



Journal of Basic & Applied Genetics

(Formerly MENDELIANA)

JOURNAL OF THE ARGENTINE SOCIETY OF GENETICS
REVISTA DE LA SOCIEDAD ARGENTINA DE GENÉTICA



BUENOS AIRES - ARGENTINA

CG
BA
BB

Included in:



Cited by:



COMITÉ EDITORIAL

Editor General:

Dra. Elsa L. Camadro

Fac. de Ciencias Agrarias, Universidad Nacional de Mar del Plata (UNMdP, y Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET)
Balcarce, Argentina
camadro.elsa@inta.gob.ar

Editores Asociados:

Citogenética Animal

Dra. Liliana M. Mola

Departamento de Ecología, Genética y Evolución, Fac. de Ciencias Exactas y Naturales, Universidad de Buenos Aires (UBA), y CONICET
Buenos Aires, Argentina
limola@ege.fcen.uba.ar; lilimola@yahoo.com.ar

Citogenética Humana

Dra. Roxana Cerretini

Centro Nacional de Genética Médica, ANLIS, "Dr Carlos G Malbrán"
Buenos Aires, Argentina
rcerretini@gmail.com

Citogenética Vegetal

Dra. Liliana M. Mola

Departamento de Ecología, Genética y Evolución, Fac. de Ciencias Exactas y Naturales, UBA, y CONICET
Buenos Aires, Argentina
limola@ege.fcen.uba.ar; lilimola@yahoo.com.ar

Dr. José Guillermo Seijo

Instituto de Botánica del Nordeste, Universidad Nacional del Nordeste (UNNE), y CONICET
Corrientes, Argentina
seijo@agr.unne.edu.ar

Genética de Poblaciones y Evolución

Dr. Jorge Cladera

Instituto de Genética "Ewald Favret", Centro de Investigación en Ciencias. Veterinarias y Agronómicas, INTA
Castelar, Argentina
cladera.jorge@inta.gob.ar

Dra. Noemí Gardenal

Fac de Ciencias Exactas, Físicas y Naturales, Universidad Nacional de Córdoba (UNC), y CONICET
Córdoba, Argentina
ngardenal@unc.edu.ar

Dr. Juan César Vilardi

Departamento de Ecología, Genética y Evolución, Fac. Ciencias Exactas y Naturales, UBA, y CONICET
Buenos Aires, Argentina
vilardi@bg.fcen.uba.ar

Genética Humana y Genética Médica

Dr. Santiago Lippold

Centro de Educación Médica e Investigaciones Clínicas (CEMIC)
Buenos Aires, Argentina
sel1@fibertel.com.ar

Genética Médica, Humana y Citogenética

Dra. María Inés Echeverría

Instituto de Genética, Facultad de Ciencias Médicas, Universidad Nacional de Cuyo (UNCu)
Mendoza, Argentina
miecheve@fcm.uncu.edu.ar

Dra. Silvia Ávila

Universidad Nacional de COMAHUE
Hospital Castro Rendón
Neuquén, Argentina
silvia347@gmail.com

Genética Molecular (Animal)

Dr. Guillermo Giovambattista

Instituto de Genética Veterinaria (IGEVET), Fac. de Ciencias Veterinarias, Universidad Nacional de La Plata (UNLP) y CONICET
La Plata, Argentina
ggiovam@fcv.unlp.edu.ar

Genética Molecular (Vegetal)

Dr. Alberto Acevedo

Centro de Investigación de Recursos Naturales, INTA
Castelar, Argentina
acevedo.alberto@inta.gob.ar

Dr. Andrés Zambelli

Fac. de Ciencias Agrarias, Universidad Nacional de Mar del Plata.
Balcarce, Argentina
andres.d.zambelli@gmail.com

Genética y Mejoramiento Animal

Ing. (M. Sc.) Carlos A. Mezzadra

Área de Investigación en Producción Animal, EEA Balcarce, INTA y Fac. de Ciencias Agrarias, UNMdP
Balcarce, Argentina
mezzadra,carlos@inta.gob.ar

FOTOGRAFÍAS Y AUTORES

Dra. Liliana A. Picardi

Cátedra de Genética, Fac. de Ciencias Agrarias,
Universidad Nacional de Rosario (UNR)
Zavalla, Argentina
lpicardi@fcagr.unr.edu.ar

Genética y Mejoramiento Genético Vegetal

Dra. Natalia Bonamico

Facultad de Agronomía y Veterinaria,
Universidad Nacional de Río Cuarto (UNRC)
Córdoba, Argentina
nbonamico@ayv.unrc.edu.ar

Dr. Ricardo W. Masuelli

EEA La Consulta, INTA
Fac. de Ciencias Agrarias, Universidad Nacional de Cuyo
(UNCu), y CONICET,
Mendoza, Argentina
rmasuelli@fca.uncu.edu.ar

Dra. Mónica Poverene

Departamento de Agronomía, Universidad Nacional del
Sur (UNS).
Bahía Blanca, Argentina
poverene@criba.edu.ar

Mutagénesis

Dr. Alejandro D. Bolzán

Laboratorio de Citogenética y Mutagénesis,
Instituto Multidisciplinario de Biología Celular (IMBICE),
y CONICET
La Plata, Argentina
abolzan@imbice.gov.ar; abolzan64@gmail.com

Mutaciones Inducidas en Mejoramiento Vegetal

Ing. Agr. (M.Sc.) Alberto R. Prina

Instituto de Genética "Ewald Favret", Centro de
Investigación en Ciencias Veterinarias y Agronómicas,
INTA
Castelar, Argentina
prina.albertoraul@inta.gob.ar

Consultor Estadístico:

Ing. Agr. Francisco J. Babinec

EEA Anguil, INTA, y
Fac. de Agronomía, Universidad Nacional de La Pampa
(UNLPam)
La Pampa, Argentina
babinec.francisco@inta.gob.ar

Secretaría de Redacción:

Dra. María de las Mercedes Echeverría

Fac. de Ciencias Agrarias, UNMdP
Balcarce, Argentina
echeverria.maria@inta.gob.ar



Alstroemeria aurea Graham

Autor: Pedro Jara Seguel



Habranthus robustus Herb. ex Sweet. $2n=2x=12$.
Metafase mitótica.

Autora: Anita Gianini Aquino



Habranthus robustus Herb. ex Sweet.

Autora: Anita Gianini Aquino

ÍNDICE

9 - 15

Article 1 - opinion

CRISPR/CAS9: REFLEXIONES BIOÉTICAS SOBRE LAS MODIFICACIONES GENÓMICAS
CRISPR/CAS 9: BIOETHICAL REFLECTIONS ON GENOME MODIFICATIONS

Lima N.S.

17 - 23

Article 2 - research

CLINICAL AND CYTOGENETIC CHARACTERIZATION OF A PATIENT WITH TETRASOMY 18P
CARACTERIZACIÓN CLÍNICA Y CITOGENÉTICA DE UN PACIENTE CON TETRASOMÍA 18P

Vazquez Cantu D.L., Gutiérrez García V.M., Cruz Camino H., Lara Díaz V.J., Garza García K., Garduño L.M., Meléndez R., Paz A., Mayén D.G., Cantú Reyna C.

25 - 36

Article 3 - research

LA ARQUITECTURA GENÉTICA COMO HERRAMIENTA DE ANÁLISIS DEL MAPA GENOTIPO-FENOTIPO
GENETIC ARCHITECTURE AS AN ANALYSIS TOOL OF THE GENOTYPE-PHENOTYPE MAP

Petino Zappala M.A., Fanara J.J.

37 - 49

Article 4 - research

CONTROL DE FALSOS DESCUBRIMIENTOS EN MAPEO ASOCIATIVO CON POBLACIONES ESTRUCTURADAS
FALSE DISCOVERY RATE CONTROL IN ASSOCIATION MAPPING WITH GENETICALLY STRUCTURED POPULATIONS

Peña Malavera A., Bruno C., Balzarini M.



51 - 64

Article 5 - research

¿FUNCIONA LA LEY DE EQUILIBRIO DE HARDY-WEINBERG EN AUTOPOLIPLOIDES IGUAL QUE EN DIPLOIDES?

DOES THE HARDY-WEINBERG LAW OF EQUILIBRIUM WORK IN AUTOPOLYPLOIDS AS IT DOES IN DIPLOIDS?

Núñez-Colín C.A.



65 - 69

Communication 6 - research

CHILEAN PLANTS CYTOGENETIC DATABASE: COVERAGE, FEATURES AND USAGES
BASE DE DATOS CITOGENÉTICOS DE PLANTAS CHILENAS: COBERTURA, CARACTERÍSTICAS Y USOS

Jara-Seguel P., Urrutia J.



71

Obituary

DR. EDUARDO E. CASTILLA



CRISPR/Cas9: REFLEXIONES BIOÉTICAS SOBRE LAS MODIFICACIONES GENÓMICAS

CRISPR/CAS 9: BIOETHICAL REFLECTIONS ON GENOME MODIFICATIONS

Lima N.S.

Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET).
Facultad de Psicología, Universidad de Buenos Aires.

lima.natacha@hotmail.com

ABSTRACT

CRISPR/Cas9 can be considered the biotechnological discovery of the century. Recent publications on the feasibility of producing permanent changes in the DNA of gametes and embryos –changes that, in modifying the germinal line, can be inherited throughout generations– has renovated the interest for the ethical reflection around the impact of the new and not so new technologies. The first reactions rapidly occurred, from a request for the complete suspension of clinical application (that is, the prohibition of implanting a genetically modified embryo with CRISPR in the uterus) till the approval of certain lines of basic research with similarly great preventions. This technology, characterized as “disruptive”, is analyzed at the light of the social phenomenon which has been generated around ethical, political, social, medical and subjective issues.

Key words: genetic edition, DNA changes, human genome

RESUMEN

CRISPR–Cas9 puede ser considerado el descubrimiento biotecnológico del siglo. Las recientes publicaciones sobre la posibilidad de producir cambios permanentes en el ADN de gametas y embriones, cambios que al modificar la línea germinal pueden ser heredados a través de las generaciones, ha renovado el interés por la reflexión ética en torno al impacto de las nuevas y no tan nuevas tecnologías. Las primeras reacciones no se han hecho esperar, desde un pedido de suspensión parcial o total de la aplicación clínica (es decir la prohibición de implantar en un útero un embrión modificado genéticamente con CRISPR) hasta la aprobación de determinadas líneas de investigación básica, con igualmente grandes reparos. Esta tecnología, caracterizada como “disruptiva”, es analizada a la luz del fenómeno social que ha generado siguiendo las aristas éticas, políticas, sociales, médicas y subjetivas.

Palabras clave: edición genética, cambios AND, genoma humano

Fecha de recepción: 18/10/2017
Fecha de aceptación de versión final: 8/12/2017

INTRODUCCIÓN

Podríamos arriesgarnos a considerar que la publicación del Proyecto Genoma Humano¹ marca el inicio de la “Era Genómica”, caracterizada por la aceleración en los descubrimientos genéticos. CRISPR/Cas9 es el último adelanto biotecnológico presentado en 2015². Se trata de “una herramienta” destinada a la “edición” genética desarrollada a partir de una bacteria y Cas9 que, al modo de un GPS enzimático, detecta el segmento del ADN dañado y lo repara. Se trata de un método rápido y de gran precisión que “edita” (es decir, altera) una secuencia específica (“target”) del ADN. La tecnología CRISPR puede utilizarse en cualquier tipo de células tanto somáticas como germinales. La edición genética de células somáticas tiene el objetivo de reparar o eliminar una mutación que podría causar una enfermedad. La diferencia está en que, en la utilización en células germinales (gametos o embriones *in vitro*), la alteración de la secuencia genómica de ese organismo podría alcanzar a su descendencia de un modo que es aún impredecible.

Las primeras publicaciones que presentan la aplicación de CRISPR/Cas9 en embriones no viables han generado un debate a escala mundial sobre distintos aspectos que reseñaremos en este escrito. Uno de los temas que ha resurgido es la alta expectativa de que estos descubrimientos se traduzcan rápidamente en beneficios clínicos para pacientes afectados por enfermedades genéticas.

Las tecnologías de edición genómica podrían lograr el avance en el tratamiento de muchas enfermedades como el HIV/SIDA, la hemofilia, la anemia de células falciformes, y muchos tipos de cáncer (Lanphier, 2015).

Antes de avanzar sobre las cuestiones éticas que genera la aplicación de CRISPR en la línea germinal (gametos o embriones) es preciso situar el llamado internacional a la cautela y a una atenta reflexión en torno a los aspectos éticos, sociales, políticos, de seguridad y de salud pública que deberían ser zanjados antes de proceder a la implantación en el útero de un embrión humano modificado con CRISPR.

Aprovechamiento compartido de los beneficios y bien común

El primer punto que ha surgido en esta indagación ha sido el alcance del principio de *aprovechamiento compartido de los beneficios y bien común*. ¿Deberían los resultados científico-tecnológicos (aplicación de CRISPR en este caso) ser parte del bien común de la sociedad que los creó o de quien financió su desarrollo o aplicación)? ¿Qué alcance tiene la noción de “*aprovechamiento compartido de los beneficios*”³ en este caso?

La Organización del Genoma Humano (*The Human Genome Organisation - HUGO*) sostiene que: “*un beneficio es un bien que contribuye al bienestar individual o de una comunidad dada. Por lo tanto, un beneficio no es idéntico a una ganancia en el sentido monetario o económico. Determinar qué constituye un beneficio depende de las necesidades, valores, prioridades y expectativas culturales. El Comité ético de HUGO recomienda que toda la humanidad comparta y tenga acceso a los beneficios de la investigación genética*” (Capps et al., 2017).

Más recientemente, el comité de ética de HUGO ha declarado que, en este ámbito, la “*solidaridad genómica*” (*genomic solidarity*) es el horizonte que podrá reunir a los individuos, las comunidades y las poblaciones, la academia, la investigación y la industria, en pos del bien común.

¹ El Proyecto Genoma Humano produjo un momento de gran visibilidad científica que atrajo un gran financiamiento público y privado. Situar la ideología que rodeó la revolución genética de entonces puede contribuir a entender mejor las controversias que un descubrimiento como CRISPR genera actualmente a nivel social. El Proyecto Genoma Humano puede equipararse con lo que fue “la carrera a la luna”, sólo que en este caso, las partes en disputa eran The International Human Genome Sequencing Consortium (IHGSC) y la compañía Celera Genomics. Lo que diferenciaba a estos grupos era una disputa ideológica sobre el acceso (público/libre) o la restricción (patentes) de la secuenciación completa del genoma humano. Mientras que el IHGSC quería publicar la secuenciación de un modo abierto y accesible a la comunidad, Celera aseguraba que, gracias a la propiedad intelectual que generaría derechos exclusivos y ganancias derivadas del descubrimiento genómico, podría realizar la secuenciación de un modo mucho más rápido y eficaz, salvando infinidad de vidas humanas a nivel mundial. Finalmente ambos grupos debieron cooperar y la publicación se hizo de un modo conjunto. Sin embargo, este ejemplo es paradigmático para reflexionar sobre qué constituye el “bien común” (public good) y qué representa el aprovechamiento compartido de los beneficios (benefit sharing) de la investigación genómica. Aspectos que retomaremos al analizar el alcance de CRISPR.

² El comienzo del debate en torno a CRISPR se remonta a los meses de marzo/abril de 2015 y a la publicación en la revista *Protein & Cell* realizada por un grupo de científicos chinos, sobre los primeros resultados de la aplicación de CRISPR en embriones humanos no viables.

³ Art. 15 Aprovechamiento compartido de los beneficios. Los beneficios resultantes de toda investigación científica y sus aplicaciones deberían compartirse con la sociedad en su conjunto y en el seno de la comunidad internacional, en particular con los países en desarrollo. Los beneficios que se deriven de la aplicación de este principio podrán revestir las siguientes formas: a) asistencia especial y duradera a las personas y los grupos que hayan tomado parte en la actividad de investigación y reconocimiento de los mismos; b) acceso a una atención médica de calidad; c) suministro de nuevas modalidades o productos de diagnóstico y terapia obtenidos gracias a la investigación; d) apoyo a los servicios de salud; e) acceso a los conocimientos científicos y tecnológicos; f) instalaciones y servicios destinados a crear capacidades en materia de investigación; g) otras formas de beneficio compatibles con los principios enunciados en la presente Declaración. Declaración Universal de Bioética y Derechos Humanos (UNESCO, 2005).

