

# 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.

"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

**T**heodosius Dobzhansky<sup>1</sup> 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<sup>2-5</sup>), 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 cruces para tratar de entender el comportamiento genético, el honor recayó en Gregor Mendel, que descubrió como segregaba la variación genética en una cruce en casos de herencia de un solo carácter, 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 rearrreglos 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 cruces en los que se obtenían moscas completamente homocigotas para cromosomas enteros que se comparaban con la viabilidad de moscas silvestres, que son heterocigotas para esos cromosomas, experimentos que sugerían la existencia de altos niveles de variación genética (ver un resumen de estos experimentos, en Hedrick<sup>6</sup> págs. 51-57).

En 1966, un ex alumno de Dobzhansky, Richard Lewontin<sup>7</sup>, 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<sup>7</sup> 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<sup>8</sup>, 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 sustituciones 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 sustitución de aminoácido. Esta estimación sesgada de la variación genética preocupaba mucho a Richard Lewontin (ver por ejemplo su libro clásico del 1974<sup>9</sup>), 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 aquella 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<sup>9</sup> 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<sup>10</sup>, de 1990, donde analizamos 22 loci en total en la palma tropical

Marcador	Características	Ventajas	Desventajas
Variación fenotípica	Variantes observables y cuantificables	<ul style="list-style-type: none"> <li>- Se pueden observar en las poblaciones</li> <li>- En muchas ocasiones son adaptativos</li> <li>- Comparable con fósiles</li> </ul>	<ul style="list-style-type: none"> <li>- No representan la variación genética en todo el genoma</li> <li>- Pueden reflejar plasticidad</li> <li>- Muchas veces son poligénicos</li> </ul>
Aloenzimas	Variación en la migración de las enzimas durante la electroforesis, por cambios de aminoácidos	<ul style="list-style-type: none"> <li>- Muchos loci son polimórficos</li> <li>- Codominantes (identifica homocigos y heterocigos)</li> </ul>	<ul style="list-style-type: none"> <li>- Poca variación</li> <li>- No detecta mutaciones silenciosas</li> <li>- Pocos marcadores a lo largo del genoma</li> <li>- Difíciles de montar</li> </ul>
RFLPs	Variación en los fragmentos generados por enzimas de restricción, debido a cambios mutacionales en los sitios de reconocimiento		<ul style="list-style-type: none"> <li>- Información limitada</li> <li>- Complicados de montar</li> <li>- Dominantes</li> </ul>
RAPDs	Variación en la amplificación de fragmentos anónimos, que depende de la presencia diferencial de sitios de unión (mutación y cambios estructurales)	<ul style="list-style-type: none"> <li>- Muchos sitios polimórficos a lo largo del genoma</li> <li>- Alta variación</li> <li>- Oligonucleótidos universales</li> </ul>	<ul style="list-style-type: none"> <li>- Dominantes</li> <li>- Cada sitio poco informativo</li> <li>- Poco reproducibles</li> <li>- Se desconocen los sitios amplificados</li> <li>- Muchos artefactos</li> <li>- Difíciles de leer</li> </ul>
ISSRs	Similar a los RAPDs, pero los oligonucleótidos son secuencias más complejas y largas en tándem (microsatélites)	<ul style="list-style-type: none"> <li>- Muchos sitios polimórficos</li> <li>- Alta variación</li> <li>- Más reproducibles que los RAPDs</li> <li>- Oligonucleótidos universales</li> </ul>	<ul style="list-style-type: none"> <li>- Dominantes</li> <li>- Cada sitio poco informativo</li> <li>- Se desconocen los sitios amplificados</li> <li>- Difíciles de montar</li> <li>- Difíciles de leer</li> </ul>
AFLPs	Variación en la amplificación de fragmentos cortados con enzimas de restricción y ligados con adaptadores	<ul style="list-style-type: none"> <li>- Genera muchísimos fragmentos</li> <li>- Alta variación</li> <li>- Muy reproducibles</li> <li>- Oligonucleótidos universales</li> <li>- Alta representación genómica</li> </ul>	<ul style="list-style-type: none"> <li>- Muy complicados de montar</li> <li>- Dominantes</li> <li>- Cada sitio poco informativo</li> <li>- Difíciles de leer</li> </ul>
Microsatélites	Variación en el número de repeticiones de secuencias en tándem de ADN por inserción o pérdida de motivos	<ul style="list-style-type: none"> <li>- Muy polimórficos</li> <li>- Codominantes</li> <li>- Se conoce la región en la que se encuentran</li> <li>- En muchos casos se conoce la tasa de mutación</li> </ul>	<ul style="list-style-type: none"> <li>- Difíciles de montar</li> <li>- Modelos de mutación complicados</li> <li>- Error en la estimación de tasas y modelos de mutación genera resultados erróneos</li> <li>- Alta homoplasia</li> <li>- Alelos nulos</li> </ul>
SNPs	Mutaciones puntuales en secuencias de ADN Se pueden obtener por medio de RAD-tags o GBS (ver texto) o micro-arreglos	<ul style="list-style-type: none"> <li>- Existen miles en el genoma</li> <li>- Permite entender la evolución a nivel genómico</li> <li>- Nuevas herramientas de secuenciación permiten la obtención de miles de SNPs</li> <li>- En sitios neutrales y codificantes</li> </ul>	<ul style="list-style-type: none"> <li>- Caros</li> <li>- Cada SNP es poco informativo por sí solo</li> <li>- Difícil de analizar</li> </ul>

Tabla 1. 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.*<sup>11</sup> 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<sup>12</sup>. 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<sup>13</sup>.

