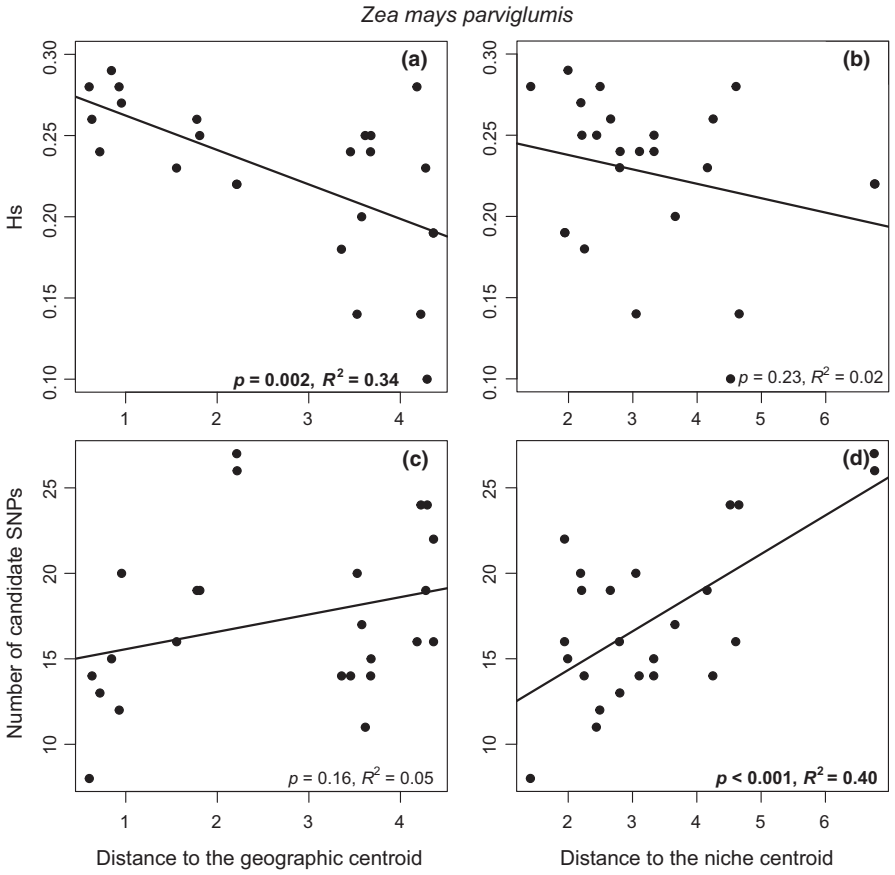


FIGURE 5 Test of the central abundance hypothesis for populations of *parviglumis*. Relation between the distance to the geographic centroid and: (a) the genetic diversity (p = .002, R^2 = .34); (c) the number of candidate SNPs (p = .16, R^2 = .05). Relation between the distance to the niche centroid and: (b) the genetic diversity (p = .23, R^2 = .02); (d) the number of candidate SNPs (p < .001, R^2 = .40)

despite being at intermediate values of the distance to the niche centroid. We therefore used the *chisq.out.test* function of the R package "Outliers" to test whether these populations have an outlier distribution (Dixon, 1950). We found that only Puruandiro was an outlier (p = .04), and therefore removed it from the analyses. In this case, the association between the distance to the niche centroid and the number of candidate SNP increased the significance of the model (*mexicana*: p = .02, R^2 = .20), while the association of the distance to the geographic centroid and the number of candidate SNP still remained nonsignificant (*mexicana*: p = .18, R^2 = .04).

As the higher number of candidate SNPs found in edge populations could depend on the probability of a SNP being called as such in those populations, we performed four tests to determine whether this interpretation could result from a statistical artefact or a circularity bias (See Materials and Methods, and Fig. S5). We first tested the pipeline outlined in Figure 1 but using geographic clines instead of environmental ones to call a SNP "adaptive" in a given population. We did not observe a geographic enrichment with this test for any of the two taxa, suggesting that SNPs are not more likely to be called adaptive in populations located at the edge of the distribution (Test 1: Fig. S5a). We then fitted nonoutlier SNPs (neutral) to the environmental clines and did not find any kind of enrichment for any taxa as populations were further away from the niche centroid. This suggests that a SNP is not more likely to be called "adaptive" only because it is surveyed in a population located at the edge of the climatic niche of a species (Test 2: Fig. S5b). The absence of similar patterns between these tests and our results suggest that there is not a statistical bias in the pipeline described in Figure 1. In a third test, we removed the effects of the cline and simply called a SNP "adaptive" in a given population when it had frequencies above 0.8 or below 0.2. We found that populations with less suitable conditions tend to have more candidate SNPs with nearly fixed frequencies than those at the centre of the environmental distribution, even when the cline analyses are not considered (See Fig. S5c, *parviglumis*: p = .001, *mexicana*: p = .078). However, when we used neutral SNPs, we did not find a significant higher number of fixed SNPs in less suitable conditions, indicating that the result of outlier SNPs is not generated by increased genetic drift in the less suitable conditions (See Fig. S5c, lower barplots and scatterplots), or else we would have found a similar pattern with neutral SNPs. In the fourth test, we analysed the effect of randomizing the environmental information, then searching for outlier SNPs and running the clines analyses (Fig. S5d). Using the same approximation as explained in Figure 1, we did not find an association between the number of candidate SNPs and the distance to the randomized niche centroid (Fig. S5d), indicating that the increased enrichment of candidate SNPs in environmentally differentiated populations is not produced by an artefactual circularity in the approximation.

Finally, we tested the effect of modifying some parameters from our pipeline. First, we modified the percentile cut with which we removed populations that are far away from the environmental cline to 60%, 75% and 90% of the distribution. Then, we modified the

limit values for removing populations with intermediate allelic frequencies between 0.1/0.9, 0.2/0.8, 0.3/0.7 and 0.4/0.6. In all cases, we still found a positive and significant association between the distance to the niche centroid and the number of candidate SNPs per population (Fig. S14).

4 | DISCUSSION

We assembled a vast set of 49 teosinte populations and gathered genotyping data of 646 individuals and 33,464 SNPs to test the central abundance hypothesis (Eckert et al., 2008; Hengeveld & Haeck, 1982; Sexton et al., 2009). By analysing genetic diversity and selection along the geographic and environmental distribution of *mexicana* and *parviglumis*, we were interested in testing whether local adaptation occurs in the centre or the edge of the niche and geographic distributions.

According to our results, we found that populations at the edge of the geographic distribution had lower H_S than those in the centre (Figures 4a and 5a), showing that both *mexicana* and *parviglumis* populations fit in broad terms the CAH, in which populations at the geographic edge have lower genetic diversity (Hengeveld & Haeck, 1982; Eckert et al., 2008). However, we did not find that these populations are significantly associated to lower number of candidate SNPs (Figures 4c and 5c), suggesting that the reduced genetic diversity at the edge is not associated to decreased adaptation, as expected in the CAH (Lee-Yaw et al., 2016). Many of the edge populations grow in intermediate or highly suitable conditions (Figure 3), and therefore, we consider that their decreased genetic diversity is more likely associated to increased genetic drift given founder events in the distribution limit (Alleaume-Benharira, Pen, & Ronce, 2006; Case, Holt, McPeek, & Keitt, 2005), or perhaps polygenic selection.

In contrast, when we used the distance to the niche centroid as a predictive variable, we did not find a significant difference in the level of genetic diversity in niche edge populations in contrast to the prediction of the CAH (Figures 4b and 5b). More important, we found a positive significant association between the distance to the niche centroid and the number of candidate SNPs, indicating that populations at the edge of the niche show increased genomic signatures of local adaptation to new unsuitable environments (Figures 4d and 5d). In other words, our results suggest that populations at the edge of the niche distribution could be actively adapting, irrespective of their geographic distribution. It is important to recall that there was no correlation between the distance to the population's niche and geographic centroid (Figures 3 and S8), which explains why the distance to the niche centroid has a better predictive value (Sagarin & Gaines, 2002a; Sagarin et al., 2006) of the number of candidate SNPs. The decoupling between the distance to the geographic and niche centroid is likely explained by the complex orography of Mexico, resulting in heterogeneous distribution that not necessarily relate to geographic distance (Lira-Noriega & Manthey, 2014; Martínez-Meyer et al., 2013; Yañez-Arenas et al., 2012). The comparison of niche and geographic distributions is interesting in the study of edge limits, as it has been described that populations at the geographic limits usually do not locally adapt (Bridle & Vines, 2006), either because they receive deleterious mutations from central adapted populations (Bridle & Vines, 2006; Kirkpatrick & Barton, 2006), or because genetic drift is too strong to allow local adaptation (Alleaume-Benharira et al., 2006; Case et al., 2005). This does not seem to be the case in teosintes, as the proximity of some climatic edge populations to the centre of the geographic distribution indicates that these should not be sink populations (these are expected to be found in the limit of the geographic distribution, as gene flow is reduced in these regions). Furthermore, genetic diversity of environmental edge populations does not correlate with the distance to the niche centroid, as it would be expected if these populations were less adapted (Lira-Noriega & Manthey, 2014; Martínez-Meyer et al., 2013; Yañez-Arenas et al., 2012). Another possibility could be that populations with high number of candidate SNPs may have recently colonized regions that are at the niche limit. However, we believe this is not the case, as the distribution of some of the environmental edge populations is predicted during the last glacial maximum (Hufford et al., 2012) and is not found in the limits of the geographic distribution (Figures 3 and S8).

The result of increased local adaptation in the edge of the niche in teosintes is in agreement with recent reciprocal transplants (Halbritter, Billeter, Edwards, & Alexander, 2015), genetic connectivity (Sexton et al., 2016) and biological interaction (O'Brien, Sawers, Ross-Ibarra, & Straus, 2015) studies in plants. Understanding the dynamics of ecological edge populations is becoming increasingly popular, as these populations grow at the limit of their environmental distribution, and therefore could be pre-adapted to global change (Hampe & Petit, 2005; Lenormand, 2002), having important evolutionary implications (Eckert et al., 2008; Sexton et al., 2009). We consider that our results of active local adaptation at the niche limit in teosintes suggest that these populations could be important for the dynamics of populations as they may be a source of adaptive variation for populations exposed to changing environmental conditions (Hampe & Petit, 2005; Takeda & Matsuoka, 2008; Warschefsky, Penmetsa, Cook, & von Wettberg, 2014). Recent results suggest that local adaptation can occur very fast as a consequence of strong selection along the genome (Egan et al., 2015; Soria-Carrasco et al., 2014). This is particularly true if it occurs from standing genetic variation (Barrett & Schluter, 2008; Hohenlohe et al., 2010). Recently, Franks, Kane, O'Hara, and Rest (2016) found evidences of very rapid evolution to drought from standing genetic variation in *Brassica rapa*, which could also be the case in teosintes, as we found that all candidate SNPs are polymorphic within species (Fig. S15). However, this is expected as the SNP50 chip was designed to identify polymorphic SNPs in maize.

As niche limits can also translate into divergent selective processes, ecological speciation could initiate at the edges if populations become reproductively isolated between them (Nosil, 2012; Nosil, Funk, & Ortiz-Barrientos, 2009; Rundle & Nosil, 2005), and therefore, local adaptation at the niche limit could have effects above the

species level. Both teosinte subspecies are adapted to divergent environments (Hufford et al., 2012; Pyhäjärvi et al., 2013a,b; Ross-Ibarra, Tenaillon, & Gaut, 2009), and inversion polymorphisms associated to altitude and climatic variables support that genetic mechanisms could be contributing to reducing reproductive success between divergently adapted populations (Bradburg et al., 2013; Fang et al., 2012; Pyhäjärvi et al., 2013a,b). For instance, from the 34 candidate SNPs detected within the *mexicana* panel, eight were found all along the inversion *Inv9e* in chromosome 9 (Pyhäjärvi et al., 2013a,b). This inversion covers around 40 MB in chromosome 9, and candidate SNPs within this inversion have been found to show strong association with altitude (Pyhäjärvi et al., 2013a,b). Chromosomal inversions suppress recombination in heterozygous individuals or polymorphic populations, which can reduce gene flow between divergently selected populations (Kirkpatrick & Barton, 2006) and generate patterns of isolation by adaptation at the genomic level (Feder, Flaxman, Egan, Comeault, & Nosil, 2013; Twyford & Friedman, 2015). As we did not find this inversion in the *parviglumis* panel, we consider that the inversion could be playing an important role as driver of incipient ecological speciation between teosinte populations. Future works combining local adaptation and phylogenetic relationship between both teosintes should be tested to determine whether local adaptation to a cooler niche could be at the base of the cold, high altitude and derived *mexicana* cluster (Ross-Ibarra et al., 2009).

Given the use of clines in this analysis to differentiate conditional neutrality, and therefore identify locus-by-population interactions (γ -effects in Beaumont & Balding, 2004), it is important to consider that genome scans are also sensitive to nonselective forces (De Mita et al., 2013; Lotterhos & Whitlock, 2014; Schoville et al., 2012). This is particularly true in expanding edge populations, where low-frequency alleles can increase in frequency by stochastic forces like gene surfing (Klopfstein, Currat, & Excoffier, 2006), potentially generating clines that could be confused with selection (Excoffier, Foll, & Petit, 2009; Excoffier & Ray, 2008). Although we acknowledge that gene surfing may occur in teosintes, it is unlikely that it is generating artefactual results in our analyses, especially because we retained only candidate SNPs that were previously detected as targets of selection by both Bayescenv and Bayenv, and for which the environmental cline had a better fit than the geographic and null clines. In fact, the retained SNPs corresponded to those in which geographically and genetically distant populations shared allelic frequencies based on climate alone (i.e., they show convergent evolution). Moreover, as geographic and environmental clines do not correlate (Fig. S8), and niche edge populations do not correspond to geographic edge populations (Figure 3), we consider that this comparison controls for gene surfing, which should occur in expanding geographic edges (Excoffier & Ray, 2008; Excoffier, Foll, et al., 2009; Klopfstein et al., 2006), and not at ecological edges.

In addition, we found that in contrast to the neutral SNPs (44% are in coding regions), a significantly higher amount of the candidate SNPs identified (62%) are within coding regions of annotated genes in maize (chi-square = 7.29, df = 1, *p*-value = .007; Tables 1 and S6), many of which have previously been found to be associated, upregulated or downregulated in response to abiotic stress in maize and other plants (Hudson, Lisch, & Quail, 2003; Zhou et al., 2013). For example, in *parviglumis*, we found a candidate SNP in locus GRMZM2G111529 that is a 1,3-beta-glucan synthase, which is involved in callose synthesis. This gene has been tightly associated to drought response in wheat (Faghani, Gharechahi, Komatsu, Mirzaei, & Hosseini, 2014; Peremarti et al., 2014). In *mexicana*, we found a SNP within locus GRMZM2G141596, which is related to day-length response due to its interactions with CONSTANS proteins (CO). Although we consider that the annotation of these genes supports that the SNPs detected in the cline analyses can have important implications in local adaptation, we consider that future experimental studies of reciprocal transplants should be also performed to confirm the results of local adaptation at the niche limits. In addition, these experiments should also allow identifying other aspects related to local adaptation, such as biological interactions (Fumagalli et al., 2011) or soil chemistry (Turner et al., 2010).

Another potential bias in our results is the use of clines to label SNPs as putatively adaptive or neutral in a given population. If the fit of the cline is dependent on extreme populations, then these populations may have a higher probability of having SNPs considered as adaptive. However, two tests conducted to discern whether this bias was present in our pipeline (see Fig. S5) showed that outliers were not more likely to be called adaptive at the geographic extremes or that nonoutliers were not likely to be labelled the same way at the environmental edges of both species niches, as it would be expected if there was a indeed a statistical bias. Moreover, in these tests, we further found that populations at intermediate geographic regions were more enriched than those in the periphery (coloured dots in Fig. S5a–b), indicating that the cline analyses can select as "candidate" populations that are not at the edge. Interestingly, our analyses also showed that LD or genetic drift cannot account for the increased fixation of SNPs at the niche limits of both teosintes, as these patterns were not shared between candidate and neutral SNPs. Also, by removing the cline step (Test 3), we still found higher number of nearly fixed SNPs in edge populations, supporting that we correctly identified the relevant SNPs and that selection would bring alleles to low or high frequencies. To verify whether increased fixation could be due to genetic drift, we tested the same with 50 runs of neutral SNPs per subspecies. None of the neutral runs had an increased number of nearly fixed SNPs in less suitable conditions, corroborating that the pattern is not related to genetic drift. Finally, randomizing the environment prior to the detection of outlier SNPs and the running the pipeline described in Figure 1 (Test 4) did not result in a higher enrichment of candidate SNPs in the new false niche distribution. This indicates that using the environment to detect outlier SNPs is not skewing the patterns detected in Figures 4d and 5d of increased signals of adaptation at the niche limit.

In summary, we found that populations at the edge of the teosintes distributions have lower genetic diversity, and are probably evolving by genetic drift, while populations at the edge of the

niche present evidences of increased local adaptation and are likely affected by selection. These results do not fit the general findings that edge limits correspond to niche limits and that adaptation limits are imposed by decreased suitability (Lee-Yaw et al., 2016). This difference could be related to the life history, population sizes, mutation rates and large standing variation in teosinte, compared to other plant species. In this sense, it will be interesting to test these methods in species with contrasting life-history traits. However, our results highlight the importance of studying local adaptation at edge populations combining analyses of environmental and geographic distances, especially in regions where these two axes are decoupled, like central Mexico (Lira-Noriega & Manthey, 2014; Martínez-Meyer et al., 2013; Sagarin & Gaines, 2002a; Sagarin et al., 2006). Defining loci by population effects remains nonetheless challenging, and the pipeline described here should be viewed as a step to detect these associations. While none of the posteriori tests pointed towards a statistical bias in our reasoning, field tests are still needed to confirm the adaptive nature of the candidates identified herein.

ACKNOWLEDGEMENTS

We thank E. Aguirre-Planter, L. Espinosa-Asuar and M. Rosas-Barrera for technical support, and Ricardo Colín, Enrique Scheinvar, Gabriel Merino and Jaime Gasca for seed collection. We thank David Romero Camarena for comments on the experimental design, and three anonymous reviewers for suggestions that improved this manuscript. This work is presented in partial fulfilment of the requirements to obtain a PhD degree by Jonás A. Aguirre-Liguori, who thank the "Programa de Doctorado en Ciencias Biomédicas, from the Universidad Nacional Autónoma de México (UNAM)" and thank the scholarship provided by the Consejo Nacional de Ciencia y Tecnología (CONACYT, grant no. 255770). This work was supported by grant CB2011/167826 (CONACYT Investigación Científica Básica) to LEE, grant CN-10-393 (UC MEXUS-CONACYT) to LEE and BSG, and grant M12-A03, CONACYT-ANUIES 207571 (SEP-CONACYT-ANUIES-ECOS Nord France) to LEE and MIT.

DATA ACCESSIBILITY

All genotypes have been deposited in the Dryad Digital repository (https://doi.org/10.5061/dryad.tf556). Any reader interested in the pipeline described in Figure 1 may write the corresponding author.

AUTHOR CONTRIBUTIONS

J.A.A.L., M.I.T., B.S.G., J.P.J.C., S.M.H., V.S. and L.E.E. conceived the experiment. M.I.T., B.S.G., S.M.H., V.S., L.E.E. collected the material. J.A.A.L. and A.V.L. conducted the laboratory work. J.A.A.L. wrote and ran the cline analysis script. J.A.A.L., M.I.T., J.P.J.C. and L.E.E. analysed the data. A.V.L. annotated the candidate SNPs. J.A.A.L. and L.E.E. wrote the first draft of the study, and all authors contributed substantially to revisions.

REFERENCES

Albrechtsen, A., Nielsen, F. C., & Nielsen, R. (2010). Ascertainment biases in SNP chips affect measures of population divergence. *Molecular Biology and Evolution*, *27*, 2534–2547.

Alleaume-Benharira, M., Pen, I. R., & Ronce, O. (2006). Geographic patterns of adaptation within a species range: interactions between drift and gene flow. *Journal of Evolutionary Biology*, *19*, 203–219.

Anderson, J. T., Inouye, D. W., McKinney, A. M., Colautti, R. I., & Mitchell-Olds, T. (2012). Genetic trade-offs and conditional neutrality contribute to local adaptation. *Molecular Ecology*, *22*, 699–708.

Anderson, J. T., Willis, J. H., & Mitchell-Olds, T. (2011). Evolutionary genetics of plant adaptation. *Trends in Genetics*, *27*, 258–266.

Barrett, R. D. H., & Schluter, D. (2008). Adaptation from standing genetic variation. *Trends in Ecology and Evolution*, *23*, 38–44.

Beaumont, M. A., & Balding, D. J. (2004). Identifying adaptive genetic divergence among populations from genome scans. *Molecular Ecology*, *13*, 969–980.

Beaumont, M. A., & Nichols, R. A. (1996). Evaluating loci for use in the genetic analysis of population structure. *Proceeding of the Royal Society of London B: Biological Sciences*, *263*, 1619–1626.

Bradburg, G. S., Ralph, P. L., & Coop, G. M. (2013). Disentangling the effects of geographic and ecological isolation on genetic differentiation. *Evolution*, *67*, 3258–3273.

Bridle, J. R., & Vines, T. H. (2006). Limits to evolution at range margins: when and why does adaptation fail? *Trends in Ecology and Evolution*, *22*, 140–147.

Case, T. J., Holt, R. D., McPeek, M. A., & Keitt, T. H. (2005). The community context of species' borders: ecological and evolutionary perspectives. *Oikos*, *108*, 28–46.

CONABIO. (2015). Retrieved from http://www.conabio.gob.mx/ Last accessed 14/04/2015.

Coop, G., Witonsky, D., Di Rienzo, A., & Pritchard, J. K. (2010). Using environmental correlations to identify loci underlying local adaptation. *Genetics*, *185*, 1411–1423.

Dai, X., Mashiguchi, K., Chen, Q., Kasahara, H., Kamiya, Y., Ojha, S., … Zhao, Y. (2013). The biochemical mechanism of auxin biosynthesis by an Arabidopsis YUCCA flavin-containing monooxygenase. *Journal of Biological Chemistry*, *288*, 1448–1457.

De Mita, S., Thuillet, A. C., Gay, L., Ahmadi, N., Manel, S., Ronfort, J., & Vigouroux, Y. (2013). Detecting selection along environmental gradients: analysis of eight methods and their effectiveness for outbreeding and selfing populations. *Molecular Ecology*, *22*, 1383–1399.

Derryberry, E., Derryberry, G. E., Maley, J. M., & Brumfield, R. T. (2014). HZAR: hybrid zone analysis using an R software package. *Molecular Ecology Resources*, *14*, 252–663.

Díez, C. M., Gaut, B. S., Meca, E., Scheinvar, E., Montes-Hernandez, S., Eguiarte, L. E., & Tenaillon, M. I. (2013). Genome size variation in wild and cultivated maize along altitudinal gradients. *New Phytologist*, *199*, 264–276.

Dixon, W. J. (1950). Analysis of extreme values. *The Annals of Mathematical Statistics*, *21*, 488–506.

Doyle, J. J., & Doyle, J. L. (1987). A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemical Bulletin*, *19*, 11–15.

Eckert, C. G., Samis, K. E., & Lougheed, S. C. (2008). Genetic variation across species' geographical ranges: the central-marginal hypothesis and beyond. *Molecular Ecology*, *17*, 1170–1180.

Egan, S. P., Ragland, G. J., Assour, L., Powell, T. H. Q., Hood, G. R., Emrich, S., … Feder, J. L. (2015). Experimental evidence of genome-wide impact of ecological selection during early stages of speciation-with-gene-flow. *Ecology Letters*, *18*, 817–825.

Excoffier, L., Foll, M., & Petit, R. J. (2009). Genetic consequences of range expansion. *Annual Review of Ecology, Evolution, and Systematics*, *40*, 481–501.

Excoffier, L., Hofer, T., & Foll, M. (2009). Detecting loci under selection in a hierarchically structured population. *Heredity*, *103*, 285–298.

Excoffier, L., & Ray, N. (2008). Surfing during population expansions promotes genetic revolutions and structuration. *Trends in Ecology & Evolution*, *23*, 347–351.

Faghani, E., Gharechahi, J., Komatsu, S., Mirzaei, M., & Hosseini, G. (2014). Comparative physiology and proteomic analysis of two wheat genotypes contrasting in drought tolerance. *Journal of Proteomics*, *114*, 1–15.

Fang, Z., Pyhäjärvi, T., Weber, A. L., Dawe, K., Glaubitz, J. C., Sánchez-González, J. J., … Ross-Ibarra, J. (2012). Megabase-scale inversion polymorphism in the wild ancestor of maize. *Genetics*, *191*, 883–894.

Feder, J. L., Flaxman, S. M., Egan, S. P., Comeault, A. A., & Nosil, P. (2013). Geographic mode of speciation and genomic divergence. *Annual Review in Ecology, Evolution and Systematics*, *44*, 73–97.

Fielding, A. H., & Bell, J. F. (1997). A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environmental Conservation*, *24*, 38–39.

Fjellheim, S., Boden, S., & Trevaskis, B. (2014). The role of seasonal flowering responses in adaptation of grasses to temperate climates. *Frontiers in Plant Science*, *5*, 1–15.

Foll, M., & Gaggiotti, O. (2008). A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: A Bayesian perspective. *Genetics*, *180*, 977–993.

Franks, S. J., Kane, N. C., O'Hara, Tittes. S., & Rest, J. S. (2016). Rapid genome-wide evolution in Brassica rapa populations following drought revealed by sequencing of ancestral and descendant gene pools. *Molecular Ecology*, *25*, 3622–3631.

Fukunaga, K., Hill, J., Vigouroux, Y., Matsuoka, Y., Sanchez, G. J., Liu, K., … Doebley, J. (2005). Genetic diversity and population structure of teosinte. *Genetics*, *169*, 2241–2254.

Fumagalli, M., Sironi, M., Pozzoli, U., Ferrer-Admettla, A., Pattini, L., & Nielsen, R. (2011). Signatures of environmental genetic adaptation pinpoint pathogens as the main selective pressure through human evolution. *PLoS Genetics*, *7*, e1002355.

Goudet, J. (2005). Hierfstat, a package for R to compute and test variance components and F-statistics. *Molecular Ecology Notes*, *5*, 184–186.

Halbritter, A. H., Billeter, R., Edwards, P. J., & Alexander, J. M. (2015). Local adaptation at range edges: comparing elevation and latitudinal gradients. *Journal of Evolutionary Biology*, *28*, 1849–1860.

Hampe, A., & Petit, R. J. (2005). Conserving biodiversity under climate change: the rear edge matters. *Ecology Letters*, *8*, 461–467.

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

Hengeveld, R., & Haeck, J. (1982). The distribution of abundance. 1. Measurements. *Journal of Biogeography*, *9*, 303–316.

Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G., & Jarvis, A. (2005). Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, *25*, 1965–1978.

Hijmans, R. J., & van Etten, J. (2015). Raster. Geographic data analysis and modeling: R package version 2.5-2.

Hohenlohe, P. A., Bassham, S., Etter, P. D., Stiffler, N., Johnson, E. A., & Cresko, W. A. (2010). Population genomics of parallel adaptation in threespine sticklebacks using sequenced RAD tags. *PLoS Genetics*, *6*, e1000862.

Hudson, M. E., Lisch, D. R., & Quail, P. H. (2003). The FHY3 and FAR1 genes encode transposase-related proteins involved in regulation of gene expression by the phytochrome A-signaling pathway. *The Plant Journal*, *34*, 453–471.

Hufford, M. B., Martínez-Meyer, E., Gaut, B. S., Eguiarte, L. E., & Tenaillon, M. (2012). Inferences from the historical distribution of wild and domesticated maize provide ecological evolutionary insight. *PLoS ONE*, *7*, e47659.

Hutchinson, G. E. (1978). *An introduction to population ecology*. New Haven, CT: Yale University Press.

Johnson, S. M., Lim, F. L., Finkler, A., Fromm, H., Slabas, A. R., & Knight, M. R. (2014). Transcriptomic analysis of Sorghum bicolor responding to combined heat and drought stress. *BMC Genomics*, *15*, 456.

Jombart, T. (2008). Adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics*, *24*, 1403–1405.

Joost, S., Bonin, A., Bruford, M. W., Després, L., Conord, C., Erhardt, G., & Taberlet, P. (2007). A Spatial analysis method (SAM) to detect candidate loci for selection: towards a landscape genomics approach to adaptation. *Molecular Ecology*, *16*, 3955–3969.

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

Kirkpatrick, M., & Barton, N. (2006). Chromosome inversions, local adaptation and speciation. *Genetics*, *173*, 419–439.

Klopfstein, S., Currat, M., & Excoffier, L. (2006). The fate of mutations surfing on the wave of a range expansion. *Molecular Biology and Evolution*, *23*, 482–490.

Lee-Yaw, J. A., Kharouba, H. M., Bontrager, M., Mahony, C., Csergo, A. M., Noreen, A. M. E., … Angert, A. L. (2016). A synthesis of transplant experiments and ecological niche models suggests that range limits are often niche limits. *Ecology Letters*, *19*, 710–722.

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

Lenormand, T. (2002). Gene flow and the limits of natural selection. *Trends in Ecology and Evolution*, *17*, 183–189.

Lewontin, R. C., & Krakauer, J. (1973). Distribution of gene frequency as a test of theory of the selective neutrality of polymorphisms. *Genetics*, *74*, 175–195.

Lira-Noriega, A., & Manthey, J. D. (2014). Relationship of genetic diversity and niche centrality: a survey and analysis. *Evolution*, *68*, 1082–1093.

Lotterhos, K. E., & Whitlock, M. C. (2014). Evaluation of demographic history and neutral parameterization on the performance of F_{ST} outlier tests. *Molecular Ecology*, *23*, 2178–2192.

Lumholtz, C. (1902). *Mexico Unknown. A record of five years' exploration among the tribes of the western Sierra Madre; in the tierra caliente of Tepic and Jalisco; and among the Tarascos of Michoacan*. New York, NY: Charles Schribner's Sons.

Maeda, H., Song, W., Sage, T., & DellaPenna, D. (2014). Role of callose synthases in transfer cell wall development in tocopherol deficient Arabidopsis mutants. *Frontiers in Plant Science*, *5*, 104–116.

Maguire, B. (1973). Niche response structure and the analytical potentials of its relationship to the habitat. *American Naturalist*, *107*, 213–246.

Martínez-Meyer, E., Díaz-Porras, D., Peterson, T. A., & Yañez-Arenas, C. (2013). Ecological niche structure and range-wide abundance patterns of species. *Biology Letters*, *9*, 20120637.

Mazzucotelli, E., Belloni, S., Marone, D., De Leonardis, A. M., Guerra, D., Di Fonzo, N., … Mastrangelo, A. M. (2006). The E3 ubiquitin ligase gene family in plants: Regulation by degradation. *Current Genomics*, *7*, 509–522.

Nei, M. (1972). Genetic distances between populations. *American Naturalist*, *106*, 283–292.

Nosil, P. (2012). *Ecological speciation*. Oxford: Oxford University Press.

Nosil, P., Funk, D. J., & Ortiz-Barrientos, D. (2009). Divergent selection and heterogeneous genomic divergence. *Molecular Ecology*, *18*, 375–402.

O'Brien, A., Sawers, R., Ross-Ibarra, J., & Straus, S. Y. (2015). Extending the Stress-Gradient hypothesis: increased local adaptation between teosinte and soil biota at the stressful end of a climate gradient. *BioRxiv*, http://biorxiv.org/content/early/2015/11/11/031195.abstract

O'Mahony, P. J., & Oliver, M. J. (1999). Characterization of a desiccation-responsive small GTP-binding protein (Rab2) from the desiccation-tolerant grass *Sporobolus stapfianus*. *Plant Molecular Biology*, *39*, 809–821.

Peremarti, A., Marè, C., Aprile, A., Roncaglia, E., Cattivelli, L., Villegas, D., & Royo, C. (2014). Transcriptomic and proteomic analyses of a pale-green durum wheat mutant shows variations in photosystem components and metabolic deficiencies under drought stress. *BMC Genomics*, *15*, 125.

Phillips, S. J., Anderson, R. P., & Schapire, R. E. (2006). Maximum entropy modeling of species geographic distributions. *Ecological Modelling*, *190*, 231–259.

Phytozome. Retrieved from https://phytozome.jgi.doe.gov Last accessed 10 August 2016.

Pinosa, F., Begheldo, M., Pasternak, T., Zermiani, M., Paponov, I. A., Dovzhenko, A., … Palme, K. (2013). The *Arabidopsis thaliana* Mob1A gene is required for organ growth and correct tissue patterning of the root tip. *Annals of Botany*, *112*, 1803–1814.

Purcell, S., Neale, B., & Todd-Brown, K. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *The American Journal of Human Genetics*, *81*, 559–575.

Pyhäjärvi, T., Hufford, M., Mezmouk, S., & Ross-Ibarra, J. (2013a). Complex patterns of local adaptation in teosinte. *Genome Biology and Evolution*, *5*, 1594–1609.

Pyhäjärvi, T., Hufford, M., Mezmouk, S., & Ross-Ibarra, J. (2013b). Data from: Complex patterns of local adaptation in teosinte. *Dryad Digital Repository*. https://doi.org/10.5061/dryad.8m648

R Development Core Team (2008). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. ISBN 3-900051-07-0, Retrieved from http://www.R.project.org

Ross-Ibarra, J., Tenaillon, M. I., & Gaut, B. S. (2009). Historical divergence and gene flow in the genus *Zea*. *Genetics*, *181*, 1399–1413.

Rundle, H. D., & Nosil, P. (2005). Ecological Speciation. *Ecology Letters*, *8*, 336–352.

Sagarin, R. D., & Gaines, S. D. (2002a). The 'abundant center' distribution: to what extent is it a biogeographical rule? *Ecology Letters*, *5*, 137–147.

Sagarin, R. D., & Gaines, S. D. (2002b). Geographical abundance distributions of coastal invertebrates: using one-dimensional ranges to test biogeographic hypothesis. *Journal of Biogeography*, *29*, 985–997.

Sagarin, R. D., Gaines, S. D., & Gaylord, B. (2006). Moving beyond assumptions to understand abundance distributions across the ranges of species. *Trends in Ecology and Evolution*, *21*, 524–530.

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

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

Schluter, D. (2001). Ecology and the origin of species. *Trends in Ecology and Evolution*, *16*, 372–380.

Schnee, F., & Thompson, J. (1984). Conditional neutrality of polygene effects. *Evolution*, *38*, 42–46.

Schoville, S. D., Bonin, A., François, O., Lobreaux, S., Melodelima, C., & Manel, S. (2012). Adaptive genetic variation on the landscape: methods and cases. *Annual Review of Ecology, Evolution, and Systematics*, *43*, 23–43.

Sexton, J. P., Hufford, M. B., Bateman, A., Lowry, D. B., Meimberg, H., Strauss, S. Y., & Rice, K. J. (2016). Climate structures genetic variation across a species' elevation range: a test of range limits hypotheses. *Molecular Ecology*, *25*, 911–928.

Sexton, J. P., McIntyre, P. J., Angert, A. L., & Rice, K. J. (2009). Evolution and ecology of species range limits. *Annual Review of Ecology, Evolution, and Systematics*, *40*, 415–436.

Sharma, S., Kaur, C., Singla-Pareek, S. L., & Sopory, S. K. (2016). OsSRO1a interacts with RNA binding domain-containing protein (OsRBD1) and functions in abiotic stress tolerance in yeast. *Frontiers in Plant Science*, *7*, 1–12.

Soria-Carrasco, V., Gompert, Z., Comeault, A. A., Farkas, T. E., Parchman, T. L., Johnston, S., … Nosil, P. (2014). Stick insect genomes reveal natural selection's role in parallel speciation. *Science*, *344*, 738–742.

Takeda, S., & Matsuoka, M. (2008). Genetic approaches to crop improvement: responding to environmental and population changes. *Nature Review Genetics*, *9*, 444–457.

Turner, T. L., Bourne, E. C., Von Wettberg, E. J., Hu, T. T., & Nushdin, S. V. (2010). Population resequencing reveals local adaptation of *Arabidopsis lyrata* to serpentine soils. *Nature Genetics*, *42*, 260–263.

Twyford, A. D., & Friedman, J. (2015). Adaptive divergence in the monkey flower *Mimulus guttatus* is maintained by a chromosomal inversion. *Evolution*, *69*, 1476–1486.

de Villemereuil, P., & Gaggiotti, O. E. (2015). A new F_{ST} -based method to uncover local adaptation using environmental variables. *Methods in Ecology and Evolution*, *6*, 1248–1258.

Warschefsky, E., Penmetsa, R. V., Cook, D. R., & von Wettberg, E. J. B. (2014). Back to the wilds: Tapping evolutionary adaptations for resilient crops through systematic hybridization with crop wild relatives. *American Journal of Botany*, *101*, 1791–1800.

Wilkes, H. G. (1997). Teosinte in Mexico: Personal retrospective and assessment. In CIMMYT (Eds.), *Gene Flow among maize landraces, improved maize varieties, and teosinte: Implications for transgenic maize* (pp. 10–17). Mexico: El Batan.

Xing, H., Pudake, R. N., Guo, G., Xing, G., Hu, Z., Zhang, Y., … Ni, Z. (2011). Genome-wide identification and expression profiling of auxin response factor (ARF) gene family in maize. *BMC Genomics*, *12*, 1–13.

Yan, Y. S., Chen, X. Y., Yang, K., Sun, Z. X., Fu, Y. P., Zhang, Y. M., & Fang, R. X. (2011). Overexpression of an F-box protein gene reduces abiotic stress tolerance and promotes root growth in rice. *Molecular Plant*, *4*, 190–197.

Yañez-Arenas, C., Martínez-Meyer, E., Mandujano, S., & Rojas-Soto, O. (2012). Modelling geographic patterns of population density of the white-tailed deer in central Mexico by implementing ecological niche theory. *Oikos*, *121*, 2081–2089.

Zhang, X., Li, J., Liu, A., Zou, J., Zhou, X., Xiang, J., … Chen, X. (2012). Expression profile in rice panicle: Insights into heat response mechanism at reproductive stage. *PLoS One*, *7*, e49652.

Zhou, X. F., Jin, Y. H., Yoo, C. Y., Lin, X. L., Kim, D. J., Yun, D. J., … Jin, J. B. (2013). CYCLIN H; 1 regulates drought stress responses and blue light-induced stomatal opening by inhibiting reactive oxygen species accumulation in Arabidopsis. *Plant Physiology*, *162*, 1030–1041.