Otros autores (Mulvihill *et al.*, 2017) han propuesto el término “ecogenética” (*Ecogenetics*) para situar cómo cada persona puede apartarse del camino saludable debido a una combinación única de variaciones genéticas intrínsecas, que son el resultado de la interacción celular y del medioambiente; una acumulación de exposiciones prenatales, de la infancia, la niñez, la adolescencia y la adultez a lo largo de toda una vida. Cosmovisión que puede contribuir a evitar dos extremos que suelen reiterarse en este ámbito de indagación: el determinismo genético ⁴ (genetic “*determinism*”) y el excepcionalismo genético ⁵ (genetic “*exceptionalism*”).

Si uno examina el desarrollo de CRISPR como un fenómeno social, rápidamente puede percibir el gran número de instituciones públicas que han contribuido a su desarrollo: instituciones educativas, instituciones dedicadas a la investigación y el desarrollo, instituciones encargadas de garantizar la seguridad y la estabilidad de sus instalaciones (desde sofisticados laboratorios de alta tecnología, hasta el desarrollo del sistema normativo y regulatorio). Así, a la hora de analizar el “bien común” hay que tener en cuenta el rol que la infraestructura pública ha desempeñado a lo largo de las distintas instancias del proceso innovador.

El fenómeno de *Big Data* combinado con los avances e innovaciones propios de la Era Genómica ha generado una red de alcance mundial que despierta retos globales. Uno de los efectos más inmediatos de este fenómeno social ha sido la aparición de una narrativa en relación al avance y a la manipulación genética que pone en cuestión importantes aspectos sociales.

No debemos olvidar, además, que la comercialización de los productos biotecnológicos funciona en muchos niveles. En el año 2013, y luego de una década de controversias, la Corte Suprema de los Estados Unidos de América determinó que *los genes humanos no pueden ser patentados porque el ADN es “un producto de la naturaleza”*. Sin embargo, la tecnología CRISPR puede no ser considerada como un “producto de la naturaleza” ya que la modificación que

introduce en células animales y vegetales, no se hubiera “producido naturalmente” (Mulvihill *et al.*, 2017).

En este punto David DeGrazia (2016) presenta un argumento interesante para pensar en qué grado las modificaciones genéticas amenazarían la “naturaleza humana”. Pero, para eso, es necesario considerar que “la naturaleza humana” existe, punto que puede ser debatible. Luego, deberíamos preguntarnos qué quiere decir que “el mejoramiento” (*enhancement*) genético podría amenazar la naturaleza humana. Para aclarar este punto DeGrazia dice que es necesario distinguir entre lo que supondría “una superación” (*surpassing*) de “una alteración” (*altering*) (DeGrazia, 2016).

Enfoque utilitario vs. principio precautorio

CRISPR ha sido catalogado como una tecnología “disruptiva” (Ledford, 2015); su uso ha dividido las aguas de la arena científica⁶. Mientras que los grandes grupos abocados a la investigación de la manipulación genética de la línea germinal están de acuerdo en que la investigación básica debe continuarse; que diversos actores deben ser involucrados en la toma de las decisiones; que actualmente el uso clínico no debería realizarse; que el uso clínico debería proceder sólo si la seguridad y las cuestiones de eficacia técnica han sido resueltas y la sociedad ha acordado sus limitaciones; otros grupos sostienen que debería existir una suspensión (*moratorium*) en la investigación (Lanphier, 2015).

El grupo de científicos que ha solicitado la **moratoria** sostiene que aunque la técnica CRISPR ha expandido de un modo exponencial la investigación genética, no existen situaciones para las cuales su utilización en embriones humanos pueda ofrecer beneficios terapéuticos superiores o distintos a los métodos consolidados (PGD o IVF + análisis genético antes de la implantación suelen ser mejores opciones para casos en que ambos padres presentan la misma mutación para una enfermedad).

⁴ El *determinismo genético* sostiene que la secuencia de ADN es la causa principal de todos los rasgos humanos, normales y anormales (de salud y de enfermedad). El error está en que la enfermedad de una persona no es únicamente el resultado de su genética. Los factores medioambientales desempeñan un papel importante, en cualquier momento de la vida de la persona, ya sea a partir de una afectación repentina o crónica.

⁵ El *excepcionalismo genético* supone que la lógica y la retórica habituales para las demás situaciones humanas no aplican en estos casos, por lo tanto las consideraciones típicas deben ser revisadas cuando se trata de cuestiones relativas a

la modificación genética. El razonamiento subyacente es simple: los resultados o las consecuencias de las mutaciones de la línea germinal afectan no sólo a una persona o a una pareja, sino también a descendientes de incontables generaciones. Para una ampliación de estos puntos ver: Ethical issues of CRISPR technology and gene editing through the lens of solidarity, Mulvihill et al. (2017).

⁶ El documento de la ASHG Position Statement (2017) contiene una tabla con las principales recomendaciones que han esbozado los grandes grupos, organizaciones y agencias gubernamentales relacionadas con la edición genética de la línea germinal humana.

La utilización de CRISPR en embriones corre el riesgo de generar un mosaico genético, además no se puede determinar con exactitud el alcance que tendrá la intervención genómica efectuada. Además está decir que en muchos países existe un gran vacío normativo, donde no se ha logrado arribar a una reglamentación, o bien, estas investigaciones han sido consideradas como experimentales y no terapéuticas. “*Los efectos precisos de la modificación genética de un embrión pueden ser imposible de conocer hasta después del nacimiento*” (Lanphier, 2015: 411). En el caso de investigaciones genéticas que implican la modificación de la línea germinal, no queda claro qué información se necesitaría - o podría obtenerse - para informar adecuadamente a los futuros padres sobre los riesgos presentes en ese hijo y/o en las generaciones futuras.

Rápidamente, otras voces salieron a responder la moratoria. La académica británica Lovell Badge dijo “*Yo no estoy de acuerdo con una moratoria de este tipo, la cual en cualquier caso es poco probable que sea efectiva. Apoyo plenamente la investigación que se está llevando a cabo en embriones humanos in vitro, especialmente en embriones que no son requeridos para reproducción, y que de lo contrario serían descartados. Si la técnica funciona, hay muchas preguntas interesantes que podrían hacerse sobre el rol de genes específicos en el desarrollo embrionario temprano*” (Camporesi y Cavaliere, 2016: 576).

Otros más arriesgados, como Julian Savulescu ⁷, sostienen que lejos de estar mal, utilizar esta tecnología de edición genómica es un “imperativo moral”, ya que la manipulación genética podría reducir significativamente la carga de la enfermedad beneficiando a billones de personas en todo el mundo. Plantea que “abstener deliberadamente de participar en investigaciones que salvan vidas es ser moralmente responsable de las muertes previsibles y evitables de quienes podrían haberse beneficiado. *La investigación sobre la edición de genes no es una opción, es una necesidad moral*” (Camporesi y Cavaliere, 2016: 576).

Este debate, una vez más, trae a colación el antiguo

planteo bioético que diferencia el “enfoque utilitario” (representado por el argumento de Savulescu, donde el criterio de la acción correcta es el principio de utilidad. Una acción es moralmente correcta si maximiza la utilidad; no seguir investigando alternativas que posiblemente podrían salvar vidas es, desde esta perspectiva, igual a matar gente), del principio precautorio (expresado en la moratoria de Lanphier).

El principio precautorio en relación a CRISPR se basa en dos elementos: el antiguo miedo a generar “bebés de diseño” ⁸ o el retorno de las prácticas eugenésicas, y el riesgo de efectos imprevistos en las generaciones futuras ⁹ (las limitaciones en el conocimiento actual con respecto a la viabilidad técnica, y a las complejas relaciones entre los genes y el ambiente, impulsan las consideraciones precautorias).

Aplicación de CRISPR en el ámbito reproductivo

Si focalizamos el análisis en la utilización de CRISPR dentro del ámbito reproductivo, vemos que existen tres beneficios: a) un incremento en el conocimiento y en el entendimiento de los procesos de desarrollo y del funcionamiento genético que podría contribuir a desarrollar o a mejorar las tecnologías médicas; b) podría servir para solucionar los problemas de infertilidad en los futuros padres; y c) podría ser la solución de enfermedades para futuros niños. En este sentido, la edición genómica aumentaría la autonomía reproductiva de las personas (Vassena et al., 2016).

Sin embargo, también encontramos argumentos en contra de la manipulación de la línea germinal: a) es imposible proporcionar el consentimiento intergeneracional; b) las consecuencias son imposibles de predecir; c) tales manipulaciones suponen una amenaza para la dignidad humana (Sugarman, 2015).

Como vimos previamente, algunos autores se han pronunciado en contra de la modificación de la línea germinal del ADN en gametas y embriones, considerando

⁷ Savulescu ha desarrollado el principio de beneficencia procreativa (*procreative beneficence*) el cual propone que los padres tienen la obligación moral de elegir al “mejor” embrión disponible luego de haber realizado el diagnóstico genético pre-implantatorio (PGD). Presentar a esta elección como una obligación moral ha cuestionado el alcance del principio del *interés superior del niño* (Convención Internacional sobre los derechos del niño) y del artículo 11 de la Declaración Universal sobre Bioética y Derechos Humanos (Unesco, 2005) sobre *no discriminación y no estigmatización*.

⁸ En discusiones bioéticas suele denominarse a estos argumentos como “*slippery*

slope argument”. Aunque este tipo de argumentación ha sido ampliamente criticado en el ámbito filosófico, suelen utilizarse a menudo en relación a los desenlaces inciertos que podrían generar algunas tecnologías reproductivas.

⁹ CRISPR-Cas9 también debe ser juzgado considerando el bien de las generaciones futuras. Es difícil expresar obligaciones para con los recién nacidos sin implicar cierto grado de mutualidad (*mutuality*) o reciprocidad. El principio de solidaridad enfatiza las vulnerabilidades compartidas por los seres humanos dondequiera y cuando sea que existan.

que aún no están dadas las condiciones para garantizar la salud y la seguridad de estas aplicaciones en las generaciones futuras. La comunidad científica se ha dividido al respecto de la aplicación de CRISPR para la modificación de la línea germinal en embriones viables. La noción de “innovación responsable” (*responsible research and innovation RRI*) y la introducción de nuevas tecnologías para el ámbito de la salud reproductiva, demuestran una vez más sus aspectos más controversiales.

Si bien la investigación con embriones humanos es un tema espinoso y que no ha sido reglamentado en muchas partes mundo (incluidos la Argentina y Chile), podría considerarse incluso que en los países donde sí está reglamentado, aún es muy pronto para iniciar investigaciones con embriones viables de buena calidad (además de que estos pueden ser considerados como un “bien escaso”). Un uso responsable implicaría contar con altas probabilidades de que la investigación arroje hallazgos valiosos. Pero, dado el limitado conocimiento actual, los resultados aún no pueden ser garantizados.

Un último argumento que suele ser esgrimido frente a la aplicación de CRISPR en este ámbito es que el desenlace será la modificación del genoma humano. Esta argumentación es un tanto ambigua por varias razones: primero pareciera que los seres humanos contamos con un “catálogo fijo” de genes que se transmiten de generación en generación. Sin embargo, toda mutación resulta en la modificación del *pool* genético. Luego, cuando la edición genética es utilizada para reparar o reinsertar un gen que ya existe en el *pool* genético, en sentido estricto no podríamos hablar de una modificación. Otra razón podría ser que, a partir de la utilización de estas tecnologías, estaríamos alterando la especie humana. Sin embargo, para que una alteración tal se produzca, necesitaríamos de la aplicación a gran escala. Aunque esto parece improbable, no es imposible. No obstante, si la técnica es utilizada para solucionar defectos o enfermedades, sus efectos serían similares a las técnicas de *screening genético* (PGD) y prenatal utilizados corrientemente ¹⁰.

¹⁰ Para una ampliación de estos puntos ver el artículo completo *Genome engineering through CRISPR/ Cas9 technology in the human germline and pluripotent stem cells de Vassena et al.* (2016).

¹¹ El límite del día 14 fue establecido sobre la base de consideraciones biológicas y filosóficas, con el objetivo de encontrar una solución favorable ante el desacuerdo del estatuto moral del embrión humano, y no se debe sólo a limitaciones técnicas. En este punto, siguiendo la *falacia naturalista* de David

El Comité Internacional de Bioética de la UNESCO ha elaborado, en el año 2015, una argumentación contra la edición genómica al considerarla como una técnica que amenaza la dignidad humana, renovando antiguas preocupaciones eugenésicas. Sin embargo, las prácticas actuales de selección embrionaria post-diagnóstico genético pre-implantatorio (PGD) pueden también ocasionar efectos eugenésicos, y esta preocupación no ha sido lo suficientemente sólida como para prohibir su aplicación (Vassena *et al.*, 2016).

Por supuesto hay que tener en cuenta el contexto de estos desarrollos. Por ejemplo en el Reino Unido, la investigación con CRISPR se incluye en el apartado de “investigación en embriones humanos hasta el límite del día 14” (cuando se produce la formación de la línea primitiva)¹¹ permitido por el Acta de Fertilidad y Embriología Humana de 1990.

Como vemos, estos escenarios traen preguntas que son recurrentes en el ámbito de las tecnologías reproductivas: ¿cuál es el riesgo que estamos dispuestos a correr para tener un hijo que comparta nuestra genética? ¿Cuál es el valor atribuido al vínculo genético en nuestras sociedades?

Por último, se han elaborado algunos documentos tendientes a sentar posición respecto de las modificaciones en la línea germinal, los cuales sostienen que antes de cualquier aplicación clínica es necesario determinar: la justificación médica adecuada al caso, teniendo en cuenta las condiciones de discapacidad y calidad de vida; la gravedad de la condición; los tratamientos y opciones disponibles; el riesgo de ocurrencia; y la disponibilidad de otras opciones establecidas como PGD y diagnóstico prenatal (Ormond *et al.*, 2017).

Aplicación de CRISPR por fuera del ámbito embrionario

El impacto que generó la noticia sobre la aplicación de CRISPR en embriones humanos (no viables) pudo haber servido de pantalla para opacar otras aplicaciones de la tecnología con efectos igualmente importantes para el ámbito de lo humano. Las autoras Silvia Camporesi

Hume, decimos que una conclusión ética “debería hacer” no puede ser deducida a partir de premisas que establecen lo que “es”. Es decir, Hume argumentó que ninguna conclusión ética (por ejemplo: “deberíamos extender el límite del día 14 para la investigación con embriones”) puede deducirse de premisas puramente fácticas (“los científicos son capaces de cultivar embriones in vitro por más de 14 días”). Para una ampliación ver el artículo completo: *Emerging ethical perspectives in the clustered regularly interspaced short palindromic repeats genome-editing debate* de Camporesi y Cavaliere (2016).

y Giulia Cavaliere (Camporesi y Cavaliere, 2016) han hecho un excelente trabajo analizando tres ámbitos de la aplicación de CRISPR por fuera del ámbito embrionario, pero que igualmente genera grandes controversias:

- La utilización de CRISPR para erradicar vectores de enfermedad.
- Para la creación de modelos animales humanizados destinados al trasplante de órganos (los genes humanos pueden ser insertados en animales no humanos en la etapa de blastocisto, siendo los cerdos los candidatos más prometedores).
- Aplicaciones destinadas a la ingeniería genética de los cultivos.

La utilización de CRISPR para erradicar vectores de enfermedad ha despertado los peligros del doble uso de la tecnología (*dual-use dilemma*). Este aspecto presenta implicaciones éticas que deben ser consideradas. Un ejemplo de este tipo ha sido la modificación del *Aedes aegypti*, mosquito portador del virus del dengue y del Zika. Cuando el mosquito modificado se encuentra con uno salvaje, la modificación introducida por CRISPR hace que la descendencia que se produzca sea sólo de mosquitos machos (que son los que no pican). El mosquito modificado genéticamente fue liberado en el ecosistema de Piracicaba, Brasil, en 2015, con resultados que indican una reducción larval de 82% para fines de ese año¹².