#### 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*<sup>6,14</sup> 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 *cuernos 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 heterocigos son los que funcionan mejor, viven más y/o dejan más hijos. En esos casos hay exceso de heterocigos 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<sup>15</sup> 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<sup>6</sup>, Eguiarte *et al.*<sup>16</sup> y Futuyma<sup>17</sup> 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<sup>18-20</sup>. Estas nuevas perspectivas han sido comparadas con la revolución que, como mencionamos arriba, ocurrió en 1966 cuando Lewontin y Hubby<sup>7</sup>, introdujeron la técnica de la electroforesis al estudio de la variación genética de las poblaciones naturales<sup>21</sup>, 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 a la 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<sup>22-24</sup>. 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<sup>24-25</sup>. 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<sup>24</sup>, 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 mRNA (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<sup>24,26</sup>.

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 aquellas 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)<sup>26</sup>. 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<sup>26</sup>. 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<sup>26,27</sup>.

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: Raddtags, 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: Raddtags, 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<sup>26</sup>.

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<sup>26</sup>.

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<sup>12</sup>), analizar la historia demográfica de las poblaciones (si las poblaciones han crecido o decrecido) y proponer genes candidatos para adaptaciones a condiciones ambientales específicas<sup>27</sup>, 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<sup>26</sup>.

Como mencionamos antes, la cobertura genética (que tantas veces en promedio se secuencian 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<sup>28</sup>. 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*<sup>29</sup> (Tabla I) o *genotyping by sequencing*<sup>28</sup> (Tabla I). Así, no se secuencian 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 especie (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<sup>30</sup>: 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<sup>31</sup>. 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.*<sup>32</sup> y Eckert *et al.*<sup>33</sup>). La estimación de la  $F_{ST}$  es más fácil de visualizar siguiendo la definición de Nei<sup>34</sup>:  $F_{ST} = (H_T - H_S) / H_T$ , 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:  $F_{ST} = \text{Varianza}(p) / (p(1-p))$ , 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<sup>33,35-37</sup>. 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<sup>35</sup>. 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<sup>38</sup> al estudio del efecto de la endogamia en plantas. Las bases genéticas de la depresión por endogamia han sido discutidas desde hace mucho tiempo<sup>10</sup>, y algunos investigadores, entre los cuales se encontraba Theodosius Dobzhansky sospechaban que se debía a la heterosis, el fenómeno por el cual los organismos heterocigos 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 heterocigos para este gen en particular, se producen individuos homocigos 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*<sup>39</sup>), 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 homocigos, 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<sup>40,41</sup>.

¿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<sup>42</sup>: *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. *huhuetaenangensis* 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<sup>43</sup>, y que la domesticación del maíz a partir de *Z. m. parviglumis* ocurrió hace unos 11 mil años<sup>44</sup>. 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<sup>43</sup>.

Veamos cómo ha avanzado el número de marcadores usados en estas especies: Buckler *et al.*<sup>45</sup> 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. huhuetaenangensis* es el taxa basal, hermano de los anteriores. Buckler *et al.*<sup>45</sup> 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.*<sup>46</sup>, 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 *huhuetaenangesis*), 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.*<sup>47</sup> 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.*<sup>43</sup> obtuvieron las 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 *Zeamays* fue de hace unos 140 mil años. Gore *et al.*<sup>48</sup> 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.*<sup>12</sup>, 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.*<sup>44</sup> 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 descende 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.*<sup>20</sup> 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.*<sup>37</sup> 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 *bl*, 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<sup>49</sup>. 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.*<sup>49</sup> 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<sup>50</sup>.

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<sup>51</sup>, 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.*<sup>27</sup> 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.*<sup>52</sup> 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 "serpentinicos", 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.*<sup>33</sup> 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.*<sup>14</sup> 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.*<sup>14</sup> 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<sup>53</sup> 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.*<sup>14</sup>, 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.*<sup>54</sup>, 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<sup>55,56</sup> y un análisis reciente por Moreno-Letelier et al.<sup>57</sup> 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<sup>21</sup>: "...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 naufragas 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. *Proceedings of the sixth International 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. 4<sup>th</sup> 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 coccineus* spp. *polyanthus*: electrophoretic evidence. *Euphytica* **37**, 199-203 (1988).
9. Lewontin, R.C. The genetical basis of evolutionary change (Columbia University 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 differentiation 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 Sunderland, 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. *Molecular 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 Chihuahua, 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., Tenailon, 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., Tenailon, 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. Tenailon, 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).