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

How to cite this article: Aguirre-Liguori JA, Tenaillon MI, Vázquez-Lobo A, et al. Connecting genomic patterns of local adaptation and niche suitability in teosintes. *Mol Ecol*. 2017;00:1–15.
https://doi.org/10.1111/mec.14203

CAPÍTULO 3: ESPECIACIÓN ECOLÓGICA DEL TEOCINTLE

Artículo terminado: Genomic differentiation and ecological speciation in teosinte

Genomic differentiation and ecological speciation in teosintes (*Zea mays parviglumis* and *Zea mays mexicana*)

Jonás A. Aguirre-Liguori[1]*, Brandon S. Gaut[2], Juan Pablo Jaramillo-Correa[1], Maud I. Tenaillon[3], Salvador Montes-Hernández[4], Felipe García-Oliva[5], Sarah Hearne[6], Luis E. Eguiarte[1]*

ABSTRACT

Ecological speciation theory predicts that reproductive isolation is more likely to be achieved if multiple niche axes are generating divergent adaptation. This has been termed the multifarious selection hypothesis. The aim of this study is to determine how many environmental axes have been responsible for the genomic differentiation of the wild maize, the teosinte (*Zea mays* ssp. *parviglumis* and ssp. *mexicana*). For this, we generated 10,000 non-ascertained DarTseqTM SNPs for 49 populations and combined two published data sets containing 33,454 in 49 populations. We used both datasets to analyze the demographic history of teosintes and determine the time and ecological frame of the divergence of both subspecies. We found that divergence initiated around 15,500 generations ago, when environment was cooler. Then, we performed genome wide scans to determine the number of "outlier SNPs" associated to 4 climatic and three soil variables. We found that candidate SNPs are mainly associated to temperature and phosphorus concentration in the soil. By comparing isolation by distance and isolation by environment patterns, we only found environmental differentiation related to neutral loci in the case of temperature. We conclude that the *mexicana* subspecies initiated their divergence in association to adaptation to reduced temperature and then during the Holocene migrated to higher altitude where they encountered soils with low phosphorus availability and adapted to it. Our results suggest that teosinte have diverged by a two step multifarious selection where populations first adapted to cold environments and then had to locally adapt to lower availability of phosphorus in the soil.

KEY WORDS

Demographic inferences; Ecological speciation; Genomic differentiation; Local adaptation; Multifarious selection hypothesis; Teosinte (*Zea mays* ssp. *mexicana*, *Zea mays* ssp. *parviglumis*)

INTRODUCTION

Divergent selection between populations may generate barriers to gene flow and lead to ecological speciation (Rundle & Nosil 2005; Schluter & Conte 2009; Nosil 2012). Speciation initiates when populations adapt to their local environments (Schluter 2000; Schluter 2001) and then diverge because selection acts against migrants and hybrids (Favre, Widmer, & Karrenberg, 2016; Jackson, Kawakami, Cooper, Galindo, & Butlin, 2012; Nosil, Egan, & Funk, 2008; Via, 2009). Genetically, ecological speciation is expected to manifest itself as blocks of differentiation across the genome. Initially, differentiation will be restricted to loci experiencing divergent selection and to regions of linkage disequilibrium (LD) around selected loci (Via, 2009; Via & West, 2008). If a second favorable variant is in the LD region, it too will increase in frequency, because gene flow will not counteract the selective effect (Feder, Flaxman, Egan, Comeault, & Nosil, 2013; Feder, Gejji, Powell, & Nosil, 2011), and the size of the differentiated region will extend as a result. Over time, differentiated regions will form 'genomic islands' that increase in both size, as they capture additional locally adapted variants, and in number, as additional traits become subject to divergent selection. Eventually, this process will trigger global differentiation along the genome, which is also known as 'genome wide congealing' (Feder et al., 2014a, 2013, 2011; Nosil, Funk, & Ortiz-Barrientos, 2009). Importantly, this process is expected to occur even when selection and LD are weak (Flaxman, Wacholder, Feder, & Nosil, 2014).

Ecological speciation is complete when divergent selection constructs reproductive barriers (Nosil, 2012a; Rundle & Nosil, 2005). However, the process may not always reach completion, because it is slowed by gene flow (Elias, Faria, Gompert, & Hendry, 2012; Nosil, 2012a; Nosil, Harmon, & Seehausen, 2009; Nosil & Sandoval, 2008). Models and empirical studies suggest that the formation of reproductive isolation (RI) is favored when divergent selection is strong (Funk, Nosil, & Etges, 2006; Nosil, 2012; Rundle & Nosil,

2005; Schluter & Conte, 2009), when there is a link between divergent selection and assortative mating (Bolnick, 2011; Servedio, Doorn, Kopp, Frame, & Nosil, 2011), and/or when divergence occurs in geographic isolation for at least a brief period (Aguilée, Lambert, & Claessen, 2011; Nosil & Feder, 2012; Surget-Groba, Johansson, & Thorpe, 2012).

Another crucial feature of ecological speciation may be multifarious selection, which is the concept that selection acts on multiple traits. The multifarious selection hypothesis argues that selection on multiple traits increases the number of genomic islands between populations, thereby elevating the potential for RI (Chevin, Decorzent, & Lenormand, 2014; Nosil, Funk, et al., 2009; Nosil, Harmon, et al., 2009; Nosil & Sandoval, 2008). While theoretical and modeling analyses predict the importance of multifarious selection (Chevin et al., 2014; Flaxman et al., 2014; Nosil, Funk, et al., 2009; Nosil, Harmon, et al., 2009; Smadja & Butlin, 2011), few empirical studies have explicitly explored the relevance of adaptation to multiple traits during ecological divergence (Arnegard et al., 2014; Liu et al., 2013; Malinsky et al., 2015; Michel et al., 2010; Nosil & Sandoval, 2008; Scholl, Nice, Fordyce, Gompert, & Forister, 2012). Before the importance of ecological speciation can be assessed, empirical studies need to characterize both the number of traits that contribute to local adaptation and their effects on genome differentiation (Elias et al., 2012; Keller & Seehausen, 2012; Nosil, Funk, et al., 2009).

Here we focus on local adaptation and potentially orthogonal selective forces in two annual teosintes: *Zea mays* ssp. *mexicana* (hereafter *mexicana*) and Z. *mays* ssp. *parviglumis* (hereafter *parviglumis*). These two subspecies constitute a model system for population and ecological genomics (Hufford, Bilinski, Pyhäjärvi, & Ross-Ibarra, 2012), in part because they are adapted to different climatic conditions. Subspecies *parviglumis* is adapted to warm, humid, lowland conditions (<1900 m), but *mexicana* grows in colder and dryer highland populations (>1500 m). Previous genetic analyses suggest that the two teosintes diverged ~60,000 years ago (Ross-ibarra, Tenaillon, & Gaut, 2009) and that their divergence has been accompanied by gene flow among populations (Fukunaga et al., 2005; Pyhäjärvi, Hufford, Mezmouk, & Ross-Ibarra, 2013; Ross-ibarra, Tenaillon, & Gaut, 2009; van Heerwaarden et al., 2011).

Despite the presence of gene flow, there is ample evidence of local adaptation among *mexicana* and *parviglumis* populations. For example, Pyhajarvi *et al.* (2013) identified hundreds of outlier SNPs that were associated with temperature, altitude and soil variables across 21 teosinte populations. Fustier *et al.* (2017) used whole genome data from six populations to identify patterns of local adaptation along two parallel *mexicana-parviglumis* environmental gradients (Díez et al., 2013). Fustier *et al.* (2017) found 47 genomic regions that cluster candidate SNPs, at 10 to 60 SNPs per cluster. Twenty-eight of the 47 genomic regions were identified in both environmenal gradients, which strongly supports convergent local adaption. Aguirre-Liguori *et al.* (2017) analyzed patterns of local adpatation among 49 *parviglumis* and *mexicana* populations, finding signals of enhanced local adaptation at the limit of environmental niches. These analyses have consistently identified candidate SNPs in chromosomal inversions that are themselves associated with bioclimatic gradients (Aguirre-Liguori et al., 2017; Fang et al., 2012; Fustier et al., 2017; Pyhäjärvi et al., 2013). Altogether, these results suggest that local adaptation is prevalent, but patterns of divergence have yet to be compared between subspecies in the context of ecological divergence in the presence of gene flow.

Fortunately, it is typically not difficult to discriminate between subspecies; most teosinte populations can be classified either as *parviglumis* or *mexicana* based on genetic markers (Aguirre-Liguori *et al.*, 2017) or on the environmental conditions in which they grow (Aguirre-Liguori, Aguirre-Planter, & Eguiarte, 2016). However, some populations in the state of Guerrero are more difficult to assign because they are genetically intermediate to *mexicana* and *parviglumis* (Fukunaga *et al.* 2005; Pyhäjärvi *et al.* 2013; Aguirre-Liguori *et al.* 2017) and grow in warm environments (Pyhäjärvi et al., 2013). At least three studies have investigated the history of Guerrero populations. Fukunaga *et al.* (2005) used SSR data to identify that Guerrero populations are intermediate between *mexicana* and *parviglumis*, but they could not differentiate between the possibility that Guerrero populations were either admixed or ancestral to both *parviglumis* and *mexicana*. Later Pyhajärvi *et al.* (2013) genotyped one Guerrero population with a 50K SNPchip. They tentatively concluded it had an admixed origin, because they found no evidence that the population was either ancestral or sister to *mexicana*, based on haplotype structures. More recently Aguirre-Liguori *et al.* (2017) surveyed five additional Guerrero populations with

the same 50K SNPchip. They found that the complex, potentially admixed, genetic structure was common across populations, and suggested that Guerrero populations represent an intermediate evolutionary step between basal *parviglumis* and derived *mexicana*. However, the origin of these Guerrero populations remains unresolved, and it is important to consider their origin more carefully if we are understand patterns of divergence and gene flow among subspecies.

Altogether, this study has three goals. The first is to better elucidate the status of the Guerrero populations by testing demographic models. Demographic inference also provides insights into the age of various taxa and the presence of gene flow. The second is to use ecological information to help identify genomic regions that may be under selection, based on correlations to both multiple climatic variables and local soil content. The third goal is to evaluate the possibility that divergent selection has led to the formation of genomic islands between populations. To reach these goals, we perform analyses with two genotyping datasets: a set of non-ascertained 9,780 SNPs from DArTSeq data from 47 populations of wild maize, including 19 populations that have not been studied previously, and a set of 33,454 SNPs from the maize 50K SNP from 49 previously described populations (Aguirre-Liguori *et al.* (2017), Pyhäjärvi *et al.* (2013)).

MATERIALS AND METHODS

Plant samples and genomic data: We used two datasets. The first was based on DarTseqTM (DTS dataset) technology (Ren et al., 2015; Sansaloni et al., 2011), which does not have an ascertainment bias and thus is suitable for demographic analyses. For this dataset, we collected seeds from 15 to 30 mothers in 47 populations of teosinte distributed along their entire geographical and environmental distribution (Figure S1, Table S1). Teosinte seedlings were established in pots in screenhouses at the CIMMYT Headquarters in El Batán, Texcoco, Mexico. Leaf tissue from 30 plants per sample were harvested, the tissue was lyophilized, and DNA was extracted from an equal area (28 mm^2) of leaf tissue from each individual plant, using a modified CTAB method (CIMMYT, 2005). DNA was quantified and diluted to equal concentration (200 ng/µl). A genomic representation of the set of samples was generated by digesting genomic DNA with a combination of two restriction enzymes, PstI (CTGCAG) and HpaII (CCGG), and ligating barcoded adapters to identify to which sample belong each DNA fragment produced. For each 96 well plate,

16% of the samples were replicated to assess reproducibility. Equimolar amounts of amplification products from each sample were pooled by plate and amplified by c-Bot (Illumina) bridge PCR, followed by fragment sequencing on Illumina Hiseq 2500 (www.illumina.com). SNPs and InSilico DArTs were identified using the DArTsoft analytical pipeline (http://www.diversityarrays.com/software.html#dartsoft).

The final DTS dataset contained allelic counts for each alternative allele of each SNP. We used *blastn* to annotate each SNP against the maize genome reference (Hapmap2, Chia et al. 2012), and removed all SNPs that had multiple hits (>=2) to remove possible duplications. We also removed all sites that were fixed for all populations analyzed, all SNPs that had missing data in more that 50% of the populations, and all SNPs that had minor allele frequency below 0.05 to remove potential sequencing errors. At the end, we retained 9,780 polymorphic SNP in 47 populations of teosintes.

The second dataset consisted of previously reported data based on the MaizeSNP50 Genotyping BeadChip (50K dataset; Aguirre-Liguori *et al.* 2017 and Pyhäjärvi *et al.* 2013). We downloaded both datasets (http://dx.doi.org/10.5061/dryad.8m648 and https://doi.org/10.5061/dryad.tf556) and concatenated them using the *merge* function in R. The final database consisted of 49 populations (Figure S1; Table S1) and more than 33,464 SNPs that covered the geographic, genetic and environmental distribution of teosintes (Pyhäjärvi *et al.* 2013; Aguirre-Liguori *et al.* 2017). These SNPs cover the entire genome in maize and are found mainly in genetic regions and separated between a median distance of 7.4 kbp in chromosome 6 to 17.7 kbp in chromosome 4 (Aguirre-Liguori *et al*. 2017). Positions of these SNPs are based on the HapMap 2 (Chia et al., 2012). Of these 47 populations, 29 were shared with the DTS dataset (Figure S1b).

Climatic and soil data: We gathered both climatic and soil data to test for associations with allele frequencies. The climatic data came from the WorldClim dataset (Hijmans, Cameron, Parra, Jones, & Jarvis, 2005), which we downloaded at a resolution of 30 arc-sec for each of 573 non-duplicated teosinte records (http://www.conabio.gob.mx/). We built a species distribution model with Maxent 3.3 (Phillips, Aneja, Kang, & Arya, 2006) based on all of the locations. We followed the same approximation as in Hufford *et al.* (2012) to build the distribution model. Then, from the distribution model, we extracted 5,000 random

points from the consensus map and all of the populations to perform a principal component analysis (PCA) of the 19-bioclimatic variables (Table S2). For each of the teosinte populations, we obtained four principal component (PC) scores (Table S1).

We also collected soil from the base of teosinte plants for a subset of 22 populations (Table S1). For each population, we selected three individuals at random and sampled soil surrounding their roots. In each soil sample, we measured three variables: pH, Nitrogen (N) concentration and Phosphorus (P) concentration. Soil pH was measured in deionized water (soil:solution ratio, 1:2 w/v and litter:solution ratio 1:5 w/v) with a pH meter equipped with a glass electrode (Corning). N and P forms were analyzed colorimetrically using a Bran-Luebbe Auto analyzer 3 (Norderstedt, Germany). An aliquot of each sample was oven-dried at 70°C and ground in a Thomas Scientific mill to pass through a 40-mesh screen (0.425 mm) to remove the plant material, and ground with a pestle and agate mortar. Total N and P were determined following acid digestion in a mixture of concentrate H_2SO_4 and K_2SO_4, plus $CuSO_4$ as a catalyst; N was determined by a micro-Kjeldahl method (Bremmer 1996) and P by the molybdate colorimetric method, following ascorbic acid reduction (Murphy and Riley 1962). For each population, we obtained the average value across the 3 replicates.

Demographic inferences:

We used both the DTS and 50K datasets to investigate the demographic history among teosinte populations. We compared the genetic structure defined by both datasets by analyzing 29 populations that were shared between them (Figure S1.b). First, we performed a principal component analysis (PCA) on allelic counts between populations for each dataset. Based on the first four principal components (PC), we used the Ward algorithm to cluster all populations (Figure S2). Then, we used Adegenet (Jombart 2008) to estimated the Nei's pairwise genetic distance between populations and Ape (Paradis 2012) to perform and compare the Neighbor Joining (NJ) trees.

Finally, for both datasets we used an Approximate Bayesian Computation to compare three models of divergence of teosinte subspecies to estimate demographic parameters. These analyses were targeted to interpreting the history of Guerrero populations. To do this, we classified the populations as *mexicana* (M), Guerrero (G) and

Balsas (B). For the demographic inferences, we removed the very divergent populations from the lowlands of Jalisco state (Aguirre-Liguori *et al.* 2017). We decided to do this to simplify the models, and because these populations represent a very differentiated group of populations, with very high F_{ST} compared to other *mexicana* and *parviglumis* populations. For each dataset, we used DADI (Gutenkunst, Hernandez, Williamson, & Bustamante, 2009) to extrapolate the final dataset into a folded multidimensional site frequency spectrum (SFS) of M, G and B groups. The observed SFS was then used to compare three models of divergence of Guerrero populations and teosintes subspecies. We used Fastsimcoal2 (Excoffier, Dupanloup, Huerta-Sánchez, Sousa, & Foll, 2013; Excoffier & Foll, 2011) to obtain the maximum likelihood of each of the three models: the Ancestral, Admixture and Intermediate models (Figure 1A). For the Ancestral model, we forced Balsas populations to have an earlier origin than *mexicana* populations. For the Admixture model, we constrained the origin of *mexicana* to occur before the origin of Guerrero populations. Finally, for the Intermediate model, we conditioned the origin of *mexicana* to be posterior to the origin of Guerrero populations.

For each model, we estimated effective population sizes (N_e), migration rates (m) and time of origin of the Guerrero populations (T). Prior information of time of origin, effective population sizes and migration events were based on Ross-ibarra et al. (2009). More specifically, we limited the time of divergence (T) of the three subspecies relative to the origin of *parviglumis* (140,000 generations ago); N_e of each taxon to 1,000,000; and migration rates (m) to 0.0001-0.1. All the priors were set to log uniform.

For each of the demographic models, we performed 15 replicates of Fastsimcoal, and allowed each replicate to run for 500,000 generations. For each model, the 15 replicates converged to similar likelihood values, indicating good model performance. We therefore combined all the parameters from each replicate and retained the parameters with the highest 5% of the maximum likelihood. Since we have the same number of parameters in each of the model, we compared the models directly based on their maximum likelihood.

Candidate selected SNPs: We used both the maize 50K and DTS datasets to test for selection by detecting outlier SNPs that have an association with each of the bioclimatic and soil variables. To test for associations, we relied on the detection of shared outlier

SNPs between two methods: *bayescenv* (Villemereuil, Frichot, Bazin, Fran, & Gaggiotti, 2014), and F_{ST} values within the top 5% of the F_{ST} distributions (Funk et al., 2016). In bayescenv, F_{ST} is measured by three components. The β component corresponds to the shared differentiation between populations (neutral SNPs), the α component correspond to specific local effects (outlier SNPs) and γ refer to environmental effects (outlier SNPs that also have an association with environment). For *bayescenv*, we set the "pr_jump" prior to π= 0.5, the "pr_pref" prior to p=0.7, and the upper bounds of γ and α to 10 and -1, respectively.

For each of the 4 climatic PCs, we tested 49 populations in the 50K dataset and 47 in the DTS datateset. For the 3 soil variables, we examined 15 populations in the 50K dataset and 22 in the DTS dataset. For all *bayescenv* analyses, we ran 20 pilot runs of 5000 iterations and a burn-in of 50,000 iterations, ultimately obtaining 5000 MCMC iterations for each environmental variable. For each *bayescenv* run, three sets of SNPs were identified based on Posterior Error Probability (PEP): the β SNPs that showed no association with environmental variables, which we denote 'neutral' SNPs; the α-SNPs that were detected as outliers but had PEP_ α < PEP_ γ; and, finally, the γ -SNPs that had had PEP_ γ < PEP_ α.

The γ -SNPs contain SNPs with potential associations with the environment, but they also likely contain many false positives. We further reduced the set of γ -SNPs to include only those that were not significantly associated with geographic distance. To identify these SNPs, we performed a multiple regression on distance matrices (MRM) using the R package *ecodist* (Goslee & Urban, 2007) for each outlier γ -SNPs detected in Bayescenv. The distance matrices for *ecodist* consisted of allele frequency distances (see Andrew & Rieseberg 2013), geographic distances and environmental distances among populations. The allele frequency distance was used as the response variable and the environmental and geographic distances were used as predictive variables. We performed 1000 iterations of the MRM model to obtain the significance of the geographic and ecological matrixes on the genetic distance. We retained the subset of γ -SNPs that had a significant effect of the environmental variable ($p \leq 0.01$), and a non-significant effect of the geographic variable ($p > 0.05$). Finally, given this last set of SNPs, we retained only those SNPs that were contained within the top 5% of the F_{ST} distributions (as in Funk *et al.*

2016). The γ -SNPs that passed both the MRM and F_{ST} tests were considered to be candidate selected SNPs.

To sum: our analyses identified four sets of SNPs: *i*) neutral SNPs (β-SNPs) that showed no association with environment in *bayescenv* analyses *ii*) outlier SNPs (α-SNPs) that were detected as outliers by *bayescenv*, *iii*) candidate-selected SNPs, which were detected in the γ model of *bayescenv*, fell within the top 5% of F_{ST} values and were also significantly correlated with environment by the MRM test, and, finally, *iv)* the remaining γ-SNPs that did not pass the MRM test. We call this last category the γ-non-candidate SNPs.

Pairwise Population Analyses: We calculated the Nei genetic distances between populations based on each of the four SNP categories. We also calculated pairwise environmental and geographic distances between populations. With these matrices in hand, we performed linear regression analyses between the geographic and genetic distances and also between the environmental and genetic distances. We reasoned that if the environmental association to genetic distance was stronger than the geographic association, then the set of SNPs was affected by Isolation by Environment, and not by Isolation by Distance. We corroborated these results with the R package BEDASSLE (Bradburd, Ralph, & Coop, 2013). BEDASSLE uses a Bayesian approximation to compare the effects of the geographic and environmental matrix on the genetic matrix, ultimately estimating the ratio of their effects (Bradburd et al., 2013). For BEDASSLE analyses, we used the over-dispersion model (MCMC_BB) as in Bradburd *et al.* (2013). We manually modified the tuning parameters, until we found acceptance rates between 0.2 and 0.8 and good posterior probability functions, as suggested by the authors (Bradburd, Ralph, & Coop, 2013). We ran the model for 15,000,000 generations for each of the samples analyzed. Finally, we confirmed that for each parameter in the model the posterior distribution had *caterpillar* trace plots, indicating good performance of the analyses.

Islands of Genomic Differentiation: We analyzed genomic differentiation between *parviglumis* and *mexicana*, with the goal of inferring whether candidate-selected SNPs were located within genomic regions of high divergence. To identify regions of high

divergence, we first plotted the locus-by-locus F_{ST} for all SNPs along the 10 teosinte chromosomes. We then used a Hidden Markov Model to identify blocks of contiguous differentiation (Hofer, Foll, & Excoffier, 2012). To run the HMM, we transformed F_{ST} values to logit(F_{ST}), assumed three states of divergence (hidden states) with a Gaussian distribution and with the mean and standard deviation estimated from the data. The HMM restricted transitions matrices between high differentiation and low differentiation states (Baum-Welch Algorithm); the states of divergence were estimated with the Viterbi algorithm (Hofer *et al.* 2012, Soria-Carrasco et al., 2014). These test were performed modifying the code of Soria-Carrasco et al. (2014) to accommodate our data and using the R package HiddenMarkov (Harte, 2016). Once we identified the blocks of high differentiation, we counted the number of blocks, estimated their mean F_{ST} and mapped them to the locus-by-locus F_{ST} plots. We defined as putatively islands of divergence those high divergence blocks that further had a high enrichment of candidate SNPs.

RESULTS

Data: To investigate genomic patterns of potentially ongoing ecological speciation between teosinte subspecies, we assembled two datasets. The DTS dataset, which was produced for this study, was based on reduced representation sequencing. After controlling for duplicated regions, missing and monomorphic data, we identified 9,780 DTS SNPs from 47 populations, including 24 of *mexicana* and 23 populations of *parviglumis* (Figure S1, Table S1). The number of SNPs per chromosome ranged from 144 on chromosome 9, to 1,751 on chromosome 1. The 50K SNP chip dataset was a synthesis of two previously published datasets (Pyhajarvi et al. 2013, Aguirre-Liguori et al. 2017). After filtering, the 50K dataset contained 33,454 SNPs from 25 populations of *mexicana* and 24 populations of *parviglumis* (Figure S1, Table S1).

The two datasets have different advantages. The DTS dataset was probably more suitable for demographic inference, because there was no ascertainment bias. Conversely, the 50K chip data had more SNPs, fewer missing data and therefore likely high power to detect selection. Fortunately, both datasets exhibited qualitative agreement for all of our analyses. For simplicity, in some sections we report results that are based on one data set, but results from the alternative data set are also available in supporting materials.

Genetic structure and demography: To begin to untangle the demographic history of teosintes, we first focused on the 29 populations that were shared between the DTS and the 50K dataset (Figure S1b). We performed PCA and built NJ trees. The analyses reveal concordant, but complex, genetic structure (Aguirre-Liguori et al., 2017; Fukunaga et al., 2005; Moeller, Tenaillon, & Tiffin, 2007; Pyhäjärvi et al., 2013). Similar to previous studies (Pyhajarvi et al. 2013, Aguirre-Liguori et al. 2017), the results indicated that teosintes can be divided into subspecies *mexicana* and *parviglumis,* (Figure S2). Subspecies *mexicana* represented one clearly differentiated genetic group, but *parviglumis* was divided into three groups. One group, which we called Balsas, consisted of populations from the Mexican states from Guerrero, Michoacán, Oaxaca and Jalisco. Another group, which we called Guerrero, was sampled exclusively from the state of Guerrero and fell intermediate between *mexicana* and the remainder of *parviglumis* populations. The final group, called Jalisco, came from the lowlands of Jalisco (Sánchez-González *et al.* 1998; Aguirre-Liguori *et al.* 2017) and was widely divergent based on PCA.

Guerrero populations had previously been noted as potentially admixed (Pyhajarvi *et al.* 2013) or perhaps ancestral to Balsas and *mexicana* populations (Fukunaga *et al.* 2005). To test these models explicitly – and to gain additional insights into the tempo of taxon divergence - we compared three demographic scenarios using Fastsimcoal. The Ancestral model treated Guerrero as the ancestral population, from which Balsas diverged earlier than *mexicana* populations (Figure 1A). The Admixture model constrained the origin of *mexicana* to occur before the origin of admixed Guerrero populations. Finally, the Intermediate model treated Balsas populations as ancestral to both Guerrero populations and *mexicana* (Figure 1A); like the Ancestral model, it constrained the origin of *mexicana* to occur after the divergence of Balsas and Guerrero. For each model, the optimized parameters were effective population sizes (N_e), migration rates (m) and times of divergence (T). For these demographic inferences, we did not consider the highly divergent lowland Jalisco populations (Figure S2).

For both the DTS and 50K datasets, we found that the Ancestral model had the highest likelihood, indicating that Guerrero populations were likely ancestral to both Balsas and *mexicana* populations (Figure 1B), as proposed by Fukunaga *et al.* (2005). While both

datasets indicate that the ancestral model has the highest likelihood, the parameter estimates differed somewhat (Table 1; Figure 1B). For example, the divergence time of *mexicana* and Guerrero populations was estimated to be 15.5k (12k-72k) generations (Table 1) for the DTS data, but 11.9k (11k-22k) generations for the 50K dataset (Table 1). Importantly, the models suggest that the divergence of groups occurred in the presence of gene flow. Migration was generally estimated to be low between groups, such that the number of migrants per generation, N_em, was < 1.0. However, N_em estimates reached or exceeded 1.0 for directional migration from Balsas into Guerrero, from *mexicana* into Balsas and from Balsas and *mexicana* based on both DTS and 50K (Figure S3). We also employed Treemix analyses to investigate gene flow (Figure S4). Treemix is based on populations, as opposed to groups, so it is difficult to compare fastsimcoal and Treemix results directly. The Treemix analyses did, however, confirm the inference that migration occurs between populations of *Mexicana* and Jalisco and Balsas and Jalisco populations, but is absent from populations of *mexicana* to Guerrero, and *mexicana* to Balsas. Overall, Treemix results support that Guerrero does not have a hybrid origin between *mexicana* and *parviglumis.*

The important point from these demographic analyses were that: *i*) three groups were detected within *parviglumis*, with one group (Jalisco) highly diverged, confirming previous results (Moeller et al. 2005; Pyhäjärvi et al. 2013; Aguirre-Liguori et al. 2017), *ii*) the divergence among some of these groups has occurred in the presence of gene flow (Fukunaga et al., 2005; van Heerwaarden et al., 2011) and *iii*) the timeframe of *mexicana* divergence from *parviglumis* populations was on the order of ~11.0k to ~15.5k years.

The temporal and ecological framework of divergence: To help understand potential climactic factors that may have played a role in the divergence of *mexicana* and *parviglumis* between 11.0k and 15.5k years ago, we used BioClim data to estimate the average mean temperature ($\bar{C}°$) of current teosinte locations. These data clearly indicate that *parviglumis* and *mexicana* inhabit regions that differ in temperature (Figure 2A). We further projected BioClim to the time of the last maximum glacial (LGM), ca. 21.0k years ago (Figure 2B). These projections suggest that the current locations of *parviglumis* were significantly cooler during the LGM (Fig. 2B, $p < 0.001$). Interestingly, the current $\bar{C}°$ of *mexicana* populations was similar to those inferred of *parviglumis* populations during the

LGM (Figure 2B). The significant overlap in $\bar{C}°$ between past Guerrero populations and current *mexicana* populations are consistent with the inference that *mexicana* diverged from the Guerrero cluster, (Figure 1B) and further suggests that *parviglumis* has had more recent adaptation to temperature in the Guerrero region during the LMG.

Since *parviglumis* and *mexicana* currently grow at different $\bar{C}°$, we investigated their association with additional climatic factors. We performed a PCA on a total of 19 climate factors and retained the first four PCs, which contained 87.95% of the environmental variance between populations (Table S1, Table S2). The first two PCs corresponded to variables associated with temperature and precipitation, respectively. The third and fourth PCs corresponded to variables associated with seasonality. Subspecies *mexicana* and *parviglumis* populations differed only for PC1 (Figure 2A, $p<0.0001$), with *mexicana* again associated with cooler temperatures. In addition to bioclimatic variables, we measured the mean pH, P and N concentration in soil samples from 11 populations of *mexicana* and 11 populations of *parviglumis* for which we have DTS data (Table S1). We found that *parviglumis* and *mexicana* populations differed significantly in P concentration (Figure 2A, $p=0.0158$), with *parviglumis* present within soils that had lower concentrations. There was a tendency for N and pH to be higher in *parviglumis* than in *mexicana* populations, but the differences were not significant (Figure S5). Overall, we found that *mexicana* likely diverged from *parviglumis* during a cooler period and that the environments typical for the two subspecies were differentiated for both $\bar{C}°$and P concentration.

Candidate Selected SNPs: Our results suggest that *mexicana* and *parviglumis* diverged in the presence of gene flow and also that divergence included adaptation to different temperatures and soil P content. In order to assess whether selection acted on a few or many regions along the genome, we first had to identify candidate-selected SNPs (see Materials and methods). We used all of the teosinte populations for these analyses.

We detected candidate-selected SNPs that have associations with the four PCs from the BioClim data and to the three soil variables (pH, N and P). We found concordant patterns in the 50K and the DTS dataset but, as expected, the number of candidate SNPs was higher for the 50K dataset than fo the DTS (253 vs. 157 SNPs, respectively; Figure 3A,

Figure S6). SNP associations occurred for all 7 variables but were skewed to PC1 and P concentration (Figure 3A). PC1 and P concentration were associated with 128 (50.6%) and 66 (26.1%) of the 50K candidate SNPs. Of these, 20 SNPs where shared between PC1 and P concentration. Accordingly, these two variables had the strongest association, measured as R^2, with allelic frequencies across associated SNPs (Figure 3B, Figure S6).

We compared the effect of geographic distance and environmental distance on genetic distances (see Materials and methods) for the 50K dataset. If the fit (R^2) of the environmental distance between populations was higher than the fit of the geographic distance between populations, we considered the SNPs as being differentiated by environment. We performed these analyses for all four classes of SNPs (i.e., α-SNPs, β-SNPs, γ–non-candidate SNPs, and candidate SNPs) with PC1, PC2 and P concentration; the other variables had too few (< 20) associated SNPs.

As expected, we found that neutral β-SNPs always display a stronger association with geography than environment (Figure 4, Table 2). In contrast, we found that candidate SNPs always had a stronger association to environment than geography (Figure 4, Table 2), which is again expected given that they were identified by their association with environmental correlates. For PC1, PC2 and P concentration we found higher mean Nei's genetic distances among populations (Table 2) for γ-candidate and outlier SNPs, followed by neutral β-SNPs and γ-non-candidate SNPs (Figure 4, Table 2).

Finally, we found in the R^2 analyses (Table 2) that environmental distance had a stronger association on γ-non-candidate SNPs (environmental outliers) for PC1 and PC2, but not for t P soil concentration. These patterns indicate that non-candidate but still outlier SNPs in PC1 and PC2 are affected by Isolation by environment (IBE). The results were supported by BEDASSLE analyses, for which we found an effect of environmental distance over the geographic distance in neutral β-SNPs (aE/aD=0.085) and in candidate SNPs (aE/aD=0.8) for PC1; but only an effect in candidate SNPs (aE/aD=76) for P concentration in the soil (compared to aE/aD=0.01 for neutral SNPs). There were too few candidate SNPs associated to PC2 to run this analyses.

Based in the comparison of the association with environment and geography, we found for the 50K dataset that 7.17%, 2.84% and 0.20% of loci had an association with PC1, PC2 and P concentration in the soil (Table 2), respectively. However, some of these

SNPs were associated to more than one climatic variable. After removing SNPs that were detected by more than one variable (*unique* function in R), we found that for the 50K dataset 8.95% of SNPs had a stronger association with environment.

Signatures of Genomic Differentiation: To assess the possibility of genomic islands, we plotted for the 50K dataset the locus-by-locus F_{ST}, as well as the chromosomal position of candidate SNPs. A qualitative interpretation shows that candidate SNPs occur all along the genome (Figure S7). Although there are regions that have a larger enrichment of candidate SNPs (Figure 5, Figure S7). To identify if these occur in "islands of differentiation", we pursued a Hidden Markov Model (HMM) to identify continuous blocks of high differentiation along the chromosomes of teosintes (Hofer et al., 2012; Riesch et al., 2017; Soria-Carrasco et al., 2014). We identified 670 block of high differentiation along the genome, covering 3.1% of the genome, and presenting a mean F_{ST} of 0.46 (the rest of the genome has mean F_{ST} of 0.23). The size of these blocks varied in size and in composition of number of candidate SNPs associated to them, suggesting that teosintes are at intermediate stages of the speciation continuum (Nosil, 2012).

Based on the overlap between blocks of high differentiation and a high enrichment of candidate SNPs, we identified interesting "islands of differentiation" in teosintes (Figure 5, grey rectangles; Figure S7). Two of these "islands" corresponded to previously described chromosomic inversions in chromosomes 1 (*Inv1n* (Fang et al., 2012; Pyhäjärvi et al., 2013)), and in chromosome 9 (*Inv9e* (Aguirre-Liguori et al., 2017; Pyhäjärvi et al., 2013)). The other two long blocks are in chromosomes 4 and 8 (Figure 5; Figure S7). The length of these regions varies in size and the composition of candidate SNPs, with larger regions presenting SNPs associated to all the environmental variables (Figure 5; Figure S7), and smaller inversions presenting SNPs associated with one environment (e.g., chromosomes 4 was associated with P concentration and chromosome 8 was associated with PC1). Overall, these long blocks of high F_{ST} cover > 85 MBP, or ~4% of the genome. The long blocks of differentiation that are not enriched with candidate SNPs, could represent high F_{ST} by changes in genomic architecture (as in chromosome 5, the block correspond to the centromere) and certainly could be under selection to other non-measured pressures (biotic or abiotic).

DISCUSSION

The teosintes have become a model system for population and ecological genomics (Hufford and Ross-Ibarra, 2012), but it has also become clear that teosinte populations have a complex demographic history (Fukunaga et al. 2005; Pyhäjärvi et al. 2013; Aguirre-Liguori et al. 2017). It is likely that not considering this complex structure in demographic inferences could be biasing our estimations of the time and mode of divergence of teosinte subspecies (Hufford et al. 2012). In this study, we used model simulations, environmental information and genome scans to study the time and mode of the divergence of teosinte subspecies.

The comparison between three models of divergence indicates that Guerrero populations were ancestral to Balsas and *mexicana* (Figure 1, Table 1) and that divergence between genetic groups occurred in the face of limited gene flow (Figure S3, Figure S4). This model was suggested, although not tested, by Fukunaga et al. (2005), and it is supported by the fact that Guerrero populations have high genetic diversity and share SNPs between *mexicana* and Balsas group. While it has been hypothesized that Guerrero populations could be admixed between both subspecies (Pyhäjärvi et al. 2013), we did not find any support from Treemix or Fastsimcoal suggesting that ancestral gene flow between *mexicana* and *parviglumis* has occurred in Guerrero (Figures S3 and S4). Furthermore, our results indicate that gene flow has been more frequent between *mexicana* and Jalisco than between *mexicana* and *parviglumis*. Perhaps the signals of admixture found by Pyhäjärvi *et al.* (2013) could correspond to shared ancestral polymorphism between groups, that has not been lost given the recent history of teosintes and their high effective population sizes (Ross-Ibarra *et al.* 2009).

In addition, our simulations indicate that the divergence of *mexicana* occurred between 11,000 and 15,500 generations ago, depending on the genetic dataset used (Table 1). This is four to six times earlier that the 60,000 generations ago suggested before (Ross-Ibarra et al. 2009), further suggesting that not considering the complex genetic structure could have biased previous demographic inferences (Hufford, Martínez-Meyer, Gaut, Eguiarte, & Tenaillon, 2012). As pointed out by Hufford *et al.* (2012b), the populations used by Ross-Ibarra et al. (2009) included many populations from the Guerrero area, but did not considered the complex genetic structure in Guerrero populations. They also

sequenced fewer regions (26), with fewer individuals (67). Intriguingly, the admixture model finds that *mexicana* was originated 60,000 years ago, supporting the argument that the time of origin could have been overestimated by not appropriately considering the genetic structure of *parviglumis* populations.

Teosintes have annual cycles, thus the time origin of *mexicana* would have occurred after the LMG, between 11,000 and 15,500 years ago. We compared the mean temperature of teosintes populations during the LGM and the current conditions. While currently, teosintes can ecologically be differentiated as warm teosintes (Balsas and Guerrero) and cold teosintes (*mexicana*), we found that during the LGM, Guerrero populations had a similar temperature to current *mexicana* populations (Figure 2B). This supports the model analyses that indicate that *mexicana* originated from Guerrero populations. According to Aguirre-Liguori *et al.* (2017), teosintes show increased signals of local adaptation to the niche limit, which could initiate ecological speciation (Eckert, Samis, & Lougheed, 2008; Sexton, McInyre, Angert, & Rice, 2009). If so, local adaptation to a cooler edge niche could have generated the divergence of teosintes during the LGM in the Guerero area (Aguirre-Liguori *et al.* 2017). In this scenario, *mexicana* populations would have adapted from Guerrero during a cooler period and migrated to higher lands during the Holocene where their initial ecological conditions existed.