Uno de los reparos más grandes frente al avance de CRISPR ha sido la eliminación de la variabilidad genética. No sabemos cómo puede resultar la alternación o eliminación de ciertos rasgos, que en un futuro, pueden resultar ventajosos en un contexto diferente, o bien ventajosos para la especie en general aunque no para el individuo en particular.

Por otro lado, si pensamos en la aplicación de CRISPR tendiente a resolver el problema de la escasez de órganos, los dilemas éticos vuelven a surgir en relación a los criterios de distribución para la asignación de un recurso escaso. Aunque el debate que rodea los *xeno-trasplantes* ha sido re-abierto a partir de CRISPR, es necesario considerar cuidadosamente todas las alternativas, por ejemplo la efectividad de las campañas públicas que fomentan la donación entre humanos o las políticas públicas tendientes

a diseñar programas de donación que alcancen a la población en general.

Difusión mediática y compromiso social

El último punto que analizaremos tiene que ver con la responsabilidad de los medios de información. Muchos estudios han indagado las diferentes reacciones del público en general sobre las bondades de los descubrimientos científicos. Otros han hecho notar que también es necesario indagar las causas que subyacen a la resistencia que las sociedades presentan frente a los nuevos retos tecnológicos. Hecho que no debe ser subestimado bajo la presunción de que el público necesita más información para ser parte activa en el proceso de la toma de las decisiones. Por un lado, es necesario considerar un manejo responsable en torno a las expectativas de curación que podrían presentar estos descubrimientos. Por el otro, conviene analizar la narrativa propia de la difusión de este tipo de noticias en los medios.

¿Cómo transmitimos al público información científicamente precisa y éticamente comprometida? ¿Cómo hablamos de las tecnologías prometedoras pero que despiertan gran incertidumbre?

En el artículo "*Editing genes: A case study about how language matters in bioethics*" (O'Keefe *et al.*, 2015), los autores analizan las metáforas que se han utilizado para transmitir la complejidad de las cuestiones éticas asociadas con el uso de CRISPR al público en general. Les interesaba analizar si estas metáforas han logrado capturar la complejidad del fenómeno o si, más bien, han "manipulado" al público restando importancia al debate en cuestión. Los medios, como grandes formadores de opinión y de representación en el nivel social, tienen un impacto alto en el modo en que las personas subjetivamos los adelantos tecnológicos. El modo en que los investigadores encontraron para analizar dicho impacto ha sido a través del análisis del uso de metáforas asociadas con CRISPR publicadas en los principales periódicos estadounidenses entre 2013 y 2015.

El uso metafórico, que tiene la habilidad de operar en ambos niveles, consciente e inconsciente, además de tener una gran influencia sobre la forma en la que las personas procesan información compleja, determinan el modo de respuesta a los "discursos establecidos".

"Las metáforas son particularmente importantes a medida que se introducen las nuevas tecnologías al público en general, ya que no sólo representan conceptos y técnicas científicas, sino que también configuran los marcos normativos y éticos utilizados para evaluar el uso

¹² Oxitec. Expansion of Oxitec's vector control solution in Brazil attacking source of Zika virus and dengue fever after positive program results. Informe de prensa: www.oxitec.com/oxitec-vector-control-solution Consultado online el 13 de noviembre de 2017.

apropiado de las mismas” (O’Keefe *et al.*, 2015: 5).

Así, luego de una cuidadosa exploración a nivel lingüístico, retórico y psicológico realizada por un equipo multidisciplinario, se extrajeron algunas conclusiones acerca de cómo el uso de las metáforas impacta en las consideraciones bioéticas del público en general.

Los investigadores concluyen que las metáforas que más se han utilizado para presentar estos desarrollos tecnológicos han sido las del genoma como un “texto”, y la intervención de CRISPR como una “edición”. Pensarlo en estos términos ha minimizado las implicancias éticas de un modo que es, para los autores, preocupante. La “edición” no suele estar asociada a un riesgo, ni transmite una necesidad de precaución. Los editores trabajan en la “mejora” del texto, corrigen, sugieren, cambian, pero siempre para mejorar la “lectura”. La edición del genoma humano no tiene nada que ver con esto (O’Keefe *et al.*, 2015).

Equipos interdisciplinarios de investigación podrían ser la clave para despejar las implicancias éticas y bioéticas y los retos que presentan el uso de estas tecnologías de cara al futuro.

BIBLIOGRAFÍA

- Camporesi S., Cavaliere G. (2016) Emerging ethical perspectives in the clustered regularly interspaced short palindromic repeats genome-editing debate. *Personalized Medicine* 13 (6): 575-586.
- Capps B., Chadwick R., Joly Y., Mulvihill J.J., Lysaght T., Zwart H. (2017) Falling giants and the rise of gene editing: ethics, private interests and the public good. *Human Genomics* 11 (1): 20.
- DeGrazia D. (2016) Ethical reflections on genetic enhancement with the aim of enlarging altruism. *Health Care Analysis* 24 (3): 180-195.
- HFEA, (1990) Human Fertilisation and Embryology Act. Her Majesty’s Stationary Office, London, UK.
- Lanphier E., Urnov F. (2015) Don’t edit the human germ line. *Nature* 519 (7544): 410.
- Ledford H. (2015) CRISPR, the disruptor. *Nature* 522: 20-4.
- Mulvihill J.J., Capps B., Joly Y., Lysaght T., Zwart H.A., Chadwick R., International Human Genome Organization (HUGO), Committee of Ethics, Law, and Society (CELS) (2017) Ethical issues of CRISPR technology and gene editing through the lens of solidarity. *British Medical Bulletin* 122 (1): 17-29.
- O’Keefe M., Perrault S., Halpern J., Ikemoto L., Yarborough M., UC North Bioethics Collaboratory for Life & Health Sciences (2015) “Editing” genes: A case study about how language matters in bioethics. *The American Journal of Bioethics* 15 (12): 3-10.
- Ormond K.E., Mortlock D.P., Scholes D.T., Bombard Y., Brody L.C., Faucett W.A. & Musunuru K. (2017). ASHG Position Statement. Human germline genome editing. *The American Journal of Human Genetics* 101(2): 167-176. Doi: <http://dx.doi.org/10.1016/j.ajhg.2017.06.012>
- Sugarman, J. (2015) Ethics and germline gene editing. *EMBO reports* (2015). <https://doi.org/10.15252/embr.201540879>
- UNESCO (2005). Declaración Universal de Bioética y Derechos Humanos. : http://portal.unesco.org/es/ev.php-URL_ID=31058&URL_DO=DO_TOPIC&URL_SECTION=201.html (Consultado el 13 de Noviembre de 2017)
- UNESCO (2015) International Bioethics Committee. Report of the IBC on Updating its reflection on the human genome and human rights. Paris, 2 October 2015.
- Vassena R., Heindryckx B., Peco R., Pennings G., Raya A., Sermon K., Veiga A. (2016) Genome engineering through CRISPR/Cas9 technology in the human germline and pluripotent stem cells. *Human reproduction update* 22 (4): 411-419.



CLINICAL AND CYTOGENETIC CHARACTERIZATION OF A PATIENT WITH TETRASOMY 18P

CARACTERIZACIÓN CLÍNICA Y CITOGENÉTICA DE UN PACIENTE CON TETRASOMÍA 18P

Vazquez Cantu D.L.^{1,2}, Gutiérrez García V.M.², Cruz Camino H.^{1,4}, Lara Díaz V.J.², Garza García K.², Garduño L.M.³, Meléndez R.³, Paz A.⁴, Mayén D.G.⁴, Cantú Reyna C.^{1,2,*}

¹Genomi-k S.A.P.I. de C.V., Monterrey, Nuevo León, México.

²Escuela de Medicina Tecnológico de Monterrey, Monterrey, Nuevo León, México.

³Unidad de Genética, Hospital Ángeles Lomas, Estado de México, México.

⁴Escuela de Biotecnología y Ciencias de la Salud, Tecnológico de Monterrey, Monterrey, México.

* Corresponding Author: cocantu@genomi-k.com

ABSTRACT

The 18p tetrasomy is a structural chromosomal abnormality with the presence of an extra isochromosome 18p, caused by a nondisjunction failure during maternal meiosis II. This additional i(18p) occurs in 1 of 180,000 live-born children worldwide, and affects males and females equally. It is characterized by craniofacial dysmorphisms; ears, nose and throat (ENT) abnormalities; musculoskeletal alterations; and global development delay. We aim to present the clinical and cytogenetic findings of a 3-year-10-month-old Latin American male with i(18p), to support the gene dosage effects, comparing his features with the ones reported in literature. This patient was product of the second pregnancy of a 39-year-old woman and the first son of a 49-year-old man. His main clinical features were microcephaly, facial dysmorphism, generalized hypotonia, and developmental delay. A blood sample of the patient was required to perform a GTG-banded karyotype and a fluorescence *in situ* hybridization (FISH) for chromosome 18 short arm. In addition, an SNP microarray analysis was carried out to detect genomic imbalances. Cytogenetic analysis revealed the presence of a metacentric supernumerary marker chromosome. The FISH study confirmed the origin of the marker chromosome by showing two signals for the 18p subtelomere and an intermediate signal for the 18 centromere. The microarray analysis showed a copy number gain of 18,385 Mb within the 18p. Tetrasomy tends to be a result of *de novo* events. The presence of the patient's isochromosome could be explained by advanced maternal age as it is known that this factor has high influence in isochromosome formation. Despite that there were no genes associated with the i(18p)'s clinical manifestations, these features are negatively correlated with dosage effects of the entire short arm. Physical and language therapy was recommended to the patient; the family received medical orientation, and awareness in family planning was raised.

Key words: tetrasomy 18p, chromosome 18, isochromosome, cytogenetic analysis, case report

RESUMEN

La tetrasomía 18p es una anomalía cromosómica estructural con la presencia de un isocromosoma extra 18p, causado por una no disyunción durante la meiosis materna II. Este adicional i(18p) ocurre en 1 de 180.000 niños nacidos vivos en todo el mundo, y afecta a hombres y mujeres por igual. Se caracteriza por dismorfismos craneofaciales; anomalías en oídos, nariz y garganta (ENT); alteraciones musculoesqueléticas y del desarrollo global. Nuestro objetivo es presentar los hallazgos clínicos y citogenéticos de un varón latinoamericano de 3 años y 10 meses de edad con i(18p), para respaldar los efectos de dosificación génica, comparando sus características con las reportadas en la literatura. Este paciente es producto del segundo embarazo de una mujer de 39 años y el primer hijo de un hombre de 49 años. Sus principales características clínicas fueron microcefalia, dismorfismo facial, hipotonía generalizada y retraso en el desarrollo. Se requirió una muestra de sangre del paciente para realizar un cariotipo con bandas GTG y una hibridación fluorescente *in situ* (FISH) para el análisis del brazo corto del cromosoma 18. Además, se llevó a cabo un análisis de microarreglos para detectar desequilibrios genómicos. El análisis citogenético reveló la presencia de un cromosoma supernumerario metacéntrico. Mientras que el estudio FISH confirma el origen del cromosoma marcador al mostrar dos señales para subtelómeros 18p y una señal intermedia para el centrómero 18. El análisis de microarreglos mostró una ganancia en el número de copias de 18,385 Mb dentro de las 18p. La tetrasomía tiende a ser el resultado de eventos *de novo*. El isocromosoma del paciente podría explicarse por la edad materna avanzada, ya que se sabe que tiene una gran influencia en la formación de isocromosomas. A pesar de que no hay genes asociados con las manifestaciones clínicas de i(18p), estas características están negativamente correlacionadas con los efectos de dosificación de todo el brazo corto. Se le recomendó terapia física y de lenguaje al paciente, la familia recibió orientación médica y se concientiza sobre la planificación familiar.

Palabras clave: tetrasomía 18p, cromosoma 18, isocromosoma, análisis citogenético, reporte de caso clínico

INTRODUCCIÓN

An 18p isochromosome results from a nondisjunction failure and a centromeric misdivision during meiosis II, first described in 1963 by Froland *et al.* When the isochromosome is supernumerary, the result is a tetrasomy. Nowadays, the incidence of the 18p tetrasomy has been reported in 1:180,000 live-born children worldwide, tending to be a result of *de novo* events (Dutra *et al.*, 2012; Plaiasu *et al.*, 2011). Based on the accumulated environmental and age-related effects on women's meiotic machinery, the increase proportion of nondisjunction in older population pregnancies can be explained (Dutra *et al.*, 2012; Froland *et al.*, 1963).

The structural arrangement on the chromosome, coupled with overexpression of genes results on craniofacial dysmorphism, characterized by ophthalmologic, high nasal bridge, long philtrum, micrognathia, prognathism, and highly arched palate. Hearing loss and recurrent otitis media are often presented, as well as orthopedic abnormalities and a global developmental delay (Callen *et al.*, 1990; Chen *et al.*, 2014; Zavala *et al.*, 2009; Wei *et al.*, 2015; Sebold *et al.*, 2010; Nusbaum *et al.*, 2015).

Despite that chromosome 18 has the lowest gene density of any human chromosome, it represents approximately 2.7% of the human genome. For a proper diagnosis, cytogenetic testing is needed, connoted by fluorescence *in situ* hybridization (FISH), multiplex ligation-dependent probe amplification (MLPA), and comparative genomic hybridization (CGH) (Wei *et al.*, 2015).

We aim to present the clinical and molecular cytogenetic findings in a 3-year-10-month-old Latin American male with i(18p), to support the genotype-phenotype dosage effects, comparing his clinical features with the ones reported in literature.

CASE REPORT

A 3-year-10-month-old male patient (Figure 1), product of the second pregnancy of a 39-year-old woman and the first son of a 49-year-old man -both healthy and non-consanguineous- was referred to genetic counseling. The obstetric antecedents showed an uncomplicated pregnancy, completed prenatal care, and an eutocic delivery; the patient's weight at birth was 2,750 g and 49 cm of length. There were no reports of repeated abortions

or malformations neither in the family history nor in close relatives.

The cephalic support was achieved at the age of 6 months, seating was achieved six months later, and the beginning of ambulation started at the age of 3 years. At the time of genetic counseling, the patient was unable to speak and could not control sphincters. He presented neonatal respiratory distress syndrome at birth, remaining in the intensive care unit for 15 days; and suffered from gastroesophageal reflux during the first 6 months of life.

At physical examination, microcephaly, narrow forehead with mild prominence in the area of the metopic suture, mild left palpebral ptosis, antimongoloid palpebral fissures, narrow and bulbous nose, long philtrum, thin lips, high palate, and macrognathia were identified. Also, he presented low implantation of auricular pavilions and short neck, multiple folds in hands and feet, thumb tending to adduction, and generalized hypotonia (Table 1, Figure 2 and 3). Ophthalmological and audiological assessment did not report an associated pathology. The electroencephalogram, echocardiogram, and renal ultrasound were reported normal.

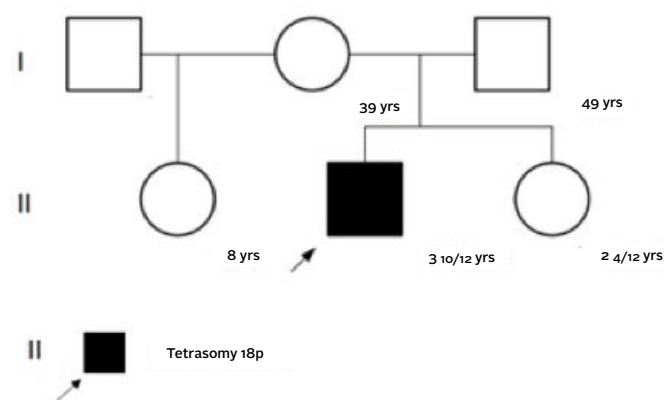


Figure 1. The family pedigree of a 3-year-10-month-old male (index case II, 2) from the second gestation of a 39-year-old woman and the first son of a 49-year-old man.

Table 1. Clinical features of patients reported with i(18p) compared with the index patient.

| Features | P1 ³ | P2 ³ | P3 ³ | P4 ³ | P5 ³ | P6 ³ | P7 ³ | P8 ³ | P9 ³ | Index P |
|--------------------------------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|---------|
| Age at review | 7 wk | 16 mo | 21 mo | 7 yr | 11 yr | 13 yr | 23 yr | 30 yr | 36 yr | 3 yr |
| Sex | F | F | M | F | F | F | F | F | F | M |
| Parental age in yrs (mother:father) | 36:22 | 31:34 | 37:43 | 34:30 | 29:28 | 26:26 | 35:35 | 39:41 | 36:46 | 38:48 |
| <i>Growth:</i> | | | | | | | | | | |
| Birth weight (g) | 2,500 | 2,250 | 2,580 | 2,860 | 3,260 | 2,130 | 2,355 | 2,950 | 3,630 | 2,750 |
| Height (per) | 3 | 10 | 3-10 | 3 | 3 | 10 | <3 | 10 | <3 | 25-50 |
| Head circumference (per) | 10 | <3 | 50 | <3 | 3 | <3 | <3 | <3 | 3 | NK |
| <i>Developmental delay</i> | MOD - SEV | MOD | MOD | SEV | MOD | MILD -SEV | MOD | MILD -MOD | SEV | MOD |
| <i>Neonatal Feeding difficulties</i> | + | + | + | - | NK | + | + | - | NK | + |
| <i>Ophthalmologic</i> | | | | | | | | | | |
| Strabismus | - | - | + | + | + | + | + | + | - | - |
| <i>ENT abnormalities</i> | | | | | | | | | | |
| Low-set ears | - | + | + | + | - | + | + | - | + | + |
| High Nasal Bridge | - | - | - | - | - | - | + | + | + | + |
| Micrognathia | + | + | + | + | - | + | - | - | - | + |
| Prognathism | - | - | - | + | - | + | + | + | + | + |
| High arched palate | + | - | - | - | + | - | - | + | - | + |
| Large philtrum | + | + | + | + | NK | + | - | - | - | + |
| <i>Neurological</i> | | | | | | | | | | |
| Seizures | + | - | - | + | + | + | + | - | - | - |
| Spasticity | + | + | + | + | + | + | + | + | + | - |
| <i>Orthopedic</i> | | | | | | | | | | |
| Camptodactyly | + | + | - | - | - | - | - | - | + | - |
| Scoliosis/kyphosis | - | - | - | - | - | + | + | - | + | + |
| Adducted thumbs | + | + | - | - | + | - | - | - | - | - |

³ P: patient, wk: weeks, mo: months, yr: years, F: female, M: male, g: grams, per: percentile, MOD: moderate, SEV: severe, NK: not known, +: present, -: not present.



Figure 2. a) Full-body image showing scoliosis and low-set ears. b) front view remarking high nasal bridge and large philtrum c) lateral view showing low-set, malformed ears, micrognathia and prognathism.



Figure 3. a) Front-view of patient's hands presenting camptodactyly and adducted thumbs, b) front-view of feet, showing adduction of toes.

METHODS

GTG-banded karyotype from peripheral blood cells cultured in RPMI-1640 supplemented and PB-Max™ GIBCO was performed. The 18 short arm FISH was done with orange spectrum (RP11-145B19) and the centromeric probe with green spectrum (CEP18). Subsequently, the patient's peripheral blood DNA was extracted to perform the SNP microarray analysis according to the CytoScan HD protocol of Affymetrix. Analyzed data, using the Chromosome Analysis Suite (ChAS) program, was reported and the informed consent was approved by the parents.

RESULTS

Cytogenetic analysis of the 25 blood cells, with a resolution of 550–600 bands, revealed the presence of a metacentric supernumerary marker chromosome. By its structure

and banding pattern, it corresponded to a short arm isochromosome in chromosome 18 (Figure 4). The FISH study corroborated the origin of the marker chromosome, by showing the two signals for the 18p subtelomere at orange spectrum, and an intermediate signal for the 18 centromere at green spectrum (Figure 5).

FISH was further analyzed in 100 interphase nuclei to verify the constitutive state of the marker chromosome; all had signals corresponding to the isochromosome. The microarray analysis showed a copy number gain of 18,385 Mb within the short arm of chromosome 18 (Figure 6). The data indicated the presence of four copies for each of the probes located between nucleotides 136,226 and 18,521,285 in the chromosomal bands of 18p11.32 to 18q11.1, which corresponded to more than 200 genes. These genomic abnormalities are identified as: 47,XY,+i(18)(p10)[25].ish;i(18)(p10)(RP11-145B19+,CEP18+,RP11-145B19+)[5].arr[hg19]; 18p11.32q11.1(136,226–18,521,285)x4.

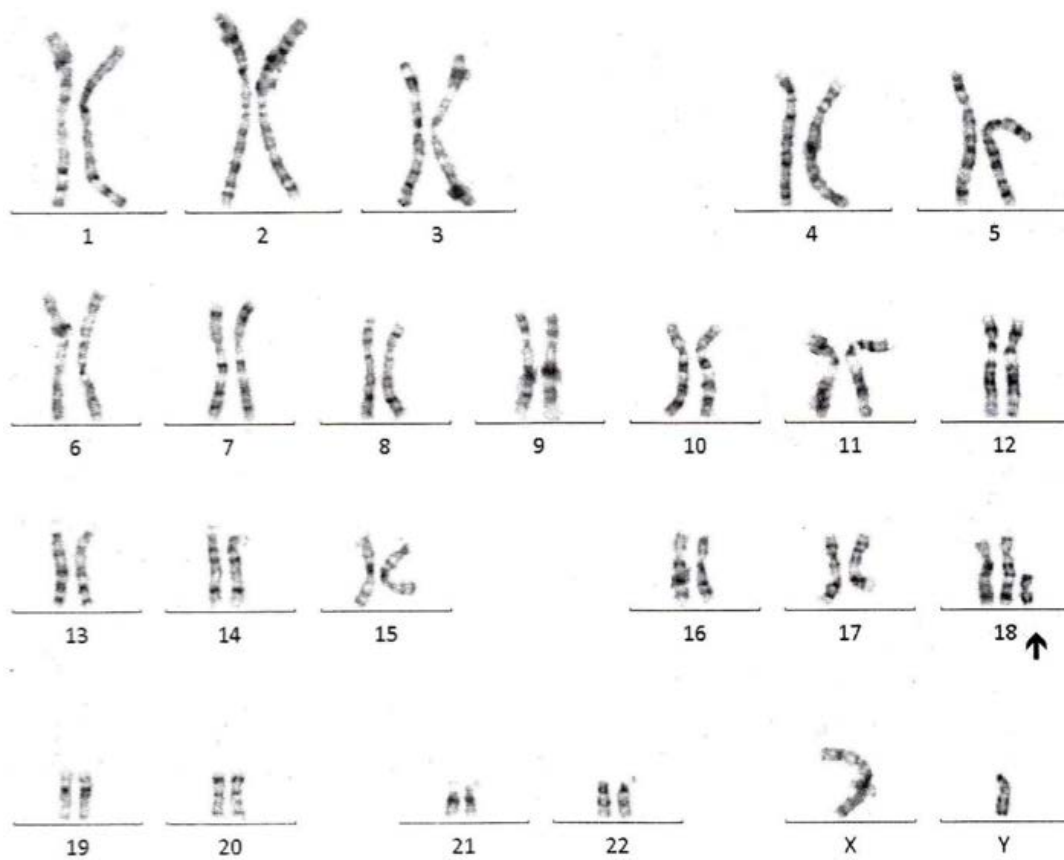


Figure 4. GTG-banded karyotype where the short arm isochromosome of chromosome 18, i(18)(p10) is shown.

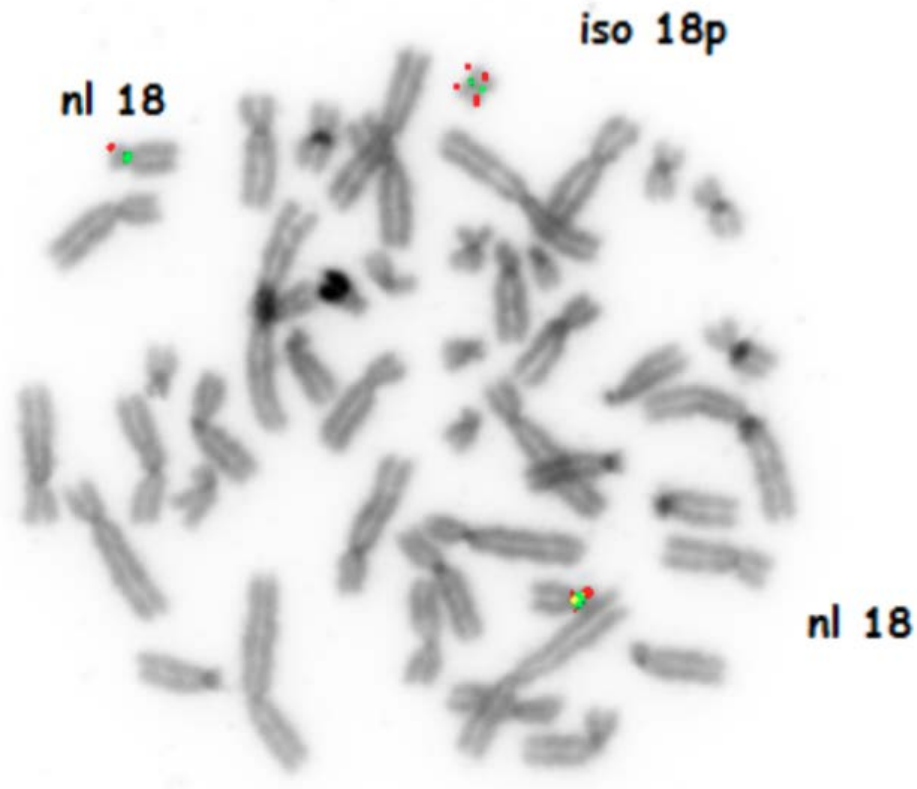


Figure 5. FISH over metaphase, isochromosome i(18p) shows two signals for short arm and one centromeric. Isochromosome 18p (iso 18p): shows centromere of 18 at green spectrum and subtelomeric region of 18p at orange spectrum. Normal chromosomes 18 (nl 18): show only a centromeric signal and a signal for short arm.

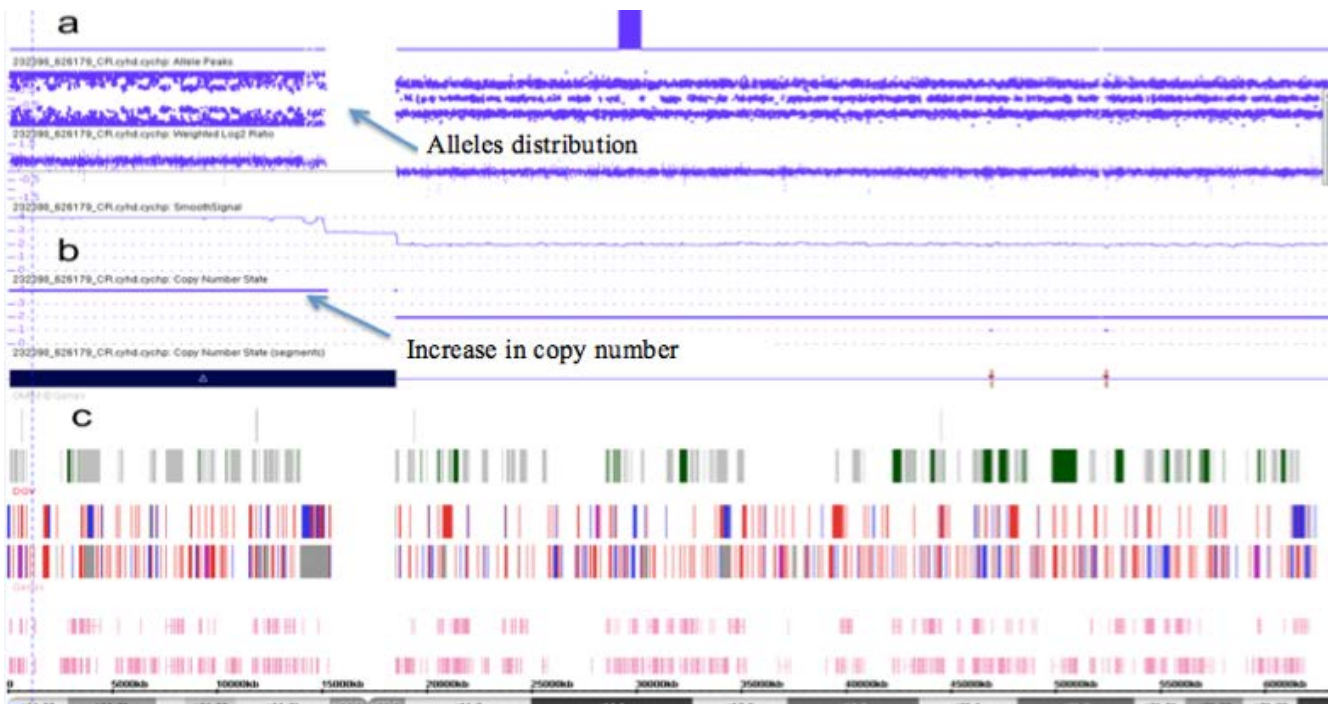


Figure 6. Analysis of the microarray in the ChAS software of chromosome 18, where the distribution of SNPs (a), increase in copy number (b), and genes and variants distributed throughout the chromosome (c) are observed.

METHODS

GTG-banded karyotype from peripheral blood cells cultured in RPMI-1640 supplemented and PB-Max™ GIBCO was performed. The 18 short arm FISH was done with orange spectrum (RP11-145B19) and the centromeric probe with green spectrum (CEP18). Subsequently, the patient's peripheral blood DNA was extracted to perform the SNP microarray analysis according to the CytoScan HD protocol of Affymetrix. Analyzed data, using the Chromosome Analysis Suite (ChAS) program, was reported and the informed consent was approved by the parents.

RESULTS

Cytogenetic analysis of the 25 blood cells, with a resolution of 550–600 bands, revealed the presence of a metacentric supernumerary marker chromosome. By its structure and banding pattern, it corresponded to a short arm isochromosome in chromosome 18 (Figure 4). The FISH study corroborated the origin of the marker chromosome, by showing the two signals for the 18p subtelomere at orange spectrum, and an intermediate signal for the 18 centromere at green spectrum (Figure 5).

FISH was further analyzed in 100 interphase nuclei to verify the constitutive state of the marker chromosome; all had signals corresponding to the isochromosome. The microarray analysis showed a copy number gain of 18,385 Mb within the short arm of chromosome 18 (Figure 6). The data indicated the presence of four copies for each of the probes located between nucleotides 136,226 and 18,521,285 in the chromosomal bands of 18p11.32 to 18q11.1, which corresponded to more than 200 genes. These genomic abnormalities are identified as: 47,XY,+i(18)(p10)[25].ish;i(18)(p10)(RP11-145B19+,CEP18+,RP11-145B19+)[5].arr[hg19]; 18p11.32q11.1(136,226-18,521,285)x4.

DISCUSSION

Chromosome 18 contains 76,117,153 bases, 243 known genes, and 45 loci implicated in genetic disorders, but there are not documented genes related to the 18p tetrasomy phenotype (Nusbaum *et al.*, 2005). Wei *et al.*, established

that aberrations of the whole short arm of chromosome 18 are negatively correlated with dosage effects in 18p⁻ (Wei *et al.*, 2015). In addition, several critical regions implicated in sensorial hearing loss and strabismus (1-1,192,031 region), as well as for scoliosis and kyphosis (1-2,931,532 region) can explain the clinical presentation.

The patient shares some of the more common clinical features in 18p tetrasomy, such as: high nasal bridge, micrognathia, prognathism, high arched palate, and low set/malformed ears. Also, feeding difficulties and musculoskeletal disorders were reported. Despite the lack? Unlike to the lack of sex predisposition, the literature reported a majority of female patients, whereas in this study, we present a male patient. In addition, the absence of spasticity and the normal height at birth differ from the reported in the literature (Table 1).