We used genome scans and environmental information to test whether the divergence of teosintes might have occurred by divergent adaptation to cold environments during and after the LGM. For both the DTS and 50K datasets, we found signals of selection along four climatic and three soil variables tested (Figure 3). Interestingly, these signals are skewed for SNPs associated with PC1, which is defined primarily by temperature, and SNPs associated to P concentration in the soil. These are the only variables that differentiate significantly both teosintes subspecies (Figure 2A), suggesting that in teosintes divergent selection is mainly driven by strong ecological divergence (Egan, Nosil, & Funk, 2008; D. J. Funk, 2010; Liu et al., 2013; Nosil & Crespi, 2006; Tobler & Carson, 2010). Furthermore, our results confirm that divergence of *mexicana* occurred mainly by adaptation to a cooler environment (during the LGM) and show that adaptation to a different availability of P in the soil has also been important. While phosphorus concentration in the soil is higher in *mexicana*, these populations grow in volcanic soils,

which have strong P-fixing capacity, resulting in a lower availability of P for the plant (Tening, Foba-Tendo, Yakum-Ntaw, & Tchuenteu, 2013; Krasilnikok et al., 2013). Fustier-Allison *et al.* (2017) used genome scans and a large genome dataset (>8,000,000 SNPs), and found also many candidate genes associated to low phosphorus availability.

Our results show that *mexicana* has diverged along two main environmental axis and other five less important axis (Figure 3). This suggests that multifarious selection has driven the ecological divergence of teosintes. The multifarious selection hypothesis argues that reproductive isolation is more likely to occur if selection occurs along multiple niche axis, increasing the number of genomic islands and favoring the congealing of the genome (Jeffrey L. Feder et al., 2013, 2012; Flaxman et al., 2013, 2014; Michel et al., 2010; Nosil, Funk, et al., 2009; Nosil, Harmon, et al., 2009). We analyzed the location of candidate SNPs along the genome to define whether they are located at single regions or clustered within specific segments. We found that candidate SNPs are distributed either individually along the entire genome, or clustered within high F_{ST} regions (Figure 5; Figure S7). In particular we found many candidate SNPs clustered in four large high F_{ST} regions identified by our HMM. These high F_{ST} regions have been previously described in teosintes and maize and have been described as putative inversions (Fang et al. 2013; Pyhajarvi et al. 2013; Aguirre-Liguori et al. 2017; Fustier et al. 2017). In particular, two of these putative inversions (*Inv1n* and *Inv9e*) have been found to be enriched for candidate SNPs associated to altitude and other climatic and soil variables (Fang *et al.* 2013, Pyhäjärvi et al. 2013; Aguirre-Liguori *et al.* 2017). Here, we found that these inversions contain 27% of the 50K candidate SNPs.

Increasing evidence shows that inversions enriched with adaptive SNPs can be important for local adaptation and ecological divergence, because they may reduce recombination between divergently adapted populations, restrict gene flow for important genes, and limit formation of mal-adapted hybrids (Fishman, Stathos, Beardsley, Williams, & Hill, 2013; He & Knowles, 2016; Kirkpatrick & Barton, 2006; Lowry & Willis, 2010; Twyford & Friedman, 2015). We analyzed patterns of LD within and between inversions or blocks to identify whether these putative inversions could be suppressing recombination (and therefore gene flow) between populations. We found that LD along differentiation blocks is very low, with mean r^2 values ranging from 0.025 in inversion 9 to 0.28 in block 8

(Figure 6, Figure S8). This is consistent with previous studies in teosintes (Aguirre-Liguori et al., 2017; Fustier et al., 2017), and suggests that recombination is common between populations adapted to similar environments is common. However, when we analyzed LD between candidate SNPs within the high differentiation regions, we found that it is always significantly stronger than LD between neutral SNPs ($p<0.0001$; Figure 6, Figure S8). These differences between neutral and candidate SNPs within the inversion suggest that within-inversion selection could be maintaining some allelic combinations (Kim & Nielsen 2004; Nosil *et al.* 2008; Andrew & Rieseberg 2013).

Such type of statistical linkage between candidate SNPs is expected to potentiate the congealing of the genome by generating longer genomic hitchhiking along the genome (Feder *et al.* 2012, 2013, 2014b, Flaxman *et al.* 2013, 2014). Therefore, we further analyzed whether selection has generated patterns of isolation by environment (IBE). IBE is concept analogous to isolation by distance, in which populations that are adapted to similar conditions show lower genetic divergence (Nosil *et al.* 2008). Moreover, it is expected that similar patterns of differentiation along neutral and selective datasets would indicate that isolation by environment is affecting the entire genome, and therefore that divergent adaptation is contributing to reproductive isolation (Jeffrey L. Feder et al., 2014; Flaxman et al., 2013, 2014; Nosil, Funk, et al., 2009; Rice et al., 2011; Shafer & Wolf, 2013). In contrast, localized signatures of differentiation would indicate that selection is occurring on individual SNPs, suggesting they are under putative local adaptation (Malinsky et al., 2015).

While a vast majority of β-SNPs (neutral category) exhibit patterns of IBD, ~9% of the genome seems to be affected by environmental variation (Figure 4, Table 2). This percentage is similar to other species that are under ecological divergence, in which neutral loci are affected by linkage to selected SNPs (see Nosil *et al.* 2008 for a review, 2009a; Funk *et al.* 2011). Overall, we found that PC1 and PC2 have affected the genetic divergence between populations for candidate and non-candidate SNPs, indicating that they are responsible for the ecological divergence between subspecies (Keller & Seehausen, 2012; Shafer & Wolf, 2013). For P concentration in the soil, we found that IBA is only evident for candidate SNPs and not extended genomic islands; it is therefore possible that adaptation to P has contributed to local adaptation. Since genomic divergence occurs in

later stages of the ecological speciation continuum, we believe that teosinte populations in Guerrero started diverging by changes associated mainly to temperature and later started adapting to different availability of phosphorus in the soil. According to potential distribution (niche) results and demographic inferences, *mexicana* populations diverged after the LGM, when conditions were cooler. Ancestral *mexicana* populations may have migrated to higher lands after the LMG and encountered volcanic soils that have reduced availability of P (Tening et al., 2013; Krasilnikov et al. 2013). Under this scenario, selection to this new environmental axis is expected to increased LD along the genome by decreasing the fitness of immigrants or hybrids and increase RI between populations (Feder et al., 2014,; Flaxman et al., 2013, 2014; Via, 2009).

There is an ongoing debate about the importance of gene flow during ecological speciation, since in theory low amounts of gene flow could be enough to counteract selection (Rundle and Nosil 2005; Nosil 2012). It has been suggested that important factors that would facilitate the completion of reproductive isolation are: 1) strong ecological divergence (D. J. Funk et al., 2006; Nosil, 2012; Rundle & Nosil, 2005; Schluter & Conte, 2009); 2) reduced geographic overlap between divergent species (Aguilée et al., 2011; Nosil, 2012; Nosil et al., 2012; Surget-Groba et al., 2012) and; 3) selection acting along multiple niche axis- the multifarious selection hypothesis (Chevin et al., 2014; Jeffrey L. Feder et al., 2014; Flaxman et al., 2014; Malinsky et al., 2015; Nosil, Funk, et al., 2009; Nosil, Harmon, et al., 2009). Teosintes meet all of these criterions and it is very likely that they could be under an ecological speciation process, as *mexicana* and *parviglumis* have well defined geographic and ecological distributions, and low areas of overlap (van Heerwaarden et al. 2011; Hufford et al. 2012). Our results further show that signals of selection occur along many environmental axis, two of which show strong ecological divergence.

Acknowledgements:

This work is presented in partial fulfillment of the requirements to obtain a PhD degree by Jonás A. Aguirre-Liguori, who thanks the "Programa de Doctorado en Ciencias Biomédicas, Universidad Nacional Autónoma de México (UNAM)" and thanks the scholarship provided by the Consejo Nacional de Ciencia y Tecnología (CONACYT, grant no. 255770). This work was supported by grant CB2011/167826 (CONACYT Investigación Científica Básica) to LE, grant CN-10-393 (UC MEXUS-CONACYT) to LE and BSG, and grant M12-A03 ECOS Nord France -CONACYT-ANUIES 207571 to LE and MIT. The paper was in part written during a sabbatical leave of LEE and in the University of Minnesota in Peter Tiffin and Michael Travisano laboratories, with support by scholarships from PASPA, DGAPA, UNAM. In addition, we are grateful to E. Aguirre-Planter, L. Espinosa-Asuar and M. Rosas-Barrera for technical support; Rodrigo Velázquez-Durán for assisting with chemical analysis; and Ricardo Colín, Enrique Scheinvar, Gabriel Merino and Jaime Gasca and Valeria Souza for seed collection. We thank David Romero Camarena for comments on the experimental design and reviewers for earlier comments of the manuscript.

LITERATURE

Aguilée, R., Lambert, A., & Claessen, D. (2011). Ecological speciation in dynamic landscapes. *Journal of Evolutionary Biology*, *24*(12), 2663–2677. doi:10.1111/j.1420-9101.2011.02392.x

Aguirre-Liguori, J. A., Aguirre-planter, E., & Eguiarte, L. E. (2016). Genetics and ecology of wild and cultivated maize: domestication and introgression. In *Ethnobotany of Mexico*. New York: Springer.

Aguirre-Liguori, J. A., Tenaillon, M. I., Vázquez-Lobo, A., Gaut, B. S., Jaramillo-Correa, J. P., Montes-Hernandez, S., … Eguiarte, L. E. (2017). Connecting genomic patterns of local adaptation and niche suitability in teosintes. *Molecular Ecology*.

Andrew, R. L., & Rieseberg, L. H. (2013). Divergence is focused on few genomic regions early in speciation: Incipient speciation of sunflower ecotypes. *Evolution*, *67*(9), 2468–2482. doi:10.1111/evo.12106

Antonovics, J. (2006). Evolution in closely adjacent plant populations X: long-term persistence of prereproductive isolation at a mine boundary. *Heredity*, *97*(1), 33–37. doi:10.1038/sj.hdy.6800835

Antonovics, J., & Bradshaw, A. D. (1970). Evolution in closely adjacent plant populations. *Heredity*, *25*(3), 349–362. doi:10.1038/hdy.1978.44

Arnegard, M. E., McGee, M. D., Matthews, B., Marchinko, K. B., Conte, G. L., Kabir, S., … Schluter, D. (2014). Genetics of ecological divergence during speciation. *Nature*, *511*(7509), 1–17. doi:10.1038/nature13301

Bolnick, D. I. (2011). Sympatric speciation in threespine stickleback: Why not? *International Journal of Ecology*, *2011*. doi:10.1155/2011/942847

Bradburd, G. S., Ralph, P. L., & Coop, G. M. (2013). Disentangling the effects of geographic and ecological isolation on genetic differentiation. *Evolution*, *67*(11), 3258–3273. doi:10.1111/evo.12193

Chevin, L. M., Decorzent, G., & Lenormand, T. (2014). Niche dimensionality and the genetics of ecological speciation. *Evolution*, *68*(5), 1244–1256. doi:10.1111/evo.12346

Chia, J.-M., Song, C., Bradbury, P. J., Costich, D., de Leon, N., Doebley, J. F., … Ware, D. (2012). Maize HapMap2 identifies extant variation from a genome in flux. *Nature Genetics*, *44*(7), 803–807. doi:10.1038/ng.2313

CIMMYT. (2005). *Laboratory Protocols: CIMMYT Applied Molecular Genetics Laboratory*.

De Mita, S., Thuillet, A. C., Gay, L., Ahmadi, N., Manel, S., Ronfort, J., & Vigouroux, Y. (2013). Detecting selection along environmental gradients: Analysis of eight methods and their effectiveness for outbreeding and selfing populations. *Molecular Ecology*, *22*(5), 1383–1399. doi:10.1111/mec.12182

Defaveri, J., Jonsson, P. R., & Merilä, J. (2013). Heterogeneous genomic differentiation in marine threespine sticklebacks: Adaptation along an environmental gradient.

Evolution, *67*(9), 2530–2546. doi:10.1111/evo.12097

Díez, C. M., Gaut, B. S., Meca, E., Scheinvar, E., Montes-Hernandez, S., Eguiarte, L. E., & Tenaillon, M. I. (2013). Genome size variation in wild and cultivated maize along altitudinal gradients. *The New Phytologist*, *199*(1), 264–76. doi:10.1111/nph.12247

Eckert, C. G., Samis, K. E., & Lougheed, S. C. (2008). Genetic variation across species' geographical ranges: The central-marginal hypothesis and beyond. *Molecular Ecology*, *17*(5), 1170–1188. doi:10.1111/j.1365-294X.2007.03659.x

Egan, S. P., Nosil, P., & Funk, D. J. (2008). Selection and genomic differentiation during ecological speciation: Isolating the contributions of host association via a comparative genome scan of Neochlamisus bebbianae leaf beetles. *Evolution*, *62*(5), 1162–1181. doi:10.1111/j.1558-5646.2008.00352.x

Egan, S. P., Ragland, G. J., Assour, L., Powell, T. H. Q., Hood, G. R., Emrich, S., … Feder, J. L. (2015). Experimental evidence of genome-wide impact of ecological selection during early stages of speciation-with-gene-flow. *Ecology Letters*, *18*(8), 817–825. doi:10.1111/ele.12460

Elias, M., Faria, R., Gompert, Z., & Hendry, A. P. (2012). Factors influencing progress toward ecological speciation. *International Journal of Ecology*, *2012*(i). doi:10.1155/2012/235010

Elser, J. J., Fagan, W. F., Kerkhoff, A. J., Swenson, N. G., & Enquist, B. J. (2010). Biological stoichiometry of plant production: Metabolism, scaling and ecological response to global change. *New Phytologist*, *186*(3), 593–608. doi:10.1111/j.1469-8137.2010.03214.x

Excoffier, L., Dupanloup, I., Huerta-Sánchez, E., Sousa, V. C., & Foll, M. (2013). Robust Demographic Inference from Genomic and SNP Data. *PLoS Genetics*, *9*(10). doi:10.1371/journal.pgen.1003905

Excoffier, L., & Foll, M. (2011). fastsimcoal: A continuous-time coalescent simulator of genomic diversity under arbitrarily complex evolutionary scenarios. *Bioinformatics*, *27*(9), 1332–1334. doi:10.1093/bioinformatics/btr124

Excoffier, L., & Ray, N. (2008). Surfing during population expansions promotes genetic revolutions and structuration. *Trends in Ecology and Evolution*, *23*(7), 347–351. doi:10.1016/j.tree.2008.04.004

Fang, Z., Pyhäjärvi, T., Weber, A. L., Dawe, R. K., Glaubitz, J. C., Sánchez González, J. de J., … Ross-Ibarra, J. (2012). Megabase-scale inversion polymorphism in the wild ancestor of maize. *Genetics*, *191*(3), 883–894. doi:10.1534/genetics.112.138578

Favre, A., Widmer, A., & Karrenberg, S. (2016). Differential adaptation drives ecological speciation in campions (Silene): evidence from a multi-site transplant experiment. *New Phytologist*. doi:10.1111/nph.14202

Feder, J. L., Flaxman, S. M., Egan, S. P., Comeault, A. a., & Nosil, P. (2013). Geographic Mode of Speciation and Genomic Divergence. *Annual Review of Ecology, Evolution, and Systematics*, *44*(October), 73–97. doi:10.1146/annurev-ecolsys-110512-135825

Feder, J. L., Gejji, R., Powell, T. H. Q., & Nosil, P. (2011). Adaptive chromosomal divergence driven by mixed geographic mode of evolution. *Evolution*, *65*(8), 2157–

2170. doi:10.1111/j.1558-5646.2011.01321.x

Feder, J. L., Gejji, R., Yeaman, S., & Nosil, P. (2012). Establishment of new mutations under divergence and genome hitchhiking. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*(1587), 461–474. doi:10.1098/rstb.2011.0256

Feder, J. L., Nosil, P., Wacholder, A. C., Egan, S. P., Berlocher, S. H., & Flaxman, S. M. (2014a). Genome-wide congealing and rapid transitions across the speciation continuum during speciation with gene flow. *Journal of Heredity*, *105*(S1), 810–820. doi:10.1093/jhered/esu038

Feder, J. L., Nosil, P., Wacholder, A. C., Egan, S. P., Berlocher, S. H., & Flaxman, S. M. (2014b). Genome-Wide Congealing and Rapid Transitions across the Speciation Continuum during Speciation with Gene Flow. *Journal of Heredity*, *105*, 810–820. doi:10.1093/jhered/esu038

Fishman, L., Stathos, A., Beardsley, P. M., Williams, C. F., & Hill, J. P. (2013). Chromosomal rearrangements and the genetics of reproductive barriers in mimulus (monkey flowers). *Evolution*, *67*(9), 2547–2560. doi:10.1111/evo.12154

Flaxman, S. M., Feder, J. L., & Nosil, P. (2013). Genetic hitchhiking and the dynamic buildup of genomic divergence during speciation with gene flow. *Evolution*, *67*(1974), 2577–2591. doi:10.1111/evo.12055

Flaxman, S. M., Wacholder, A. C., Feder, J. L., & Nosil, P. (2014). Theoretical models of the influence of genomic architecture on the dynamics of speciation. *Molecular Ecology*, *23*, 4074–4088. doi:10.1111/mec.12750

Fukunaga, K., Hill, J., Vigouroux, Y., Matsuoka, Y., Sanchez G, J., Liu, K., … Doebley, J. (2005). Genetic diversity and population structure of teosinte. *Genetics*, *169*(4), 2241–54. doi:10.1534/genetics.104.031393

Funk, D. J. (2010). Does strong selection promote host specialisation and ecological speciation in insect herbivores? Evidence from Neochlamisus leaf beetles. *Ecological Entomology*, *35*(SUPPL. 1), 41–53. doi:10.1111/j.1365-2311.2009.01140.x

Funk, D. J., Egan, S. P., & Nosil, P. (2011). Isolation by adaptation in Neochlamisus leaf beetles: Host-related selection promotes neutral genomic divergence. *Molecular Ecology*, 4671–4682. doi:10.1111/j.1365-294X.2011.05311.x

Funk, D. J., Nosil, P., & Etges, W. J. (2006). Ecological divergence exhibits consistently positive associations with reproductive isolation across disparate taxa. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(9), 3209–3213. doi:10.1073/pnas.0508653103

Funk, W. C., Lovich, R. E., Hohenlohe, P. A., Hofman, C. A., Morrison, S. A., Sillett, T. S., … Andelt, W. F. (2016). Adaptive divergence despite strong genetic drift: Genomic analysis of the evolutionary mechanisms causing genetic differentiation in the island fox (Urocyon littoralis). *Molecular Ecology*, (July). doi:10.1111/mec.13605

Fustier, M.-A., Bradenburg, J.-T., Boitard, S., Lapeyronnie, J., Eguiarte, L. E., Vigouroux, Y., … Tenaillon, M. I. (2017). Local adaptation of teosintes along altitudinal gradients using whole genome sequencing of pooled samples. *Molecular Ecology*.

Gompert, Z., Comeault, A. A., Farkas, T. E., Feder, J. L., Parchman, T. L., Buerkle, C. A.,

& Nosil, P. (2014). Experimental evidence for ecological selection on genome variation in the wild. *Ecology Letters*, *17*, 369–379. doi:10.1111/ele.12238

Goslee, S. C., & Urban, D. L. (2007). The ecodist package for dissimilarity-based analysis of ecological data. *Journal Of Statistical Software*, *22*(7), 1–19. doi:citeulike-article-id:12008924

Gutenkunst, R. N., Hernandez, R. D., Williamson, S. H., & Bustamante, C. D. (2009). Inferring the Joint Demographic History of Multiple Populations from Multidimensional SNP Frequency Data. *PLoS Genetics*, *5*(10). doi:10.1371/journal.pgen.1000695

Harte, D. (2016). HiddenMarkov: Hidden Markov Models. R package version 1.8-7. Wellington: Statistics Research Associates.

He, Q., & Knowles, L. L. (2016a). Identifying targets of selection in mosaic genomes with machine learning: applications in Anopheles gambiae for detecting sites within locally adapted chromosomal inversions. *Molecular Ecology*, *25*, 2226–2243. doi:10.1111/mec.13619

He, Q., & Knowles, L. L. (2016b). Identifying targets of selection in mosaic genomes with machine learning: applications in Anopheles gambiae for detecting sites within locally adapted chromosomal inversions. *Molecular Ecology*, *25*, 2226–2243. doi:10.1111/mec.13619

Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G., & Jarvis, A. (2005). Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, *25*(15), 1965–1978. doi:10.1002/joc.1276

Hofer, T., Foll, M., & Excoffier, L. (2012a). Evolutionary forces shaping genomic islands of population differentiation in humans. *BMC Genomics*, *13*, 107. doi:10.1186/1471-2164-13-107

Hofer, T., Foll, M., & Excoffier, L. (2012b). Evolutionary forces shaping genomic islands of population differentiation in humans. *BMC Genomics*, *13*, 107. doi:10.1186/1471-2164-13-107

Hufford, M. B., Bilinski, P., Pyhäjärvi, T., & Ross-Ibarra, J. (2012). Teosinte as a model system for population and ecological genomics. *Trends in Genetics*, *28*(12), 606–615. doi:10.1016/j.tig.2012.08.004

Hufford, M. B., Martínez-Meyer, E., Gaut, B. S., Eguiarte, L. E., & Tenaillon, M. I. (2012). Inferences from the Historical Distribution of Wild and Domesticated Maize Provide Ecological and Evolutionary Insight. *PLoS ONE*, *7*(11). doi:10.1371/journal.pone.0047659

Jackson, B., Kawakami, T., Cooper, S., Galindo, J., & Butlin, R. K. (2012). A Genome Scan and Linkage Disequilibrium Analysis among Chromosomal Races of the Australian Grasshopper Vandiemenella viatica. *PLoS ONE*, *7*(10), 1–10. doi:10.1371/journal.pone.0047549

Kawecki, T. J., & Ebert, D. (2004). Conceptual issues in local adaptation. *Ecology Letters*, *7*, 1225–1241. doi:10.1111/j.1461-0248.2004.00684.x

Keller, I., & Seehausen, O. (2012). Thermal adaptation and ecological speciation.

Molecular Ecology, *21*(4), 782–799. doi:10.1111/j.1365-294X.2011.05397.x

Kim, Y., & Nielsen, R. (2004). Linkage Disequilibrium as a Signature of Selective Sweeps. *Genetics*, *1524*(July), 1513–1524. doi:10.1534/genetics.103.025387

Kirkpatrick, M., & Barton, N. (2006). Chromosome Inversions, Local Adaptation and Speciation. *Genetics*, *434*(May), 419–434. doi:10.1534/genetics.105.047985

Klopfstein, S., Currat, M., & Excoffier, L. (2006). The fate of mutations surfing on the wave of a range expansion. *Molecular Biology and Evolution*, *23*(3), 482–490. doi:10.1093/molbev/msj057

Lenormand, T. (2012). From local adaptation to speciation: Specialization and reinforcement. *International Journal of Ecology*, *2012*. doi:10.1155/2012/508458

Linhart, Y. B., & Grant, M. C. (1996). EVOLUTIONARY SIGNIFICANCE OF LOCAL GENETIC DIFFERENTIATION IN PLANTS\rdoi:10.1146/annurev.ecolsys.27.1.237. *Annual Review of Ecology and Systematics*, *27*(1), 237–277. doi:10.1146/annurev.ecolsys.27.1.237

Liu, J., Möller, M., Provan, J., Gao, L. M., Poudel, R. C., & Li, D. Z. (2013). Geological and ecological factors drive cryptic speciation of yews in a biodiversity hotspot. *New Phytologist*, *199*(4), 1093–1108. doi:10.1111/nph.12336

Lowry, D. B., & Willis, J. H. (2010). A Widespread Chromosomal Inversion Polymorphism Contributes to a Major Life-History Transition , Local Adaptation , and Reproductive Isolation. *PLoS Biology*, *8*(9). doi:10.1371/journal.pbio.1000500

MacColl, A. D. C. (2011). The ecological causes of evolution. *Trends in Ecology and Evolution*, *26*(10), 514–522. doi:10.1016/j.tree.2011.06.009

Malinsky, M., Challis, R. J., Tyers, A. M., Schiffels, S., Terai, Y., Ngatunga, B. P., … Turner, G. F. (2015a). Genomic islands of speciation separate cichlid ecomorphs in an East African crater lake. *Science*, *350*(6267), 1493–1498. doi:DOI: 10.1126/science.aac9927

Malinsky, M., Challis, R. J., Tyers, A. M., Schiffels, S., Terai, Y., Ngatunga, B. P., … Turner, G. F. (2015b). Genomic islands of speciation separate cichlid ecomorphs in an East African crater lake. *Science*, *350*(6267), 1493–1498. doi:DOI: 10.1126/science.aac9927

Michel, A. P., Sim, S. B., Powell, T. H. Q., Taylor, M. S., Nosil, P., & Feder, J. L. (2010). Widespread genomic divergence during sympatric speciation. *Proceedings of the National Academy of Sciences of the United States of America*, *107*, 9724–9729. doi:10.1073/pnas.1000939107

Moeller, D. A., Tenaillon, M. I., & Tiffin, P. L. (2007). Population structure and its effects on patterns of nucleotide polymorphism in teosinte (Zea mays ssp. parviglumis). *Genetics*, *176*(3), 1799–1809. doi:10.1534/genetics.107.070631

Namroud, M. C., Beaulieu, J., Juge, N., Laroche, J., & Bousquet, J. (2008). Scanning the genome for gene single nucleotide polymorphisms involved in adaptive population differentiation in white spruce. *Molecular Ecology*, *17*(16), 3599–3613. doi:10.1111/j.1365-294X.2008.03840.x

Nosil, P. (2012a). *Ecological Speciation*. (P. H. Harvey, R. M. May, H. C. J. Godfray, &

Jennifer A. Dunne, Eds.). Oxford: Oxford University Press.

Nosil, P. (2012b). *Ecological Speciation*. (P. H. Harvey, R. M. May, H. C. J. Godfray, & Jennifer A. Dunne, Eds.). Oxford: Oxford University Press.

Nosil, P., & Crespi, B. J. (2006). Experimental evidence that predation promotes divergence in adaptive radiation. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(24), 9090–5. doi:10.1073/pnas.0601575103

Nosil, P., Egan, S. P., & Funk, D. J. (2008). Heterogeneous genomic differentiation between walking-stick ecotypes: "Isolation by adaptation" and multiple roles for divergent selection. *Evolution*, *62*(2), 316–336. doi:10.1111/j.1558-5646.2007.00299.x

Nosil, P., & Feder, J. L. (2012a). Genomic divergence during speciation: causes and consequences. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*(1587), 332–342. doi:10.1098/rstb.2011.0263

Nosil, P., & Feder, J. L. (2012b). Genomic divergence during speciation: causes and consequences. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*(1587), 332–342. doi:10.1098/rstb.2011.0263

Nosil, P., Funk, D. J., & Ortiz-Barrientos, D. (2009). Divergent selection and heterogeneous genomic divergence. *Molecular Ecology*, *18*, 375–402. doi:10.1111/j.1365-294X.2008.03946.x

Nosil, P., Gompert, Z., Farkas, T. E., Comeault, A. A., Feder, J. L., Buerkle, C. A., & Parchman, T. L. (2012). Genomic consequences of multiple speciation processes in a stick insect. *Proceedings of the Royal Society B: Biological Sciences*, (June), 5058–5065. doi:10.1098/rspb.2012.0813

Nosil, P., Harmon, L. J., & Seehausen, O. (2009). Ecological explanations for (incomplete) speciation. *Trends in Ecology and Evolution*, *24*(January), 145–156. doi:10.1016/j.tree.2008.10.011

Nosil, P., & Sandoval, C. P. (2008). Ecological niche dimensionality and the evolutionary diversification of stick insects. *PLoS ONE*, *3*(4). doi:10.1371/journal.pone.0001907

Nosil, P., & Schluter, D. (2011). The genes underlying the process of speciation. *Trends in Ecology and Evolution*, *26*(4), 160–167. doi:10.1016/j.tree.2011.01.001

Peischl, S., & Excoffier, L. (2015). Expansion load: Recessive mutations and the role of standing genetic variation. *Molecular Ecology*, *24*(9), 2084–2094. doi:10.1111/mec.13154

Phillips, S. B., Aneja, V. P., Kang, D., & Arya, S. P. (2006). Modelling and analysis of the atmospheric nitrogen deposition in North Carolina. *International Journal of Global Environmental Issues*, *6*(2–3), 231–252. doi:10.1016/j.ecolmodel.2005.03.026

Pickrell, J. K., & Pritchard, J. K. (2012). Inference of Population Splits and Mixtures from Genome-Wide Allele Frequency Data. *PLoS Genetics*, *8*(11). doi:10.1371/journal.pgen.1002967

Pyhäjärvi, T., Hufford, M. B., Mezmouk, S., & Ross-Ibarra, J. (2013). Complex patterns of local adaptation in teosinte. *Genome Biology and Evolution*, *5*(9), 1594–1609. doi:10.1093/gbe/evt109

Ren, R., Ray, R., Li, P., Xu, J., Zhang, M., Liu, G., … Yang, X. (2015). Construction of a high-density DArTseq SNP-based genetic map and identification of genomic regions with segregation distortion in a genetic population derived from a cross between feral and cultivated-type watermelon. *Molecular Genetics and Genomics*, *290*(4), 1457–1470. doi:10.1007/s00438-015-0997-7

Rice, A. M., Rudh, A., Ellegren, H., & Qvarnström, A. (2011). A guide to the genomics of ecological speciation in natural animal populations. *Ecology Letters*, *14*(1), 9–18. doi:10.1111/j.1461-0248.2010.01546.x

Riesch, R., Muschick, M., Lindtke, D., Villoutreix, R., Comeault, A. A., Farkas, T. E., … Nosil, P. (2017). Transitions between phases of genomic differentiation during stick-insect speciation. *Nature Ecology & Evolution*, *1*(FEbruary), 82. doi:10.1038/s41559-017-0082

Rodríguez F, J. G., Sánchez G, J. J., Baltazar, B. M., De la Cruz L, L., Santacruz-Ruvalcaba, F., Ron P, J., & Schoper, J. B. (2006). Characterization of floral morphology and synchrony among Zea species in Mexico. *Maydica*, *51*(2), 383–398. Retrieved from http://cat.inist.fr/?aModele=afficheN&cpsidt=17987262

Rolland, J., Condamine, F. L., Jiguet, F., & Morlon, H. (2014). Faster Speciation and Reduced Extinction in the Tropics Contribute to the Mammalian Latitudinal Diversity Gradient. *PLoS Biology*, *12*(1). doi:10.1371/journal.pbio.1001775

Ross-ibarra, J., Tenaillon, M., & Gaut, B. S. (2009). Historical Divergence and Gene Flow in the Genus Zea. doi:10.1534/genetics.108.097238

Rundle, H. D., & Nosil, P. (2005). Ecological speciation. *Ecology Letters*, *8*, 336–352. doi:10.1111/j.1461-0248.2004.00715.x

Sánchez-González, J. J., Kato-Yamamake, T. A., Aguilar-Sanmiguel, M. A., Hernández-Casillas, J. M., López-Rodriguez, A., & Ruiz-Corral, J. A. (1998). *Distribución y caracterización del teocintle*. (INIFAP, Ed.). Guadalajara.

Sansaloni, C., Petroli, C., Jaccoud, D., Carling, J., Detering, F., Grattapaglia, D., & Kilian, A. (2011). Diversity Arrays Technology (DArT) and next-generation sequencing combined: genome-wide, high throughput, highly informative genotyping for molecular breeding of Eucalyptus. *BMC Proceedings*, *5*(Suppl 7), P54. doi:10.1186/1753-6561-5-S7-P54

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

Schluter, D. (2001). Ecology and the origin of species. *Trends in Ecology and Evolution*, *16*(7), 372–380. doi:10.1016/S0169-5347(01)02198-X

Schluter, D., & Conte, G. L. (2009a). Genetics and ecological speciation. *Proceedings of the National Academy of Sciences of the USA*, *106 Suppl*(October 2016), 9955–62. doi:10.1073/pnas.0901264106

Schluter, D., & Conte, G. L. (2009b). Genetics and ecological speciation. *Proceedings of the National Academy of Sciences of the USA*, *106 Suppl*(October 2016), 9955–62. doi:10.1073/pnas.0901264106

Scholl, C. F., Nice, C. C., Fordyce, J. A., Gompert, Z., & Forister, M. L. (2012). Larval

performance in the context of ecological diversification and speciation in lycaeides butterflies. *International Journal of Ecology*, *2012*. doi:10.1155/2012/242154

Schoville, S. D., Bonin, A., François, O., Lobreaux, S., Melodelima, C., & Manel, S. (2012). Adaptive Genetic Variation on the Landscape: Methods and Cases. *Annual Review of Ecology, Evolution, and Systematics*, *43*(1), 23–43. doi:10.1146/annurev-ecolsys-110411-160248

Servedio, M. R., Doorn, G. S. Van, Kopp, M., Frame, A. M., & Nosil, P. (2011). Magic traits in speciation: "magic" but not rare? *Trends in Ecology and Evolution*, *26*(8), 389–397. doi:10.1016/j.tree.2011.04.005

Sexton, J. P., McInyre, P. J., Angert, A. L., & Rice, K. J. (2009). Evolution and Ecology of Species Range Limits . *Annual Review of Ecology, Evolution, and Systematics*, *40*(1), 415–436. doi:doi:10.1146/annurev.ecolsys.110308.120317

Shafer, A. B. A., & Wolf, J. B. W. (2013). Widespread evidence for incipient ecological speciation: A meta-analysis of isolation-by-ecology. *Ecology Letters*, *16*(7), 940–950. doi:10.1111/ele.12120

Smadja, C. M., & Butlin, R. K. (2011). A framework for comparing processes of speciation in the presence of gene flow. *Molecular Ecology*, *20*(24), 5123–5140. doi:10.1111/j.1365-294X.2011.05350.x

Soria-Carrasco, V., Gompert, Z., Comeault, A. A., Farkas, T. E., Parchman, T. L., Johnston, J. S., … Nosil, P. (2014). Stick Insect Genomes Reveal Natural Selection's Role in Parallel Speciation. *Science*, *344*, 738–742. doi:10.1126/science.1172133

Surget-Groba, Y., Johansson, H., & Thorpe, R. S. (2012). Synergy between allopatry and ecology in population differentiation and speciation. *International Journal of Ecology*, *2012*. doi:10.1155/2012/273413

Tening, a. S., Foba-Tendo, J. N., Yakum-Ntaw, S. Y., & Tchuenteu, F. (2013). Phosphorus fixing capacity of a volcanic soil on the slope of mount Cameroon. *Agriculture and Biology Journal of North America*, *4*(1990), 166–174. doi:10.5251/abjna.2013.4.3.166.174

Tobler, M., & Carson, E. W. (2010). Environmental variation, hybridization, and phenotypic diversification in Cuatro Ciénegas pupfishes. *Journal of Evolutionary Biology*, *23*(7), 1475–1489. doi:10.1111/j.1420-9101.2010.02014.x

Twyford, A. D., & Friedman, J. (2015). Adaptive divergence in the monkey flower Mimulus guttatus is maintained by a chromosomal inversion. *Evolution*, *69*(6), 1476–1486. doi:10.1111/evo.12663

van Heerwaarden, J., Doebley, J. F., Briggs, W. H., Glaubitz, J. C., Goodman, M. M., De Jesús Sánchez González, J., & Ross-Ibarra, J. (2011). Genetic signals of origin, spread, and introgression in a large sample of maize landraces. *Proceedings of the National Academy of Sciences*, *108*(3), 1088–1092. doi:10.1073/pnas.1013011108

Via, S. (1999). Reproductive Isolation between Sympatric Races of Pea Aphids . I . Gene Flow Restriction and Habitat Choice Author. *Evolution*, *53*(5), 1446–1457.