CONCLUSIONS

Advanced maternal age is known as a risk factor in the formation of an 18p tetrasomy 18. At the suspicion of this etiology, it is important to detect small supernumerary i(18p) by cytogenetic analysis, FISH, PCR, MLPA and, in some cases, array techniques. Parental testing is highly recommended.

Future studies? reports will allow to establish accurate data on its incidence and prognosis, raising awareness of its relevance to physicians. The latter will allow the patient to receive early and proper treatment, specifically for the ophthalmological and audiological features. Also, neurological, cardiac and gastrointestinal follow-ups would prevent the described clinical implications. It is important to give genetic counseling to the guardians.

BIBLIOGRAPHY

- Callen D.E, Freemantie C.J., Ringenbergs M.L., Baker E., Eyre H.J. (1990) The isochromosome 18p syndrome: Confirmation of cytogenetic diagnosis in nine cases by *in situ* hybridization. American Journal of Human Genetics 47, 3rd ser.: 493–498.
- Chen C., Lin C., Ko T., Chern S., Chen Y., Wu P., Wang W. (2014) Interphase FISH on uncultured amniocytes at repeat amniocentesis for rapid confirmation of low-

- level mosaicism for tetrasomy 18p. *Taiwanese Journal of Obstetrics and Gynecology* 53 (1): 126-128.
- Dutra A.N., Mancini T., Takeno S.S., Oliveira M.M., Kim C., Perez A.A., Melaragno M. (2012) Different Conformation of Two Supernumerary 18p Isochromosomes, One with a Concomitant Partial 18q Trisomy. *Cytogenetic and Genome Research Cytogenet Genome Res.* 138 (1): 1-4.
- Froland A., Holst G., Terslev E. (1963) Multiple anomalies associated with an extra small autosome. *Cylogenetics* 85: 99-106.
- Nusbaum C., Zody M.C., Borowsky M.L., Langer E.S. (2005) DNA sequence and analysis of human chromosome 18. *Nature* 438 (7068): 696.
- Plaiasu V., Ochiana D., Motei G., Georgescu A. (2011) A rare Chromosome Disorder- Isochromosome 18p syndrome. *MAEDICA a Journal of Clinical Medicine* 6 (2): 132-136.
- Zavala J., Ramirez M., Medina R., Heard P., Carter E., Crandall A., Escamilla M. (2009) Psychiatric syndromes in individuals with chromosome 18 abnormalities. *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics*, 9999B.
- Wei J., Xie Y., He W., Liu W., Jian W., Chen M., Sun X. (2015) Clinical Outcome: A Monosomy 18p is Better than a Tetrasomy 18p. *Cytogenetic and Genome Research Cytogenet Genome Res* 144 (4): 294-298.
- Sebold C., Roeder E., Zimmerman M., Soileau B., Heard P., Carter E., Cody J.D. (2010) Tetrasomy 18p: Report of the molecular and clinical findings of 43 individuals. *American Journal of Medical Genetics Part A*, 152A (9): 2164-2172.
- Takeda K., Okamura T., Hasegawa T. (1989) Sibs with tetrasomy 18p born to a mother with trisomy 18p. *Journal of Medical Genetics* 26 (3): 195-197. doi:10.1136/jmg.26.3.195.



LA ARQUITECTURA GENÉTICA COMO HERRAMIENTA DE ANÁLISIS DEL MAPA GENOTIPO-FENOTIPO

GENETIC ARCHITECTURE AS AN ANALYSIS TOOL OF THE GENOTYPE-PHENOTYPE MAP

Petino Zappala M.A.^{1*}, Fanara J.J.¹

¹Universidad de Buenos Aires, Facultad de Ciencias Exactas y Naturales, Departamento de Ecología, Genética y Evolución, Laboratorio de Evolución, CONICET-Universidad de Buenos Aires, Instituto de Ecología, Genética y Evolución de Buenos Aires (IEGEB), Buenos Aires, Argentina.

* Autor correspondiente: mapz@ege.fcen.uba.ar

ABSTRACT

Studying the relationship between genotype and phenotype is of great importance for genetics and life science studies in general. In contrast with the traditional view of this relationship as an invariant set of parameters, the current approach incorporates the concept of genetic architecture, a realistic and dynamic tool that allows to elucidate the genotype-phenotype map, which is now regarded as an evolving structure. From the complex relationships between the elements in the genotype-phenotype map several emergent properties arise that can explain different evolutionary phenomena. Moreover, some of these properties promote the accumulation of genetic variability in natural populations, which constitutes the substrate to evolutionary processes such as natural selection. The characterization and analysis of the genetic architecture of adaptive traits constitutes a powerful tool to understand the genetics underpinnings of evolution.

Key words: plasticity, canalization, modularity, cryptic genetic variability, evolvability

RESUMEN

El estudio de la relación entre el genotipo y el fenotipo es de gran importancia para las investigaciones en genética y en las ciencias de la vida en general. A diferencia de la concepción tradicional de esta relación como un conjunto invariante de parámetros, el enfoque actual utiliza la arquitectura genética, una herramienta realista y dinámica que permite elucidar el mapa genotipo-fenotipo, ahora considerado una estructura en evolución. De las complejas relaciones entre los elementos del mapa genotipo-fenotipo surgen diversas propiedades emergentes que pueden explicar distintos fenómenos evolutivos. Además, algunas de estas propiedades promueven la acumulación de variabilidad genética en poblaciones naturales, la cual constituye el sustrato de procesos evolutivos como la selección natural. La caracterización y análisis de la arquitectura genética de caracteres adaptativos constituye una herramienta eficaz para comprender los procesos genéticos subyacentes al cambio evolutivo.

Palabras clave: plasticidad, canalización, modularidad, variabilidad genética críptica, "evolvabilidad"

Fecha de recepción: 06/02/2017
Fecha de aceptación de versión final: 27/04/2018

INTRODUCCIÓN

La biodiversidad es el resultado de una larga historia de complejas interacciones entre las especies y su ambiente, a lo largo de la cual aquéllas han evolucionado desarrollando adaptaciones manifestadas a través del fenotipo, que les permiten vivir y reproducirse. Comprender las bases del surgimiento y mantenimiento de esta diversidad es uno de los desafíos con los que se enfrentan la genética, la ecología y la biología evolutiva (Badyaev, 2011; Futuyma, 2013).

Determinar las relaciones entre genotipo y fenotipo y cómo éstas se modifican a lo largo de la evolución constituye un objetivo fundamental para la genética en particular y para las ciencias de la vida (biología, biomedicina, veterinaria, etc.) en general. El enfoque tradicional apuntaba a un vínculo estático en el que el genotipo era un “plano” (*blueprint*) para el fenotipo, y en el que las asociaciones entre estos dos factores estaban restringidas para un tiempo y espacio dado. Un ejemplo de esto es, para el caso de la genética de poblaciones, la heredabilidad, un parámetro poblacional circunscripto a las condiciones en las cuales dicho parámetro fue estimado. Sin embargo en 1991 Alberch planteó el concepto de mapa genotipo-fenotipo (Alberch, 1991; Houle *et al.*, 2010), el cual incorpora niveles intermedios de organización de la información como el transcriptoma, proteoma y metaboloma, conectados mediante relaciones que se modifican en función del ambiente. A diferencia del modelo anterior, el mapa genotipo-fenotipo no es determinista, sino que se considera que el genotipo codifica un espectro de posibles fenotipos para cada carácter. Los distintos caracteres serán a su vez interdependientes y además tendrán una mayor o menor capacidad de permanecer inalterables frente a distintas perturbaciones, *i.e.* robustez (Kitano, 2004; Felix, 2015; DeWitt, 2016). Esta perspectiva de la relación entre la información genética (genotipo) y la manifestación de tal información (fenotipo) es capaz de explicar fenómenos relevantes en campos de creciente desarrollo experimental y conceptual como la biología evolutiva del desarrollo (evo-devo), la genómica ecológica, la genética cuantitativa, entre otros, que no podrían derivarse del enfoque tradicional.

ARQUITECTURA GENÉTICA

El concepto del mapa genotipo-fenotipo se usa como metáfora para comprender las múltiples formas en las

que la información genética influye en el fenotipo de un organismo (Pigliucci, 2010). Una forma de estudiar este sistema de relaciones genotipo-fenotipo (en cuanto a sus componentes y mecanismos de cambios) es a través del estudio de la arquitectura genética de caracteres complejos (Hansen, 2006). En este sentido, los procesos de adaptación, el surgimiento de variación a nivel poblacional y los mecanismos de especiación sólo pueden ser abordados cabalmente mediante la comprensión de la arquitectura genética de los caracteres adaptativos y estudiando su evolución (Badyaev, 2011).

Dilucidar la arquitectura genética de un carácter no sólo consiste en identificar los *loci* involucrados en los procesos subyacentes a la manifestación fenotípica de dicho carácter, sino también caracterizar sus propiedades variacionales, es decir, los efectos contexto-dependientes que contribuyen y/o modifican la expresión del carácter. Según Hansen (2006), la principal razón para estudiar la arquitectura genética es que permite comprender cómo las propiedades variacionales determinan el potencial evolutivo de los caracteres en cuestión.

Un análisis pormenorizado de la arquitectura genética requiere de la colaboración de distintas ramas de la biología, como la genética, la genómica, la bioinformática, la ecología y la biología evolutiva. Desde el punto de vista genético poblacional y evolutivo, descifrar la arquitectura genética de caracteres complejos implica determinar (Mackay, 2001; Boyle *et al.*, 2017): i) la proporción de los *loci* involucrados en el carácter que son variables; ii) las bases moleculares de la variación entre alelos (polimorfismos de nucleótidos individuales o SNPs [*single nucleotide polymorphisms*]); iii) la plasticidad fenotípica y los efectos contexto-dependientes de caracteres adaptativos; iv) las fuerzas evolutivas que han gobernado el cambio evolutivo; y v) si la arquitectura genética varía a nivel filogenético. Si bien esta es una tarea compleja y que requiere de mucho esfuerzo, promete en el trayecto, nuevas miradas sobre el estudio de la variación natural (Anholt *et al.*, 2003; Manolio *et al.*, 2009; Mackay y Moore, 2014).

CARACTERIZACIÓN DE LA ARQUITECTURA GENÉTICA

Base genética

La base genética de un carácter se refiere a todos los *loci* involucrados en los procesos que subyacen al establecimiento

de ese carácter y sus efectos en él. Muchos de los caracteres de relevancia para estos campos son continuos y su base genética está constituida por muchos *loci* de efecto pequeño (Flint y Mackay, 2009; Paaby y Gibson, 2016). Este hecho dificulta su estudio, ya que los análisis de mutagénesis tradicionalmente utilizados involucran variantes de gran efecto, que usualmente no se presentan en la naturaleza en condiciones en las que sería esperable que un gran cambio fenotípico resultase deletéreo. Una de las técnicas que constituyó un avance sustancial en estos estudios fue el mapeo por QTL (*Quantitative Trait Locus mapping*), que buscaba identificar regiones específicas del genoma que pudiesen explicar las diferencias fenotípicas entre líneas, poblaciones o especies (Mackay, 2001; Fanara *et al.*, 2002; Buckler *et al.*, 2009; Carreira *et al.*, 2016). Actualmente, los problemas de detección de variabilidad y su resolución, inherentes al protocolo de mapeo por QTL, se resuelven en gran medida con las técnicas de asociación del genoma completo (*Genome Wide Association Studies*, GWAS). Los protocolos metodológicos que se desarrollaron a partir del GWAS involucran una gran cantidad de líneas cuyo ADN está completamente secuenciado, proveyendo un mapa muy denso de sitios polimórficos que constituye la variabilidad genética (genómica) natural (Huang *et al.*, 2014; Korte y Farlow, 2013). Con estos elementos y considerando el gran poder de cómputo disponible en la actualidad, se pueden realizar análisis de asociación entre las variabilidades genética y fenotípica, obteniéndose como resultado *loci* candidatos responsables de la variación fenotípica detectada, los que posteriormente pueden ser corroborados mediante la utilización de otras técnicas (Morozova *et al.*, 2014).

Propiedades variacionales

Tanto una búsqueda (*screening*) de mutantes puntuales como la realización de un GWAS permiten llegar a una lista de variantes nucleotídicas asociadas a un cierto cambio fenotípico. Sin embargo, esta lista no aporta información acerca de los procesos que subyacen a dicho cambio de las relaciones entre las distintas variantes, ni acerca de los efectos que un cambio fenotípico pueda tener en otros caracteres, es decir, las propiedades variacionales de la arquitectura genética. Éstas involucran a lo que se denomina la arquitectura funcional, es decir la colección de caminos metabólicos (*pathways*) que conducen de la base genética al carácter mediante el mapa genotipo-fenotipo. Estas relaciones pueden caracterizarse por medio del estudio de

distintos fenómenos emergentes de la estructura del mapa, los cuales se describen a continuación.

Pleiotropía y modularidad

Una variante genética es pleiotrópica cuando afecta a la vez a más de un carácter. Las complejas y flexibles redes génicas que actúan a lo largo de distintos procesos del desarrollo que subyacen a varios caracteres explican la generalidad de la pleiotropía que puede afectar, paralelamente, a distintos estadios ontogenéticos (Flint y Mackay, 2009; Lavagnino *et al.*, 2008). El estudio de la interacción entre estos procesos (también conocido como integración fenotípica) permite una visión realista de los organismos estudiados y las dinámicas que siguen las poblaciones a lo largo de la evolución (Goswami *et al.*, 2014). Cheverud (1996) afirma que esta integración fenotípica estructura la pleiotropía, de forma que la selección favorecerá una arquitectura del desarrollo caracterizada por la respuesta coordinada de caracteres relacionados funcionalmente. Por otro lado, si dos caracteres presentan combinaciones que resultan en óptimos de *fitness*, puede favorecerse una integración entre ellos por medio de alguna forma de pleiotropía como consecuencia de la selección correlacional (Schluter y Nychka, 1994; Sinervo y Svensson, 2002).

Los procesos pleiotrópicos que ocurren dentro de los individuos también pueden producir soluciones de compromiso (*trade-offs*), debido a condicionamientos y limitaciones de los caracteres por interacciones negativas (Lailvaux y Husak, 2014). Éstos se generan cuando para un genotipo determinado un aumento del *fitness*, como consecuencia de un cambio en un carácter, se asocia a una disminución del *fitness* por un cambio concomitante en otro carácter. Según el modelo de adquisición y asignación de recursos (*acquisition-allocation model*) (van Noordwijk y de Jong, 1986), estos compromisos entre caracteres se deberían a la necesidad de “repartir” los recursos disponibles entre varias funciones vitales, lo cual hace imposible optimizar todos los caracteres a la vez.

Si bien el efecto de los *loci* pleiotrópicos puede determinar la existencia de correlaciones significativas entre caracteres, una falta de correlación no implica que no existan sectores del genoma afectándolos a la vez. El modelo de adquisición y asignación de recursos permite explicar las posibles faltas de correlación estadística entre caracteres, dado que tanto la adquisición como la asignación de recursos presentan variaciones dependiendo de factores genéticos, ambientales y de la relación entre ambos

(Glazier, 1999; Roff y Fairbairn, 2007). La multiplicidad de *loci* afectando a los caracteres (muchas veces a más de dos de ellos) y las interacciones entre ellos, complejizan aún más estas relaciones (Chippindale *et al.*, 2003; Roff y Fairbairn, 2007; Paaby y Rockman, 2014a). Por otro lado, los *loci* pleiotrópicos que tengan efectos distintos sobre dos caracteres no causarán necesariamente una correlación significativa entre ellos (Hansen, 2006). También debe considerarse que los efectos sobre los distintos caracteres pueden deberse a distintos polimorfismos dentro del mismo gen pleiotrópico, los que pueden no estar en desequilibrio de ligamiento (Flint y Mackay, 2009). Esto último constituye una forma de modularidad.