Via, S. (2009). Natural selection in action during speciation. *Proceedings of the National Academy of Sciences of the United States of America*, *106 Suppl*, 9939–9946.

doi:10.1073/pnas.0901397106

Via, S., & West, J. (2008). The genetic mosaic suggests a new role for hitchhiking in ecological speciation. *Molecular Ecology*, *17*(19), 4334–4345. doi:10.1111/j.1365-294X.2008.03921.x

Villemereuil, P., Frichot, E., Bazin, E., Fran, O., & Gaggiotti, O. E. (2014). Genome scan methods against more complex models: when and how much should we trust them? *Arxiv*. doi:10.1111/mec.12

Whitlock, M. C., & Lotterhos, K. E. (2015). Reliable Detection of Loci Responsible for Local Adaptation: Inference of a Null Model through Trimming the Distribution of F ST *. *The American Naturalist*, *186*(October), S000–S000. doi:10.1086/682949

TABLES

Table 1. Model parameters for the highest maximum likelihood values estimated with Fastsimcoal for the 50K and DTS datasets

Model	Ancestral		Admixture		Intermediate	
Dataset	DTS	50K	DTS	50K	DTS	50K
Maximum Likelihood	-12513.88	-58084.91	-13388.63	-61076.6	-12788.97	-60268.42
2 N_e Balsas	2,342	2,164	2160	2176	2152	2300
2 N_e Guerrero	2,284	2,160	4222	2732	2162	4904
2 N_e *mexicana*	6,338	5802	2160	2198	3,872	2646
Time of origin: *mexicana* generations (years)	15,550	11,774	77739	108457	36,459	12284
Time of origin generations (years)	Balsas: 87,609	Balsas: 51,682	Guerrero: 42171	Guerrero: 26592	Guerrero: 101,218	Guerrero: 51690

Table 2. Environmental and geographic association for loci with different levels of significance in the *bayescenv* analyses for the 50K dataset

	PC1			PC2			P		
	$\overline{\Delta Nei}$	Env(R^2)	Geo(R^2)	$\overline{\Delta Nei}$	Env(R^2)	Geo(R^2)	$\overline{\Delta Nei}$	Env(R^2)	Geo(R^2)
Neutral SNPs (β-SNPs)	0.103	0.102	0.205	0.103	0.053	0.213	0.088	0.002	0.422
Outlier SNPs (α-SNPs)	0.380	**0.155**	0.074	0.370	0.007	0.079	0.491	0.008	0.012
γ-non-candidate SNPs	0.154	**0.455**	0.230	0.180	**0.154**	0.105	0.134	0.015	0.453
γ-candidate-selected SNPs	0.343	**0.6**	0.023	0.313	**0.226**	0.005	0.471	**0.410**	0.011
Percentage of SNPs that contribute to environmental divergence	7.17%			2.84%			0.20%		
Total Percentage	8.95%**								

* Bold values indicate that the environmental association was stronger than the geographic association

** The total is not the sum of the three environments because some of the SNPs are shared between environments or categories

$\overline{\Delta Ne}$: Mean Nei's genetic distance between populations

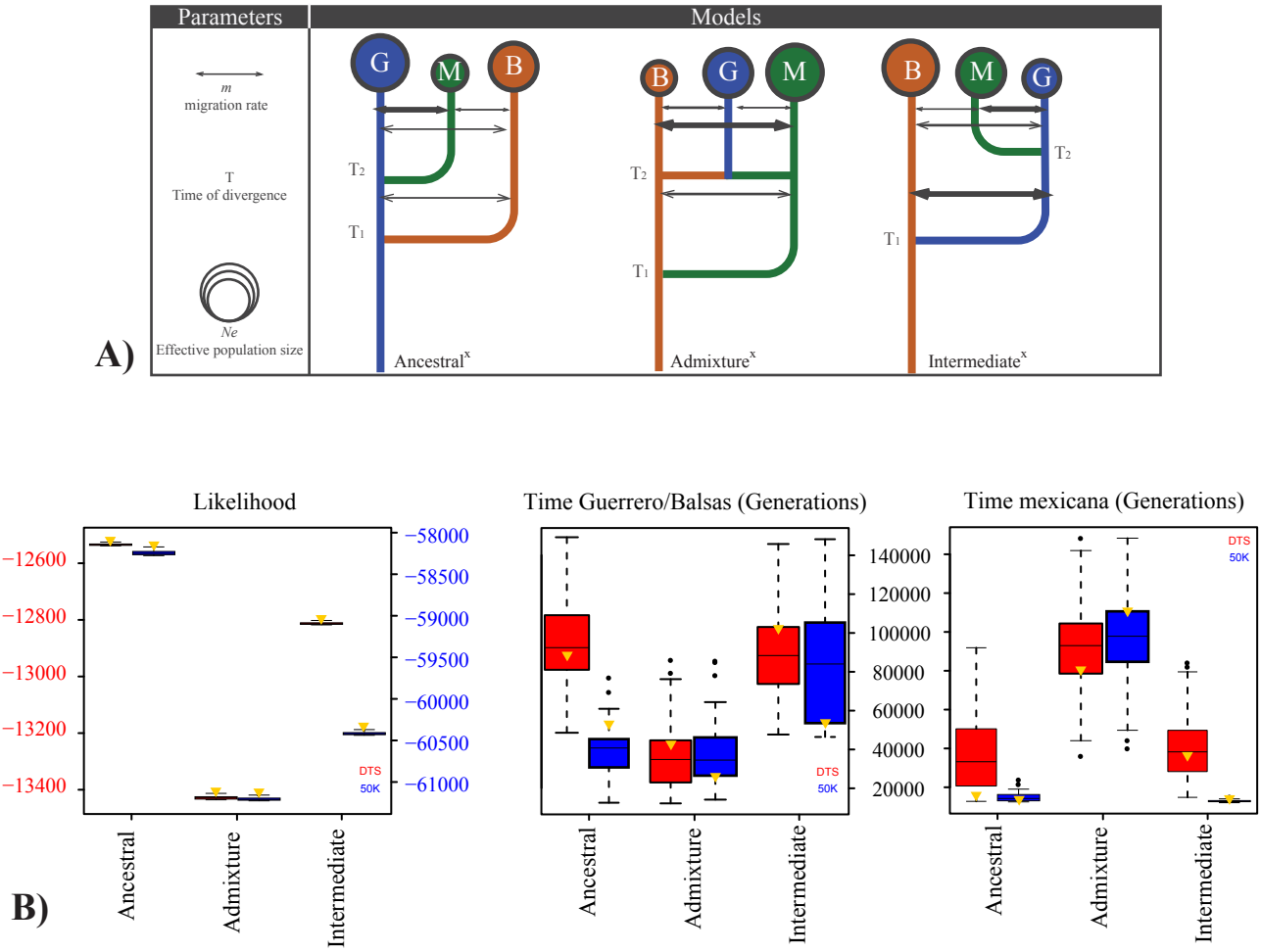


Figure 1. A) Models of divergence of Balsas, mexicana and Guerrero. The Ancestral model suggest that Guerrero gives origin to Balsas and mexicana; The Admixture model suggest that Balsas gives origin to mexicana and they both hybridize and give origin to Guerrero; The Intermediate model suggest that Balsas gives origin to Guerrero, and this one gives origin to mexicana. Parameters estimated by Fastsimcoal are indicated in grey. Different sizes of the parameters indicate that parameters are not fixed between genetic groups. B) Likelihood and parameter range for the Intermediate, Ancestral and Admixture models for the DTS dataset (red) and DTS (blue). Yellow triangles indicate the estimates for the highest likelihood value of the model (See Supporting information for *Nem* estimates between genetic groups).

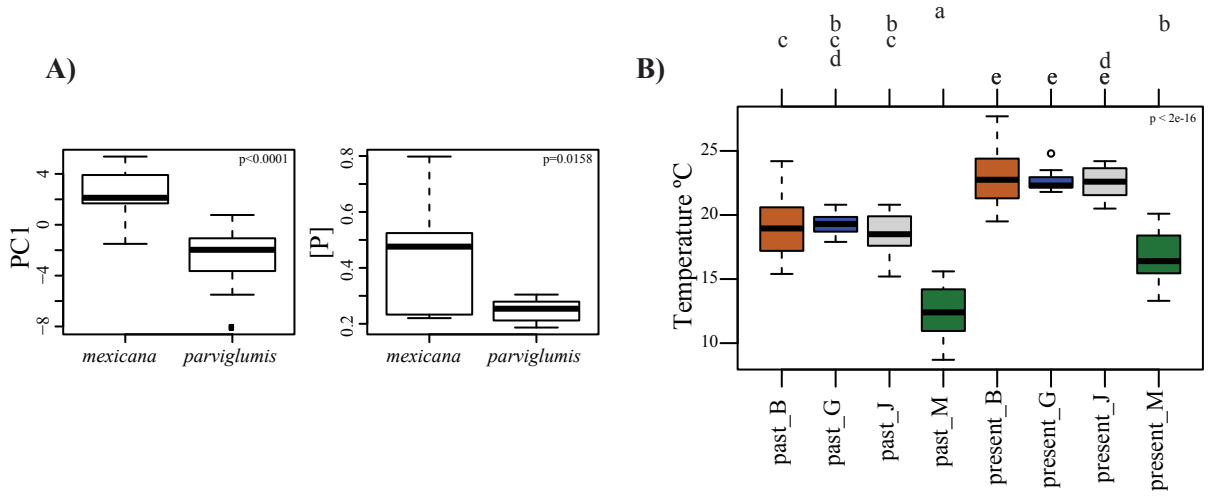


Figure 2: A) Variation between PC1 (temperature) and P concentration in the soil for mexicana and parviglumis populations (For all the variables tested see Figure S5); B) Mean annual temperature for mexicana, Balsas, Guerrero and Jalisco populations during the present or during the LGM, 21,000 years ago. Same letters above the boxplots indicate temperatures that statistically do not differ. Colors indicate the genetic groups as defined in Figure 1 and the Ward algorithm Figure S2.B).

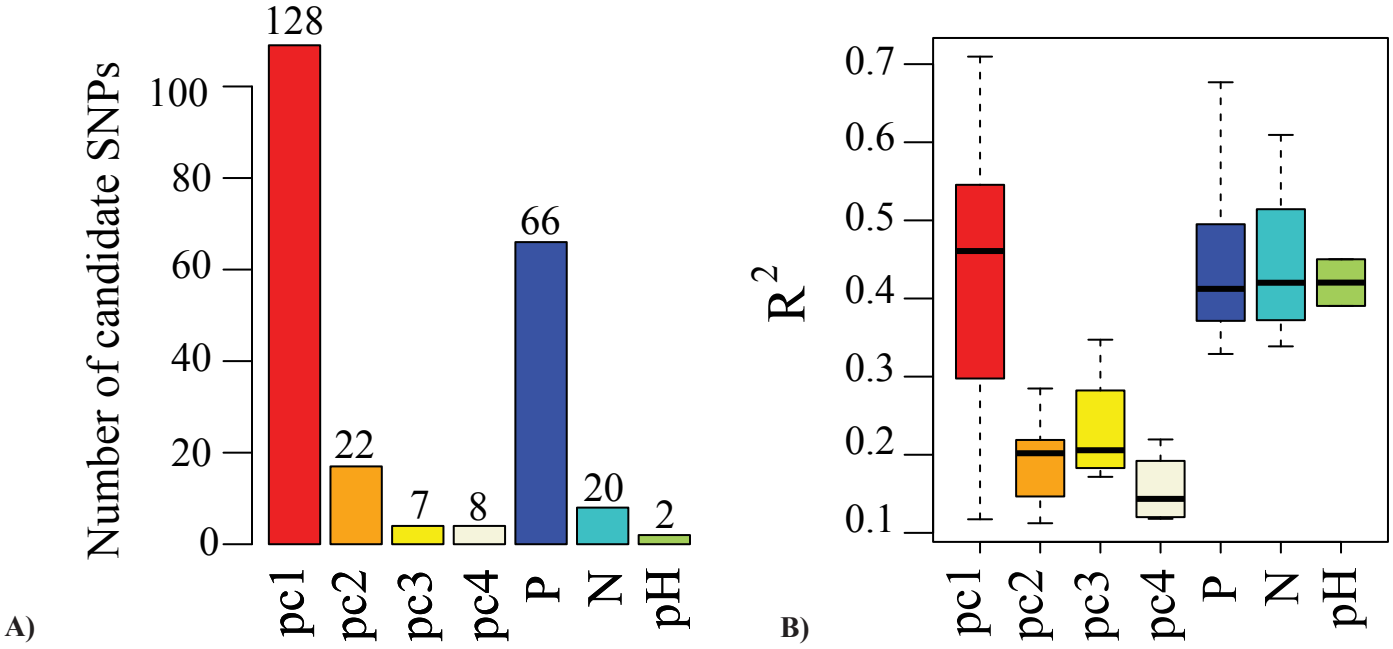


Figure 3. A) Number of candidate SNPs (50K dataset) associated to 7 environmental axis and B) association (R2) between allelic frequencies and environment for each candidate SNP

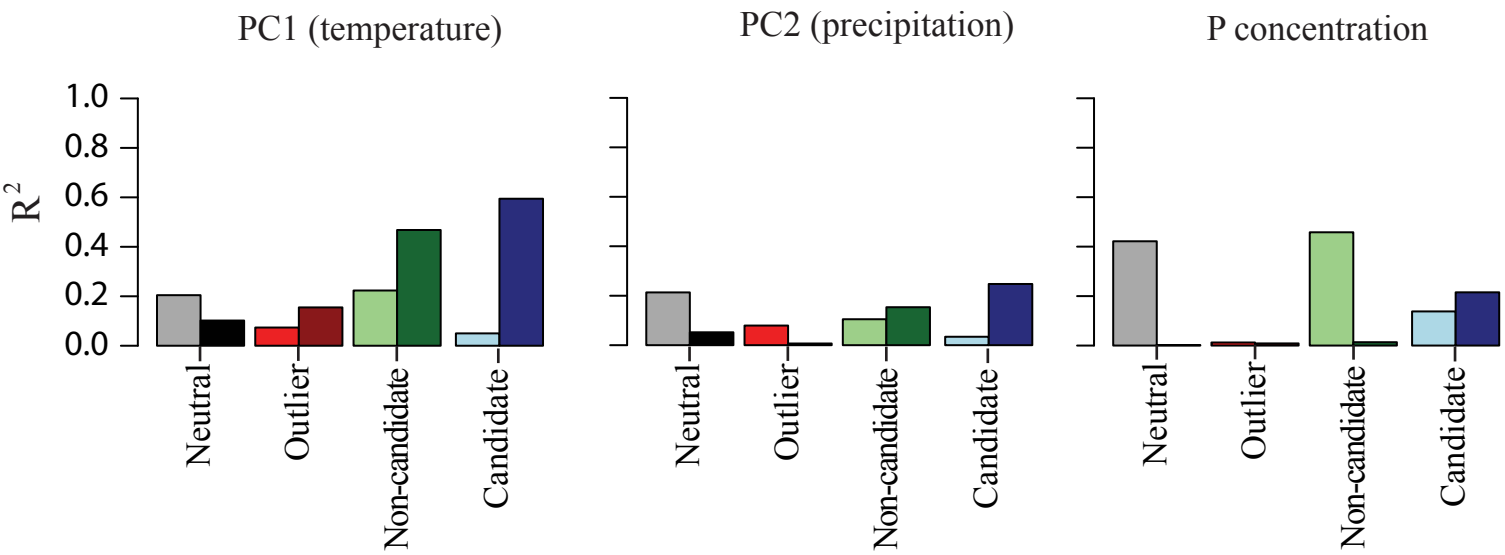


Figure 4. Environmental and geographic effect on genetic distance between populations for three neutral SNP categories for the 50K dataset (Neutral □-SNPs, Outlier □-SNPs, □-non-candidate SNPs) and for the □-candidate-selected). Lighter color shows the geographic association, while the Darker color shows the environmental association. Each model tested the Nei's genetic distance between populations of a set of loci, with different level of significance in the bayescenv analyses, as response variables, and geographic and environmental distances between populations as independent variables.

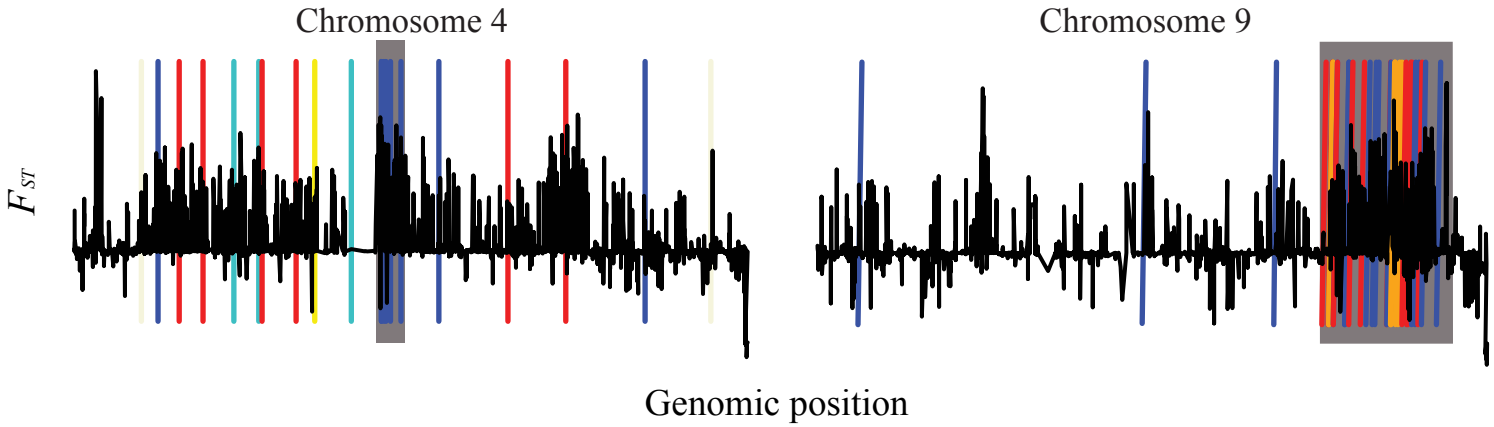


Figure 5. Locus-by-Locus FST along two chromosomes with high FST blocks enriched with candidate SNPs (for all chromosomes see Figure S7). Vertical colored lines correspond to the position of candidate loci (colors correspond to those in Figure 3). Grey vertical rectangles correspond to the areas where there is an overlap between high enrichment of candidate SNPs and long blocks of differentiation according to the HMM. We chose to present chromosome 4 and 9, since chromosome 4 one has a small inversion associated only to P concentration in the soil and chromosome 9 has a large inversion enriched with SNPs associated to PC1, PC2 and P concentration in the soil.

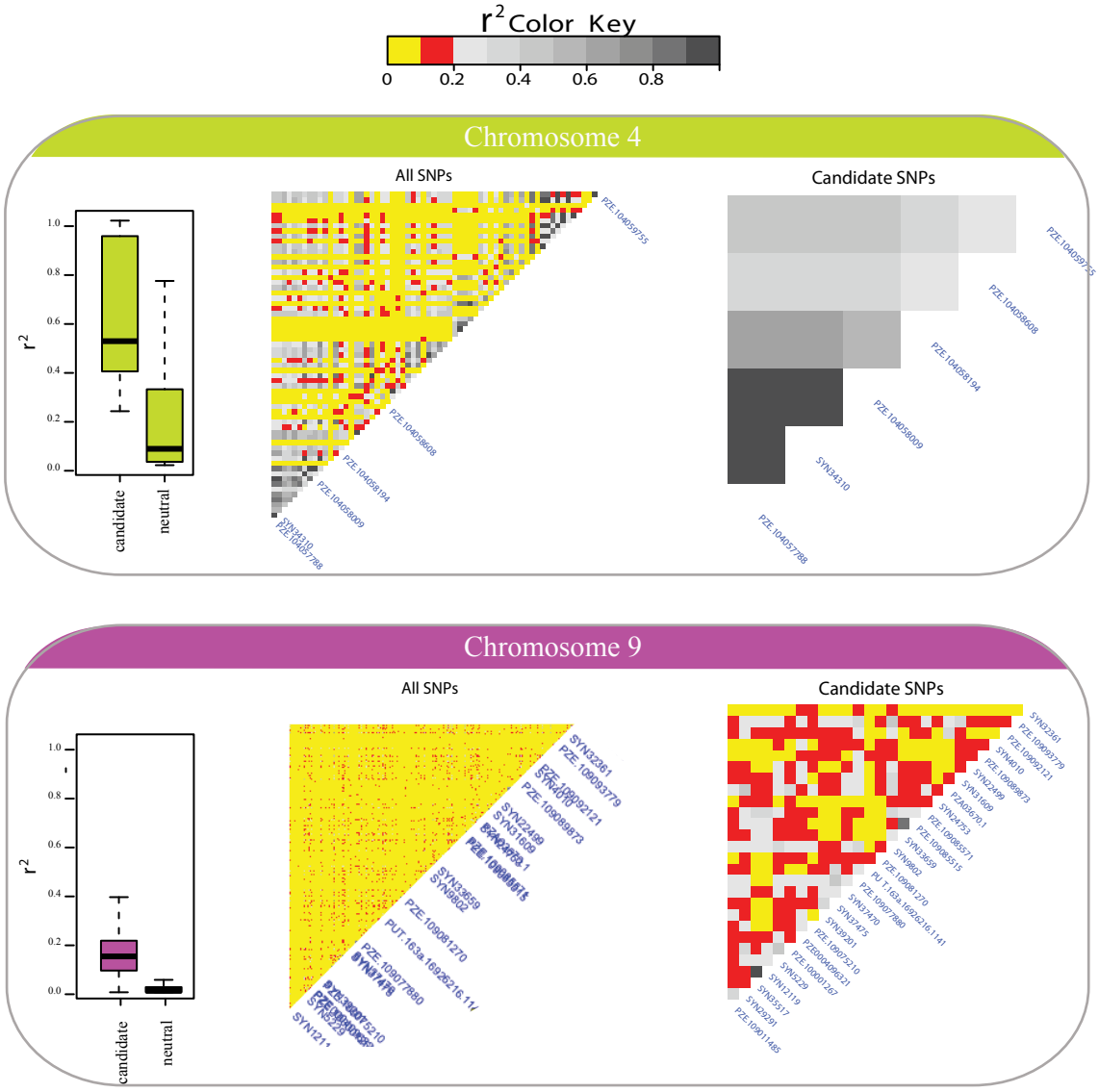


Figure 6. LD along neutral and candidate SNPs in the long high FST regions of chromosomes 4 and 9 (For chromosome 1 and 8 see supporting information). In all cases candidate SNPs present significantly higher LD than neutral SNPs ($p<0.001$) even if they are separated by regions with high recombination (as shown by the heatmaps).

DISCUSIÓN Y CONCLUSIONES

En este trabajo desarrollamos el estudio a nivel genómico y poblacional más extenso que se ha realizado hasta la fecha en teocintles. Inicialmente, analizamos la adaptación local a escala poblacional y encontramos que los teocintles se están adaptando al límite de nicho (donde las condiciones son más divergentes), irrespectivamente de la distribución geográfica de las poblaciones (Aguirre-Liguori *et al.* 2017). Posteriormente, revisamos si la adaptación local al límite de nicho podría tener efectos sobre la divergencia de las especies (Eckert, Samis, and Lougheed 2008; Sexton *et al.* 2009). Encontramos que la adaptación al límite de nicho está generando la divergencia de las dos subespecies de teocintle (Aguirre-Liguori *et al.* en preparación). Pocos estudios han mostrado hasta el momento que la adaptación local puede ocurrir en el límite de nicho, y menos son los que han mostrado su efecto sobre la especiación ecológica. Dada la detección de genes candidatos a estar bajo selección, este trabajo tiene una importancia para la biología aplicada, ya que estos genes podrían ser usados para el mejoramiento del maíz cultivado (Warschefsky *et al.* 2014).

Teocintles: un modelo interesante, pero complejo, para estudiar procesos de selección

En este trabajo analizamos 49 poblaciones de teocintles que cubren toda su distribución ecológica y geográfica. Asimismo, analizamos estas poblaciones utilizando dos métodos genómicos que nos permitieron obtener 34,000 SNPs utilizando el chip MaizeSNP50 de Illumina y cerca de 10,000 SNPs de marcadores Dartseq (similares a GBS). Este trabajo supera en número de marcadores y/o de poblaciones los trabajos que se han realizado hasta este momento. Hasta la fecha, los estudios clásicos de teocintles han sido realizado con un número limitado de marcadores moleculares, poblaciones y en accesiones (Fukunaga *et al.* 2005; Moeller, Tenaillon, and Tiffin 2007; van Heerwaarden *et al.* 2011).

Al incrementar el número de marcadores y poblaciones, encontramos que la estructura genética de los teocintles es muy compleja (Figura 2 del Capítulo 2, "*Connecting genomic patterns of local adaptation and niche suitability in teosintes*"), especialmente en *parviglumis*, (Pyhäjärvi *et al.* 2013; Aguirre-Liguori *et al.* 2017; Moreno-Letelier *et al.* sometido) y que tiene una asociación muy estrecha con el ambiente (Aguirre-Liguori *et al.* En preparación). Encontramos que los teocintles se dividen en dos subespecies ecológicas, las de tierras altas y frías (*mexicana*) y las de tierras bajas y calientes (*parviglumis*). Sin

embargo, *parviglumis* incluye a su vez tres grupos genéticos altamente diferenciados. Los grupos más complejos son las poblaciones de Jalisco divergentes (Figura 2, Capítulo 2; *Genetic groups II and III*), que presentan baja diversidad genética y altísima diferenciación genética, alcanzando valores de F_{ST} de hasta 0.4 con otras poblaciones de *parviglumis*. Aunque Pyhäjärvi *et al.* (2013), ya habían mostrado que los teocintles presentan una estructura compleja, al incrementar nuestro muestreo, encontramos un grupo genético extra en Jalisco. Todo esto es relevante, porque altos niveles de diferenciación genética complican los estudios de adaptación local, ya que se vuelve muy difícil determinar el umbral a partir de cual la F_{ST} en un locus es neutral o candidato a estar bajo selección (de Mita *et al.* 2013).

Asimismo, encontramos que la población de Guerrero que Pyhäjärvi *et al.* (2013) denominaron admixta (ya que aparece intermedia en sus análisis de Structure), en realidad forma parte de muchas otras poblaciones en Guerrero que presentan la misma estructura genética compleja. Encontramos que la mayoría de poblaciones de Guerrero que colectamos forman un grupo genético intermedio el cual analizamos detalladamente (Aguirre-Liguori *et al.* en preparación). En este estudio extendimos el muestreo de poblaciones centrales de Guerrero de una a siete poblaciones. Esto es importante, ya que no considerar esta estructura genética compleja podría generar sesgos importantes (Hufford *et al.* 2012a) en los estudios de demografía histórica del genero *Zea* (Ross-Ibarra, Tenaillon, and Gaut 2009) y podría incrementar el número de falsos positivos en la identificación de genes bajo selección (Excoffier *et al.* 2009a; De Mita *et al.* 2013; de Villemereuil y Gaggiotti 2015).

Cabe mencionar que en el Capítulo 3 encontramos que las poblaciones intermedias de Guerrero parecen ser ancestrales a *mexicana* y al resto de *parviglumis* (Figura 1, Capítulo 3). Nuestras simulaciones sugieren que la subespecie *mexicana* se originó después del LGM, hace 15,500 años, mucho después del tiempo propuesto inicialmente (60,000 generaciones; Ross-Ibarra *et al.* 2009). Hasta donde sabemos, son pocos los estudios que han propuesto que el origen de *mexicana* podría ser mucho más reciente que lo propuesto por Ross-Ibarra *et al.* (2009). Buckler y Holtsford (1996) utilizaron ITS para analizar la filogenia del género *Zea* y encontraron que *parviglumis* parece tener una estructura compleja y que distintos clados pueden estar emparentados con el maíz y con la subespecie

mexicana. También encontraron que el origen de *mexicana* podría ser más reciente que la domesticación del maíz. Estos resultados concuerdan con nuestros datos y sugieren que *mexicana* podría tener un origen reciente y asociado al final del Pleistoceno o al Holoceno.

De igual forma, nuestros resultados y los de Buckler y Holtsford (1996) corroboran la importancia de analizar muestreos extensos que abarquen toda la estructura genética de los teocintles para entender mejor como han divergido las especies y subespecies del género *Zea*. En particular, tomar en cuenta la compleja estructura de los teocintles permitirá determinar con mayor precisión el tiempo y lugar de la domesticación del maíz. Justamente, al incluir varias poblaciones de Guerrero y de Jalisco de *parviglumis*, Moreno-Letelier *et al.* (en preparación) encontró evidencias que sugieren que la domesticación del maíz podría haber sido más antigua y que podría haber ocurrido en el estado de Jalisco.

Adaptación al límite de nicho, una oportunidad ecológica nueva

Los teocintles presentan una estructura genética compleja y esto dificulta la correcta detección genes que sean candidatos a estar bajo selección (Excoffier *et al.* 2009a; Schoville *et al.* 2012; De Mita *et al.* 2013; Eguiarte *et al.* 2013; de Villemereuil y Gaggiotti 2015). La mejor manera de solucionar este problema es construyendo hipótesis específicas y explorar las hipótesis con diseños experimentales o de análisis de datos adecuados.

Inicialmente, intentamos realizar comparaciones entre gradientes ambientales con el objetivo de encontrar paralelismos en señales de selección (sensu Hohenlohe *et al.* 2010). Sin embargo, los genomas de los teocintles evolucionan muy rápidamente (Díez *et al.* 2013; Tenaillon *et al.* 2011; Zerjal *et al.* 2012), tienen alta recombinación (Aguirre-Liguori, Aguirre-planter, and Eguiarte 2016), y estas especies tienen además una estructura filogeográfica compleja (Pyhäjärvi *et al.* 2013; Aguirre-Liguori *et al.* 2017; Aguirre-Liguori *et al.* en preparación; Moreno-Letelier *et al.* sometido), por lo que fue complicado realizar este tipo de observaciones. Buscando estos paralelismos, con base en Bayescan encontramos que existen relativamente pocos SNPs anómalos, cuyas frecuencias alélicas muestran patrones claros de paralelismo con el ambiente. Creemos que esto se debe a que en teocintles puede ser común la neutralidad condicional, que se refiere a que un locus puede estar bajo selección en ciertas poblaciones, pero ser neutral en otras (Anderson y Mitchell-Olds 2011; Anderson *et al.* 2011, 2013; Aguirre-Liguori *et al.* 2017). La

naturaleza de la neutralidad condicional sugiere que puede haber variación en el número de genes candidatos a estar bajo selección en distintas poblaciones y esto puede estar relacionado a la intensidad de la selección y/o a la calidad del nicho en la que crecen las poblaciones.

Seguimos el proceso ilustrado en la Figura 1 del Capítulo 2, para identificar qué SNPs son neutrales o candidatos en las poblaciones de teocintle. Encontramos que hay variación en el número de genes candidatos entre poblaciones y que existe una asociación entre el número de genes candidatos y la distancia del centro de la distribución ecológica (Lira-Noriega y Manthey 2014), pero no con la geográfica (Eckert, Samis, and Lougheed 2008). A diferencia de los estudios que se han realizado con marcadores neutrales (Eckert *et al.* 2008, Lira-Noriega y Manthey 2014), encontramos que las poblaciones de teocintle se están adaptando al límite de nicho, donde las condiciones son más adversas (Aguirre-Liguori *et al.* 2017). Estos resultados contrastan con la hipótesis de abundancia central (HAC), que sugiere que las poblaciones centrales deberían de estar mejor adaptadas, ya que las condiciones son óptimas, y que las periféricas deberían de ser regiones sumidero, donde hay mucha extinción local (Eckert, Samis, and Lougheed 2008; Sexton *et al.* 2009). Específicamente, encontramos que las poblaciones se están adaptando a los ambientes contrastantes y que por selección divergente (Lenormand 2002; Hampe y Petit 2005; Eckert *et al.* 2008) podrían estar adaptándose al límite de nicho.

Es este estudio, utilizamos señales genéticas y datos climáticos para determinar la lista de genes candidatos a estar bajo selección en distintas poblaciones. Por lo tanto, realizamos muchas pruebas para poder controlar la identificación de falsos positivos. Primero, la elección de los marcadores se basó en la identificación con dos métodos de detección de genes anómalos, que asumen distintas hipótesis adaptativas. Aunque esto puede reducir el número de verdaderos candidatos (Pyhäjärvi *et al.* 2013), el uso de dos métodos distintos aumenta la probabilidad de detectar los SNPs que realmente estén bajo selección (Eguiarte *et al.* 2013). Así mismo, realizamos métodos posteriores muy rigurosos, para quedarnos con aquellos SNPs que estuvieran estrechamente asociados a las clinas ecológicas y no geográficas, para eliminar aquellos que se estuvieran fijando por *gene surfing* en algunas poblaciones, particularmente en la periferia de la distribución de los teocintles (Excoffier y Ray 2008; Excoffier *et al.* 2009b). De igual manera, realizamos

varias pruebas para verificar que los resultados del método mostrado en la Figura 1 no se debieran a sesgos estadísticos. Finalmente, la anotación de los genes candidatos indica que muchos de ellos en efecto están asociados a respuesta a presiones abióticas, lo que sugiere fuertemente que podrían estar bajo selección. No obstante, será muy importante realizar diseños experimentales, que comprueben que los genes candidatos están realmente bajo selección. Por ejemplo, se podrían realizar transplantes recíprocos, y ver si las poblaciones en efecto se comportan como se espera. Así mismo se podrían hacer experimentos en jardines comunes permitiendo que las poblaciones evolucionen por algunas generaciones y luego determinar si los genes candidatos cambiaron de acuerdo a lo esperado (Por ejemplo, ver Soria-Carrasco *et al.* 2014; Egan *et al.* 2015). Finalmente, sería interesante hacer *knockouts* de los genes candidatos y ver si la adecuación de los individuos se reduce con la ausencia de las variantes adaptativas.

Otra consideración importante que hay que tomar en cuenta es que los teocintles son especies anuales, que tiene tamaños efectivos grandes y altos niveles de diversidad genética (Ross-Ibarra *et al.* 2009; Hufford *et al.* 2012b; Aguirre-Liguori *et al.* 2016). Estas características promueven la adaptación local y por lo tanto es posible que los resultados de adaptación al límite de nicho sean exclusivos de especies con características similares a los teocintles. Será muy interesante realizar estudios similares en especies con otras historias de vida (por ejemplo, perenes, baja diversidad, autógamas) y en otros grupos taxonómicos para analizar si la adaptación al límite de nicho es común o a qué características está asociada. Con el creciente desarrollo de la secuenciación masiva y las bases de datos disponibles en NCBI y Dryad, será posible hacer meta-análisis probando estas hipótesis. Estos resultados y discusiones son importantes, ya que en el límite de nicho las condiciones son divergentes y en algunos casos pueden llegar a asemejarse a aquellas pronosticadas con el cambio climático (Hampe y Petit 2005). En particular, los genes que confieren adaptación al límite de nicho caliente pueden ser muy interesantes e importantes en el escenario de cambio climático (Lenormand 2002; Hampe y Petit 2005). Será relevante analizar detalladamente estas poblaciones y genes para diseñar estrategias de flujo genético adaptativo para rescatar poblaciones adaptadas a frío. Para ello habrá que modelar las frecuencias alélicas de estos genes candidatos al futuro (Fitzpatrick y Keller 2015) y definir estrategias de conservación o mejoramiento. De ser así, este mecanismo permitiría la

conservación de especies o entender mejor los mecanismos que generan diversidad.

Finalmente, un aspecto importante que no analizamos con mayor detalle son las bases genéticas de las adaptaciones. Encontramos SNPs que presentan una asociación estrecha con el ambiente, pero muchos de ellos son cambios sinónimos y otros se encuentran en regiones intergénicas. Será importante analizar si las bases genéticas de las adaptaciones en teocintles se deben a cambios puntuales en los genes, o bien a cambios en regulación u otros mecanismos. Los teocintles y maíces presentan alta variación en elementos móviles (Díez *et al.* 2013) y algunos de ellos podrían estar relacionados con adaptación local (Tenaillon *et al.* 2011; Zerjal *et al.* 2012). Para caracterizar las bases genéticas de las adaptaciones se podrían generar mejores datos genómicos en teosintes, comparar transcriptomas entre poblaciones que crecen en condiciones distintas, y/o realizar *knockouts* para determinar si los genes candidatos son los importantes.

La adaptación al límite de nicho podría haber promovido la divergencia de los teocintles

La adaptación al límite de nicho puede tener implicaciones importantes en el origen de nuevas especies. Esto se debe a que en el límite de nicho las condiciones son contrastantes y por lo tanto las poblaciones podrían adaptarse a nuevas condiciones ecológicas y si se generan barreras reproductivas podría iniciarse un proceso de especiación ecológica (Eckert, Samis, and Lougheed 2008; Sexton *et al.* 2009).

Para poder comprobar si las subespecies de teocintles podrían estar divergiendo por especiación ecológica, re-analizamos la demografía histórica de estas dos subespecies y posteriormente realizamos análisis de selección y divergencia genómica para determinar si el ambiente ha jugado un papel importante en su divergencia y posible asilamiento reproductivo. Como mencioné anteriormente, la subespecie *mexicana* se originó después del LGM, cuando las condiciones eran más frías (hace unos 15,500 años, Figura 1 en el capítulo 3). Así mismo, utilizando análisis genómicos y ecológicos encontramos que la divergencia de los teocintles ha ocurrido principalmente a lo largo de dos ejes ecológicos: temperatura y disponibilidad de fósforo en el suelo (Figura 2 en el capítulo 3).

Aunque nuestros resultados sugieren que el enriquecimiento ocurre principalmente a lo largo de dos ejes climáticos, es importante considerar que existen muchos marcadores

que presentan una elevada F_{ST} y que son identificados como anómalos por bayescenv (α-outliers). Es probable que muchos de estos marcadores sean falsos positivos (no presentan asociación con el ambiente). Sin embargo, también es posible que otros sean verdaderos genes bajo selección que se correlacionan con otras variables bióticas y abióticas no consideradas. En otros organismos se ha encontrado que la mayoría de las señales de selección se relacionan con respuestas a defensas contra patógenos (Fumagalli *et al.* 2011). También hay que considerar que los teocintles crecen a lo largo de un gradiente altitudinal, generando diversos gradientes climáticos y físicos. El índice de UV es mayor en poblaciones elevadas. Usando otros métodos para detectar genes candidatos, Pyhäjärvi *et al.* (2013) encontraron un SNP en el gen *b1* que participa en la síntesis de antocianina y genera un cambio en la pigmentación de las vainas entre *mexicana* y *parviglumis.* Estos cambios en la coloración se han propuesto como respuestas a cambios en la incidencia de UV. Otras variables que podrían estar asociados con cambios en la altitud son la presión atmosférica y la presión parcial de O_2 y CO_2. Nosotros no incluimos altitud en nuestras variables climáticas, ya que esta puede estar correlacionada con muchas otras variables (incluida distancia geográfica entre poblaciones). No obstante será interesante usar la altitud como una variable y anotar genes anómalos para identificar este tipo de genes. Finalmente, en este mismo aspecto, cabe resaltar que los suelos volcánicos en los que crece *mexicana*, son ricos en otros metales pesados (Vielle-Calzada *et al.* 2009; Krasilnikok *et al.*, 2013) y es posible que se estén adaptando a estas variables. En estudios futuros, sería muy deseable incorporar esta información para identificar genes que puedan estar bajo selección en respuesta a metales pesados. Usando bayescenv y esta información sería una aproximación poderosa para detectarlos, pero para ello es necesario tener buenos datos químicos del suelo.

Aunque la especiación ecológica puede iniciarse múltiples veces, pocas veces se completa el desarrollo de barreras al flujo genético (Nosil y Sandoval 2008; Soria-Carrasco *et al.* 2014; Riesch *et al.* 2017). Esto ocurre porque aún eventos de migración relativamente raros pueden ser suficientes para homogenizar las poblaciones (Hedrick 2011). Un mecanismo que ha sido poco estudiado empíricamente, pero se ha trabajado con cuidado teóricamente, y que puede ayudar a explicar nuestros resultados y ayudar a proponer hipótesis para trabajos futuros, es el de la selección multifaria (Nosil y Sandoval 2008;

Nosil *et al.* 2009a; b; MacColl 2011; Lenormand 2012; Chevin *et al.* 2014). Esta sugiere que la divergencia genómica y la reducción en el flujo genético será más probables si las poblaciones se están adaptando a diversos ejes del nicho, ya que más partes del genoma divergirán simultáneamente (Nosil, Funk, and Ortiz-Barrientos 2009; Nosil, Harmon, and Seehausen 2009). Nuestros resultados indican que la divergencia de teocintles ha ocurrido en respuesta a por lo menos dos ejes del nicho (Figura 3, Capítulo 3): la temperatura y la disponibilidad de fósforo en el suelo. Consideramos que la divergencia en ambos factores está estrechamente relacionada con el último máximo glacial. Así mismo consideramos que su relación podría estar promoviendo el aislamiento reproductivo entre las subespecies como explico a continuación.