Un sistema es modular si se puede dividir en múltiples conjuntos de partes fuertemente interactivas, si bien las partes (módulos) son relativamente autónomas entre sí (Melo *et al.*, 2016). La modularidad impone un límite a la pleiotropía: mediante el surgimiento de módulos a distintos niveles (anatómicos, genéticos, ontogenéticos, etc.) se evita que las alteraciones en un módulo comprometan a otros, dado que la pleiotropía ocurre mayormente de forma intramodular (Wagner y Zhang, 2011). Este fenómeno conocido como pleiotropía modular (Figura 1), puede ser visto como un compromiso entre una integración completa e inflexible y caracteres completamente descoordinados e independientes (Goswami *et al.*, 2014).

Al favorecer o restringir ciertas combinaciones de caracteres, los efectos de la pleiotropía determinarán qué trayectorias evolutivas puede seguir una población o especie (Roff y Fairbairn, 2007). Contrariamente, la existencia de módulos relaja estos condicionamientos, promueve el mantenimiento de variabilidad genética y fomenta la divergencia fenotípica (Raff, 1996; Yang, 2001; Melo *et*

al., 2016). Además, la modularidad facilita la aparición de novedades evolutivas pues los módulos pueden ser cooptados para nuevas funciones (Wilkins, 2002; Parter *et al.*, 2008; Shubin *et al.*, 2009; Badyaev, 2011) mediante cambios en la expresión de los genes que constituyen el módulo (Stern, 2000).

Plasticidad Fenotípica e Interacción Genotipo-Ambiente

La necesidad de los organismos de responder a cambios ambientales es uno de los factores que promueven la evolución. Para un dado carácter, cuando distintos ambientes se asocian a óptimos fenotípicos diferentes, distintas estrategias serán seleccionadas en función de los desafíos que presupone la heterogeneidad ambiental. Una de ellas es la plasticidad fenotípica, es decir la capacidad de un genotipo de producir diferentes fenotipos en respuesta a la variación ambiental (Schlichting y Pigliucci 1998; van Gestel y Weissing 2016).

Esta plasticidad tiene base genética y además puede presentar variabilidad, lo que se denomina “interacción genotipo-ambiente” (Figura 2) (Callahan, 2005; Mensch *et al.*, 2008; GTExConsortium, 2015); esta variabilidad puede sustentar una eventual evolución en la plasticidad fenotípica (DeWitt y Scheiner, 2004).

La alteración del fenotipo en relación al ambiente favorece una mayor diversificación, teniendo el organismo la capacidad de explotar distintos nichos y soportar un mayor rango de condiciones ambientales (Chevin *et al.*, 2013; Forsman y Wennersten, 2016). En un organismo con una alta tasa migratoria, como por ejemplo la que se ha informado en *Drosophila melanogaster* (Lange *et al.*, 1990), la plasticidad fenotípica y su variación podrían formar parte de una estrategia que permite la supervivencia y

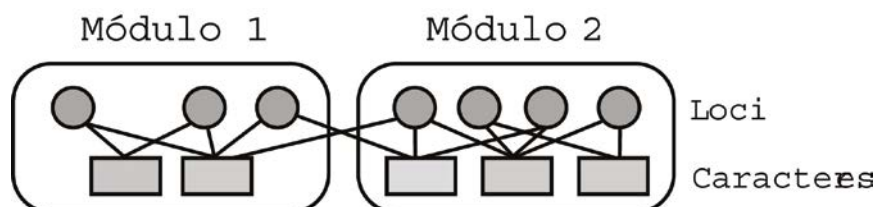


Figura 1. Esquema de pleiotropía modular. Los módulos representan grupos de caracteres co-afectados por grupos de *loci* intramodularmente pleiotrópicos. Las conexiones intermodulares son menos frecuentes que las intramodulares. Adaptado de Wagner y Zhang (2011).

adecuación en distintos ambientes sin la necesidad de cambios genéticos (Meyers y Bull, 2002; Bellard *et al.*, 2012). En los últimos años se ha puesto en relevancia la importancia de considerar la plasticidad fenotípica y la interacción genotipo-ambiente en estudios relacionados con la genética y la biología evolutiva: *e.g.* vinculados a la adaptación (Davidson *et al.*, 2011; Del Pino *et al.*, 2012; Richter *et al.*, 2012; Carreira *et al.*, 2013; Nunes *et al.*, 2014; DeWitt, 2016), resistencia a condiciones de estrés (Fallis *et al.*, 2014; Biddle *et al.*, 2016), capacidad invasora (Richards *et al.*, 2006; Keller y Taylor, 2008) así como en procesos de especiación (Pfennig *et al.*, 2010; Wennersten y Forsman, 2012; Levis y Pfennig, 2016) entre otros aspectos. Por otro lado, un alelo puede tener efectos en el *fitness* en algunos ambientes, y en otros ser “invisible” a la selección natural (lo que se conoce como “neutralidad condicional”), lo cual favorece la acumulación de variabilidad genética (Paaby y Gibson, 2016).

Epístasis, canalización y variabilidad genética críptica

El papel de la epístasis (Wright, 1931) en la arquitectura genética de los caracteres cuantitativos ha sido controvertido, si bien en los últimos años se incrementó notablemente la importancia que la epístasis detenta para explicar las interacciones moleculares no lineales que sustentan el mapa genotipo-fenotipo (Phyllips, 2008; Mackay, 2014). En este sentido, la generalización de las interacciones epistáticas llevó a cambiar el paradigma previo que consideraba a los genes como miembros de cascadas genéticas o *pathways* lineales por una idea de redes génicas susceptibles al ambiente, las cuales otorgan a la vez flexibilidad y robustez a la relación entre genotipo y fenotipo (Greenspan, 2001; Sambandan *et al.*, 2006).

Dados los efectos del acervo genético en la relación entre genotipo y fenotipo, es evidente que este factor puede tener a nivel poblacional consecuencias importantes desde el punto de vista genético y evolutivo. Ciertamente, podría condicionarse el destino de una mutación en función del acervo (*background*) genético poblacional y el contexto ambiental (Woods *et al.*, 2011; Chandler *et al.*, 2013; Miton y Tokuriki, 2016), lo cual aumentaría la posibilidad de incompatibilidades entre poblaciones divergentes (Mackay, 2014). Las eventuales interacciones podrían condicionar los fenotipos que se manifiestan y, por lo tanto, las trayectorias evolutivas futuras (Blount *et al.*, 2008; Chandler *et al.*, 2013).

Las interacciones epistáticas pueden ser seleccionadas

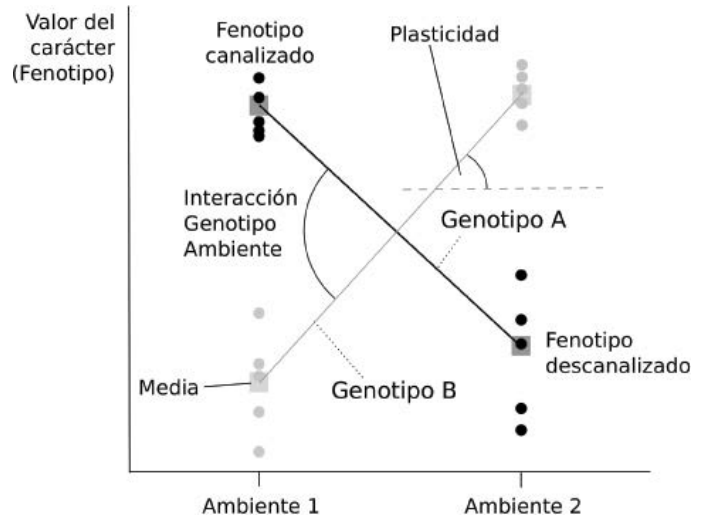


Figura 2. Esquema de los distintos aspectos del fenotipo considerados en este trabajo. La figura representa dos normas de reacción (línea oscura y línea clara) que indican el valor del carácter (fenotipo) frente a dos ambientes para los genotipos A y B. En el caso del genotipo A, la expresión del carácter (fenotipo) en el ambiente 2 muestra un efecto de descanalización, puesto que en este ambiente se observa un cambio tanto en la media fenotípica como en la variabilidad del carácter, que se representa mediante la mayor dispersión de puntos respecto al ambiente 1. Para el genotipo B se observa el patrón opuesto. Ambas normas de reacción son plásticas (plasticidad fenotípica) dado que los fenotipos que se expresan en los dos ambientes son diferentes. La falta de plasticidad está representada por la línea punteada para el genotipo B que revelaría que la expresión fenotípica es independiente del ambiente. Dado que se están analizando dos genotipos es posible establecer si hay interacción genotipo-ambiente mediante una prueba de paralelismo de las normas de reacción.

como un mecanismo de *buffer* ante condiciones ambientales relativamente estables, provocando un efecto de canalización del fenotipo (Waddington, 1942; Yamamoto *et al.*, 2009) alrededor de un óptimo adaptativo (Figura 2). Como resultado de estas interacciones, es posible determinar dos grupos de variantes resultantes: las que forman parte de la expresión del carácter y que están bajo proceso selectivo y otras cuya manifestación es nula o despreciable respecto a los parámetros adaptativos en un determinado tiempo y espacio. Este segundo tipo de variantes puede acumularse debido a que su efecto permanece oculto de la selección natural, generando una variabilidad genética críptica u oculta (Gibson y Wagner, 2000; Gibson y Dworkin, 2004; Paaby y Rockman, 2014b). Bajo diferentes condiciones genéticas o ambientales que causen modificaciones en la red de interacciones epistáticas, la variabilidad genética críptica podría expresarse y por lo tanto, estar sujeta a los efectos de la selección natural (Flatt, 2005; Gibson, 2009). Durante este proceso, se produce un efecto de descanalización, es decir, aumento de la varianza fenotípica y cambio en la media fenotípica como producto de la “liberación” de variabilidad genética que estaba oculta (Gibson y Wagner, 2000; Dworkin, 2005; Chandler *et al.*, 2013; Lavagnino y Fanara, 2016). El estudio del grado de canalización de distintos fenotipos puede realizarse a través del cálculo del Coeficiente de Variación Ambiental (CVA) que surge de la división entre la varianza del carácter y su media; al ser una variable estandarizada, pueden compararse incluso CVAs para distintos caracteres (Dworkin, 2005; Lavagnino y Fanara, 2016).

La plasticidad en el nivel de canalización podría surgir en combinación con distintos niveles de plasticidad de las medias como estrategia ecológica, especialmente si las funciones de *fitness* difieren entre ambientes (DeWitt y Scheiner, 2004). Podría pensarse que un caso extremo sería el de las especies invasoras, las cuales enfrentan un importante grado de incertidumbre en cuanto a las nuevas condiciones ambientales. La descanalización y/o la plasticidad en el nivel de canalización sería, en el caso de las especies invasoras, una estrategia que resultaría adaptativa (Carroll, 2008).

Los fenómenos de canalización y descanalización son un mecanismo que permitiría explicar la aparente paradoja entre la gran robustez que presentan los sistemas biológicos frente a las perturbaciones de distinto tipo y los numerosos ejemplos de radiaciones adaptativas abruptas con cambios morfológicos drásticos en tiempos evolutivos relativamente

cortos (Stanley, 1998; Futuyma, 2013). Estos fenómenos entraban en contradicción con los modelos clásicos, que postulaban una disminución de la variabilidad genética poblacional por selección estabilizadora interrumpidos por períodos de evolución direccional ante cambios de nicho (Paaby y Gibson, 2016). El modelo alternativo propone una sustancial acumulación de variabilidad genética críptica frente a un proceso de selección, generándose un fenotipo canalizado. Ante una modificación en el ambiente o drásticos cambios en el *background* genético (por ejemplo, por efectos de deriva) puede expresarse esta variabilidad genética escondida, conllevando una rápida aparición de novedades evolutivas (Pigliucci, 2008). Este modelo conecta los procesos macro y microevolutivos sin necesidad de incorporar hipótesis *ad-hoc*, respetando los principios y mecanismos propuestos por la genética de poblaciones y biología del desarrollo, entre otras áreas del conocimiento (Paaby y Gibson, 2016).

“Evolvabilidad” (del inglés, “*evolvability*”)

Se define como “evolvabilidad” a la capacidad de evolucionar. Hansen (2006) y Schlichting (2008) la relacionan con la habilidad del sistema genético para producir y mantener variantes genéticas que le permitirán seguir trayectorias evolutivas diversas. La “evolvabilidad” es, por lo tanto, una propiedad emergente del mapeo entre genotipo y fenotipo. El estudio de la arquitectura genética, es decir de todos los elementos genéticos que contribuyen o podrían contribuir en el futuro a la expresión fenotípica (Figura 3), es el único medio para abordar las preguntas relativas a este concepto.

Resulta necesario en este punto diferenciar los conceptos de “evolvabilidad” y heredabilidad, que refieren a aspectos distintos de la arquitectura genética de un carácter. La heredabilidad, ya sea en sentido amplio o reducido, refiere a la habilidad de una población de responder a la selección desde la variabilidad genética preexistente (Figura 3). La “evolvabilidad” expresa el potencial evolutivo de una dada arquitectura genética, involucrando tanto la variabilidad genética (visible, críptica y neutra) como las características que favorecen su generación y mantenimiento (Le Rouzic y Carlborg 2008). En consecuencia, todas las propiedades variacionales que mencionamos en las secciones anteriores condicionan la “evolvabilidad”. La base genética de un carácter, definida en principio por las regiones del genoma cuya variabilidad implica un efecto sobre el fenotipo, se organiza en redes de interacciones epistáticas con distinta

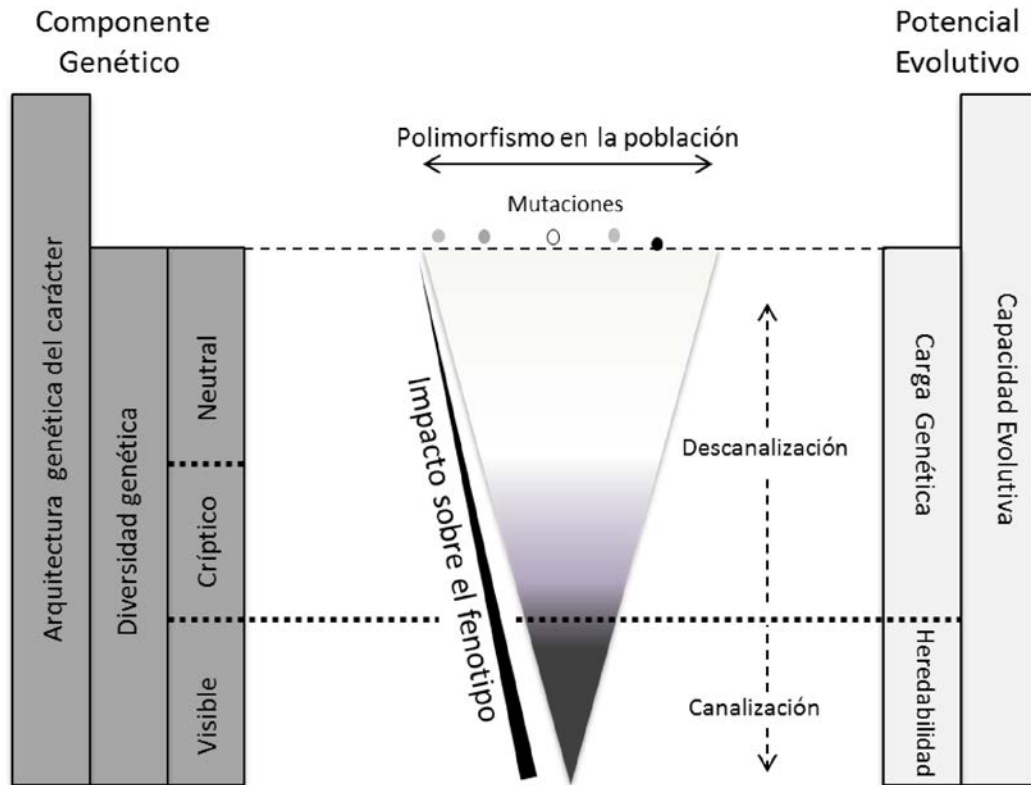


Figura 3. Representación de la arquitectura genética de un carácter. El triángulo representa la base genética que conforma la arquitectura genética de un carácter en una especie o población. Las variantes alélicas ingresan por mutación (parte superior del triángulo); las que presentan un efecto inmediato (beneficioso o perjudicial) sobre el fenotipo se colorean en negro, mientras que las que no tienen un efecto inmediato sobre el fenotipo se encuentran representadas en color claro. Las variantes con efectos beneficiosos deben ingresar y propagarse (debido a su baja frecuencia alélica) a través de la población antes de ser fijadas (vértice del triángulo) por selección natural. La mayor parte de la variabilidad genética no se traduce en variabilidad fenotípica y se considera “neutral”. A nivel fenotípico sólo una parte de la variación es “visible” a los efectos inmediatos de la selección natural, mientras que el resto (denominado “carga genética”) es “críptico” o “neutro”, aunque este límite no es estático. Por efecto de perturbaciones tanto ambientales como genéticas puede producirse un cambio de diversidad genética visible (que está bajo efecto de la selección natural) a críptica y/o neutra y viceversa. Modificado de Le Rouzic y Carlborg (2008).

modularidad que determinan el grado de canalización y plasticidad del fenotipo. Estas propiedades favorecen el mantenimiento de variabilidad genética (operando a distintos niveles de organización de la información genética) que luego será la materia prima del cambio evolutivo (Le Rouzic y Carlborg, 2008; Paaby y Gibson, 2016). De hecho, muchos trabajos demuestran que las respuestas evolutivas rápidas que pueden observarse en la

naturaleza dependen no sólo de esta variabilidad genética (en sentido amplio, ver Figura 3) (Barrett y Schluter, 2008; Hayden *et al.*, 2011; Rohner *et al.*, 2013) sino también de la capacidad de responder en forma plástica ante cambios ambientales (Davidson *et al.*, 2011; Lin *et al.*, 2016) y de los procesos de canalización y descanalización que pudieran acontecer (Flatt, 2005; Le Rouzic y Carlborg, 2008; Paaby y Gibson, 2016), involucrando alteraciones en la función

y/o expresión de genes a causa de factores epigenéticos, es decir, no relacionados a variaciones en la secuencia del ADN (Richards, 2006; Feinberg, 2007; Zhang *et al.*, 2013; Duncan *et al.*, 2014).