Las simulaciones que desarrollamos y los datos ecológicos sugieren que *mexicana* se originó durante un periodo frío en regiones cercanas a las poblaciones de Guerrero (Figura 1, Figura 2ª, Capítulo 3). Tomando en cuenta que las poblaciones limítrofes de teosintes presentan altas señales de adaptación local (Aguirre-Liguori *et al.* 2017), consideramos que la divergencia habría iniciado por adaptación al límite frío del nicho. Posteriormente, durante el interglacial, estas poblaciones habrían migrado a tierras altas donde las condiciones son similares a su clima original. Esto habría promovido su aislamiento geográfico, reduciendo así el flujo genético entre poblaciones y, por lo tanto, acelerando el proceso de divergencia ecológica (Aguilée *et al.* 2011; Nosil y Feder 2012; Surget-Groba *et al.* 2012; Nosil *et al.* 2017). *Mexicana* crece en tierras donde ocurren heladas en invierno. Por lo tanto, *mexicana* ha evolucionado a presentar una floración más temprana que *parviglumis*. Los suelos en los que crece *mexicana* son volcánicos y aunque presentan mayor concentración de fósforo en el suelo que las poblaciones de *parviglumis* (Figura 2, Capítulo 3), estos suelos suelen retener más fuertemente el fósforo (Tening *et al.* 2013; Krasilnikok *et al.*, 2013). En términos prácticos, esto resulta en una menor disponibilidad de fósforo para la planta y se han encontrado genes que regulan el uso de fósforo en poblaciones de *mexicana* (Fustier *et al.* 2017). La disponibilidad del fósforo en el suelo es importante para el crecimiento y floración de las plantas (Elser *et al.* 2010) por lo que consideramos que es probable que *mexicana* se haya adaptado a extraer eficientemente el fósforo, el cual es retenido más fuertemente en los suelos volcánicos donde crece (Tening *et al.* 2013; Krasilnikok *et al.*, 2013). En efecto, esta retención de

fósforo en suelos volcánicos es limitante para la producción de plantas cultivables como el maíz, frijol, trigo y otros (Krasilnikok *et al.*, 2013).

Cambios en la fenología floral han sido reportados en poblaciones simpátricas que crecen en ambientes altamente contrastantes (Antonovics y Bradshaw 1970; Linhart y Grant 1996; Antonovics 2006; Via 2009) y se ha propuesto que este desplazamiento en la floración reduciría la posibilidad de formar híbridos mal adaptados a las condiciones contrastantes (Via 2009). Similarmente, consideramos que el cambio en la floración entre *mexicana* y *parviglumis* está promoviendo el aislamiento reproductivo entre las dos subespecies (Sánchez-González *et al.* 1998; Rodríguez *et al.* 2006; Hufford *et al.* 2012b). Adicionalmente, consideramos que este aislamiento reproductivo tiene una base ecológica relacionada con cambios en temperatura y disponibilidad de fósforo en el suelo. Aunque la relación entre selección y apareamiento selectivo es una etapa importante durante el proceso de especiación ecológica (Nosil y Schluter 2011; Servedio *et al.* 2011), es importante que exista una base genética que ligue estos dos aspectos. De acuerdo a nuestros datos genómicos existen cuatro regiones de alta diferenciación altamente enriquecidas en genes candidatos. Tres de estas regiones han sido descritas como inversiones cromosómicas (Fang *et al.* 2012, Pyhäjärvi *et al.* 2013), y podrían estar reduciendo el flujo genético entre subespecies, ya que impiden la recombinación en individuos con distintos polimorfismos (Kirkpatrick y Barton 2006; Lowry y Willis 2010; Fishman *et al.* 2013; Twyford y Friedman 2015; He y Knowles 2016). Paradójicamente, encontramos que a lo largo de estas inversiones el DL es bajo, por lo que creemos que existe mucha recombinación entre poblaciones que presentan el mismo polimorfismo. Sin embargo, es interesante que los marcadores candidatos a estar bajo selección presentan mayor DL dentro de las inversiones (Figura 3 en Capítulo 3). Es posible que la selección natural pudiera estar manteniendo combinaciones alélicas específicas. El conjunto de datos sugiere que dentro de estas inversiones cromosómicas podría haber un efecto pleiotrópico entre los genes candidatos que confieren adaptación a frío y absorción de fósforo en el suelo. De esta forma, las inversiones cromosómicas enriquecidas en estos genes candidatos podrían generar el puente entre adaptación divergente y aislamiento reproductivo (cambio en la floración). Aunque las señales genéticas sugieren que el ambiente podría estar promoviendo el aislamiento reproductivo entre las dos subespecies será importante realizar experimentos que

comprueben el aislamiento reproductivo entre las dos subespecies y determinar si las posibles inversiones cromosómicas juegan un papel importante en el aislamiento reproductivo. Para ello se podrían hacer cruzas en jardines comunes y utilizando poblaciones con distintas frecuencias de las inversiones cromosómicas.

Aunque la hipótesis de selección multifaria podría ser importante para finalizar el proceso de especiación ecológica, han sido pocos los estudios que la han probado empíricamente (Nosil y Sandoval 2008; Michel *et al.* 2010; Scholl *et al.* 2012; Liu *et al.* 2013; Arnegard *et al.* 2014; Malinsky *et al.* 2015). Considero interesante que la adaptación a los dos ejes climáticos en teocintles (temperatura y disponibilidad de fósforo en el suelo) no ha sido simultánea, ya que la adaptación a menor disponibilidad de fósforo en el suelo habría ocurrido hasta que *mexicana* migrara a tierras altas. En *Timema cristinae* se ha propuesto que la falta de más ejes ecológicos podría explicar que la especiación ecológica nunca se haya completado (Nosil y Sandoval 2008). Estas especies únicamente difieren en su capacidad de camuflajearse en respuesta a depredadores, pero los distintos ecotipos no se han adaptado a alimentarse más eficientemente en los hospederos donde presentan mayor cripsis (Nosil y Sandoval 2008). Por el contrario, las especies *T. podura* y *T. chumash* ya han evolucionado a lo largo de dos nichos, primero adaptándose a ocultarse de los depredadores y después adaptándose a alimentarse preferentemente en las plantas en las que se esconden (Nosil y Sandoval 2008). Será interesante comparar distintas especies y ver si la adaptación a múltiples nichos ocurre al mismo tiempo o tiene que ser progresiva. Asimismo será interesante analizar si la divergencia en teocintles ocurre a lo largo de otros ejes del nicho. Nosotros analizamos la divergencia a lo largo de variables abióticas, pero será muy importante analizar el efecto de parásitos y herbívoros, los cuales pueden estar correlacionados también con cambios en la altitud.

Estudios recientes apoyados en las nuevas tecnología indican que la especiación ecológica parece ser más común de lo que se creía (Nosil 2012; Shafer y Wolf 2013). Este proceso podría ser el responsable de generar las radiaciones adaptativas y el origen de nuevas especies aprovechando las oportunidades ecológicas (Schluter 2016). La diversidad de especies a lo largo del planeta es heterogénea y justamente la divergencia de nicho y la heterogeneidad ambiental podrían ser elementos importantes que expliquen estos cambios en las tasas de diversificación (Kozak y Wiens 2010; Ramírez-Barahona *et al.* 2016). En

países como México, que son tan heterogéneos ambientalmente, la especiación ecológica podría ser más importante de lo que pensamos y este trabajo contribuirá a plantear mejor esta teoría.

Conservación, mejoramiento y cambio climático

El calentamiento global es una de las principales amenazas para la biodiversidad (Pecl *et al.* 2017). En el caso de las especies cultivadas, esto es alarmante, ya que la alimentación y gran parte de la economía depende de estos cultivos (Godfray *et al.* 2010). En el caso del maíz, se ha proyectado que el cambio climático tendrá repercusiones en su diversidad y distribución (Ureta *et al.* 2012). Por lo tanto, es importante desarrollar estrategias de mitigación, y en el caso de las especies cultivadas, se ha propuesto que se podrían introducir genes adaptativos a partir de las especies silvestres (Warschefsky *et al.* 2014; Dempewolf *et al.* 2017; Palmgren *et al.* 2014).

Tomando en cuenta lo anterior, nuestros resultados podrían ser interesantes para proponer estrategias de mejoramiento, conservación y mitigación al cambio climático en el maíz. En los capítulos dos y tres encontramos señales de selección positiva en genes asociados principalmente con temperatura, precipitación y capacidad para extraer fósforo del suelo. Sin embargo, como menciono antes, nuestro muestreo y desarrollo experimental, solo nos permitió identificar asociaciones con características específicas analizadas. Existen muchos marcadores que parecen presentar una F_{ST} anómala, y que no se asociaron a ninguno de nuestros genes. Muchos de ellos podrían ser verdaderos candidatos a estar bajo selección a otras características. Las subespecie *mexicana* crece en suelos altos y volcánicos, por lo que pueden estar adaptadas a condiciones bióticas y abióticas que no consideramos (mayor índice de UV, menor presión atmosférica, menor presión parcial de O_2 y CO_2, metales pesados, parásitos). Siguiendo aproximaciones similares a la de esta tesis, en principio, sería posible identificar genes de interés para conservación. Actualmente, estamos trabajando en analizar el efecto que genes candidatos a estar bajo selección a mayor temperatura, podrían tener sobre la vulnerabilidad de las poblaciones en el futuro. Usando un método de aprendizaje automático, que permite construir un modelo que predice el cambio en las frecuencias alélicas a lo largo de un paisaje geográfico (Fitzpatrick and Keller 2015), hemos identificado poblaciones que serán menos vulnerables

a cambio climático. Estas poblaciones presentan ciertas combinaciones de SNPs, que podrían buscarse en razas de maíz, y eventualmente diseñar estrategias para responder al cambio climático.

No obstante antes de hacer manipulaciones hay tres aspectos fundamentales que tenemos que considerar. Primero, los análisis que realizamos durante este doctorado, se limitaron a identificar SNPs con señales de selección. Las anotaciones de estos SNPs generalmente son limitadas, y la mayoría de los genes candidatos corresponden a proteínas predichas (aunque en el capítulo 2, si tenemos una lista de genes interesantes). Será importante realizar anotaciones más finas (por ejemplo, buscando dominios conservados), y alrededor de los SNPs detectados, para tener información más fidedigna de los genes interesantes. Vielle-Calzada *et al.* (2009) y Fustier *et al.* (2017), realizando anotaciones más finas, detectaron genes relacionados con la detoxificación de metales pesados y con la utilización de fósforo en el suelo. Una segunda consideración importante, es que nosotros solo realizamos asociaciones entre frecuencias alélicas y ambiente, por lo que será necesario realizar experimentos de asociación, para corroborar que nuestros genes son buenos candidatos. Finalmente, es necesario considerar que las especies domesticadas, por definición, están estrechamente ligadas a los humanos y no necesariamente responden a las presiones ecológicas naturales. En el caso del maíz, se ha encontrado que el manejo puede ser incluso más importante que la adaptación local (Jorge Nieto, comunicación personal). Esto hace sentido, ya que los maíces son cuidados intensamente por los cultivadores y en función de sus necesidades. Tomando en cuenta que existe la posibilidad de que los maíces y los teocintles no respondan similarmente a las presiones de selección, será fundamental hacer estudios de asociación en maíz con los SNPs candidatos.

REFERENCIAS

Aguilée, R., Amaury Lambert, and D. Claessen. 2011. "Ecological Speciation in Dynamic Landscapes." *Journal of Evolutionary Biology* 24 (12):2663–77. https://doi.org/10.1111/j.1420-9101.2011.02392.x.

Aguirre-Liguori, Jonas A., Erika Aguirre-planter, and Luis E. Eguiarte. 2016. "Genetics and Ecology of Wild and Cultivated Maize: Domestication and Introgression." In *Ethnobotany of Mexico*. New York: Springer.

Aguirre-Liguori, Jonas A., Brandon S. Gaut, Maud I. Tenaillon, Sarah Hearne, Felipe García-Oliva, Juan Pablo Jaramillo-Correa, and Luis E. Eguiarte. 2017. "Ecological Speciation of Teosintes." *Proceedings of the National Academy of Sciences*.

Aguirre-Liguori, Jonas A., Maud I. Tenaillon, Alejandra Vázquez-Lobo, Brandon S. Gaut, Juan Pablo Jaramillo-Correa, Salvador Montes-Hernandez, Valeria Souza, and Luis E. Eguiarte. 2017. "Connecting Genomic Patterns of Local Adaptation and Niche Suitability in Teosintes." *Molecular Ecology*.

Allendorf, Fred W., Paul A. Hohenlohe, and Gordon Luikart. 2010. "Genomics and the Future of Conservation Genetics." *Nature Reviews. Genetics* 11 (10). Nature Publishing Group:697–709. https://doi.org/10.1038/nrg2844.

Anderson, Jill T., Cheng Ruei Lee, Catherine A. Rushworth, Robert I. Colautti, and Thomas Mitchell-Olds. 2013. "Genetic Trade-Offs and Conditional Neutrality Contribute to Local Adaptation." *Molecular Ecology* 22 (3):699–708. https://doi.org/10.1111/j.1365-294X.2012.05522.x.

Anderson, Jill T., and Thomas Mitchell-Olds. 2011. "Ecological Genetics and Genomics of Plant Defences: Evidence and Approaches." *Functional Ecology* 25 (2):312–24. https://doi.org/10.1111/j.1365-2435.2010.01785.x.

Anderson, Jill T., John H. Willis, and Thomas Mitchell-Olds. 2011. "Evolutionary Genetics of Plant Adaptation." *Trends in Genetics* 27 (7). Elsevier Ltd:258–66. https://doi.org/10.1016/j.tig.2011.04.001.

Andrew, Rose L., and Loren H. Rieseberg. 2013. "Divergence Is Focused on Few Genomic Regions Early in Speciation: Incipient Speciation of Sunflower Ecotypes." *Evolution* 67 (9):2468–82. https://doi.org/10.1111/evo.12106.

Antonovics, Janis. 2006. "Evolution in Closely Adjacent Plant Populations X: Long-Term

Persistence of Prereproductive Isolation at a Mine Boundary." *Heredity* 97 (1):33–37. https://doi.org/10.1038/sj.hdy.6800835.

Antonovics, Janis, and A. D. Bradshaw. 1970. "Evolution in Closely Adjacent Plant Populations." *Heredity* 25 (3):349–62. https://doi.org/10.1038/hdy.1978.44.

Arnegard, Matthew E., Matthew D. McGee, Blake Matthews, Kerry B. Marchinko, Gina L. Conte, Sahriar Kabir, Nicole Bedford, *et al.* 2014. "Genetics of Ecological Divergence during Speciation." *Nature* 511 (7509). Nature Publishing Group:1–17. https://doi.org/10.1038/nature13301.

Baird, Nathan A., Paul D. Etter, Tressa S. Atwood, Mark C. Currey, Anthony L. Shiver, Zachary A. Lewis, Eric U. Selker, William A. Cresko, and Eric A. Johnson. 2008. "Rapid SNP Discovery and Genetic Mapping Using Sequenced RAD Markers." *PLoS ONE* 3 (10):1–7. https://doi.org/10.1371/journal.pone.0003376.

Beaumont, Mark A. 2005. "Adaptation and Speciation: What Can Fst Tell Us?" *Trends in Ecology and Evolution* 20 (8):435–40. https://doi.org/10.1016/j.tree.2005.05.017.

Beaumont, Mark A., and Richard A. Nichols. 1996. "Evaluating Loci for Use in the Genetic Analysis of Population Structure." *Proceedings of the Royal Society of London Series B: Biological Sciences* 263 (1377):1619–26. https://doi.org/10.1098/rspb.1996.0237.

Bolnick, Daniel I. 2011. "Sympatric Speciation in Threespine Stickleback: Why Not?" *International Journal of Ecology* 2011. https://doi.org/10.1155/2011/942847.

Chevin, Luis Miguel, Guillaume Decorzent, and Thomas Lenormand. 2014. "Niche Dimensionality and the Genetics of Ecological Speciation." *Evolution* 68 (5):1244–56. https://doi.org/10.1111/evo.12346.

Chia, Jer-Ming, Chi Song, Peter J. Bradbury, Denise Costich, Natalia de Leon, John F. Doebley, Robert J. Elshire, *et al.* 2012. "Maize HapMap2 Identifies Extant Variation from a Genome in Flux." *Nature Genetics* 44 (7). Nature Publishing Group:803–7. https://doi.org/10.1038/ng.2313.

Coyne, J.A., and H.A. Orr. 2004. *Speciation*. Edited by Sinauer Associates. Sunderland.

Darwin, Charles. 1859. *On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life*. Edited by John Murray. London. http://www.uruguayeduca.edu.uy/Userfiles/P0001%5CFile%5COrigen de las

especies.pdf.

Dempewolf, Hannes, Gregory Baute, Justin Anderson, Benjamin Kilian, Chelsea Smith, and Luigi Guarino. 2017. “Past and Future Use of Wild Relatives in Crop Breeding.” https://doi.org/10.2135/cropsci2016.10.0885.

Díez, Concepción M, Brandon S. Gaut, Esteban Meca, Enrique Scheinvar, Salvador Montes-Hernandez, Luis E. Eguiarte, and Maud I. Tenaillon. 2013. “Genome Size Variation in Wild and Cultivated Maize along Altitudinal Gradients.” *The New Phytologist* 199 (1):264–76. https://doi.org/10.1111/nph.12247.

Donoghue, Michael J., and Erika J. Edwards. 2014. “Biome Shifts and Niche Evolution in Plants.” *Annual Review of Ecology, Evolution, and Systematics* 45:547–72. https://doi.org/10.1146/annurev-ecolsys-120213-091905.

Eckert, C. G., K. E. Samis, and S. C. Lougheed. 2008. “Genetic Variation across Species’ Geographical Ranges: The Central-Marginal Hypothesis and beyond.” *Molecular Ecology* 17 (5):1170–88. https://doi.org/10.1111/j.1365-294X.2007.03659.x.

Egan, Scott P., Gregory J. Ragland, Lauren Assour, Thomas H Q Powell, Glen R. Hood, Scott Emrich, Patrik Nosil, and Jeffrey L. Feder. 2015. “Experimental Evidence of Genome-Wide Impact of Ecological Selection during Early Stages of Speciation-with-Gene-Flow.” *Ecology Letters* 18 (8):817–25. https://doi.org/10.1111/ele.12460.

Eguiarte, Luis E., Jonas A. Aguirre-Liguori, Lev Jardón-barbolla, Erika Aguirre-planter, and Valeria Souza. 2013. “Genómica de Poblaciones: Nada En Evolución va a Tener Sentido Si No Es a La Luz de La Genómica, Y Nada En Genómica Tendrá Senntido Si No Es a La Luz de La Evolución.” *TIP Revista Especializada En Ciencias Químico-Biológicas* 16 (1):42–56.

Ekblom, Robert, and Juan Galindo. 2010. “Applications of next Generation Sequencing in Molecular Ecology of Non-Model Organisms.” *Heredity* 107 (1). Nature Publishing Group:1–15. https://doi.org/10.1038/hdy.2010.152.

Elser, J. J., W. F. Fagan, A. J. Kerkhoff, N. G. Swenson, and B. J. Enquist. 2010. “Biological Stoichiometry of Plant Production: Metabolism, Scaling and Ecological Response to Global Change.” *New Phytologist* 186 (3):593–608. https://doi.org/10.1111/j.1469-8137.2010.03214.x.

Elshire, Robert J., Jeffrey C. Glaubitz, Qi Sun, Jesse A. Poland, Ken Kawamoto, Edward S.

Buckler, and Sharon E. Mitchell. 2011. “A Robust, Simple Genotyping-by-Sequencing (GBS) Approach for High Diversity Species.” *PLoS ONE* 6 (5):1–10. https://doi.org/10.1371/journal.pone.0019379.

Excoffier, Laurent, Matthieu Foll, and Rémy J. Petit. 2009. “Genetic Consequences of Range Expansions.” *Annual Review of Ecology, Evolution, and Systematics* 40 (1):481–501. https://doi.org/10.1146/annurev.ecolsys.39.110707.173414.

Excoffier, Laurent, T Hofer, and Matthieu Foll. 2009. “Detecting Loci under Selection in a Hierarchically Structured Population.” *Heredity* 103 (4). Nature Publishing Group:285–98. https://doi.org/10.1038/hdy.2009.74.

Excoffier, Laurent, and Nicolas Ray. 2008. “Surfing during Population Expansions Promotes Genetic Revolutions and Structuration.” *Trends in Ecology and Evolution* 23 (7):347–51. https://doi.org/10.1016/j.tree.2008.04.004.

Fang, Zhou, Tanja Pyhäjärvi, Allison L. Weber, R. Kelly Dawe, Jeffrey C. Glaubitz, José de Jesus Sánchez González, Claudia Ross-Ibarra, John F. Doebley, Peter L. Morrell, and Jeffrey Ross-Ibarra. 2012. “Megabase-Scale Inversion Polymorphism in the Wild Ancestor of Maize.” *Genetics* 191 (3):883–94. https://doi.org/10.1534/genetics.112.138578.

Feder, Jeffrey L., Samuel M. Flaxman, Scott P. Egan, Aaron A. Comeault, and Patrik Nosil. 2013. “Geographic Mode of Speciation and Genomic Divergence.” *Annual Review of Ecology, Evolution, and Systematics* 44 (1):73–97. https://doi.org/10.1146/annurev-ecolsys-110512-135825.

Feder, Jeffrey L., Patrik Nosil, Aaron C. Wacholder, Scott P. Egan, Stewart H. Berlocher, and Samuel M. Flaxman. 2014. “Genome-Wide Congealing and Rapid Transitions across the Speciation Continuum during Speciation with Gene Flow.” *Journal of Heredity* 105:810–20. https://doi.org/10.1093/jhered/esu038.

Fishman, Lila, Angela Stathos, Paul M. Beardsley, Charles F. Williams, and Jeffrey P. Hill. 2013. “Chromosomal Rearrangements and the Genetics of Reproductive Barriers in Mimulus (Monkey Flowers).” *Evolution* 67 (9):2547–60. https://doi.org/10.1111/evo.12154.

Fitzpatrick, Matthew C., and Stephen R. Keller. 2015. “Ecological Genomics Meets Community-Level Modelling of Biodiversity: Mapping the Genomic Landscape of

Current and Future Environmental Adaptation." *Ecology Letters* 18 (1):1–16. https://doi.org/10.1111/ele.12376.

Flaxman, Samuel M., Jeffrey L. Feder, and Patrik Nosil. 2012. "Spatially Explicit Models of Divergence and Genome Hitchhiking." *Journal of Evolutionary Biology* 25 (12):2633–50. https://doi.org/10.1111/jeb.12013.

———. 2013. "Genetic Hitchhiking and the Dynamic Buildup of Genomic Divergence during Speciation with Gene Flow." *Evolution* 67 (1974):2577–91. https://doi.org/10.1111/evo.12055.

Flaxman, Samuel M., Aaron C. Wacholder, Jeffrey L. Feder, and Patrik Nosil. 2014. "Theoretical Models of the Influence of Genomic Architecture on the Dynamics of Speciation." *Molecular Ecology* 23:4074–88. https://doi.org/10.1111/mec.12750.

Fukunaga, Kenji, Jason Hill, Yves Vigouroux, Yoshihiro Matsuoka, Jesus Sanchez G., Kejun Liu, Edward S. Buckler, and John F. Doebley. 2005. "Genetic Diversity and Population Structure of Teosinte." *Genetics* 169 (4):2241–54. https://doi.org/10.1534/genetics.104.031393.

Funk, Daniel J., Scott P. Egan, and Patrik Nosil. 2011. "Isolation by Adaptation in Neochlamisus Leaf Beetles: Host-Related Selection Promotes Neutral Genomic Divergence." *Molecular Ecology*, 4671–82. https://doi.org/10.1111/j.1365-294X.2011.05311.x.

Funk, Daniel J., Patrik Nosil, and William J. Etges. 2006. "Ecological Divergence Exhibits Consistently Positive Associations with Reproductive Isolation across Disparate Taxa." *Proceedings of the National Academy of Sciences of the United States of America* 103 (9):3209–13. https://doi.org/10.1073/pnas.0508653103.

Fustier, Margaux-Alison, Jean-Tristan Bradenburg, Simon Boitard, Jason Lapeyronnie, Luis E. Eguiarte, Yves Vigouroux, Domenica Manicacci, and Maud I. Tenaillon. 2017. "Local Adaptation of Teosintes along Altitudinal Gradients Using Whole Genome Sequencing of Pooled Samples." *Molecular Ecology*.

Futuyma, Douglas J. 2005. *Evolution*. Sunderland: Sinauer associates.

Godfray, H Charles J., John R. Beddington, Ian R. Crute, Lawrence Haddad, David Lawrence, James F. Muir, Jules Pretty, Sherman Ronbinson, Sandy M. Thomas, and Camilla Toulmin. 2010. "Food Security: The Challenge of Feeding 9 Billion People."

Science 812. https://doi.org/10.1126/science.1185383.

Hampe, Arndt, and Rémy J. Petit. 2005. "Conserving Biodiversity under Climate Change: The Rear Edge Matters." *Ecology Letters* 8 (5):461–67. https://doi.org/10.1111/j.1461-0248.2005.00739.x.

He, Qixin, and L. Lacey Knowles. 2016. "Identifying Targets of Selection in Mosaic Genomes with Machine Learning: Applications in Anopheles Gambiae for Detecting Sites within Locally Adapted Chromosomal Inversions." *Molecular Ecology* 25:2226–43. https://doi.org/10.1111/mec.13619.

Hedrick, Philip W. 2011. *Genetics of Populations*. Jones and Bartlett Learning.

Heerwaarden, Joost van, John F. Doebley, William H Briggs, Jeffrey C. Glaubitz, Major M. Goodman, Jose De Jesús Sánchez González, and Jeffrey Ross-Ibarra. 2011. "Genetic Signals of Origin, Spread, and Introgression in a Large Sample of Maize Landraces." *Proceedings of the National Academy of Sciences* 108 (3):1088–92. https://doi.org/10.1073/pnas.1013011108.

Hendry, Andrew P., Patrik Nosil, and Loren H. Rieseberg. 2007. "The Speed of Ecological Speciation." *Functional Ecology* 21:455–64. https://doi.org/10.1111/j.1365-2435.2006.01240.x.The.

Hohenlohe, Paul A., Susan Bassham, Paul D. Etter, Nicholas Stiffler, Eric A. Johnson, and William A. Cresko. 2010a. "Population Genomics of Parallel Adaptation in Threespine Stickleback Using Sequenced RAD Tags." *PLoS Genetics* 6 (2). https://doi.org/10.1371/journal.pgen.1000862.

———. 2010b. "Population Genomics of Parallel Adaptation in Threespine Stickleback Using Sequenced RAD Tags." *PLoS Genetics* 6 (2):e1000862. https://doi.org/10.1371/journal.pgen.1000862.

Hufford, Matthew B., Paul Bilinski, Tanja Pyhäjärvi, and Jeffrey Ross-Ibarra. 2012. "Teosinte as a Model System for Population and Ecological Genomics." *Trends in Genetics* 28 (12):606–15. https://doi.org/10.1016/j.tig.2012.08.004.

Hufford, Matthew B., Enrique Martínez-Meyer, Brandon S. Gaut, Luis E. Eguiarte, and Maud I. Tenaillon. 2012a. "Inferences from the Historical Distribution of Wild and Domesticated Maize Provide Ecological and Evolutionary Insight." *PLoS ONE* 7 (11). https://doi.org/10.1371/journal.pone.0047659.

———. 2012b. “Inferences from the Historical Distribution of Wild and Domesticated Maize Provide Ecological and Evolutionary Insight.” *PLoS ONE* 7 (11). https://doi.org/10.1371/journal.pone.0047659.

Jones, Matthew R., Brenna R. Forester, Ashley I. Teufel, Rachael V. Adams, Daniel N. Anstett, Betsy A. Goodrich, Erin L. Landguth, Stéphane Joost, and Stéphanie Manel. 2013. “Integrating Landscape Genomics and Spatially Explicit Approaches to Detect Loci under Selection in Clinal Populations.” *Evolution* 67 (12):3455–68. https://doi.org/10.1111/evo.12237.

Joost, Stéphane, Aurélie Bonin, M. W. Bruford, L. Després, C. Conord, G. Erhardt, and Pierre Taberlet. 2007. “A Spatial Analysis Method (SAM) to Detect Candidate Loci for Selection: Towards a Landscape Genomics Approach to Adaptation.” *Molecular Ecology* 16 (18):3955–69. https://doi.org/10.1111/j.1365-294X.2007.03442.x.

Kirkpatrick, Mark, and Nick Barton. 2006. “Chromosome Inversions, Local Adaptation and Speciation.” *Genetics* 434 (May):419–34. https://doi.org/10.1534/genetics.105.047985.

Kozak, Kenneth H., and John J. Wiens. 2010. “Accelerated Rates of Climatic-Niche Evolution Underlie Rapid Species Diversification.” *Ecology Letters* 13 (11):1378–89. https://doi.org/10.1111/j.1461-0248.2010.01530.x.

Lenormand, Thomas. 2002. “Gene Flow and the Limits to Natural Selection \r.” *Trends in Ecology and Evolution* 17 (4):183–89.

———. 2012. “From Local Adaptation to Speciation: Specialization and Reinforcement.” *International Journal of Ecology* 2012. https://doi.org/10.1155/2012/508458.

Lewontin, Richard C., and Jesse Krakauer. 1973. “DISTRIBUTION OF GENE FREQUENCY AS A TEST OF THE THEORY O F THE SELECTIVE NEUTRALITY OF and LEWONTIN.” *Genetics* 74:175–95.

Linhart, Yan B, and Michael C Grant. 1996. “EVOLUTIONARY SIGNIFICANCE OF LOCAL GENETIC DIFFERENTIATION IN PLANTS\rdoi:10.1146/annurev.ecolsys.27.1.237.” *Annual Review of Ecology and Systematics* 27 (1):237–77. https://doi.org/10.1146/annurev.ecolsys.27.1.237.

Lira-Noriega, Andrés, and Joseph D. Manthey. 2014. “Relationship of Genetic Diversity and Niche Centrality: A Survey and Analysis.” *Evolution* 68 (4):1082–93.

https://doi.org/10.1111/evo.12343.

Liu, Jie, Michael Möller, Jim Provan, Lian Ming Gao, Ram Chandra Poudel, and De Zhu Li. 2013. “Geological and Ecological Factors Drive Cryptic Speciation of Yews in a Biodiversity Hotspot.” *New Phytologist* 199 (4):1093–1108. https://doi.org/10.1111/nph.12336.

Lowry, David B., and John H. Willis. 2010. “A Widespread Chromosomal Inversion Polymorphism Contributes to a Major Life-History Transition , Local Adaptation , and Reproductive Isolation.” *PLoS Biology* 8 (9). https://doi.org/10.1371/journal.pbio.1000500.

MacColl, Andrew D C. 2011. “The Ecological Causes of Evolution.” *Trends in Ecology and Evolution* 26 (10):514–22. https://doi.org/10.1016/j.tree.2011.06.009.

Malinsky, M, R J Challis, A M Tyers, S Schiffels, Y Terai, B P Ngatunga, E A Miska, R Durbin, M J Genner, and G F Turner. 2015. “Genomic Islands of Speciation Separate Cichlid Ecomorphs in an East African Crater Lake.” *Science* 350 (6267):1493–98. https://doi.org/DOI: 10.1126/science.aac9927.

Matsuoka, Yoshihiro, Yves Vigouroux, Major M. Goodman, Jesus Sanchez G, Edward S. Buckler, and John F. Doebley. 2002. “A Single Domestication for Maize Shown by Multilocus Microsatellite Genotyping.” *Proceedings of the National Academy of Sciences of the United States of America* 99 (9):6080–84. https://doi.org/10.1073/pnas.052125199.

Mayr, E. 1942. *Systematics and the Origin of Species*. Edited by Columbia University Press. New York.

Metzker, Michael L. 2010. “Sequencing Technologies - the next Generation.” *Nature Reviews. Genetics* 11 (1). Nature Publishing Group:31–46. https://doi.org/10.1038/nrg2626.

Michel, Andrew P, Sheina B. Sim, Thomas H Q Powell, Michael S Taylor, Patrik Nosil, and Jeffrey L. Feder. 2010. “Widespread Genomic Divergence during Sympatric Speciation.” *Proceedings of the National Academy of Sciences of the United States of America* 107:9724–29. https://doi.org/10.1073/pnas.1000939107.

Mita, Stéphane De, Anne Céline Thuillet, Laurène Gay, Nourollah Ahmadi, Stéphanie Manel, Joëlle Ronfort, and Yves Vigouroux. 2013. “Detecting Selection along

Environmental Gradients: Analysis of Eight Methods and Their Effectiveness for Outbreeding and Selfing Populations." *Molecular Ecology* 22 (5):1383–99. https://doi.org/10.1111/mec.12182.

Moeller, David A., Maud I. Tenaillon, and Peter L. Tiffin. 2007. "Population Structure and Its Effects on Patterns of Nucleotide Polymorphism in Teosinte (Zea Mays Ssp. Parviglumis)." *Genetics* 176 (3):1799–1809. https://doi.org/10.1534/genetics.107.070631.

Moreno-Letelier, Alejandra, Jonas A. Aguirre-Liguori, Maud I. Tenaillon, Daniel Piñero, Alejandra Vázquez-Lobo, and Luis E. Eguiarte. 2017. "Maize Domestication: Genetic Divergence, Gene Flow and Ancestral Introrgession between Teosinte and Maize Show a Complex History and an Origin in Jalisco." *Proceedings of the National Academy of Sciences*, 1–14.

Nosil, Patrik. 2012. *Ecological Speciation*. Edited by Paul H. Harvey, Robert M. May, H. Charles J. Godfray, and Jennifer A. Dunne. Oxford: Oxford University Press.

Nosil, Patrik, and Bernard J. Crespi. 2006. "Experimental Evidence That Predation Promotes Divergence in Adaptive Radiation." *Proceedings of the National Academy of Sciences of the United States of America* 103 (24):9090–95. https://doi.org/10.1073/pnas.0601575103.

Nosil, Patrik, Scott P. Egan, and Daniel J. Funk. 2008. "Heterogeneous Genomic Differentiation between Walking-Stick ecotypes: 'Isolation by Adaptation' and Multiple Roles for Divergent Selection." *Evolution* 62 (2):316–36. https://doi.org/10.1111/j.1558-5646.2007.00299.x.

Nosil, Patrik, and Jeffrey L. Feder. 2012. "Genomic Divergence during Speciation: Causes and Consequences." *Philosophical Transactions of the Royal Society B: Biological Sciences* 367 (1587):332–42. https://doi.org/10.1098/rstb.2011.0263.

Nosil, Patrik, Jeffrey L. Feder, Samuel M. Flaxman, and Zachariah Gompert. 2017. "Tipping Points in the Dynamics of Speciation." *Nature Publishing Group* 1 (January). Macmillan Publishers Limited:1–8. https://doi.org/10.1038/s41559-016-0001.

Nosil, Patrik, Daniel J. Funk, and Daniel Ortiz-Barrientos. 2009. "Divergent Selection and Heterogeneous Genomic Divergence." *Molecular Ecology* 18:375–402.

https://doi.org/10.1111/j.1365-294X.2008.03946.x.

Nosil, Patrik, Zachariah Gompert, Timothy E. Farkas, Aaron A. Comeault, Jeffrey L. Feder, C. Alex Buerkle, and Thomas L. Parchman. 2012. "Genomic Consequences of Multiple Speciation Processes in a Stick Insect." *Proceedings of the Royal Society B: Biological Sciences*, no. June:5058–65. https://doi.org/10.1098/rspb.2012.0813.

Nosil, Patrik, Luke J. Harmon, and Ole Seehausen. 2009. "Ecological Explanations for (Incomplete) Speciation." *Trends in Ecology and Evolution* 24 (January):145–56. https://doi.org/10.1016/j.tree.2008.10.011.

Nosil, Patrik, and Cristina P. Sandoval. 2008. "Ecological Niche Dimensionality and the Evolutionary Diversification of Stick Insects." *PLoS ONE* 3 (4). https://doi.org/10.1371/journal.pone.0001907.

Nosil, Patrik, and Dolph Schluter. 2011. "The Genes Underlying the Process of Speciation." *Trends in Ecology and Evolution* 26 (4). Elsevier Ltd:160–67. https://doi.org/10.1016/j.tree.2011.01.001.

Palmgren, Michael G, Anna Kristina Edenbrandt, Suzanne Elizabeth Vedel, Martin Marchman Andersen, Xavier Landes, Jeppe Thulin Østerberg, Janus Falhof, *et al.* 2014. "Are We Ready for Back-to-Nature Crop Breeding ?" *Trends in Plant Science*. Elsevier Ltd, 1–10. https://doi.org/10.1016/j.tplants.2014.11.003.

Pecl, Gretta T, Miguel B Araújo, Johann D Bell, Julia Blanchard, Timothy C Bonebrake, I-ching Chen, Timothy D Clark, *et al.* 2017. "Biodiversity Redistribution under Climate Change: Impacts on Ecosystems and Human Well-Being" 9214. https://doi.org/10.1126/science.aai9214.

Pickrell, Joseph K., and Jonathan K. Pritchard. 2012. "Inference of Population Splits and Mixtures from Genome-Wide Allele Frequency Data." *PLoS Genetics* 8 (11). https://doi.org/10.1371/journal.pgen.1002967.

Pyhäjärvi, Tanja, Matthew B. Hufford, Sofiane Mezmouk, and Jeffrey Ross-Ibarra. 2013. "Complex Patterns of Local Adaptation in Teosinte." *Genome Biology and Evolution* 5 (9):1594–1609. https://doi.org/10.1093/gbe/evt109.

Ramírez-Barahona, Santiago, L Barrera-Redondo, and Luis E. Eguiarte. 2016. "Rates of Ecological Divergence and Body Size Evolution Are Correlated with Species Diversification in Scaly Tree Ferns." *Proceedings of the Royal Society B*

283:20161098. https://doi.org/10.1098/rspb.2016.1098.

Ren, Runsheng, Rumiana Ray, Pingfang Li, Jinhua Xu, Man Zhang, Guang Liu, Xiefeng Yao, Andrzej Kilian, and Xingping Yang. 2015. "Construction of a High-Density DArTseq SNP-Based Genetic Map and Identification of Genomic Regions with Segregation Distortion in a Genetic Population Derived from a Cross between Feral and Cultivated-Type Watermelon." *Molecular Genetics and Genomics* 290 (4). Springer Berlin Heidelberg:1457–70. https://doi.org/10.1007/s00438-015-0997-7.