En resumen, para dilucidar la evolución de la “evolabilidad” necesariamente se debe caracterizar y comprender los mecanismos y procesos subyacentes a la evolución de la arquitectura genética de caracteres adaptativos (Pigliucci, 2008).

CONCLUSIONES

A diferencia del enfoque clásico de la genética de poblaciones que considera una arquitectura genética invariante, el estudio de las propiedades variacionales y su evolución permite definir relaciones más realistas entre genotipo y fenotipo. Este tipo de estructura puede explicar fenómenos complejos que son relevantes para comprender las dinámicas macro y microevolutivas, y entender los cambios en la capacidad de evolucionar (Hansen, 2006).

La concepción de la arquitectura genética descansa en las intrincadas estructuras que la subyacen y su estudio requiere considerar aspectos del fenotipo que aún son habitualmente relegados o simplificados. Dada la gran complejidad inherente a los fenómenos biológicos, siguen siendo usuales los estudios que involucran líneas mutagenizadas en *background* genéticos controlados o mutantes de gran efecto que no representan la situación esperable en la naturaleza (Rockman, 2012). Además, la mayoría de los trabajos se enfocan exclusivamente en el estudio de las medias fenotípicas, obviando información sobre otros aspectos del fenotipo, como su varianza o plasticidad (DeWitt y Scheiner, 2004; DeWitt, 2016). Estas metodologías no sólo limitan la posibilidad de inferencia, sino que eliminan la variabilidad que representa el objeto mismo de estudio (Gasch *et al.*, 2016). Sin embargo, en la actualidad es posible abordar estudios más realistas y complejos a partir del desarrollo de métodos de secuenciación de genomas completos y el aumento del poder de cómputo. Sólo considerando a la variabilidad como objeto de estudio y no como mero ruido estadístico se podrán caracterizar las complejas estructuras que moldean el mapa genotipo-fenotipo y su evolución para comprender cabalmente el surgimiento de la diversidad biológica y los procesos inherentes a los mecanismos de adaptación a los diferentes escenarios ecológicos.

BIBLIOGRAFÍA

- Alberch P. (1991) From genes to phenotype: dynamical systems and evolvability. *Genética* 84: 5-11.
- Anholt R.R.H., Dilda C.L., Chang S., Fanara J.J., Kulkarni N.H., Ganguly I., Rollmann S.M., Kamdar K.P., Mackay T.F.C. (2003) The genetic architecture of odor-guided behavior in *Drosophila*: epistasis and the transcriptome. *Nat. Genet.* 35: 180-184.
- Badyaev A.V. (2011) Origin of the fittest: link between emergent variation and evolutionary change as a critical question in evolutionary biology. *Proc. R. Soc. B* 278: 1921-1929.
- Barrett R.D., Schluter D. (2008) Adaptation from standing genetic variation. *Trends Ecol. Evol.* 23: 38-44.
- Bellard C., Bertelsmeier C., Leadley P., Thuiller W., Courchamp F. (2012) Impacts of climate change on the future of biodiversity. *Ecol. Lett.* 15: 365-377.
- Biddle A., Gammon L., Liang X., Costea D.E., Mackenzie I.C. (2016) Phenotypic plasticity determines cancer stem cell therapeutic resistance in oral squamous cell carcinoma. *EBioMedicine* 4: 138-145.
- Blount Z.D., Borland C.Z., Lensk R.E. (2008) Historical contingency and the evolution of a key innovation in an experimental population of *Escherichia coli*. *PNAS* 105: 7899-7906.
- Boyle E.A., Li Y.I., Pritchard J.K. (2017) An expanded view of complex traits: from polygenic to omnigenic. *Cell* 169: 1177-1186.
- Buckler E.S., Holland J.B., Bradbury P.J., Acharya C.B., Brown P.J., Browne C., Ersoz E., Flint-Garcia S., Garcia A., Glaubitz J.C., Goodman M.M., Harjes C., Guill K., Kroon D.E., Larsson S., Lepak N.K., Li H., Mitchell S.E., Pressoir G., Peiffer J.A., Rosas M.O., Rocheford T.R., Romay M.C., Romero S., Salvo S., Sanchez Villeda H., da Silva H.S., Sun Q., Tian F., Upadyayula N., Ware D., Yates H., Yu J., Zhang Z., Kresovich S., McMullen M.D. (2009) The genetic architecture of maize flowering time.

- Science 325 (5941): 714–718.
- Callahan H.S. (2005) Using artificial selection to understand plastic plant phenotypes. *Integr. Comp. Biol.* 45: 475–485.
- Carreira V.P., Imberti M., Mensch J., Fanara J.J. (2013) Gene-by-temperature interactions and candidate plasticity genes for morphological traits in *Drosophila melanogaster*. *Plos One*. 8: e70851.
- Carreira V.P., Mensch J., Hasson E., Fanara J.J. (2016) Natural genetic variation and candidate genes for morphological traits in *Drosophila melanogaster*. *Plos One* 11: e0160069.
- Carroll S.P. (2008) Facing change: forms and foundations of contemporary adaptation to biotic invasions. *Mol. Ecol.* 17: 361–372.
- Chandler C.H., Chari S., Dworkin I. (2013) Does your gene need a background check? How genetic background impacts the analysis of mutations, genes, and evolution. *Trends Genet.* 29: 358–366.
- Cheverud J.M. (1996) Developmental integration and the evolution of pleiotropy. *Integr. Comp. Biol.* 36: 44–50.
- Chevin L.M., Collins S., Lefèvre F. (2013) Phenotypic plasticity and evolutionary demographic responses to climate change: taking theory out to the field. *Funct. Ecol.* 27: 967–979.
- Chippindale A.K., Ngo A.L., Rose M.R. (2003) The devil in the details of life-history evolution: instability and reversal of genetic correlations during selection on *Drosophila* development. *J. Genet.* 82: 133–145.
- Davidson A.M., Jennions M., Nicotra A.B. (2011) Do invasive species show higher phenotypic plasticity than native species and, if so, is it adaptive? A meta-analysis. *Ecol. Lett.* 14: 419–43.
- Del Pino F., Salgado E., Godoy-Herrera R. (2012) Plasticity and genotype×environment interactions for locomotion in *Drosophila melanogaster* larvae. *Behav. Genet.* 42: 162–169.
- DeWitt T.J. (2016) Expanding the phenotypic plasticity paradigm to broader views of traits space and ecological function. *Curr. Zool.* 62: 463–473.
- DeWitt T.J., Scheiner S.M. (2004) Phenotypic plasticity: functional and conceptual approaches. Oxford University Press, New York, USA.
- Duncan E.J., Gluckman P.D., Dearden P.K. (2014) Epigenetics, plasticity, and evolution: How do we link epigenetic change to phenotype? *J. Exp. Zool. B. Mol. Dev. Evol.* 322 (4): 208–20.
- Dworkin I. (2005) A study of canalization and developmental stability in the sternopleural bristle system of *Drosophila melanogaster*. *Evolution* 59: 1500–1509.
- Fallis L.C., Fanara J.J., Morgan T.J. (2014) Developmental thermal plasticity among *Drosophila melanogaster* populations. *J. Evol. Biol.* 27: 557–564.
- Fanara J.J., Robinson K.O., Rollmann S.M., Anholt R.R.H., Mackay T.F.C. (2002) *Vanaso* is a candidate Quantitative trait gene for *Drosophila* olfactory behavior. *Genetics* 162: 1321–1328.
- Feinberg A.P. (2007) Phenotypic plasticity and the epigenetics of human disease. *Nature* 447: 433–440.
- Félix M.A. (2015) Pervasive robustness in biological systems. *Nat. Rev. Genet.* 16: 483–496.
- Flatt T. (2005) The evolutionary genetics of canalization. *Quart. Rev. Biol.* 80: 287–316.
- Flint J., Mackay T.F.C. (2009) Genetic architecture of quantitative traits in mice, flies, and humans. *Genome Res.* 19: 723–733.
- Forsman A., Wennersten L. (2016) Inter-individual variation promotes ecological success of populations and species: evidence from experimental and comparative studies. *Ecography* 39: 630–648.
- Futuyma D.J. (2013) *Evolution*. Sinauer Associates, Sunderland, MA, USA.
- Gasch A.P., Payseur B.A., Pool J.E. (2016) The power of natural variation for model organism biology. *Trends Genet.* 32: 147–154.

- Gibson G. (2009) Decanalization and the origin of complex disease. *Nat. Rev. Genet.* 10: 134-140.
- Gibson G., Dworkin I. (2004) Uncovering cryptic genetic variation. *Nat. Rev. Genet.* 5: 681-690.
- Gibson G., Wagner G. (2000) Canalization in evolutionary genetics: a stabilizing theory? *Bioessays* 22: 372-380.
- Glazier D.S. (1999) Trade-offs between reproductive and somatic (storage) investments in animals: A comparative test of the Van Noordwijk and De Jong model. *Evol. Ecol.* 13: 539-555.
- Goswami A., Smaers J.B., Soligo C., Polly P.D. (2014) The macro evolutionary consequences of phenotypic integration: from development to deep time. *Phil. Trans. R. Soc. B* 369: 20130254.
- Greenspan R.J. (2001) The flexible genome. *Nat. Rev. Genet.* 2: 383-387.
- GTEX Consortium (2015) The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science* 348: 648-660.
- Hansen T.F. (2006) The evolution of genetic architecture. *Annu. Rev. Ecol. Evol. Syst.* 37: 123-157.
- Hayden E.J., Ferrada E., Wagner A. (2011) Cryptic genetic variation promotes rapid evolutionary adaptation in an RNA enzyme. *Nature* 474: 92-95.
- Houle D., Govindaraju D.R., Omholt S. (2010) Phenomics: the next challenge. *Nat. Rev. Genet.* 11: 855-866.
- Huang W., Massouras A., Inoue Y., Peiffer J., Ràmia M., Tarone A.M., Turlapati L., Zichner T., Zhu D., Lyman R.F., Magwire M.M., Blankenburg K., Carbone M.A., Chang K., Ellis L.L., Fernandez S., Han Y., Highnam G., Hjelman C.E., Jack J.R., Javadi M., Jayaseelan J., Kalra D., Lee S., Lewis L., Munidasa M., Ongeri F., Patel S., Perales L., Perez A., Pu L., Rollmann S.M., Ruth R., Saada N., Warner C., Williams A., Wu Y.Q., Yamamoto A., Zhang Y., Zhu Y., Anholt R.R., Korbel J.O., Mittelman D., Muzny D.M., Gibbs R.A., Barbadilla A., Johnston J.S., Stone E.A., Richards S., Deplancke B., Mackay T.F.C. (2014) Natural variation in genome architecture among 205 *Drosophila melanogaster* Genetic Reference Panel lines. *Genome Res.* 24: 1193-1208.
- Keller S.R., Taylor D.R. (2008) History, chance and adaptation during biological invasion: separating stochastic phenotypic evolution from response to selection. *Ecol. Lett.* 11: 852-866.
- Kitano H. (2004) Biological robustness. *Nat. Rev. Genet.* 5: 826-837.
- Korte A., Farlow A. (2013) The advantages and limitations of trait analysis with GWAS: a review. *Plant Methods* 9: 29.
- Lailvaux S.P., Husak J.F. (2014) The life history of whole-organism performance. *Quart. Rev. Biol.* 89: 285-318.
- Lange B.W., Langley C.H., Stephan W. (1990) Molecular evolution of *Drosophila* metallothionein genes. *Genetics* 126: 921-932.
- Lavagnino N., Anholt R.R.H., Fanara J.J. (2008) Variation in genetic architecture of olfactory behaviour among wild-derived populations of *Drosophila melanogaster*. *J. Evol. Biol.* 21: 988-996.
- Lavagnino N., Fanara J.J. (2016) Changes across development influence visible and cryptic natural variation of *Drosophila melanogaster* olfactory response. *Evolutionary Biology* 43: 96-108.
- Le Rouzic A., Carlborg O. (2008) Evolutionary potential of hidden genetic variation. *Trends Ecol. Evol.* 23: 33-37.
- Levis N.A., Pfennig D.W. (2016) Evaluating “plasticity-first” evolution in nature: key criteria and empirical approaches. *Trends Ecol. Evol.* 31: 563-574.
- Lin Y., Chen Z.X., Oliver B., Harbison S.T. (2016) Microenvironmental gene expression plasticity among individual *Drosophila melanogaster*. *G3: Genes, Genomes, Genetics* 6: 4197-4210.
- Mackay T.F.C. (2001) The genetic architecture of quantitative traits. *Ann. Rev. Genet.* 35: 303-339.