Rice, Amber M., Andreas Rudh, Hans Ellegren, and Anna Qvarnström. 2011. "A Guide to the Genomics of Ecological Speciation in Natural Animal Populations." *Ecology Letters* 14 (1):9–18. https://doi.org/10.1111/j.1461-0248.2010.01546.x.

Riesch, Rüdiger, Moritz Muschick, Dorothea Lindtke, Romain Villoutreix, Aaron A. Comeault, Timothy E. Farkas, Kay Lucek, *et al.* 2017. "Transitions between Phases of Genomic Differentiation during Stick-Insect Speciation." *Nature Ecology & Evolution* 1 (FEbruary). Macmillan Publishers Limited, part of Springer Nature.:82. https://doi.org/10.1038/s41559-017-0082.

Rodríguez F, J. G., J. J. Sánchez G, Baltazar M. Baltazar, L. De la Cruz L, Fernando Santacruz-Ruvalcaba, J. Ron P, and J. B. Schoper. 2006. "Characterization of Floral Morphology and Synchrony among Zea Species in Mexico." *Maydica* 51 (2). Maydica:383–98. http://cat.inist.fr/?aModele=afficheN&cpsidt=17987262.

Ross-Ibarra, Jeffrey, Peter L. Morrell, and Brandon S. Gaut. 2007. "Plant Domestication, a Unique Opportunity to Identify the Genetic Basis of Adaptation." *Proceedings of the National Academy of Sciences of the United States of America* 104 Suppl (suppl_1):8641–48. https://doi.org/10.1073/pnas.0700643104.

Ross-Ibarra, Jeffrey, Maud I. Tenaillon, and Brandon S. Gaut. 2009. "Historical Divergence and Gene Flow in the Genus Zea." *Genetics* 181 (4):1399–1413. https://doi.org/10.1534/genetics.108.097238.

Rundle, Howard D., and Patrik Nosil. 2005. "Ecological Speciation." *Ecology Letters* 8:336–52. https://doi.org/10.1111/j.1461-0248.2004.00715.x.

Sánchez-González, José J., Takeo A. Kato-Yamamake, Mario A. Aguilar-Sanmiguel, Juan M. Hernández-Casillas, Angel López-Rodriguez, and José A. Ruiz-Corral. 1998. *Distribución Y Caracterización Del Teocintle*. Edited by INIFAP. Guadalajara.

Sansaloni, Carolina, Cesar Petroli, Damian Jaccoud, Jason Carling, Frank Detering, Dario Grattapaglia, and Andrzej Kilian. 2011. "Diversity Arrays Technology (DArT) and next-Generation Sequencing Combined: Genome-Wide, High Throughput, Highly Informative Genotyping for Molecular Breeding of Eucalyptus." *BMC Proceedings* 5 (Suppl 7):P54. https://doi.org/10.1186/1753-6561-5-S7-P54.

Schluter, Dolph. 2000. *The Ecology of Adaptive Radiation*. Edited by Oxford University Press. Oxford.

———. 2001. "Ecology and the Origin of Species." *Trends in Ecology and Evolution* 16 (7):372–80. https://doi.org/10.1016/S0169-5347(01)02198-X.

———. 2016. "Speciation, Ecological Opportunity, and Latitude." *The American Naturalist* 187 (1):1–18. https://doi.org/10.1086/684193.

Schnable, Patrick S., Ware Doreen, Robert S. Fulton, Joshua C. Stein, Fusheng Wei, Shiran Pasternak, Chengzhi Liang, *et al.* 2009. "The B73 Maize Genome: Complexity, Diversity, and Dynamics." *Science* 326:1112–15.

Scholl, Cynthia F., Christopher C. Nice, James A. Fordyce, Zachariah Gompert, and Matthew L. Forister. 2012. "Larval Performance in the Context of Ecological Diversification and Speciation in Lycaeides Butterflies." *International Journal of Ecology* 2012. https://doi.org/10.1155/2012/242154.

Schoville, Sean D., Aurélie Bonin, Olivier Francois, Stéphane Lobreaux, Christelle Melodelima, and Stéphanie Manel. 2012. "Adaptive Genetic Variation on the Landscape: Methods and Cases." *Annual Review of Ecology, Evolution, and Systematics, Vol 43* 43:23–43. https://doi.org/10.1146/annurev-ecolsys-110411-160248.

Schoville, Sean D., Aurélie Bonin, Olivier François, Stéphane Lobreaux, Christelle Melodelima, and Stéphanie Manel. 2012. "Adaptive Genetic Variation on the Landscape: Methods and Cases." *Annual Review of Ecology, Evolution, and Systematics* 43 (1):23–43. https://doi.org/10.1146/annurev-ecolsys-110411-160248.

Servedio, Maria R., G. Sander Van Doorn, Michael Kopp, Alicia M. Frame, and Patrik Nosil. 2011. "Magic Traits in Speciation: 'Magic' but Not Rare?" *Trends in Ecology and Evolution* 26 (8):389–97. https://doi.org/10.1016/j.tree.2011.04.005.

Sexton, Jason P., Patrick J. McInyre, Amy L. Angert, and Kevin J. Rice. 2009. "Evolution

and Ecology of Species Range Limits .” *Annual Review of Ecology, Evolution, and Systematics* 40 (1):415–36. https://doi.org/doi:10.1146/annurev.ecolsys.110308.120317.

Shafer, Aaron B A, and Jochen B W Wolf. 2013. “Widespread Evidence for Incipient Ecological Speciation: A Meta-Analysis of Isolation-by-Ecology.” *Ecology Letters* 16 (7):940–50. https://doi.org/10.1111/ele.12120.

Soria-Carrasco, Víctor, Zachariah Gompert, Aaron A. Comeault, Timothy E. Farkas, Thomas L. Parchman, J. Spencer Johnston, C. Alex Buerkle, *et al.* 2014. “Stick Insect Genomes Reveal Natural Selection’s Role in Parallel Speciation.” *Science* 344:738–42. https://doi.org/10.1126/science.1172133.

Stapley, Jessica, Julia Reger, P. G D Feulner, Carole M. Smadja, Juan Galindo, Robert Ekblom, Clair Bennison, Alexander D. Ball, Andrew P. Beckerman, and Jon Slate. 2010. “Adaptation Genomics: The next Generation.” *Trends in Ecology and Evolution* 25 (12). Elsevier Ltd:705–12. https://doi.org/10.1016/j.tree.2010.09.002.

Surget-Groba, Yann, Helena Johansson, and Roger S. Thorpe. 2012. “Synergy between Allopatry and Ecology in Population Differentiation and Speciation.” *International Journal of Ecology* 2012. https://doi.org/10.1155/2012/273413.

Tenaillon, Maud I., Matthew B. Hufford, Brandon S. Gaut, and Jeffrey Ross-Ibarra. 2011. “Genome Size and Transposable Element Content as Determined by High-Throughput Sequencing in Maize and Zea Luxurians.” *Genome Biology and Evolution* 3 (1):219–29. https://doi.org/10.1093/gbe/evr008.

Tening, a. S., J. N. Foba-Tendo, S. Y. Yakum-Ntaw, and F. Tchuenteu. 2013. “Phosphorus Fixing Capacity of a Volcanic Soil on the Slope of Mount Cameroon.” *Agriculture and Biology Journal of North America* 4 (1990):166–74. https://doi.org/10.5251/abjna.2013.4.3.166.174.

Twyford, Alex D., and Jannice Friedman. 2015. “Adaptive Divergence in the Monkey Flower Mimulus Guttatus Is Maintained by a Chromosomal Inversion.” *Evolution* 69 (6):1476–86. https://doi.org/10.1111/evo.12663.

Ureta, Carolina, Enrique Martínez-Meyer, Hugo R. Perales, and Elena R. Álvarez-Buylla. 2012. “Projecting the Effects of Climate Change on the Distribution of Maize Races and Their Wild Relatives in Mexico.” *Global Change Biology* 18 (3):1073–82.

https://doi.org/10.1111/j.1365-2486.2011.02607.x.

Via, Sara. 2009. “Natural Selection in Action during Speciation.” *Proceedings of the National Academy of Sciences of the United States of America* 106 Suppl:9939–46. https://doi.org/10.1073/pnas.0901397106.

Villemereuil, Pierre de, and Oscar E. Gaggiotti. 2015. “A New FST-Based Method to Uncover Local Adaptation Using Environmental Variables.” *Methods in Ecology and Evolution* 6 (11):1248–58. https://doi.org/10.1111/2041-210X.12418.

Warschefsky, Emily, R. Varma Penmetsa, Douglas R. Cook, and Eric J. Von Wettberg. 2014. “Back to the Wilds: Tapping Evolutionary Adaptations for Resilient Crops through Systematic Hybridization with Crop Wild Relatives.” *American Journal of Botany* 101 (10):1791–1800. https://doi.org/10.3732/ajb.1400116.

Zerjal, Tatiana, Agnès Rousselet, Corinne Mhiri, Valérie Combes, Delphine Madur, Marie-Angèle Grandbastien, Alain Charcosset, and Maud I. Tenaillon. 2012. “Maize Genetic Diversity and Association Mapping Using Transposable Element Insertion Polymorphisms.” *TAG. Theoretical and Applied Genetics. Theoretische Und Angewandte Genetik* 124 (8):1521–37. https://doi.org/10.1007/s00122-012-1807-9.

ANEXOS

Anexo 1, Artículo: Genómica de poblaciones: Nada en evolución va a tener sentido si no es a la luz de la genómica, y nada en genómica tendrá sentido si no es a la luz de la evolución.

D.R. © TIP Revista Especializada en Ciencias Químico-Biológicas, 16(1):42-56, 2013.

Genómica de Poblaciones: Nada en Evolución va a tener sentido si no es a la luz de la Genómica, y nada en Genómica tendrá sentido si no es a la luz de la Evolución

***Luis E. Eguiarte, Jonás A. Aguirre-Liguori, Lev Jardón-Barbolla, Erika Aguirre-Planter y Valeria Souza**

Lab. de Evolución Molecular y Experimental, Depto. de Ecología Evolutiva, Instituto de Ecología, Universidad Nacional Autónoma de México. Ciudad Universitaria, C.P. 04510, Deleg. Coyoacán, México, D.F. E-mail: *fruns@servidor.unam.mx

Resumen

La teoría de la genética de poblaciones surgió hace más de 80 años y nos permite explicar los patrones de variación genética dentro y entre las poblaciones que forman a las especies en términos de las fuerzas evolutivas. Este programa de investigación generó las preguntas que se han abordado empíricamente mediante marcadores moleculares desde hace medio siglo. Una pregunta fundamental ha sido hasta dónde un conjunto reducido de loci es o no representativo del efecto de las fuerzas evolutivas, sobre todo el genoma de una especie. Esto ha llevado al desarrollo creciente de aproximaciones que permitan conocer de manera representativa los niveles de variación genética en las poblaciones naturales, dando origen a la genómica de poblaciones. En años recientes, las técnicas de secuenciación masiva, llamadas *Next generation sequencing*, o *next-gen*, han permitido obtener datos de grandes secciones del genoma de diferentes especies, sin que sea un requisito conocer marcadores previos. Así, al comparar los genomas de muchos individuos de diferentes poblaciones, tenemos acceso al archivo de su historia evolutiva, que nos habla del complejo y dinámico balance en el tiempo entre la selección natural y las otras fuerzas evolutivas de carácter neutral, como la deriva y el flujo génico. La existencia de enormes cantidades de información ha requerido el desarrollo de nuevas herramientas estadísticas y bioinformáticas para su análisis. Diversas disciplinas se han visto beneficiadas de estos desarrollos. Para la biología evolutiva se abre la posibilidad de estudiar de manera más precisa y clara los patrones adaptativos de la variación. Tener genomas anotados y loci bien mapeados es relevante y arduo, pero el desarrollo técnico hace que lo anterior sea cada vez más plausible, y el reto será ser capaces de plantear preguntas adecuadas para hacer inferencias del mar de información disponible. El uso de una perspectiva evolutiva y de genética de poblaciones, enriquecerá a la genómica, de la misma manera que los datos genómicos nos ayudarán a avanzar en el desarrollo del programa iniciado por Theodosius Dobzhansky a mediados del siglo pasado.

Palabras Clave: Adaptación, genética de poblaciones, maíz, *next generation sequencing*, selección natural, teosinte.

Abstract

The theory of population genetics originated over 80 years ago and allowed to explain, in terms of the evolutionary forces, the patterns of genetic variation within and between the populations that conform species. This research program generated the questions that have been empirically analyzed with the use of molecular markers for the last 50 years. A fundamental question within population genetics is if a reduced number of genes are representative of the evolutionary forces that affect the total genome of a species. This question has led to the development of molecular methods that allow the study of large sections of the genome in natural populations, giving rise to the field of population genomics. In recent years, techniques that are able to sequence DNA massively, usually called "Next generation sequencing" or "next-gen", are helping us to obtain genome wide data in many species, without needing previous molecular information. Comparing the genomes of many individuals from different populations, now we have access to an archive of their evolutionary history that narrates the complex and dynamic balance in time between natural selection and other evolutionary forces, such as genetic drift and gene flow, which act mainly in neutral regions of the genomes. The amount of information that is being produced has required the development of new statistical and bioinformatics tools for their analyses. Diverse disciplines have profited from these new developments. In particular in evolutionary biology it is now possible to study in a more precise way the adaptive patterns of variation. The annotation of genomes and the mapping of traits are important and complicated, but recent technical developments are making these goals easier, and thus the future challenge will be in asking the right questions to make relevant inferences from the sea of information these new methods generate. The evolutionary and population genetics perspective will enrich genomics, in the same way that the genomic data will help us advance in the development of the program initiated by Theodosius Dobzhansky several decades ago.

Key Words: Adaptation, population genetics, maize, natural selection, next generation sequencing, teosinte.

Nota: Artículo recibido el 01 de febrero de 2013 y aceptado el 11 de marzo de 2013.

"Me dijo: Más recuerdos tengo yo de los que habrán tenido todos los hombres desde que el mundo es mundo."
"… le molestaba que el perro de las tres y catorce (visto de perfil) tuviera el mismo nombre que el perro de las tres y cuarto (visto de frente)"
Jorge Luis Borges, *Funes el Memorioso*

Genética de poblaciones, la joya de la biología evolutiva

Theodosius Dobzhansky[1] señaló acertadamente que nada tiene sentido en la biología si no es a la luz de la evolución. Cuando Dobzhansky hizo su propuesta en 1973, apenas se comenzaba a realizar el estudio genético de las poblaciones naturales, lo que podemos llamar la genética de poblaciones empírica. La genética de poblaciones explora los niveles de variación genética dentro y entre las poblaciones que forman a las especies y explica sus patrones en términos de las fuerzas evolutivas.

Podemos pensar que las especies están conformadas por una serie de poblaciones conectadas por flujo génico y el conjunto de todas las poblaciones entre las que hay flujo génico constituye a una especie. Si las poblaciones tienen altos niveles de variación genética, implica que son (y han sido) muy grandes y/o que existe elevado flujo génico entre ellas; si al contrario las poblaciones que forman una especie son muy diferentes entre sí, es posible que sean pequeñas y/o que hay muy poco flujo génico entre ellas. La genética de poblaciones busca entender a los procesos evolutivos a una escala ecológica, misma que implica relativamente pocas generaciones - en otras palabras, cambios a corto plazo-, y cómo las diferentes fuerzas evolutivas afectan a esta variación.

Ya mencionamos una fuerza evolutiva, el *flujo génico*, e implícitamente comentamos sobre el tamaño de las poblaciones, que nos habla de qué tan importante es otra fuerza evolutiva, la *deriva génica*. La deriva génica es el cambio al azar en las proporciones de los alelos (i.e., las diferentes formas que puede tener un gen) en una población, debido a que por puro azar algunos individuos dejan más hijos y copias de sus genes que otros; entre menos individuos dejen descendientes, mayores van a ser estos cambios azarosos. Adicionalmente podemos mencionar a otra fuerza, la *mutación* que genera toda la variación genética. Pero la fuerza evolutiva que más nos interesa es la *selección natural*. Así, los individuos que tienen los alelos ventajosos, dejan más hijos y estos alelos incrementan su proporción en las pozas génicas, mientras que los alelos que no funcionan bien son removidos rápidamente por la selección natural.

En términos de la genética de poblaciones, si encontramos alelos con diferentes proporciones en las poblaciones, pueden deberse al balance entre la deriva génica y el flujo génico; a mayor deriva génica, van a ser más diferentes las poblaciones, mientras que entre más alto sea el flujo génico, serán más parecidas. Pero si encontramos diferencias entre las poblaciones en sólo ciertos genes, es posible que estemos observando los efectos de la selección natural, especialmente si sucede de manera paralela: si en ciertas condiciones siempre encontramos unos alelos (por ejemplo, en las poblaciones más secas), y otros alelos en las condiciones contrastantes (por ejemplo, sólo se encuentra ese alelo en las poblaciones más húmedas).

La genética de poblaciones, el estudio de los genes en las poblaciones, aunque tuvo un origen modesto en diferentes artículos y libros teóricos de varios investigadores en Inglaterra y Estados Unidos en los años 30 del siglo pasado (básicamente por R.A. Fisher, S. Wright y J.B.S. Haldane[2-5]), ha demostrado ser la herramienta más poderosa para el estudio de la evolución. La evolución a su vez es el estudio de los patrones y procesos que han producido el cambio de los organismos en el tiempo, y es el resultado de las fuerzas evolutivas operando sobre la variación genética por mucho tiempo. Este proceso evolutivo ha generado tanto la adaptación (el ajuste de los organismos a su medio ambiente), como la gran diversidad de especies que hay y han existido en el pasado en la tierra (¡millones de especies!).

Ahora, la genética de poblaciones inicialmente consideraba explorar, en un solo gen, el comportamiento de dos alelos que segregan de acuerdo a las leyes de Mendel. Este caso se analiza de manera sencilla, y el famoso Equilibrio de Hardy-Weinberg describe cómo se comporta la variación en ausencia de cualquier fuerza evolutiva y es una especie *modelo nulo* de que sucede si no opera ningún proceso evolutivo. De esta manera se explora el efecto de cada fuerza evolutiva… pero, ¿qué pasa al considerar a modelos más realistas (y por lo tanto más complicados) de la genética?

Actualmente, sabemos mucho más sobre el material genético del que se sabía cuando Dobzhansky inició sus estudios empíricos de la genética de las poblaciones naturales en los años 30 del siglo pasado. Sabemos que el material genético es el ADN, y cómo operan las mutaciones que lo cambian. Es obvio que no sólo tenemos un gen (locus) con dos alelos, sino que tenemos miles de genes en los genomas (unos 23 mil en el caso del humano, por ejemplo), cada uno con muchísimas posibles versiones (i.e., alelos) y puede haber mutaciones a lo largo de toda la secuencia del gen, por lo que, teóricamente, es posible que exista un número casi infinito de formas de cada uno de los genes, que a su vez están rodeados e incluyen grandes cantidades

de material genético no codificante, a veces con funciones de regulación de la expresión de los genes, pero otras veces sin una función clara.

Primeros marcadores genéticos

La exploración del material genético en las poblaciones tiene una larga y prestigiosa historia. Aunque muchos biólogos, incluyendo a Charles Darwin, habían hecho cruzas para tratar de entender el comportamiento genético, el honor recayó en Gregor Mendel, que descubrió como segregaba la variación genética en una cruza en casos de herencia de un solo caracter, determinado por un solo gen con dos formas alélicas, y en otros casos de herencia sencilla, con formas claramente determinadas por los genes. Así, durante mucho tiempo el trabajo genético dependió de encontrar a esta variación y a estos mutantes. En tiempos de Dobzhansky, además de trabajar con los mutantes clásicos de la mosca de la fruta (*Drosophila*), se comenzó a trabajar con otros marcadores genéticos (ver Tabla I). Por un lado, se descubrieron los cromosomas gigantes de las glándulas salivales de las *Drosophila*, que permiten el análisis de diferentes inversiones y rearreglos en los cromosomas. Así, Dobzhansky comenzó el estudio de la variación genética en una especie de *Drosophila*, *D. pseudoobscura*, en las montañas del sur de California, Arizona y norte de México, encontrando claros patrones estacionales que sugerían adaptación temporal. También se iniciaron complicados experimentos de cruzas en los que se obtenían moscas completamente homócigas para cromosomas enteros que se comparaban con la viabilidad de moscas silvestres, que son heterócigas para esos cromosomas, experimentos que sugerían la existencia de altos niveles de variación genética (ver un resumen de estos experimentos, en Hedrick[6] págs. 51-57).

En 1966, un ex alumno de Dobzhansky, Richard Lewontin [7], junto con un biólogo molecular, John L. Hubby, decidieron aplicar una sencilla técnica molecular para el estudio de la variación en las poblaciones naturales, llamada electroforesis de isoenzimas o alozimas (ver Tabla I), que básicamente consiste en separar proteínas según su carga y su tamaño usando un campo eléctrico; estas proteínas son codificadas por genes de herencia mendeliana y se pueden analizar usando la teoría más básica de la genética de poblaciones. En su estudio Lewontin y Hubby [7] analizaron 18 loci (genes), 9 de los cuales resultaron polimórficos en *D. pseudoobscura*. El estudio reveló altos niveles de variación genética, abriendo de manera democrática el campo de la genética de poblaciones empírica: por primera vez se podría analizar la variación genética y así empezar el estudio de la genética de poblaciones en varias especies y a conocer el papel relativo de las poblaciones naturales de cualquier especie, a un costo relativamente moderado.

Los biólogos rápidamente tomamos ventaja de esta herramienta y se comenzó a analizar esto en todo tipo de especies de plantas, bacterias, hongos y animales en todo el mundo con electroforesis de proteínas. En México tomó un tiempo, más de 20 años, y hasta donde sabemos el primer estudio evolutivo publicado con isoenzimas hecho en el país fue el de Piñero y Eguiarte[8], estudiando poblaciones en un frijol que supuestamente es un híbrido entre el frijol común (*Phaseolus vulgaris*) y el ayocote (*Phaseolus coccineus*), usando ocho loci.

Pero estos análisis de electroforesis de proteínas tenían varias desventajas. Una es que las isoenzimas/ alozimas son proteínas en las que se pueden detectar algunos cambios en el ADN, que se reflejan como diferencias en la movilidad en los geles al cambiar los aminoácidos que forman a la proteína que codifica ese gen, pero no todos los cambios en el ADN se reflejan como diferencias en la movilidad de las bandas en los geles, ya que puede cambiar uno o varios aminoácidos sin que se modifique la movilidad de la proteína. Además, puede haber muchos cambios en el ADN que no necesariamente se reflejan como substituciones en los aminoácidos, dado que el código genético es degenerado, especialmente un cambio en la tercera posición de un codón en la mayoría de los casos no resulta en una substitución de aminoácido. Esta estimación sesgada de la variación genética preocupaba mucho a Richard Lewontin (ver por ejemplo su libro clásico del 1974[9]), y no descansó hasta que pudo analizarse la variación a nivel molecular, a nivel del ADN. Pero tal vez el problema más importante es que con el método de las isoenzimas/ alozimas sólo se podían analizar unas decenas de genes, rara vez más de 20 (Tabla I). ¿Qué tan representativos del total del genoma son estas decenas de genes? Los genomas de los procariontes tienen alrededor de cinco mil genes, mientras que los eucariontes tienen entre 20 y 40 mil genes diferentes, y además el resto del genoma tiene amplias secciones que pueden o no ser neutras (se entiende por una región neutra del genoma aquélla en la que las diferentes formas o alelos funcionan igual y no son afectadas por la selección natural), que pueden tener diferentes funciones reguladoras, ser genes egoístas, genes móviles (transposones), duplicaciones más o menos degeneradas de genes y de secciones completas del ADN, etc.

Esta pregunta que se hizo Richard Lewontin en 1974[9] nos ha torturado a los biólogos evolutivos por mucho tiempo, ¿qué tan variable es realmente el ADN en una población y en una especie?, considerando TODO el genoma, no sólo muestreando unos cuantos genes. Así, actualmente se busca analizar la variación genética a nivel del genoma; es decir, los biólogos evolutivos queremos hacer *genómica de poblaciones*.

La lucha por medir la variación genética

La historia de los estudios empíricos de genética de poblaciones nos muestra cómo ha sido una lucha constante la búsqueda de medir cada vez una mayor cantidad de genes. Podemos ilustrar este esfuerzo con los estudios hechos en nuestro Instituto. Después del análisis con los frijoles (*Phaseolus* spp.) que mencionamos arriba, otro trabajo que podemos mencionar es el de la tesis de doctorado del primer autor de este artículo[10], de 1990, donde analizamos 22 loci en total en la palma tropical

Marcador	Características	Ventajas	Desventajas
Variación fenotípica	Variantes observables y cuantificables	- Se pueden observar en las poblaciones - En muchas ocasiones son adaptativos - Comparable con fósiles	- No representan la variación genética en todo el genoma - Pueden reflejar plasticidad - Muchas veces son poligénicos
Aloenzimas	Variación en la migración de las enzimas durante la electroforesis, por cambios de aminoácidos	- Muchos loci son polimórficos - Codominantes (identifica homócigos y heterócigos)	- Poca variación - No detecta mutaciones silenciosas - Pocos marcadores a lo largo del genoma - Difíciles de montar
RFLPs	Variación en los fragmentos generados por enzimas de restricción, debido a cambios mutacionales en los sitios de reconocimiento		- Información limitada - Complicados de montar - Dominantes
RAPDs	Variación en la amplificación de fragmentos anónimos, que depende de la presencia diferencial de sitios de unión (mutación y cambios estructurales)	- Muchos sitios polimórficos a lo largo del genoma - Alta variación - Oligonucleótidos universales	- Dominantes - Cada sitio poco informativo - Poco reproducibles - Se desconocen los sitios amplificados - Muchos artefactos - Difíciles de leer
ISSRs	Similar a los RAPDs, pero los oligonucleótidos son secuencias más complejas y largas en tándem (microsatélites)	- Muchos sitios polimórficos - Alta variación - Más reproducibles que los RAPDs - Oligonucleótidos universales	- Dominantes - Cada sitio poco informativo - Se desconocen los sitios amplificados - Difíciles de montar - Difíciles de leer
AFLPs	Variación en la amplificación de fragmentos cortados con enzimas de restricción y ligados con adaptadores	- Genera muchísimos fragmentos - Alta variación - Muy reproducibles - Oligonucleótidos universales - Alta representación genómica	- Muy complicados de montar - Dominantes - Cada sitio poco informativo - Difíciles de leer
Microsatélites	Variación en el número de repeticiones de secuencias en tándem de ADN por inserción o pérdida de motivos	- Muy polimórficos - Codominantes - Se conoce la región en la que se encuentran - En muchos casos se conoce la tasa de mutación	- Difíciles de montar - Modelos de mutación complicados - Error en la estimación de tasas y modelos de mutación genera resultados erróneos - Alta homoplasia - Alelos nulos
SNPs	Mutaciones puntuales en secuencias de ADN Se pueden obtener por medio de RAD-tags o GBS (ver texto) o micro-arreglos	- Existen miles en el genoma - Permite entender la evolución a nivel genómico - Nuevas herramientas de secuenciación permiten la obtención de miles de SNPs - En sitios neutrales y codificantes	- Caros - Cada SNP es poco informativo por sí solo - Difícil de analizar

Tabla I. Descripción de los principales marcadores moleculares utilizados en genética de poblaciones.

Astrocaryum mexicanum, y nos concentramos en los cinco loci polimórficos con una lectura más sencilla. Luego logramos concluir estudios con más marcadores, por ejemplo, usando marcadores relacionados con el ADN. En el 2003, Navarro-Quezada *et al*.[11] analizamos 41 loci polimórficos de RAPDs (ver Tabla I) en el complejo de *Agave deserti*, encontrando claras señales de diferenciación genética entre poblaciones y altos niveles de variación dentro de las poblaciones. También se demostró que las tres especies del complejo están muy relacionadas, más de lo que se había supuesto dado la biogeografía del grupo, que se encuentra en el desierto Sonorense, incluyendo la península de Baja California. Hace pocos años, en un estudio detallado de la diferenciación genética dentro de dos poblaciones del maíz silvestre, el teosinte *Zea mays* ssp. *parviglumis*, logramos con técnicas moleculares más sofisticadas analizar 468 SNPs, *single nucleotide polymorphisms* (Tabla I), que indican cambios en una sola base del ADN[12]. Más recientemente, en un manuscrito aún no publicado de laboratorio del Dr. Daniel Piñero, María Artega *et al.* analizan en una colección de maíces de diferentes razas de México 47 mil marcadores SNPs, homogéneamente distribuidos a lo largo de los 10 cromosomas del maíz usando el chip MaizeSNP50 BeadChip de Illumina[13].

Genómica de poblaciones ¿necesaria, pura obsesión o una fantasía?

Actualmente en nuestro laboratorio estamos realizando análisis de genomas completos de organismos tomados de poblaciones naturales, tanto de plantas como de bacterias, pero ¿para qué nos sirve tener estos estudios? La idea básica es que Richard Lewontin tenía razón de preocuparse de sólo tener unos cuantos marcadores moleculares, ya desde los primeros análisis genómicos en humanos y en la bacteria *Escherichia coli*[6,14] revelaron que hay regiones en un genoma más o menos variables, debido al efecto diferencial de las fuerzas evolutivas mencionadas arriba. Las regiones promedio del genoma, con variación genética intermedia (*regiones "neutras"*), nos hablan de la historia evolutiva general de las poblaciones: como han sido sus tamaños de población promedio históricos, los *cuellos de botella* (momentos en los que se han reducido mucho los tamaños de las poblaciones), las expansiones poblacionales que han sufrido, etc. Las partes del genoma con menor variación genética sugieren eventos de *selección natural purificadora o direccional*: si hay genes que resultan desfavorables en ciertas condiciones, al ser eliminados por la *selección natural purificadora*, se reduce la variación genética; o al revés, si de repente surge un mutante ventajoso, la *selección natural direccional* hace que ese gen se vuelva el más abundante en la población, y también se reduce la variación genética. Pero puede haber regiones con mayor variación genética que otras partes del genoma. Esto podría deberse a puro azar, pero generalmente implica un tipo de selección natural interesante, llamada *selección balanceadora*, en donde los individuos heterócigos son los que funcionan mejor, viven más y/o dejan más hijos. En esos casos hay exceso de heterócigos en esos genes y en los genes que están cercanos a ellos en el cromosoma, selección que se refleja en altos niveles de variación genética.

Así, el genoma completo de los organismos es un importante y fascinante archivo de toda su historia evolutiva, que nos habla del complejo y dinámico balance en el tiempo entre la selección natural y las otras fuerzas evolutivas. El desarrollo técnico ha hecho posible en los últimos años la ampliación de las capacidades de secuenciación genética de manera vertiginosa. Este desarrollo ha sido en parte el resultado de la promesa de que contar con más y más genomas completos permitiría la respuesta a toda clase de preguntas, no sólo evolutivas, sino médicas, agronómicas, veterinarias, etc., y que según nuestro amigo Lewontin[15] se le ha tratado de hacer creer al público (cuando menos en Estados Unidos y Europa) que es una especie de "santo grial", que una vez encontrado resolverá todos los problemas de salud, alimentación y biológicos. Como el propio Lewontin señalaba hace casi 13 años, buena parte de las respuestas que podamos elaborar a partir de conocer las secuencias de genomas completos dependerán más de la comprensión de los niveles de variación entre poblaciones e individuos a diferentes escalas, de su relación con el medio ambiente y de la manera en que esas diferencias genéticas se expresan o no en la fisiología, forma, conducta y ecología a lo largo de la historia de vida de los organismos.

El desarrollo de esta capacidad molecular para producir datos genómicos, información que consiste en millones de pares de bases de ADN de decenas, cientos y aún miles de individuos, rápidamente abruma nuestros sentidos y capacidad de análisis. Metafóricamente hablando, las secuencias genéticas son interesantes por las historias que nos cuentan, pero no como un conjunto inconexo de anécdotas que nos cuenta cada uno de los genes de forma individual, sino por los patrones generales que podemos extraer de ellas, por las comparaciones que podemos hacer desde una perspectiva histórica y de la relación trenzada genes-organismo-ambiente. Es allí donde debemos ser capaces de plantear las preguntas adecuadas y contar con el marco teórico y técnico de referencia que nos da la teoría de la genética de poblaciones y la biología evolutiva de Fisher, Wright, Haldane, Dobzhansky y Lewontin, entre otros (ver por ejemplo Hedrick[6], Eguiarte *et al*.[16] y Futuyma[17] para resúmenes accesibles y actualizados), nos permitirá aprovechar de mejor manera la enorme cantidad de datos que la secuenciación de siguiente generación nos lanza encima. De lo contrario corremos el riesgo de, como el personaje Funes el Memorioso de Borges de nuestro epígrafe, nombrar cada detalle, cada recuerdo, cada par de bases de un mar venidero de genomas, sin ser capaces de articular ideas o consideraciones y dar respuestas a problemas generales y útiles.

Pero no hay que angustiarnos antes de tiempo. La posibilidad de comparar genomas o fragmentos muy grandes de éstos nos abre la posibilidad de analizar de manera extensiva el efecto relativo

de diferentes procesos evolutivos sobre diferentes partes del genoma de una misma especie, identificar regiones asociadas a patrones de cambio fenotípico y de esta manera extraer la información realmente útil para aplicaciones médicas y de mejoramiento de plantas, bacterias y animales[18-20]. Estas nuevas perspectivas han sido comparadas con la revolución que, como mencionamos arriba, ocurrió en 1966 cuando Lewontin y Hubby[7], introdujeron la técnica de la electroforesis al estudio de la variación genética de las poblaciones naturales[21], aunque nosotros creemos que las repercusiones de esta nueva revolución van a ser más importantes, saliendo del ámbito de la biología evolutiva. Pero hasta donde este proceso podrá tener una relevancia similar o mayor al de la electroforesis estará determinado entre otras cosas por la articulación de un programa de investigación coherente que pueda realizar una síntesis adecuada entre las nuevas tecnologías moleculares, con las herramientas bioinformáticas capaces de manejar estos mares de información, junto con la teoría evolutiva, ecológica y de genética de poblaciones ya disponible. Esto con el fin de generar poder heurístico (en otras palabras, procedimientos prácticos para resolver problemas de forma más rápida que con los métodos tradicionales) y así resolver de manera adecuada las preguntas y problemas relevantes.

Los nuevos métodos moleculares

Recientemente se han desarrollado diferentes herramientas moleculares que permiten la secuenciación masiva a un precio accesible, técnicas llamadas en inglés secuenciación *next generation*, o que llamaremos *next -gen* en el resto del artículo[22-24]. Estos nuevos métodos se resumen en la Tabla II, y nos permiten secuenciar rápidamente y a precios razonables genomas enteros o gran cantidad de loci y así realizar análisis muy detallados de genética de poblaciones utilizando miles de marcadores distribuidos a lo largo de todo el genoma[24-25]. Los detalles de cómo funcionan las distintas plataformas de secuenciación son técnicos y haría esta revisión muy larga, por lo que sugerimos a los interesados revisarlos en artículos citados arriba. Aquí nos enfocaremos principalmente en las consideraciones que se tienen que tomar en cuenta cuando se quieran utilizar.

Lo que diferencia los métodos de próxima generación respecto a los métodos tradicionales es que permiten secuenciar múltiples partes del genoma sin que sea un requisito conocer marcadores previos. Dado que no se necesitan marcadores particulares, es posible estudiar genomas de organismos no-modelo, esto es, organismos para los cuales no se han hecho previamente estudios detallados de su genética, y así nos permite comenzar estudios sofisticados con organismos silvestres o pobremente estudiados[24], como la inmensa mayoría de nuestra flora y fauna.

Brevemente, los pasos a seguir en estos métodos *next-gen* son:

1) La ruptura del genoma total en fragmentos pequeños de ADN de entre unos 200 a 1000 pares de bases (pb de aquí en adelante),
2) El montaje de las librerías con sitios de unión de oligos (primers) que permitan iniciar la replicación del ADN.
3) La amplificación (obtener muchas copias) de los fragmentos anteriores y, finalmente,
4) La secuenciación mediante distintos métodos de los fragmentos amplificados (Tabla II).

La obtención de fragmentos de ADN se realiza con diferentes técnicas clásicas de la biología molecular, como son el corte con enzimas de restricción (que cortan el ADN en regiones específicas) o métodos físicos (sonificación, nebulización) o la retrotranscripción de mARN (en vez de aislar el ADN, se aísla el ARN mensajero, es decir, se analizan los genes que se están expresando, en vez del genoma completo).

Una gran ventaja de los métodos *next-gen*, es que durante la preparación de las librerías es posible añadir códigos genéticos individuales ("tags") a los fragmentos de ADN o ARN, lo que permite secuenciar múltiples individuos en una sola corrida (*multiplex*). Para ecólogos moleculares y genetistas de poblaciones esto es muy relevante, ya que estos métodos van a permitir hacer estudios poblacionales a precios por individuo relativamente bajos, aunque se reduce la cobertura (número de veces que se secuencia el genoma en promedio, ver más abajo) por individuo[24,26].

Las plataformas *next-gen* que existen en la actualidad (Tabla II, pero esto cambia rápidamente) se pueden dividir en dos grandes grupos. Por un lado están aquéllas que generan lecturas largas pero en menor cantidad (por ejemplo, 454 genera fragmentos de 400 a 700 pb) y por el otro las plataformas que generan lecturas cortas en mayor cantidad (por ejemplo, SOLiD genera fragmentos cercanos a 50 pb)[26]. Así, la selección de plataformas depende de la pregunta u objetivo que se quiere responder, de los costos y de la disponibilidad del equipo. Por un lado, las lecturas largas son óptimas para cualquier proyecto que involucre caracterizar un genoma que no se conoce, ya que se ensamblarán de manera más eficiente[26]. Por su lado, las lecturas cortas y abundantes suelen ser más económicas y pueden servir muy bien para trabajar genomas conocidos y generar múltiples marcadores[26,27].

Un aspecto importante que hay que considerar es que las distintas plataformas presentan distintas tasas de errores de secuenciación. Si la plataforma genera muchos errores, pero se quiere ensamblar y se tiene un genoma de referencia, no es tan importante, ya que no se están buscando polimorfismos. Por otro lado, si son de interés los polimorfismos particulares (i.e., SNPs) es mejor utilizar plataformas que generen pocos errores de secuenciación o que generen mucha cobertura. La cobertura se refiere al número de veces que se secuencia un mismo fragmento, por lo que los polimorfismos reales se detectarán más de una vez.

Plataforma	Tiempo corrida	Millones de lecturas por corrida	Bases por lectura	Ventajas	Desventajas	Aplicaciones
454	~ 10hrs	0.10 - 1	400 - 700	Lecturas largas	Alto costo por Mb Errores en homopolímeros	Conveniente para secuenciar y resecuenciar genomas Caracterización de transcriptomas Metagenómica
Illumina	~ 8 días	3.4-329	~100-150	Muy barato Alta cobertura	Lecturas cortas	Resecuenciación de genomas Caracterización y conteo de transcriptomas Obtención de SNPs: Radtags, GBS Metagenómica
SOLiD	~ 8 días	700 - 1400	~50	Pocos errores de secuenciación Barato Mucha información	Lecturas cortas Tardado Complicado analizar las secuencias (código)	Resecuenciación de genomas Caracterización y conteo de transcriptomas Obtención de SNPs: Radtags, GBS
Ion Torrent	2 hrs	1-8	~100 - 400	Muy barato Mejoramiento continuo Fragmentos grandes	Preparación compleja Caro	Secuenciación y resecuenciación de genomas Obtención de SNPs Caracterización de transcriptomas
PacBio	2 hrs	0.01	860-1100	Secuenciación en tiempo real (una molécula) Fragmentos largos	Altos errores de secuenciación Pocas lecturas Alto costo por Mb	Secuenciación y resecuenciación de genomas Metagenómica

Tabla II. Algunos de los principales métodos de secuenciación recientemente desarrollados, conocidos como métodos de *next-generation*, o *next-gen*. Para más datos y las diferentes variantes, consultar las revisiones citadas en el texto, en particular Glenn[26].

También es importante considerar si existen recursos genéticos previos, por ejemplo, genomas ya secuenciados y anotados para los organismos que queremos estudiar o para especies cercanas, es decir saber si se estudia a un organismo modelo, o hay alguna especie cercana que sea modelo, ya que eso facilita mucho el ensamble y anotación de los fragmentos. En caso de no tener esas referencias, utilizar dos tipos de plataformas simultáneamente puede facilitar mucho el trabajo de ensamble de genomas. Al utilizar fragmentos grandes se puede generar un genoma de referencia *de novo* y con los fragmentos pequeños se puede aumentar la cobertura de las secuencias[26].

Un punto crítico es que los métodos *next-gen* permiten la obtención de cantidades antes insospechadas de datos, lo que hace que los análisis informáticos sean muy complicados, superando los niveles que usualmente usamos los biólogos, agrónomos o médicos. Se necesitan computadoras potentes, contar con colaboradores y estrategias bioinformáticas eficientes que puedan analizar todos estos datos de manera adecuada, con una perspectiva evolutiva.

La secuenciación de *next-gen* se puede utilizar para obtener análisis detallados de la estructura genética de las poblaciones, o sea ver cómo cambian las frecuencias de los alelos en el espacio (ver por ejemplo van Heerwaarden[12]), analizar la historia demográfica de las poblaciones (si las poblaciones han crecido o decrecido) y proponer genes candidatos para adaptaciones a condiciones ambientales específicas[27], reduciendo a su vez los falsos positivos, como explicamos más abajo. Muchos estudios de ecología molecular buscan analizar centenas de loci particulares en cientos de individuos. Utilizando los métodos multiplex (donde se estudia a varios individuos en un solo análisis, usando los *tags* mencionados arriba) y amplificando específicamente esos marcadores con PCR, es posible secuenciar una gran cantidad de fragmentos en pocas corridas, lo que reduce mucho el costo[26].

Como mencionamos antes, la cobertura genética (que tantas veces en promedio se secuencia un genoma) es importante, ya que permite distinguir entre errores de secuenciación y los polimorfismos verdaderos en la secuencia del ADN, aunque estén en muy bajas frecuencias[28]. El aumento en la cobertura se logra, o secuenciando más veces los genomas, que es muy caro, o mediante la "reducción" del genoma a analizar con enzimas de restricción y preparación de las librerías, método utilizado por ejemplo en las estrategias llamadas *radtags*[29] (Tabla I) o *genotyping by sequencing*[28] (Tabla I). Así, no se secuencia todo el genoma, sino sólo el comienzo o ciertas partes de algunas secuencias, de preferencia de los genes que se expresan, y de los que se expresan más intensamente. Estos métodos tienen la ventaja adicional de que con ellos se obtienen menos tipos de secuencias, facilitando el análisis informático.

Ideas básicas del análisis estadístico de la genómica de las poblaciones

Con los enfoques *next-gen* podemos no sólo conocer con precisión la historia evolutiva de la especies (con los genes neutros), sino que en principio podemos detectar los genes relacionados con las adaptaciones o con las enfermedades, en caso de la medicina que revisamos más adelante. En muchos casos se podrán detectar estos genes adaptados a condiciones locales debido a que van a presentar frecuencias alélicas contrastantes entre poblaciones. La idea es sencilla, y proviene, otras vez, de nuestro amigo Richard Lewontin y un colaborador, J. Krakauer[30]: los loci que tengan alelos que estén bajo selección diferencial en las distintas poblaciones deben tener niveles de diferenciación genética más elevada que los genes neutrales, o sea que los alelos se deben de encontrar en proporciones diferentes entre poblaciones sujetas a distintas presiones selectivas. Por el contrario, genes que estén bajo selección balanceadora, deberían de tener niveles de diferenciación más bajos que los neutrales.

Estas diferencias en las frecuencias alélicas se pueden analizar usando el estadístico F_{ST} de Wright[31]. La F_{ST} es una medida de la diferenciación genética entre poblaciones, va de 0 si tienen exactamente los mismos alelos en las mismas frecuencias, hasta 1 si no comparten ningún alelo (ver ejemplos de su uso en este tipo de análisis en Namroud *et al.*[32] y Eckert *et al.*[33]). La estimación de la F_{ST} es más fácil de visualizar siguiendo la definición de Nei[34]: $F_{st} = (H_t - H_s) / H_t$, donde H_t es el promedio de la heterocigosis esperada en la población total, para todos los loci y H_s es el promedio de la heterocigosis esperada dentro de subpoblaciones para todos los loci. F_{ST} mide la reducción en la heterocigosis debida a la diferenciación genética entre poblaciones. También se puede definir la F_{ST} en términos de las varianzas en las frecuencias alélicas entre las poblaciones, lo cual puede ser más intuitivo: F_{ST} = Varianza $(p)/ (p(1-p))$, entre mayor haya sido la deriva génica, mayor será la varianza de (las diferencias entre) las frecuencias alélicas entre poblaciones.

Aunque estos métodos son conceptualmente muy atractivos, son susceptibles de generar falsos positivos, debido a que otras fuerzas evolutivas además de la selección pueden también cambiar las frecuencias alélicas. Por otro lado, problemas de muestreo y otros artefactos estadísticos pueden sugerir asociaciones falsas entre los alelos y la adaptación, como veremos más adelante[33,35-37]. Dada la variación genética, la capacidad para detectar genes bajo selección dependerá de la intensidad e historia de la selección, de las tasas de mutación y recombinación y de la historia demográfica de la población. Sin embargo, se ha encontrado con simulaciones de computadora, que el principal problema relacionado con la obtención de falsos positivos es estimar incorrectamente la historia demográfica de las poblaciones (i.e., si las poblaciones han crecido, decrecido o se han mantenido estables en el pasado). Por lo tanto, para realizar este tipo de métodos, es necesario analizar muchos sitios para poder hacer las comparaciones y ajustar el modelo demográfico que ha sufrido la población en muestras grandes[35]. En cualquier estudio de selección es fundamental conocer bien la historia demográfica de las poblaciones, definir los modelos demográficos que se ajusten a los datos observados, y hacer las pruebas de selección asumiendo esos modelos. Con el desarrollo de los métodos estadísticos e informáticos, ha sido posible desarrollar métodos coalescentes más robustos que pueden incluir muchos factores demográficos como recombinación, crecimiento poblacional e introgresión (que entren genes de otra población o especie).

Para comenzar, el ejemplo de la medicina

Para la medicina, estas ideas sobre los niveles de variación a lo largo del genoma son efectivamente muy importantes y atractivas ya que, en teoría, de sus análisis se pueden conocer las bases genéticas de susceptibilidades a enfermedades e infecciones o resistencia a éstas. Estos estudios en humanos se llamaron inicialmente GWAS - *genome wide association studies* - y en estos trabajos usualmente se analizan en poblaciones sanas y enfermas muchos marcadores genéticos y se intenta descubrir los marcadores diferentes que se encuentran en los individuos enfermos, en comparación con los individuos sanos. Estos estudios, aunque atractivos en principio, requieren de muestras grandes y buenos controles estadísticos y metodológicos. Es común que se encuentren muchos falsos positivos, o sea marcadores genéticos que parecen asociados a una enfermedad, pero lo son de forma espuria, falsa, debido a que si se tienen muchos marcadores, algunos parecen estar asociados simplemente por un error estadístico, por azar. También es posible que las bases genéticas para una enfermedad sean diferentes entre distintas poblaciones: que la enfermedad parezca ser igual o parecida, pero en realidad sea generada por diferentes procesos moleculares. Otra alternativa es que se encuentre una asociación estadística entre la enfermedad y algunos marcadores genéticos, pero que estos marcadores no son los causantes de la enfermedad, sino que solamente están ligados (cerca) en el cromosoma, y que el gen verdadero que causa la enfermedad se

encuentre a muchos miles (o millones) de pares de bases del gen sugerido por el marcador genético. Por otro lado, es muy común que las enfermedades tengan bases genéticas complejas, y que en buena parte estén determinadas por el ambiente, y que además estén determinadas por no uno, sino muchos genes, que pueden ser decenas o cientos de ellos, todos con pequeños efectos, por lo que es muy difícil detectarlos con estos métodos.

Las ciencias forestales

Frente al cambio climático global, se espera que los árboles tengan problemas graves para adaptarse comparados con otras plantas, debido a sus largos ciclos de vida, o comparados con animales, dado que éstos pueden moverse. Por este motivo, se están usando métodos de genómica de poblaciones para poder identificar alelos que permitan a las especies sobrevivir en diferentes escenarios de cambio global, no sólo al aumento de temperatura, ya que se esperan fuertes cambios en términos de humedad y sequías, y en los extremos en temperatura, que pueden llegar a ser muy bajos en algunas condiciones. Afortunadamente, los estudios previos con otros marcadores genéticos indican que la mayoría de los árboles, tanto angiospermas como coníferas, tienen niveles muy altos de variación genética dentro de las poblaciones. En México se están iniciando en nuestro Instituto estudios en diferentes especies de oyamel (*Abies*) y pinos, junto con detallados análisis de sus distribuciones actuales, y de las distribuciones potenciales en diferentes escenarios de cambio global, y tratando de describir la distribución actual de diferentes alelos en el genoma. Igual que se ha intentado en los estudios GWAS en salud humana, se busca identificar con estos métodos alelos en las poblaciones que viven, por ejemplo, en los lugares más calientes y secos, que puedan servir para hacer plantaciones de árboles que resistan las futuras condiciones ambientales extremas, y así mantener el potencial forestal de nuestro país.

Mejoramiento de especies de interés agronómico y veterinario

Obviamente, una posible aplicación muy importante de los métodos genómicos es para el mejoramiento de especies. La idea es que, más que usar la ingeniería genética para introducir en los genomas genes provenientes de otras especies -como usualmente se hace-, se trata de una estrategia en la que se exploran poblaciones silvestres y criollas de la especie de interés. Estas poblaciones tienen amplios reservorios de variación genética y adaptaciones a diferentes condiciones climáticas, de tipo de suelo, de resistencia a herbívoros y patógenos, y de variación en características importantes para la agronomía, como pueden ser diferentes tipos y colores de frutos, forma de crecimiento de la planta, diferentes fenologías y estrategias de desarrollo, etc., que pueden ser usados exitosamente para mejorar las variedades cultivadas modernas. Esta idea es atractiva, ya que implica que se usen genes ya probados y adaptados en la especie que se quiere mejorar. Por esta razón se están explorando diferentes plantas de interés comercial en México, junto con sus parientes silvestres, especialmente el maíz, pero se está avanzando en proyectos genómicos con las calabazas (género *Cucurbita*) en nuestro laboratorio, y en el LANGEBIO del CINVESTAV, Irapuato, se están explorando, entre otras especies, los genomas del frijol, del chile y del aguacate, con diferentes colaboraciones.

Biología de la conservación

Una de las aplicaciones más interesantes de la genómica de poblaciones es en relación a la biología de la conservación. Actualmente, por las actividades humanas, gran cantidad de especies se encuentran en fuerte peligro de extinción por diversas razones. Tal vez la más importante es la destrucción del hábitat, que a su vez reduce el tamaño de las poblaciones y su conectividad, reduciendo el flujo génico, el tamaño efectivo y la variación genética; si a estos problemas sumamos el cambio climático global, como comentamos arriba para los árboles, es claro que las especies necesitan de su variación genética para adaptarse. Así, usando enfoques genómicos de poblaciones, pueden evaluarse los niveles de variación de las poblaciones, se pueden describir las diferencias en composición genética entre las mismas y tratar de proponer métodos para generar patrones de flujo génico que mantengan la variación en estas poblaciones y que minimicen la deriva génica y los problemas de sobrevivencia de éstas. Con esto en mente, es posible desarrollar programas de manejo que minimicen la pérdida de variación genética y encontrar y mantener a los genes que permitan a las especies adaptarse a las nuevas condiciones.

Uno de los principales efectos de la reducción de las poblaciones y de su manejo, especialmente en zoológicos y jardines botánicos, es el aumento de los apareamientos entre parientes, cruzamientos llamados endogámicos o consanguíneos. La endogamia, además de conducir a la pérdida de la variación genética en cada uno de los grupos familiares endogámicos, tiene efectos dañinos en el funcionamiento (adecuación) de la progenie, efecto que se denomina actualmente como "depresión por endogamia", y que se conoce desde hace mucho tiempo. Por ejemplo, Charles Darwin dedicó uno de sus libros completo[38] al estudio del efecto de la endogamia en plantas. Las bases genéticas de la depresión por endogamia han sido discutidas desde hace mucho tiempo[10], y algunos investigadores, entre los cuales se encontraba Theodosius Dobzhansky sospechaban que se debía a la heterosis, el fenómeno por el cual los organismos heterócigos funcionan mejor, que describimos arriba cuando hablamos de la selección balanceadora. Otros investigadores creían que se debía simplemente a la expresión de genes recesivos deletéreos: por mutación, como explicamos arriba, surgen nuevas variantes, pero la mayor parte de las variantes van a ser dañinas (¡siempre es más fácil descomponer algo que arreglarlo!). La mayor parte de estas mutaciones no se expresan, son recesivas, pero si se cruzan dos individuos heterócigos para este gen en particular, se producen individuos homócigos para estos genes deletéreos, se expresa el gen y la progenie tiene

una baja adecuación, funciona muy mal. Un buen ejemplo es la hemofilia, la enfermedad de las casas reales europeas. La reina Victoria de Inglaterra era heteróciga para ese gen, y se lo heredó a sus descendientes, que cuando se casaban endogámicamente, producían hijos hemofílicos, que se morían desangrados muy jóvenes. Así, varios de sus descendientes, famosamente los herederos al trono de Rusia tuvieron estos problemas genéticos, que en parte desencadenaron la revolución Rusa.

Recientemente, con métodos genómicos, se ha encontrado en especies como la papa (*Solanum tuberosum*) y en el maíz (*Zea mays*[39]), que buena parte de la depresión por endogamia se debe a que es común que los genomas pierdan secciones completas, con muchos genes e información. Un heterocigoto funciona bien, ya que en alguna de las dos copias de los genomas que tiene conserva la variación genética que necesita, pero si se obtienen individuos homócigos, no tiene esos genes, y se mueren en diferentes etapas del desarrollo, o su funcionamiento es menos eficiente, ya que cuando hay cualquier estrés no tienen la información genética para enfrentarlo. Así, con métodos genómicos, se pueden ver cuáles son los individuos que no tienen esas deleciones en sus genomas y usarlos como el pie de cría para comenzar linajes que, aunque no tengan mucha variación genética, no tengan esos problemas.

Desafortunadamente, aunque los precios para estos estudios están y siguen bajando drásticamente, involucra mucho trabajo y tiempo para el muestreo de los organismos, así como en los análisis moleculares implicados en su estudio y el análisis informático (que va a ser el principal cuello de botella en estos estudios), por lo que sólo se pueden analizar algunas especies emblemáticas, interesantes ya sea por su abundancia, por ser especies clave, especies carismáticas, por su importancia simbólica, o por ser parientes de especies económicamente relevantes, y así funcionan, como mencionamos arriba, como reservorios de variación genética útil para continuar con el mejoramiento de las especies económicamente de interés.

También con métodos genómicos se podrá decidir cuáles son las especies o poblaciones que son genéticamente más diferentes, más interesantes por diferentes razones ecológicas, evolutivas o de usos, o más variables genéticamente para que sean el foco de los esfuerzos de conservación, como enunciamos en nuestro trabajo para el *Agave victoriae-reginae* y para varias especies de pinos[40,41].

¿Y la adaptación? El caso del maíz y sus parientes silvestres, los teosintes

Obviamente, los métodos *next-gen* nos sirven para entender finamente las bases de la selección natural y la adaptación, que era lo que buscaban inicialmente Fisher, Haldane, Wright y Dobzhansky desde sus estudios en los años 30 del siglo pasado. Veamos algunos resultados relevantes obtenidos en diferentes organismos:

Una serie de ejemplos del análisis de genética de poblaciones y recientemente de genómica de poblaciones usando numerosos marcadores lo ofrece el estudio del maíz y su ancestro silvestre, el teosinte y consideramos que es interesante, porque redondea lo que hemos comentado sobre la "lucha por medir la variación genética" e ilustra cómo ha mejorado nuestro entendimiento del comportamiento evolutivo de los genomas. El género *Zea*, además del maíz cultivado, incluye otros taxa divididos en dos secciones[42]: *Luxuriantes* y *Zea*. La sección *Luxuriantes* incluye a *Z. luxurians*, una especie diploide anual que crece en Guatemala y Nicaragua (aunque más recientemente las poblaciones en Nicaragua fueron reclasificadas como *Z. nicaraguensis*), a *Z. diploperennis*, una especie diploide perenne que se encuentra principalmente en el estado de Jalisco, y a *Z. perennis*, que es una especie tetraploide perenne, también de Jalisco. En la sección *Zea* todos los taxa son diploides, e incluye al maíz cultivado, *Z. mays* ssp. *mays* y a tres taxa de teosinte, *Z. mays* ssp. *mexicana* del centro de México, *Z. mays* ssp. *parviglumis* de las tierras bajas del centro-oeste de México y a *Z. mays* ssp. *huehuetenangensis* del oeste de Guatemala. A partir de datos moleculares, se ha calculado que la divergencia entre *Z. luxurians* y *Z.m. parviglumis* sucedió hace unos 140,000 años, mientras que la separación entre *Z. m. mexicana* y *Z. m. parviglumis* es más reciente, de unos 60,000 años[43], y que la domesticación del maíz a partir de *Z.m. parviglumis* ocurrió hace unos 11 mil años[44]. En *Z. mays mexicana* y *Z. mays parviglumis*, el análisis de secuencias de genes nucleares de copia única muestran altos niveles de variación genética y expansiones demográficas recientes[43].

Veamos cómo ha avanzado el número de marcadores usados en estas especies: Buckler *et al*.[43] re-analizaron 21 loci isoenzimáticos de estudios previos, en 9 a 50 plantas por población de un total de 61 poblaciones. *Z.m. parviglumis* resulta basal aunque parafilético, mientras que *Z.m. mexicana* es monofilético y derivado de *Z.m. parviglumis*. *Z. m. huhuetenangensis* es el taxa basal, hermano de los anteriores. Buckler *et al*.[45] concluyen que un modelo filogeográfico no-adaptativo de dispersión se ajusta a los datos, aunque hay aislamiento por distancia y cierto efecto de la altitud sobre el nivel del mar. Para la filogeografía del cloroplasto, encontraron 5 haplotipos, 4 en parviglumis, 3 en mexicana y 2 de éstos compartidos por ambos linajes.

Fukunaga *et al*.[46], analizaron 93 loci de microsatélites nucleares de 237 plantas, de un total de 172 accesiones (sitios de colecta). Además de las tres subspecies silvestres de *Z. mays* (ssp. *mexicana*, *parviglumis* y *huhuetenangesis*), incluyeron varias accesiones de *Z. diploperennis*, *Z. luxurians* y *Z. perennis*. La heterocigosis esperada (la medida estándar de variación genética) fue un poco más alta en *Z.m. parviglumis* que en *Z. m. mexicana*. Sus análisis filogenéticos indican que *mexicana* + *parviglumis* son efectivamente un grupo monofilético. Curiosamente, ni en éste ni en el estudio anterior se incluyeron muestras de maíz cultivado (*Z. m. mays*).

Moeller *et al*.[47] obtuvieron la secuencia de ADN de cinco genes nucleares y dos regiones del cloroplasto en 84 plantas de siete poblaciones de *Z.m. parviglumis*, incluyendo cuatro poblaciones de Jalisco y tres del "Balsas" (de Guerrero y del Estado de México). Usando un análogo de la F_{ST} (AMOVA), no detectaron diferenciación genética entre las dos regiones geográficas (Jalisco *vs.* Balsas), pero sí entre las poblaciones de cada región. Las poblaciones del Balsas en general tienen mayor variación que las de Jalisco, tal vez como resultado de hibridación. La mayor parte de las poblaciones sólo tienen un haplotipo de cloroplasto (el mismo en cada región geográfica), lo que sugiere cierta estructura genética, pero otros datos indican algo de flujo génico por semilla entre las zonas (ya que el cloroplasto se hereda exclusivamente por vía materna). El estudio muestra procesos filogeográficos complejos entre y dentro de cada área en el teosinte, pero el número de genes, individuos y poblaciones utilizados para el análisis es bastante bajo.

Ross-Ibarra, *et al*.[43] obtuvieron la secuencias de ADN de 26 loci nucleares para 13 individuos por taxa, incluyendo el maíz cultivado, *Z.m. parviglumis*, *Z.m. mexicana* y *Z. luxurians*. Encontraron que la variación genética a nivel secuencia de ADN fue más alta en *parviglumis*, luego en *mexicana* y en *luxurians* y el maíz cultivado es el que tiene menor variación a nivel nucleotídico. También detectaron claras señales de expansión demográfica en *parviglumis* y un poco más débiles en *mexicana*. F_{ST} pareadas entre taxa indican que *mexicana* se parece un poco más al maíz cultivado que *parviglumis*, y diferentes análisis confirman que la separación entre *mexicana* y *parviglumis* fue hace unos 60 mil años y entre *luxurians* y los diferentes taxa de *Zea mays* fue de hace unos 140 mil años. Gore *et al.*[48] secuenciaron el 20% del genoma del maíz cultivado que es de baja copia en 27 líneas endógamas fundadoras del MAM (*maize nested association mapping*), que representan diversas accesiones del maíz. Obtuvieron 93 millones de pares de bases de copia baja que se encontraron en más de 13 líneas del estudio. De todas las regiones que secuenciaron, algunas no se alinearon al genoma de referencia, o sea, ¡no se encuentran en el genoma de referencia! Estas regiones podrían corresponder a regiones únicas de esas líneas y que podrían estar asociadas a adaptaciones únicas de estos organismos a los ambientes en los que crecen. El estudio muestra el poder de los métodos de nueva generación para encontrar nuevos genes relacionados con las adaptaciones, sobre todo en organismos con genomas que evolucionan rápidamente, como es el caso del maíz. Encontraron 148 regiones de menor diversidad que el gen *tb1*, supuesto gen principal de la domesticación de los maíces, que indica importante selección en la domesticación en otros genes y regiones. Como comentamos arriba, estos métodos de secuenciación pueden utilizarse para identificar a otros genes asociados con la domesticación de diferentes plantas. Finalmente, encontraron que el 43% de los sitios presentaron diferenciación genética elevada entre todas las accesiones ($P<0.05$) y 183 regiones presentaron una F_{ST} muy elevada y altamente significativa. Estos autores concluyeron que los genes con alta diferenciación genética podrían ser genes involucrados en la adaptación a ambientes templados y tropicales. El conjunto de los datos obtenidos muestran que los métodos de secuenciación de próxima generación han sido muy importantes en entender la domesticación de los maíces, y posiblemente las bases genéticas de las adaptaciones en maíz.

Van Heerwaarden *et al*.[12], como ya comentamos arriba, analizó la estructura genética fina en dos poblaciones de *Z.m. parviglumis* en la cuenca del Balsas. Utilizaron 468 SNPs en 389 individuos de un sitio y en 575 de otro sitio, separados por 6.3 km. La diversidad genética con estos marcadores en ambos sitios fue la misma, y la diferenciación medida como F_{ST} fue baja. Sin embargo, el gran número de marcadores permitió detectar estructura genética (diferenciación) significativa dentro de cada sitio, a pesar de que las plantas estaban a menos de 350 metros de distancia. Esta diferenciación tal vez esté correlacionada con la heterogeneidad en las condiciones ambientales de cada sitio.

En un segundo estudio, van Heerwaarden *et al*.[44] trabajaron con 1,127 accesiones de razas nativas de maíz, más 100 de *Z.m. parviglumis* y 96 de *Z.m. mexicana*. En total analizaron 964 SNPs provenientes de 547 genes, incluyendo los del estudio anterior y concluyen que el maíz cultivado desciende del teosinte *Z.m parviglumis*, y que la similitud que tiene *Z.m. mexicana* con el maíz de las tierras altas de México se debe a flujo génico entre ambos taxa. Sin embargo, debido a que los muestreos son de accesiones (donde usualmente se analiza una sola semilla o muy pocas para cada población) y no poblacionales, las inferencias evolutivas finas que se pueden hacer son limitadas.

La conclusión de los estudios anteriores es que el maíz cultivado (*Z. m. mays*) fue domesticado a partir del teosinte de las tierras bajas (*Z.m. parviglumis*) primero por un proceso muy intenso de selección artificial, seguido de un proceso de mejoramiento continuo a lo largo de mucho tiempo, para responder a las diferentes condiciones ambientales de donde se cultiva ahora el maíz y los diferentes usos que se le da. Con el objetivo de entender como ha operado la selección en estas dos etapas, Hufford *et al*.[20] secuenciaron y compararon con más de 21 millones de SNPs en total, los genomas de 35 líneas mejoradas de maíz, 23 razas de maíz tradicionales y 17 de teosinte *Z. m. parviglumis*. Encontraron evidencias genómicas asociadas a la domesticación y al mejoramiento que incluyen reducciones de diversidad y aumento en bloques de ligamiento (regiones del cromosoma con baja o nula recombinación, que se heredan entonces como un paquete), resultado de cuellos de botella (reducciones drásticas en el tamaño de la población, tal vez asociadas a la domesticación y/o selección artificial y natural intensa, o sólo fuertes reducciones en el número de individuos que se usó en la domesticación). Por otro lado, encontraron evidencias de introgresión (flujo génico proveniente de otra variedad o especie) del maíz con la subespecie *Z.m. mexicana*, el teosinte de tierras altas, que respaldan el papel importante que

ha tenido esta planta durante la domesticación. Detectaron que la domesticación se ha caracterizado por un proceso selectivo más intenso que en el posterior mejoramiento. Así definieron los "rasgos" (bloques de ligamiento que incluyen varios genes) que presentaron mayor evidencia de selección, con lo que definieron 484 rasgos de domesticación y 695 rasgos de mejoramiento. Los rasgos de domesticación tuvieron en promedio 3.4 genes y cubrieron en promedio 322 miles de pares de bases y 7.6% del genoma de maíz. En promedio encontraron que estos rasgos presentaron una intensidad de la selección elevada (s=0.015), un orden de magnitud mayor que el resto de los genes. Por su lado, los rasgos de mejoramiento fueron menores en promedio, involucraron menos genes y presentaron una intensidad de selección menor que para los rasgos de la domesticación (s=0.003). Finalmente, encontraron que el 23% de los rasgos de domesticación han sufrido selección adicional posterior durante el mejoramiento, mostrando que muchos rasgos contribuyen a características fenotípicas de importancia agronómica.

Pyhajarvi *et al.*[37] analizaron cerca de 36,000 SNP en 250 individuos provenientes de 21 poblaciones de teosintes, tanto de las tierras bajas (*parviglumis*) como del altiplano (*mexicana*). Encontraron distintas bases genéticas para las adaptaciones. Al estudiar los patrones de desequilibrio de ligamiento (baja recombinación), identificaron 4 regiones de alto desequilibrio de ligamiento (de más de 10 millones de pares de bases), que identificaron como inversiones. Aunque no se ha determinado la función de estas inversiones, los autores encontraron que son ricas en SNPs asociados con características ambientales de temperatura y altitud, lo que indica que son candidatos a estar bajo selección. Encontraron que la inversión *Inv1n* presenta una clina altitudinal, con frecuencias medias en altitudes bajas (*parviglumis*) y frecuencias bajas en altitudes altas (*mexicana*). Por su parte, otra inversión (*Inv4n*) se encontró en altas frecuencias en *parviglumis* y estuvo restringida a las poblaciones distribuidas a mayores altitudes de *parviglumis*, indicando que su presencia es importante a elevadas altitudes. Finalmente, otra inversión (*Inv9d*) sólo se encontró en las poblaciones de mayor altitud de *mexicana*. Por otro lado, encontraron más de mil SNPs asociados con condiciones ambientales y, particularmente, con altitud y temperatura, y algunos asociados con el tiempo a la floración y adaptación a suelos. La mayoría de los SNPs candidatos a estar bajo selección, se encontraron en regiones no génicas y no fueron mutaciones no-sinónimas (o sea, fueron mutaciones que no cambian aminoácidos, ya sean sinónimas o en regiones no codificantes). Concluyen que probablemente la complejidad del genoma del maíz, que tiene 85% de elementos móviles, está permitiendo la evolución de elementos funcionales no codificantes. Estos resultados son congruentes con otros estudios, donde se ha encontrado que muchos QTL se encuentran en regiones génicas cerca de los genes y muestran que las bases genéticas de las adaptaciones pueden ser complejas. Sin embargo, si se encontraron algunos SNPs génicos candidatos a estar bajo selección, como en el gen *b1* , que se asoció con altitud y temperatura y cuya función está involucrada en la síntesis de antocianina. Estos pigmentos se han asociado a la adaptación de plantas a ambientes más fríos y cambios en la iluminación.

Por otro lado, con las herramientas de nueva generación, ha sido posible estudiar en maíz cómo evolucionan los elementos móviles y su relación con las adaptaciones y su domesticación[49]. Los elementos móviles ocupan el 85 % del genoma del maíz y son responsables de cambios importantes en el tamaño de estos genomas. Existen diversas familias de elementos móviles y, en general, se han asociado con selección deletérea. Sin embargo, los elementos móviles a veces discriminan donde se insertan y hay una familia que tiene preferencia por regiones génicas, lo que podría tener algunos efectos selectivos sobre los maíces. Un ejemplo son los *Class 2 miniature inverted repeat elements* (MITEs). Tenaillon *et al*.[49] compararon, utilizando métodos *next-gen*, el tamaño de los genomas de *Z. mays* y *Z. luxurians*. Encontraron una reducción en el tamaño de los genomas asociados al origen de los teosintes y otra reducción que sucedió alrededor de la domesticación de los maíces, concluyendo que podría haber un efecto selectivo a la reducción del genoma, lo que podría conferir cambios fisiológicos, fenológicos y de historia de vida[50].

Actualmente se están estudiando en el país con métodos *next-gen* tanto el genoma de las variedades criollas del maíz, como es el caso del detallado estudio de María Arteaga *et al*. (no publicado) que mencionamos arriba, como diferentes estudios en proceso en nuestro laboratorio con las poblaciones silvestres del teosinte.

Otros ejemplos de genómica de poblaciones y adaptación

La literatura de genómica de poblaciones está creciendo muy rápidamente. Vamos a ver sólo cuatro ejemplos más o menos arbitrarios que nos parecen interesantes.

a) El pez espinoso

El pez espinoso *Gasterosteus aculeatus*, tiene una larga tradición como sistema de estudio científico. Niko Tinbergen[51], el fundador del estudio científico moderno de la conducta animal, desarrolló en parte sus ideas analizando los complejos patrones de apareamiento y de cuidado parental en estos organismos. Existen poblaciones de las especies que viven en los mares del norte, en Europa, Asia y Norte América, y han sido especialmente estudiados, ya que aunque en general son marinos, en diferentes sistemas de ríos se han adaptado a vivir en agua dulce. La mayor parte de los peces y organismos acuáticos no pueden moverse libremente entre agua dulce y marina, debido a que las diferencias en salinidad generan diferentes problemas osmóticos que pueden matar a los organismos. Así, de manera independiente, en diferentes ríos las poblaciones de este pez se han adaptado recientemente a vivir en agua dulce. Hohenlohe *et al*.[27] estudiaron

gran cantidad de marcadores genéticos tipo *rad-tags* en la plataforma Illumina (ver Tabla II), y detectaron 45,000 SNPs en 20 individuos en cada una de cinco poblaciones del pez en Alaska: dos poblaciones marinas y tres de ríos. Con análisis de F_{ST} pareadas entre poblaciones encontraron, que aunque cada población es diferente en las frecuencias de ciertos alelos, existen zonas del genoma con claras diferencias en las frecuencias alélicas entre las poblaciones dulce-acuáticas y marinas. Estas diferencias son el resultado de selección natural divergente entre los ambientes y representan adaptaciones fisiológicas convergentes al agua dulce. Los genes adaptativos están involucrados a sistemas de regulación asociados con la osmoregulación, así como con el desarrollo de los huesos y la morfología del esqueleto. Detectaron 31 genes candidatos, ocho relacionados a la respuesta al estrés osmótico y desarrollo de órganos de osmoregulación y 23 loci relacionados con patrones y homeostasis del esqueleto. Estas regiones del genoma corresponden a sitios adaptativos, algunos de los cuales ya se habían detectado con diseños experimentales de genética cuantitativa y análisis fisiológicos, ya que como explicamos arriba, es un organismo muy estudiado. También encontraron regiones genómicas con baja diferenciación, que corresponderían a genes bajo selección balanceadora, que se asociaron a genes implicados en la defensa contra patógenos, incluyendo genes relacionados con la vía de la inflamación y la respuesta inmune.

b) Arabidopsis lyrata

Turner *et al*.[52] compararon 100 plantas de cuatro poblaciones de *Arabidopsis lyrata*, un pariente silvestre exógamo cercano a la famosa *Arabidopsis thaliana*, el organismo modelo por excelencia en plantas. Estudiaron dos poblaciones nativas de suelos "normales" y dos de suelos "serpentinos", que tiene elevados niveles de metales pesados y un bajo cociente de calcio: magnesio. Usando la plataforma Illumina (Tabla II), encontraron más de 8 millones de polimorfismos, 96 de los cuales tuvieron claras diferencias en las frecuencias alélicas entre los dos tipos de suelo. Estos genes incluyeron loci que confieren ventajas adaptativas entre los dos tipos de ambientes, como son transportadores de calcio y de magnesio y loci involucrados en la detoxificación de metales pesados. Asimismo, se encontraron claras evidencias de evolución paralela entre las diferentes poblaciones de la especie que viven en ambientes similares.

c) Pinus taeda

Eckert *et al*.[33] buscaron genes relacionados con las adaptaciones al estrés hídrico del pino *P. taeda*, utilizando cerca de 7000 SNPs. Estos autores utilizaron métodos de detección de loci "outliers," o sea genes que tienen una diferenciación mucho más alta que la esperada dados los niveles de variación genética y modelos demográficos adecuados. Con los genes identificados se van a poder hacer estudios que eventualmente permitan usar los genes para generar poblaciones resistentes a la sequía que se puede esperar que ocurra con el cambio global.

d) Estudios en bacterias

Como un estudio pionero en genómica de poblaciones podemos mencionar el trabajo de Castillo -Cobián *et al*.[14] en 2005, donde estudiamos cuidadosamente la isla de patogenicidad *LEE*, que incluye 52 genes ligados en seis cepas de *Escherichia coli*. Esta isla de patogenicidad es responsable en convertir bacterias comensales, que no son dañinas, en bacterias patógenas que producen enfermedades, en particular diarreas. Estas *E. coli* son llamadas enteropatógenas o *EPEC*. Castillo-Cobián *et al*.[14] encontraron que la selección natural es capaz de ajustar finamente los diferentes genes dentro de la isla de patogenicidad (ya que la recombinación ha "liberado" del ligamiento a las diferentes secciones de la isla; cada gen puede ser seleccionado de manera más o menos independiente). Esto es particularmente importante en los genes conocidos como *eae* y *tir*, que son los responsables de la unión entre la bacteria y la célula del intestino a la que parasitan. En estos genes se observa una clara señal de selección direccional positiva que es una respuesta a la "persecución" del sistema inmune del hospedero (los humanos y otros mamíferos donde es patógena). Este estudio cambió la visión que tenían los médicos de la enfermedad como resultado simplemente de "un cassette" (la isla *LEE*) que entra o sale de las bacterias.

En *E. coli* y otras bacterias se ha avanzado mucho en estudios genómicos, en parte debido al pequeño tamaño de sus genomas, por lo que existen ya muchos genomas completos. Por ejemplo, en nuestro laboratorio, Luna Sánchez-Reyes[53] en su tesis de licenciatura analizó el genoma de 12 cepas de *E. coli*, bacteria que puede ser patógena o comensal y aún de vida libre. En este estudio se comparó la dinámica evolutiva del genoma central, que tienen todas las cepas, contra la del llamado genoma flexible, que sólo se encuentra en algunas y corresponde a diferentes adaptaciones incluyendo los genes que las vuelven patógenas. Se encontró que las cepas patógenas de ave y extraintestinales e intestinales de humano presentan mayor diversidad genética que las cepas no-patógenas de vida libre y comensales de humano, no sólo a nivel del genoma flexible, sino también al del genoma central. Al igual que lo que reportamos en el estudio de Castillo-Cobián *et al*.[14], las cepas patógenas mostraron señales de selección positiva en genes con funciones variadas (como en genes de transporte/ patogenicidad, metabolismo energético y de transcripción), a diferencia de las cepas no- patógenas, en las cuales predominó la selección purificadora. Podemos concluir que la adaptación de *E. coli* a diferentes nichos no ocurre solamente por la adquisición horizontal de genes, sino que la evolución del genoma central, así como la regulación de la expresión génica de esta parte del genoma, juegan un papel importante en este proceso. En una muestra más grande, en González-González *et al*.[54], analizamos 128 cepas para ocho genes de *E. coli*, y corroboramos que la recombinación homóloga juega un papel muy importante en la diversificación de esta bacteria. Sin embargo, esta recombinación no ha homogeneizado tanto a los genomas como se hubiera esperado. Esto se debe a que se encuentra una gran estructura poblacional asociada tanto

a la filogenia de *E. coli* como a la de los hospederos. Más estudios son necesarios para entender mejor el papel de la selección natural en dicha subestructura.

También hemos trabajado con el genoma de diferentes *Bacillus* endémicos a la región de Cuatro Ciénegas en Coahuila[55,56] y un análisis reciente por Moreno-Letelier *et al*.[57] reveló que algunas cepas divergieron de sus ancestros hace mucho tiempo, tal vez en el Jurásico, cuando el mar penetró a lo que ahora es el valle. Una de esta cepas de *Bacillus* (llamada m3-13) es un mosaico genético aún mas ancestral, ya que divergió de su especie hermana al final del Precámbrico (hace unos 800 millones de años). Lo anterior apoya nuestra idea de que el oasis de Cuatro Ciénegas es un "mundo perdido", una especie de máquina del tiempo donde los descendientes directos de microorganismos muy antiguos han persistido.

Conclusiones y perspectivas

En el futuro, tener genomas de referencia bien anotados y con numerosos marcadores morfológicos bien mapeados va a ser importante para avanzar mejor en estos estudios, pero, aunque hay que hacerlo con cuidado, los pasos son actualmente claros, y los costos cada vez son menores. En la Tabla II mostramos los estimados relativos de los costos de las principales técnicas *next-gen*. Pero adicionalmente debemos de tener en mente que aunque los métodos de próxima generación han mostrado indudable utilidad, no siempre son la mejor herramienta, ya que depende de la pregunta; otros métodos moleculares, como los indicados en la Tabla I, tal vez sean más eficientes, más sencillos (tanto experimentalmente como en el análisis informático) y especialmente más económicos.

Por otro lado, todas las plataformas de secuenciación tienen diferentes ventajas y desventajas, por lo que es muy importante conocer muy bien cómo funcionan y saber cuáles son las más apropiadas para los objetivos particulares; el problema es que las plataformas cambian todo el tiempo y se están mejorando cada día, por lo que es difícil conocer todas las posibilidades, además, usar una tecnología muy nueva puede conducirnos a desagradables y caras sorpresas al descubrir sesgos y los problemas que tienen.

En particular a nosotros nos interesa mucho cumplir un sueño Darwiniano, que compartimos con Fisher, Wright, Haldane, Dobzhansky y Lewontin: explorar la diversidad genética relacionada a las adaptaciones. Creemos que el futuro en este campo es muy brillante, pero el problema es indudablemente complicado por diferentes razones técnicas y evolutivas, como nos recuerda Tonsor[21]: *"…a truly general understanding of the patterns in and causes of spatial genetic structure across the genome remains elusive. To what extent is spatial structure driven by drift and phylogeography vs. geographical differences in environmental sources of selection? What proportion of the genome participates?"*.

Retomando la narración de Borges de nuestro epígrafe, es importante no sólo tener "anécdotas" de miles de genes, sino ¡poder encontrar las historias relevantes náufragas en el mar de datos genómicos! Así, parafraseando a Theodosius Dobzhansky, consideramos que nada de la genómica va a tener sentido si no se hace bajo el marco de referencia de la rica teoría de la genética de poblaciones que se ha desarrollado por cerca de 80 años, y que todos los estudios ecológicos y evolutivos del futuro deberán de considerar a la genómica de poblaciones como una herramienta fundamental.

Agradecimientos

Este estudio se realizó con apoyo del proyecto CONACyT, Investigación Científica Básica CB2011/167826 "Genómica de poblaciones: estudios en el maíz silvestre, el teosinte (*Zea mays* ssp. *parviglumis* y *Zea mays* ssp. *mexicana*)", otorgado a Luis E. Eguiarte, del proyecto CONABIO "Avances y perspectivas del uso de métodos de secuenciación próxima generación en el estudio de la genética de poblaciones y su empleo en problemas ambientales", otorgado a los Drs. César Domínguez y Luis E. Eguiarte, y del proyecto PAPIIT, UNAM. Clave: IN202712, otorgado a Luis E. Eguiarte.

Referencias

1. Dobzhansky, T. Nothing in Biology Makes Sense except in the Light of Evolution. *The American Biology Teacher* **35**, 125-129 (1973).
2. Fisher, R.A. The genetic theory of natural selection (Oxford University Press, Oxford, UK, 1930). 318 págs.
3. Wright, S. Evolution in mendelian populations. *Genetics* **16**, 97- 159 (1931).
4. Wright, S. The evolution of mutation, inbreeding, crossbreeding and selection in evolution. *Proccedings of the sixth Internationa congress of genetics*, 356-366 (1932).
5. Haldane, J.B.S. The causes of evolution (Longmans, Green & Co. London, UK, 1932). 222 págs.
6. Hedrick, P.W. Genetics of populations. 4ª edition (Jones and Bartlett publishers. Sudbury, Massachusetts, 2011). 700 págs.
7. Lewontin, R.C. & Hubby, J.L. A molecular approach to the study of genic heterozygosity in natural populations. II. Amount of variation and degree of heterozygosity in natural populations of *Drosophila pseudoobscura*. *Genetics* **54**, 595-609 (1966).
8. Piñero, D. & Eguiarte, L. The origin and biosystematic status of *Phaseolus coccienus* spp. *polyanthus*: electrophoretic evidence. *Euphytica* **37**, 199-203 (1988).
9. Lewontin, R.C. The genetical basis of evolutionary change (Columbia Universiyt Press, New York, EUA, 1974). 346 págs.
10. Eguiarte, L.E. Genética de poblaciones de *Astrocaryum mexicanum* Liebm. en Los Tuxtlas, Veracruz. Tesis de Doctorado. UNAPyP del CCH, Centro de Ecología, UNAM, México, D.F. (1990).
11. Navarro-Quezada, A. *et al*. Genetic differentation in the *Agave deserti* (Agavaceae) complex in the Sonoran Desert. *Heredity* **90**, 220-227 (2003).
12. Van Heerwaarden, J. *et al*. Fine scale genetic structure in the wild ancestor of maize (*Zea mays* ssp. *parviglumis*). *Molecular Ecology* **19**, 1162-1173 (2010).
13. Illumina. MaizeSNP50 BeadChip. Data Sheet: Genotyping. San Diego, EUA (2010).
14. Castillo-Cobián, A., Eguiarte, L.E. & Souza, V. A genomic

population genetics analysis of the pathogenic enterocyte effacement island in *Escherichia coli*: The search of the unit of selection. *Proceedings of the National Academy of Sciences* **102**, 1542-1547 (2005).
15. Lewontin, R.C. El sueño del genoma humano y otras ilusiones (Barcelona, España, Ediciones Paidós, 2001). 206 págs.
16. Eguiarte, L.E., Souza, V. & Aguirre, X. Ecología Molecular (INE, SEMARNAT, CONABIO, UNAM. México, D.F., 2007). 594 págs.
17. Futuyma, D.J. Evolution (Sinauer Suderland, Mass., EUA, 2009). 633 págs.
18. Boyko, A.R. *et al*. A simple genetic architecture underlies morphological variation in dogs. *PLoS Biology* **8**, e1000451 (2010).
19. VonHoldt, B.M. *et al*. Genome-wide SNP and haplotype analyses reveal a rich history underlying dog domestication. *Nature* **464**, 898-902 (2010).
20. Hufford, M.B. *et al*. Comparative population genomics of maize domestication and improvement. *Nature Genetics* **44**, 808-813 (2012).
21. Tonsor, S.J. Population genomics and the causes of local differentiation. *Molecular Ecology* **21**, 5393-5395 (2012).
22. Harismendy, O. *et al*. Evaluation of next generation sequencing platforms for population targeted sequencing studies. *Genome Biology* **10**, R32 (2009).
23. Allendorf, F.W., Hohenlohe, P.A. & Luikart, G. Genomics and the future of conservation genetics. *Nature Reviews Genetics* **11**, 697-709 (2010).
24. Ekblom, R. & Galindo, J. Applications of next generation sequencing in molecular ecology of non-model organisms. *Heredity* **107**,1-15 (2011).
25. Metzker, M.L. Sequencing technologies -the next generation. *Nature Reviews* **11**, 31-46 (2010).
26. Glenn, T.C. Fieldguide to next-generation DNA sequencers. *Molecuar Ecology Resources* **11**, 759-769 (2011).
27. Hohenlohe, P.A. *et al*. Population genomics of parallel adaptation in threespine sticklebacks using sequenced RAD tags. *PLos Genetics* **6**, e1000862 (2010).
28. Elshire, R.J. *et al*. A Robust, Simple Genotyping-by-Sequencing (GBS) Approach for High Diversity Species. *PLosOne* **6**, e19379 (2011).
29. Baird, N.A. *et al*. Rapid SNP Discovery and Genetic Mapping Using Sequenced RAD Markers. *PLoSOne* **3**, e3376 (2008).
30. Lewontin, R.C. & Krakauer, J. Testing the Heterogeneity of F Values. *Genetics* **80**, 397-398 (1975).
31. Wright, S. The genetical structure of populations. *Annals Eugenics* **15**, 323-354 (1951).
32. Namroud, M.C. *et al*. Scanning the genome for gene single nucleotide polymorphisms involved in adaptive population differentiation in white spruce. *Molecular Ecology* **17**, 3599-3613 (2008).
33. Eckert, A.J. *et al*. Patterns of Population Structure and Environmental Associations to Aridity Across the Range of Loblolly Pine (*Pinus taeda* L., Pinaceae). *Genetics* **185**, 969-982 (2010).
34. Nei, M. Analysis of gene diversity in subdivided populations. *Proceedings of the National Academy of Sciences* **70**, 3321-3323 (1973).
35. Beaumont, M.A. Adaptation and speciation: what can FST tell us? *Trends in Ecology and Evolution* **20**, 435-440 (2005).
36. Ross-Ibarra, J., Morrell, P.L. & Gaut, B.S. Plant domestication, a unique opportunity to identify the genetic basis of adaptation. *Proceedings of the National Academy of Sciences* **104**, 8641-8648 (2007).
37. Pyhäjärvi, T. *et al*. Complex patterns of local adaptation in teosinte. http://arxiv.org/abs/1208.0634 (2012).
38. Darwin, C. The Effects of Cross and Self Fertilisation in the Vegetable Kingdom (J. Murray, London, 1876). 482 págs.
39. Chia, J.M. *et al*. Maize HapMap2 identifies extant variation from a genome in flux. *Nature Genetics* **44**, 803- 807 (2012).
40. Eguiarte, L.E. *et al*. Diversidad filogenética y conservación: ejemplos a diferentes escalas y una propuesta a nivel poblacional para *Agave victoria-reginae* en el desierto de Chihuahaua, México. *Revista Chilena de Historia Natural* **72**, 475-492 (1999).
41. Delgado, P. *et al*. Using phylogenetic, genetic and demographic evidence for setting conservation priorities for Mexican rare pines. *Biodiversity and Conservation* **17**, 121-137 (2008).
42. Kato, T.A. *et al*. Origen y diversificación del maíz: una revisión analítica (UNAM, CONABIO. México, D.F, 2009) 115 págs.
43. Ross-Ibarra, J., Tenaillon, M.I. & Gaut, B.S. Historical divergence and gene flow in the genus *Zea*. *Genetics* **181**, 1399-1413 (2009).
44. Van Heerwaarden, J. *et al*. Genetic signals of origin, spread, and introgression in a large sample of maize landraces. *Proceedings of the National Academy of Sciences* **108**, 1088-1092 (2011).
45. Buckler, E.S. *et al*. Phylogeography of the wild subspecies of *Zea mays*. *Maydica* **51**, 123-134 (2006).
46. Fukunaga, K. *et al*. Genetic diversity and population structure of teosinte. *Genetics* **169**, 2241-2254 (2005).
47. Moeller, D.A., Tenaillon, M.I. & Tiffin, P. Population structure and its effects on patterns of nucleotide polymorphism in the teosinte (*Zea mays* ssp. *parviglumis*). *Genetics* **176**, 1799-1809 (2007).
48. Gore, M.A. *et al*. A First-Generation Haplotype Map of Maize. *Science* **326**, 1115-1117 (2009).
49. Tenaillon, M.I. *et al*. Genome Size and Transposable Element Content as Determined by High-Throughput Sequencing in Maize and Zea luxurians. *Genome Biology Evolution* **3**, 219-229 (2011).
50. Gaut, B.S. & Ross-Ibarra, J. Perspective-selection on major components of angiosperm genomes. *Science* **320**, 484-486 (2008).
51. Tinbergen, N. The study of instinct (Clarendon Press, Oxford, UK, 1951) 256 págs.
52. Turner, T.L. *et al*. Population resequencing reveals local adaptation of *Arabidopsis lyrata* to serpentine soils. *Nature Genetics* **42**, 260-263 (2010).
53. Sánchez-Reyes, L. Genómica de poblaciones asociada a los nichos ecológicos de *Escherichia coli*. Tesis de licenciatura. Facultad de Ciencias, UNAM. México, D.F. (2010).
54. González-González, A. *et al*. Hierarchical clustering of genetic diversity associated to different levels of mutation and recombination in *Escherichia coli*: a study based on Mexican isolates. *Infection, Genetics and Evolution*, doi: http://dx.doi.org/10.1016/j.meegid.2012.09.003 (2012).
55. Alcaraz, L.D. *et al*. The genome of *Bacillus* coahuilensis reveals adaptations essential for survival in the relic of an ancient marine environment. *Proceedings of the National Academy of Sciences* **105**, 5803-5808 (2008).
56. Alcaraz, L.D. *et al*. Understanding the evolutionary relationships and major traits of *Bacillus* through comparative genomics. *BMC Genomics* **11**, 332, doi:10.1186/1471-2164-11-332 (2010).
57. Moreno-Letelier, A. *et al*. Divergence and phylogeny of Firmicutes from the Cuatro Ciénegas Basin, Mexico: a window to an ancient ocean. *Astrobiology* **12(7)**, 674-684 (2012).

Anexo 2, Información suplementaria del capítulo 3: Genomic differentiation and ecological speciation in teosintes (*Zea mays parviglumis* and *Zea mays mexicana*).

Supporting information for: *Genomic differentiation and ecological speciation in teosintes (Zea mays parviglumis and Zea mays mexicana)*

Authors: Jonás A. Aguirre-Liguori, Brandon S. Gaut, Juan Pablo Jaramillo-Correa, Maud I. Tenaillon, Salvador Montes-Hernández, Felipe García-Oliva, Sarah Hearne, Luis E. Eguiarte

SUPPORTING TABLES

Table S1. Geographic, environmental and genomic information for the populations sampled (See Supporting_table.1csv)

Table S2. Environmental variation of the teosintes populations (a list of the bioclim variables is in Table S1)

PC	Bioclim	Environment	% of variance
PC1	1,10,11,14,17,5,6,8,9	Temperature + Precipitation	48.07
PC2	12,13,16,7	Precipitation	21.13
PC3	15,19,3,4	Seasonality + precipitation	12.71
PC4	18,2	Diurnal range + precipitation	6.94
			Total: 87.95

SUPPORTING FIGURES

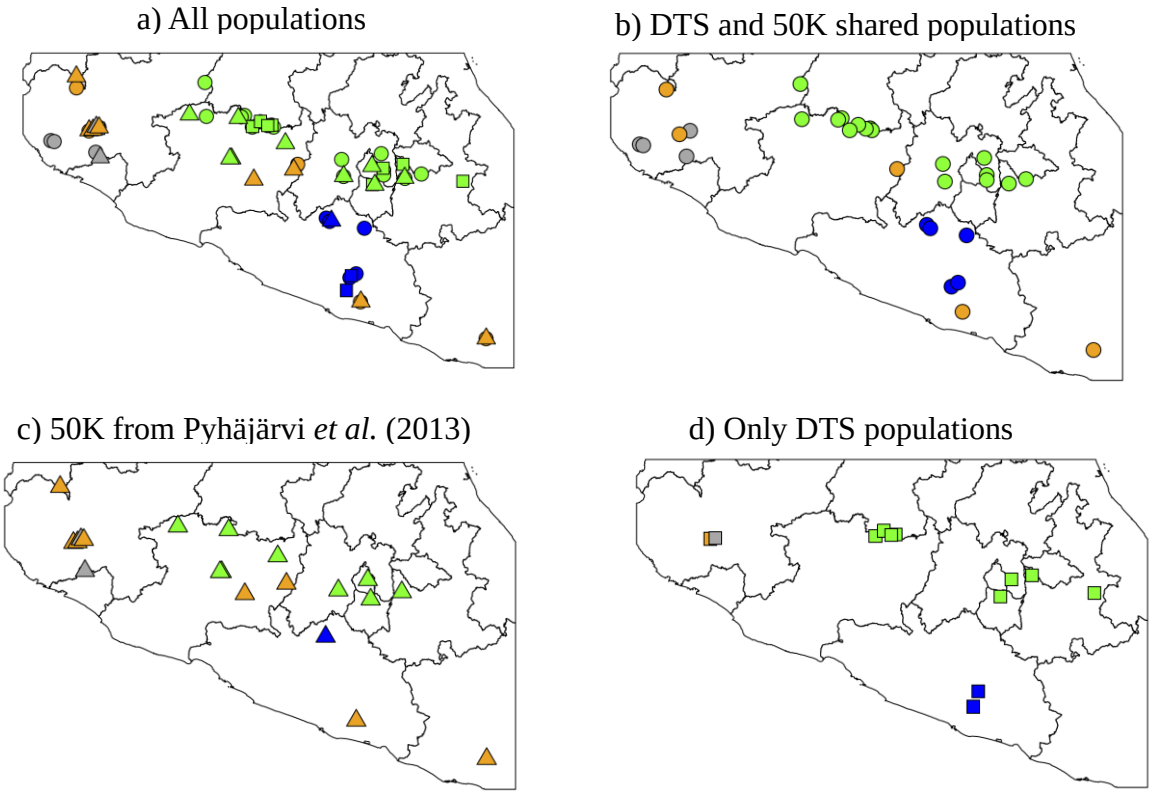


Figure S1. Studied teosinte populations. Colors correspond to the genetic groups identified by the PCA, Ward algorithM and based on Aguirre-Liguori *et al.* (2017). Triangles indicate populations for which we only have 50K data, circles for those that we have shared data, and squares for those that we only have DTS data. For simplicity reasons we separated each type of data into a single map. a) All analyzed teosinte populations, including the DTS (this paper) and 50K datasets (Pyhäjärvi *et al.* (2013) + Aguirre-Liguori *et al.* (2017). b) 29 shared populations between the DTS and 50K populations; c) 50K populations from Pyhäjärvi *et al.* (2013); d) only DTS populations, not analyzed in the other studies.

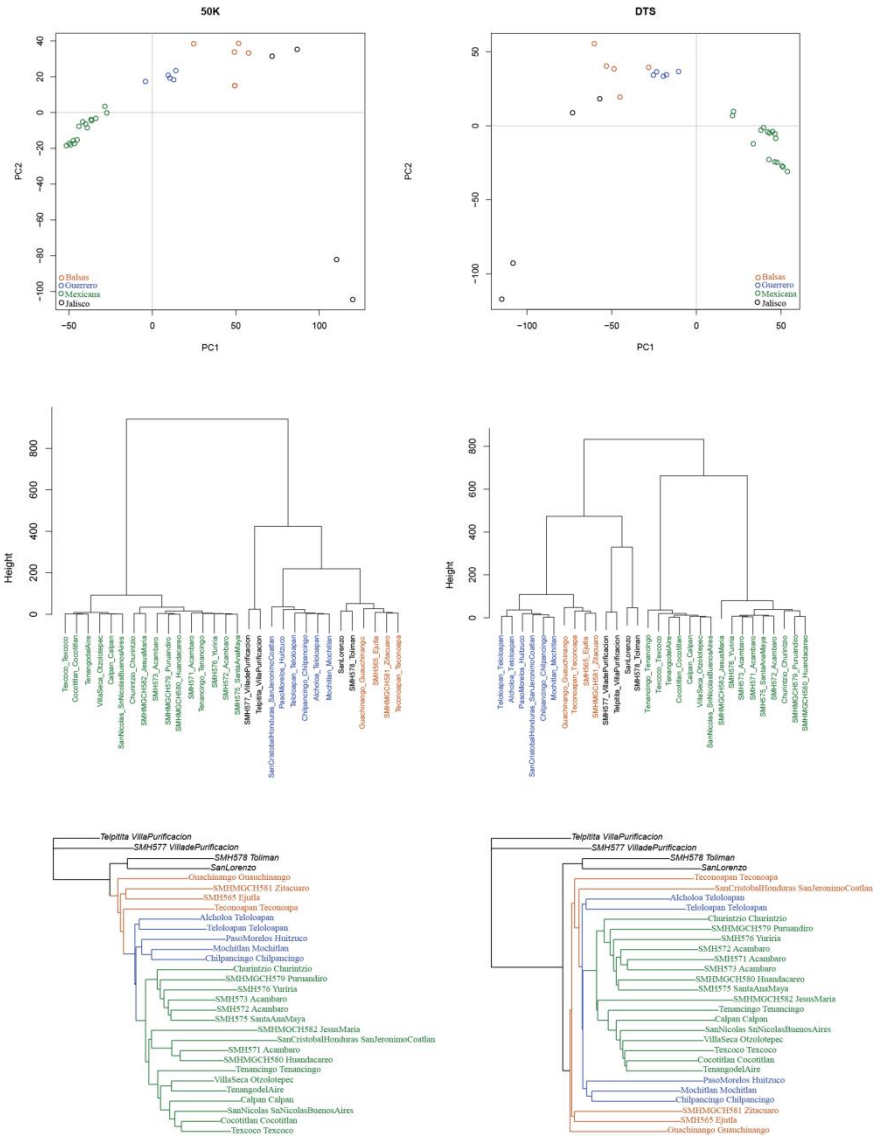


Figure S2. Principal component analyses (above), Ward clustering algorithm (center) and NJ trees (below) of allelic counts for 29 teosintes populations shared between the 50K and the DTS datasets. Green, blue, orange and black correspond to *mexicana*, Guerrero, Balsas and Jalisco groups. This clustering is based on Aguirre-Liguori *et al.* (2017) and ward clustering algorithm (Figure S1).

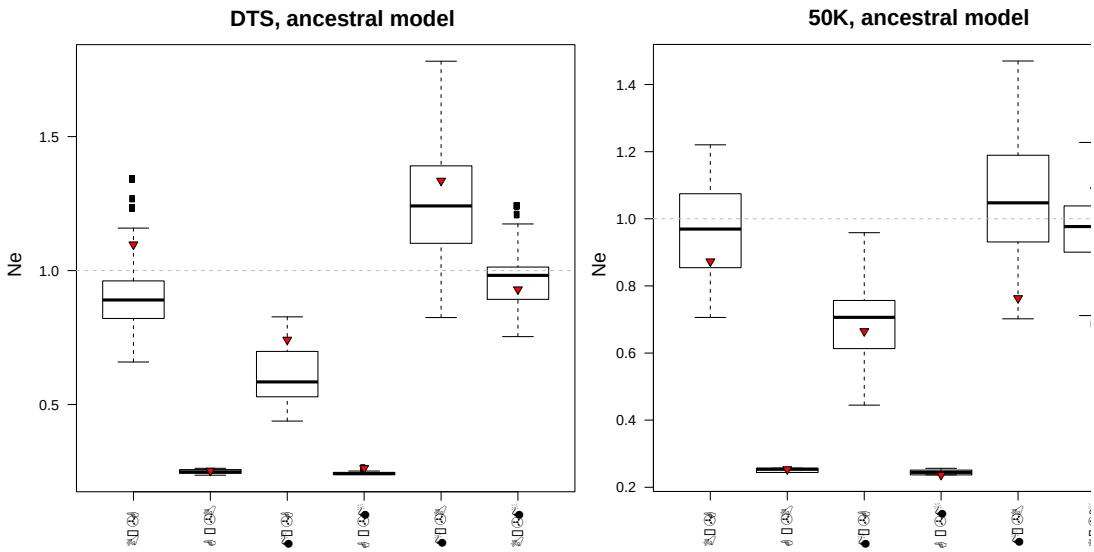
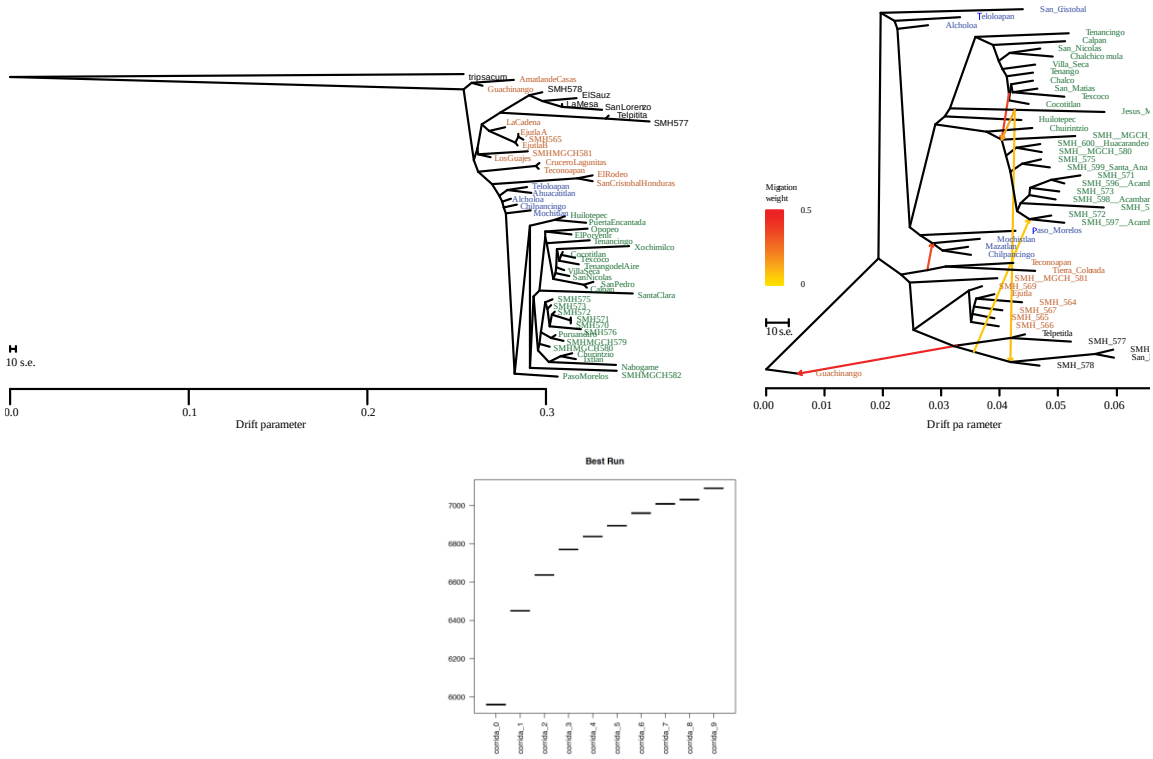


Figure S3. Number of migrants per generation between genetic groups for the Ancestral model (highest likelihood). a) DTS dataset; b) 50K dataset. B, G and M indicate Balsas, Guerrero and Mexicana. The first letter indicates the source of migrants and the second letter the reception of migrants. Red square indicate the $2N_em$ for the highest likelihood value.

Figure S4. Treemix analyses. We used Treemix (Pickrell y Pritchard, 2012) to determine the phylogenetic relationships among the 28 shared populations populations and to estimate ancestral and recent events of gene flow. Treemix uses genome-wide allele frequency data to first estimate the maximum likelihood tree of the populations. Then, based on populations that have poorer fits to the tree model, the program infers the presence and magnitude of migration events between populations that maximizes the likelihood of the tree (Pickrell y Pritchard, 2012). Since we only have an outgroup for the 50K dataset, we used this data to obtain the initial topology. We ran treemix for the 50K dataset with no migration events and *Tripsacum* as an outgroup. We then used the DTS dataset to infer gene flow events, since this dataset is not ascertained and therefore is not expected to include a large proportion of high frequency SNPS shared between subspecies. For this run, we used the 29 shared populations and tested K=1 to K=10 migration events, and set as root the Guauchinango population, from Jalisco. We retained the run for which the likelihood of the trees reached an asymptote, while the inferred migration rates became lower (Figure S3).



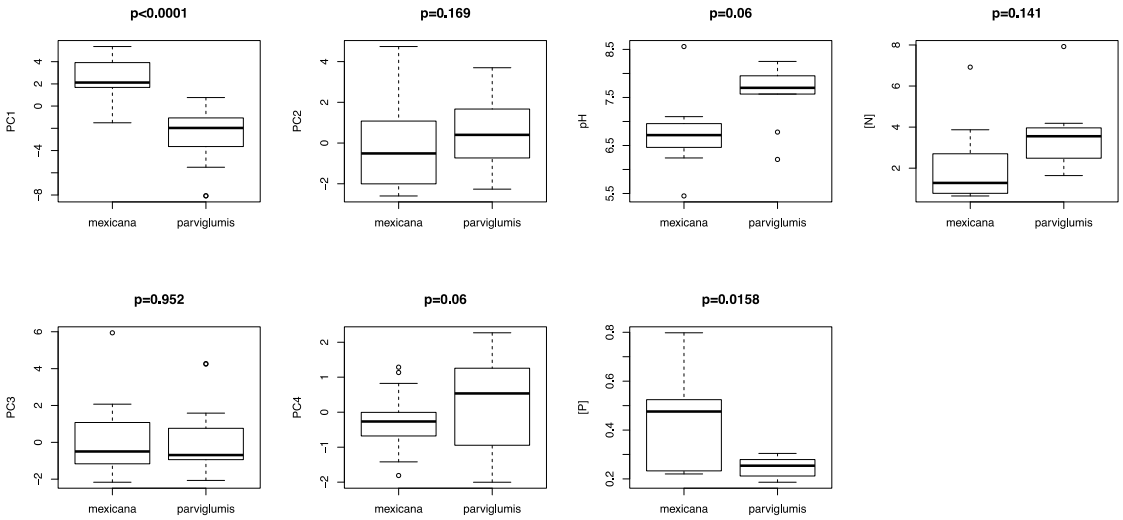


Figure S5. Environmental differences between *mexicana* and *parviglumis* for the variables tested by Bayescenv. The title of each plot gives the p-value of the ANOVA test.

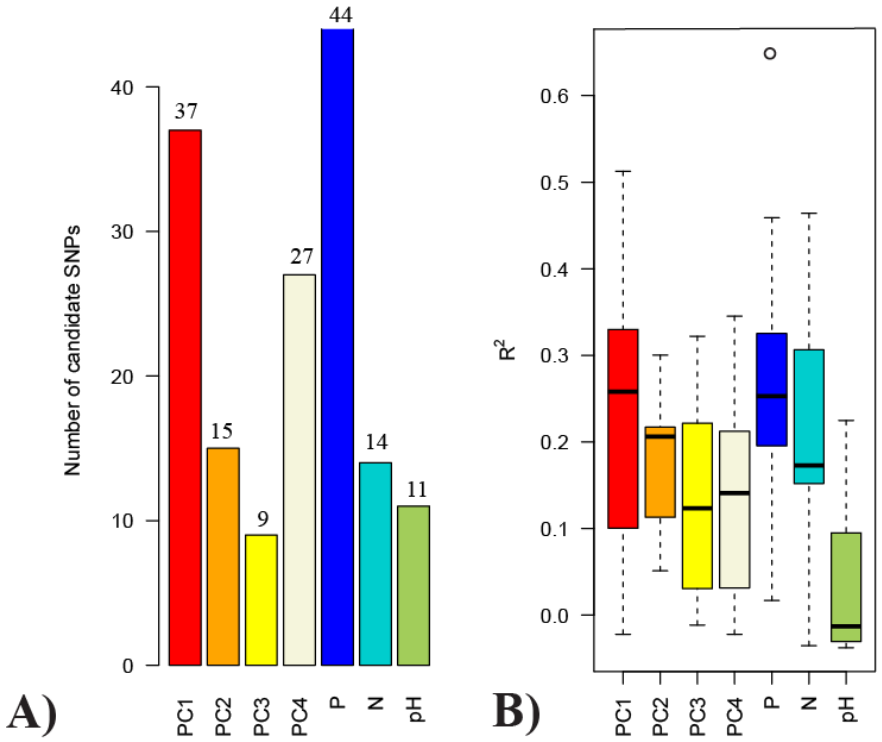


Figure S6- Number of SNPs for the 7 variables tested with the DTS dataset. Both the DTS and 50K datasets show similar results, indicating that P and PC1 are the most important variables associated to ecological differentiation.

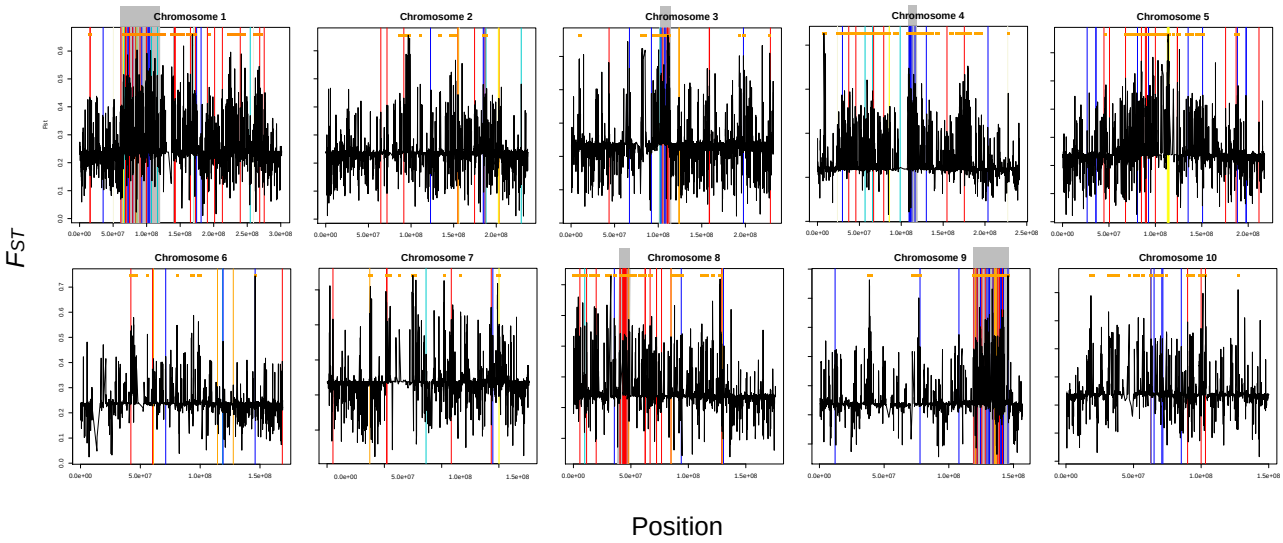


Figure S7. Locus by locus F_{ST} values along the 10 chromosomes of teosintes. Vertical color lines correspond to candidate SNPs. The colors indicate to which environmental variable the SNPs are associated to (Same colors as in Figure 3). The horizontal dark orange rectangles at the top of the figures represent the continuous blocks of high differentiation obtained by the HMM (See Methods and Results). Vertical grey rectangles correspond to regions that have large blocks of high differentiation and that are enriched with candidate SNPs.

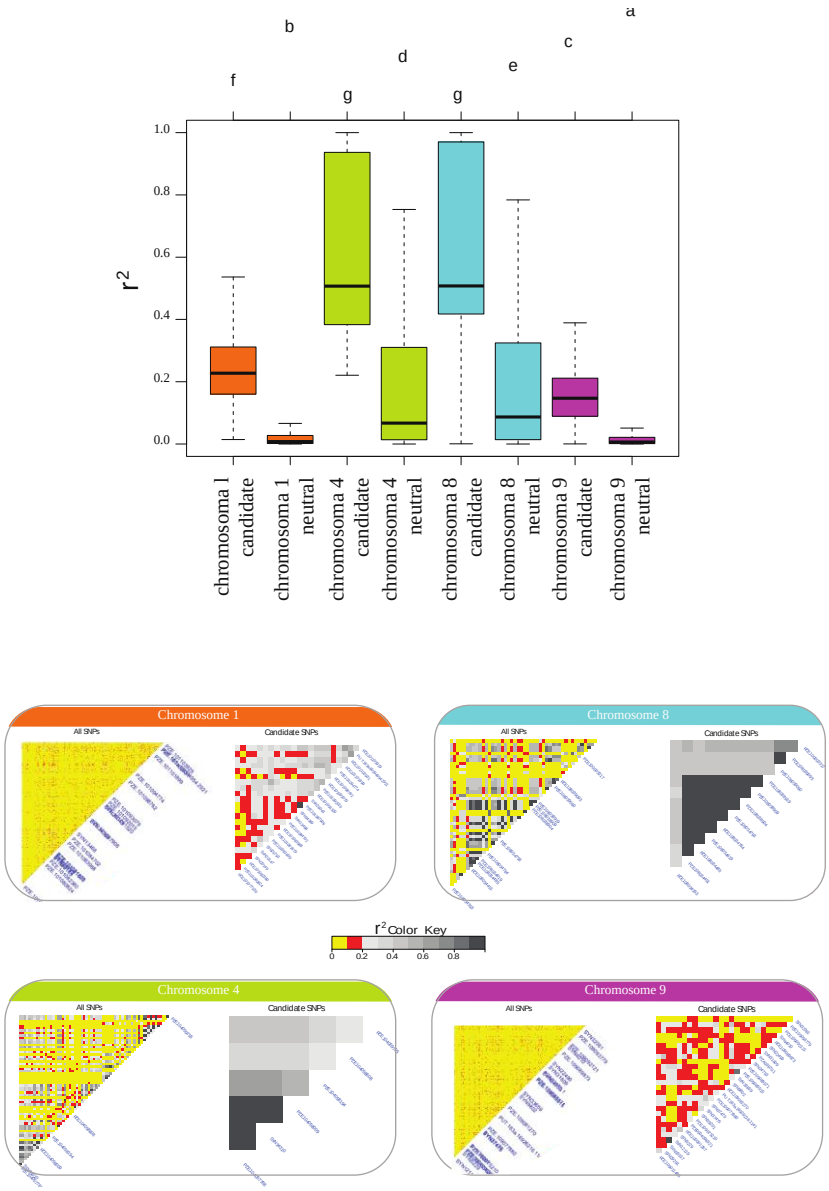


Figure S8. LD along neutral and candidate SNPs in the four long high F_{ST} regions. For all inversions, candidate SNPs present higher LD even if they are separated by regions with high recombination (as shown by the heatmaps below). Same letters above the boxplots indicate r^2 values that statistically do not differ