- Mackay T.F.C. (2014) Epistasis and quantitative traits: using model organisms to study gene-gene interactions. *Nat. Rev. Genet.* 15: 22-33.
- Mackay T.F.C., Moore J.H. (2014) Why epistasis is important for tackling complex human disease genetics. *Genome Med.* 6: 42.
- Manolio T.A., Collins F.S., Cox N.J., Goldstein D.B., Hindorf L.A., Hunter D.J., McCarthy M.I., Ramos E.M., Cardon L.R., Chakravarti A., Cho J.H., Guttmacher A.E., Kong A., Kruglyak L., Mardis E., Rotimi C.N., Slatkin M., Valle D., Whittemore A.S., Boehnke M., Clark A.G., Eichler E.E., Gibson G., Haines J.L., Mackay T.F.C., McCarroll S.A., Visscher P.M. (2009) Finding the missing heritability of complex diseases. *Nature* 461: 747-753.
- Melo D., Porto A., Cheverud J.M., Marroig G. (2016) Modularity: genes, development, and evolution. *Annu. Rev. Ecol. Evol. Syst.* 47: 463-486.
- Mensch J., Lavagnino N., Carreira V.P., Massaldi A., Hasson E., Fanara J.J. (2008) Identifying candidate genes affecting developmental time in *Drosophila melanogaster*: pervasive pleiotropy and gene-by-environment interaction. *BMC Dev. Biol.* 8: 78.
- Meyers L.A., Bull J.J. (2002) Fighting change with change: adaptive variation in an uncertain world. *Trends Ecol. Evol.* 17: 551-557.
- Miton C.M., Tokuriki N. (2016) How mutational epistasis impairs predictability in protein evolution and design. *Protein Sci.* 25: 1260-1272.
- Morozova T.V., Mackay T.F.C., Anholt R.R.H. (2014) Genetics and genomics of alcohol sensitivity. *Mol. Genet. Genom.* 289: 253-269.
- Nunes A.L., Orizaola G., Laurila A., Rebelo R. (2014) Rapid evolution of constitutive and inducible defenses against an invasive predator. *Ecology* 95: 1520-1530.
- Paaby A., Gibson G. (2016) Cryptic Genetic Variation in Evolutionary Developmental Genetics. *Biology* 5: 28.
- Paaby A.B., Rockman M.V. (2014a) The many faces of pleiotropy. *Trends Genet.* 29: 66-73.
- Paaby A.B., Rockman M.V. (2014b) Cryptic genetic variation, evolution's hidden substrate. *Nat. Rev. Genet.* 15: 247-258.
- Parter M., Kashtan N., Alon U. (2008) Facilitated variation: How evolution learns from past environments to generalize to new environments. *PLoS Comput. Biol.* 4: e1000206.
- Pfennig D.W., Wund M.A., Snell-Rood E.C., Cruickshank T., Schlichting C.D., Moczek A.P. (2010) Phenotypic plasticity's impacts on diversification and speciation. *Trends Ecol. Evol.* 25: 459-467.
- Phillips P.C. (2008) Epistasis the essential role of gene interactions in the structure and evolution of genetic systems. *Nat. Rev. Genet.* 9: 855-867.
- Pigliucci M. (2008) Isevolvability evolvable? *Nat. Rev. Genet.* 9: 75-82.
- Pigliucci M. (2010) Genotype-phenotype mapping and the end of the genes as blueprint metaphor. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 365: 557-566.
- Raff R.A. (1996) *The shape of life: Genes, development, and the evolution of animal form.* University of Chicago, Chicago, IL, USA.
- Richards E.J. (2006) Inherited epigenetic variation-revisiting soft inheritance. *Nat. Rev. Genet.* 7: 395-401.
- Richards C.L., Bossdorf O., Muth N.Z., Gurevitch J., Pigliucci M. (2006) Jack of all trades, master of some? On the role of phenotypic plasticity in plant invasions. *Ecol. Lett.* 9: 981-993.
- Richter S., Kipfer T., Wohlgemuth T., Guerrero C.C., Ghazoul J., Moser B. (2012) Phenotypic plasticity facilitates resistance to climate change in a highly variable environment. *Oecologia* 169: 269-279.
- Rockman M.V. (2012) The QTN program and the alleles that matter for evolution: All that's gold does not glitter. *Evolution* 66: 1-17.

- Roff D.A., Fairbairn D.J. (2007) The evolution of trade-offs: where are we? *J. Evol. Biol.* 20: 433-447.
- Rohner N., Jarosz D.F., Kowalko J.E., Yoshizawa M., Jeffery W.R., Borowsky R.L., Lindquist S., Tabin C.J. (2013) Cryptic variation in morphological evolution: *HSP90* as a capacitor for loss of eyes in cavefish. *Science* 342: 1372-1375.
- Sambandan D., Yamamoto A., Fanara J.J., Mackay T.F.C., Anholt R.R.H. (2006) Dynamic genetic interactions determine odor-guided behavior in *Drosophila melanogaster*. *Genetics* 174: 1349-1363.
- Schlichting C.D. (2008) Hidden reaction norms, cryptic genetic variation, and evolvability. *Ann. N. Y. Acad. Sci.* 1133: 187-203.
- Schlichting C.D., Pigliucci M. (1998) *Phenotypic Evolution: a reaction norm perspective*. Sinauer Associates, Sunderland, MA, USA.
- Schluter D., Nychka D. (1994) Exploring fitness surfaces. *Am. Nat.* 143: 597-616.
- Shubin N., Tabin C., Carroll S. (2009) Deep homology and the origins of evolutionary novelty. *Nature* 457: 818-823.
- Sinervo B., Svensson E. (2002) Correlational selection and the evolution of genomic architecture. *Heredity* 89: 329-338.
- Stanley S.M. (1998) *Macroevolution: pattern and process*. Johns Hopkins University Press, Baltimore, MD, USA.
- Stern D.L. (2000) Perspective: evolutionary developmental biology and the problem of variation. *Evolution* 54: 1079-1091.
- van Gestel J., Weissing F.J. (2016) Regulatory mechanisms link phenotypic plasticity to evolvability. *Sci. Rep.* 6: 24524.
- van Noordwijk A.J., de Jong G. (1986) Acquisition and allocation of resources: Their influence on variation in Life History tactics. *Am. Nat.* 128: 137-142.
- Waddington C.H. (1942) Canalization of development and the inheritance of acquired characters. *Nature* 150: 563-565.
- Wagner G.P., Zhang J. (2011) The pleiotropic structure of the genotype-phenotype map: the evolvability of complex organisms. *Nat. Rev. Genet.* 12: 204-213.
- Wennersten L., Forsman A. (2012) Population-level consequences of polymorphism, plasticity and randomized phenotype switching: A review of predictions. *Biol. Rev.* 87: 756-767.
- Wilkins A.S. (2002) *The evolution of developmental pathways*. Sinauer Associates, Sunderland, MA, USA.
- Woods R.J., Barrick J.E., Cooper T.F., Shrestha U., Kauth M.R., Lenski R.E. (2011) Second-order selection for evolvability in a large *Escherichia coli* population. *Science* 331: 1433-1436.
- Wright S. (1931) *Evolution in mendelian populations*. *Genetics* 16: 97-159.
- Yamamoto A., Anholt R.R.H., Mackay T.F.C. (2009) Epistatic interactions attenuate mutations affecting startle behaviour in *Drosophila melanogaster*. *Genet. Res.* 91: 373-382.
- Yang A.S. (2001) Modularity, evolvability, and adaptive radiations: A comparison of the hemi- and holometabolous insects. *Evol. Dev.* 3: 59-72.
- Zhang Y.Y., Fischer M., Colot V., Bossdorf O. (2013) Epigenetic variation creates potential for evolution of plant phenotypic plasticity. *New Phytologist* 197: 314-322.

AGRADECIMIENTOS

Expresamos nuestro agradecimiento a los Lic. Victoria Ortiz, Ignacio Satorre y Nicolás Flaibani por la lectura y sugerencias de este artículo, así como a los editores del *Journal of Basic and Applied Genetics* y a dos revisores anónimos por sus correcciones. Agradecemos el financiamiento del FONCyT (proyecto PICT 2012-0640).



CONTROL DE FALSOS DESCUBRIMIENTOS EN MAPEO ASOCIATIVO CON POBLACIONES ESTRUCTURADAS

FALSE DISCOVERY RATE CONTROL IN ASSOCIATION MAPPING WITH GENETICALLY STRUCTURED POPULATIONS

Peña Malavera A.^{1,2}, Bruno C.^{1,2}, Balzarini M.^{1,2,*}

¹ Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET).

² Estadística y Biometría, Facultad de Ciencias Agropecuarias, Universidad Nacional de Córdoba, Av. Valparaíso s/n, Ciudad Universitaria, CP: 5000 (509) Córdoba, Argentina.

*Autor correspondiente: mbalzari@agro.unc.edu.ar

ABSTRACT

The association tests between molecular markers and phenotypic traits are crucial for the Quantitative Trait Loci (QTL) identification. Biotechnological advances increased the molecular marker information; consequently, the number of genotype-phenotype association tests required incremented too. The multiple statistical inferences (multiplicity) demand corrections of the p-values obtained for each comparison in order to keep limited the error rates for the family of association tests. However, classic statistical correction methods such as Bonferroni, False Discovery Rate (FDR) and the Effective Number of Independent Test (M_{eff}) were developed in the context of independent data. Wherever, when the population genetic structure is present, the data are no longer independent. In this paper, we propose a method of correction for multiplicity based on estimation of the effective number of tests from a model that adjust for the underlying correlation structure. We evaluate the performance of the proposed procedure in the estimation of p-values for a set of simulated QTL. The results suggest that the proposed method provides control of FDR and has more power than other methods for multiplicity correction used in association mapping.

Key words: multiplicity, association studies, effective number of hypothesis test, linear models.

RESUMEN

Las pruebas de asociación entre marcadores moleculares y variables fenotípicas son cruciales para la identificación de QTL (*Quantitative Trait Loci*). Los avances biotecnológicos incrementaron la disponibilidad de marcadores genéticos y consecuentemente el número de pruebas de la asociación fenotipo-genotipo. El incremento de pruebas de significancia estadística a realizar en simultaneo (multiplicidad) demanda correcciones de los valores-p obtenidos para cada prueba de hipótesis de manera de mantener acotada las tasas de error para la familia de pruebas de asociación. Las correcciones estadísticas clásicas para el problema de multiplicidad, como Bonferroni, el método de control de la tasa de falsos descubrimientos (FDR) y el número efectivo de pruebas (M_{eff}), son ampliamente usadas, pero fueron desarrolladas para datos independientes. Sin embargo, cuando las poblaciones de mapeo están genéticamente estructuradas los datos dejan de ser independientes. En este trabajo, proponemos un método de corrección por multiplicidad basado en estimación del número efectivo de pruebas desde un modelo que ajusta por la estructura de correlación subyacente. Se evalúa el desempeño del procedimiento propuesto a través del análisis de los valores-p obtenidos para un conjunto de QTL simulados. Los resultados sugieren que el método propuesto provee control de la tasa de falsos positivos y presenta mayor potencia que otros métodos de corrección por multiplicidad usados en mapeo asociativo.

Palabras clave: multiplicidad, estudios de asociación, número efectivo de pruebas de hipótesis, modelos lineales

Fecha de recepción: 20/02/2017
Fecha de aceptación de versión final: 12/04/2018

INTRODUCCIÓN

El análisis conjunto de la información de marcadores moleculares del genoma e información fenotípica permite inferir sobre la existencia de asociaciones entre *loci* de marcadores y expresiones de caracteres cuantitativos de interés agronómico (Spindel *et al.*, 2015; Tadesse *et al.*, 2015; Yan *et al.*, 2011; Zhou *et al.*, 2016). En genética vegetal, la presencia de asociaciones estadísticamente significativas entre el estado del marcador y la variante fenotípica permite identificar los QTL subyacentes en la población de mapeo (Bressegheo y Sorrells, 2006; Parisseaux y Bernardo, 2004; Pers *et al.*, 2015). Sin embargo, el análisis de asociaciones bajo estructura genética poblacional (EGP) requiere de conceptos y métodos biológicos y estadísticos específicos orientados a disminuir los descubrimientos de falsos QTL, *i.e.* asociaciones que resultan significativas sólo por azar debido a las correlaciones que genera la estructuración genética de la población de mapeo (Malosetti *et al.*, 2007).

Además, los estudios sobre modelos estadísticos para mapeo asociativo (MA) se realizan con propuestas metodológicas que se encuentran aún en desarrollo, y no es una decisión trivial la elección del modelo de análisis más apropiado para un escenario particular (Bernardo, 2013; Cappa *et al.*, 2013; Gutiérrez *et al.*, 2015; Gutiérrez *et al.*, 2011; Locatelli *et al.*, 2013). La selección de uno u otro modelo debe contemplar aspectos estadísticos como el tamaño de la muestra y número de variables indicadoras o marcadores, y aspectos biológicos, entre ellos el nivel de divergencia genética entre subpoblaciones de la población de mapeo, cuando estas existen (Peña-Malavera, 2015).

Luego de ajustado un modelo de MA, será necesario realizar múltiples pruebas de hipótesis estadísticas sobre la asociación de cada uno de los marcadores con el carácter agronómico de interés. En el caso de los modelos de regresión usados en MA, en H_0 (hipótesis nula) se plantea que el coeficiente de regresión asociado al efecto del marcador sobre el fenotipo es nulo, *i.e.* el marcador no se encuentra ligado a un QTL. El segundo paso de la prueba de hipótesis se corresponde con la selección de un estadístico cuya distribución sea conocida cuando H_0 es cierta y que se desvíe de modo predecible de dicha distribución cuando H_0 no es cierta; el estadístico T de Student es apropiado para evaluar la significancia estadística de un coeficiente de regresión (Draper y Smith, 1998). Luego, es necesario calcular el valor del estadístico en la muestra que se tenga. Si el valor de dicho estadístico es

diferente de lo que se espera bajo H_0 , se rechazará H_0 . El nivel de significación empírico o valor-p asociado al valor observado del estadístico es la probabilidad de obtener en el muestreo (bajo H_0) valores tan o más raros que el obtenido. Este valor-p representa una medida del acuerdo (o desacuerdo) de la evidencia muestral con la hipótesis nula. Valores-p pequeños habrán de entenderse como evidencia en contra de la hipótesis nula objeto de contraste. En MA, valores-p pequeños llevan al rechazo de la hipótesis nula que establece que no existe ligamiento entre marcador y QTL, y por tanto sugieren la presencia de una variante genética informativa. Para juzgar si un valor-p es pequeño o no, éste se compara con un nivel de significación pre-especificado, α .

Dos criterios de evaluación cobran importancia para evaluar una prueba de hipótesis estadística: la capacidad de mantener su tamaño nominal o nivel de significación α , y la potencia de la prueba para detectar una hipótesis nula falsa. El primero está relacionado con el error tipo I, el cual tiene probabilidad de ocurrencia denotada por α y el segundo, con el error tipo II con probabilidad de ocurrencia denotada por β (Balzarini *et al.*, 2008). Estos errores pueden ser analizados mediante tasas de error por comparación que representan el valor esperado del cociente entre el número de inferencias erróneas y el número de inferencias realizadas o por experimento; estas últimas estiman la probabilidad de obtener al menos un error dentro de una familia de pruebas de hipótesis.

En estudios de MA, la hipótesis de interés es la hipótesis nula de falta de asociación marcador-fenotipo. Una prueba estadística con baja tasa de error tipo II es aquella con capacidad (o potencia) para detectar asociaciones verdaderas. La mayor potencia de un modelo de MA con respecto a otro no se asocia con un incremento de la tasa de error tipo I, *i.e.*, un incremento en la probabilidad de concluir que existe asociación cuando en realidad no está presente, sino con el tamaño de la población de mapeo y con la cantidad de marcadores o pruebas de hipótesis que se realizan sobre el mismo conjunto de datos.

En MA se ajustan modelos de regresión por cada marcador que se evalúa y por tanto hay múltiples hipótesis a contrastar sobre el mismo conjunto de datos. Este procedimiento debe realizarse siendo consciente de que algunas hipótesis serán objeto de rechazo sólo por azar, con una probabilidad mucho mayor que el nivel de significación nominal empleado para contrastar cada una de ellas. Para una prueba de hipótesis sobre uno de M coeficientes del modelo

de MA y bajo H_0 , hay probabilidad tan sólo α de que el estadístico T calculado exceda en valor absoluto del cuantil $\alpha/2$ de una distribución T de Student con $N-M$ grados de libertad. Pero la probabilidad de que algún estadístico T, desde una miriada de valores T (correspondientes a los M marcadores moleculares), exceda de $t_{\alpha/2, N-M}$, asumiendo independencia, es mayor con Prob (algún $\beta_i \neq 0$) = $1 - (1 - \alpha)^m$. Luego, con probabilidad mucho mayor a α , algún coeficiente de marcador molecular puede resultar significativo sólo por azar. Esta probabilidad depende de M, es decir, aumenta a medida que se incrementa el número de marcadores moleculares evaluados. Este problema de inferencia simultánea demanda, consecuentemente, de la corrección de los valores-p y debe ser atendido en el contexto de MA para no perder potencia (Xiao *et al.*, 2013). Para el contexto de datos independientes existen métodos de corrección de valores-p por multiplicidad que garantiza que la tasa de falsos positivos sea menor o igual que un valor pre-seleccionado. El método de control del error tipo I más conocido es la aproximación de Bonferroni (Bonferroni, 1935). Sin embargo, este método es excesivamente conservador cuando las pruebas de hipótesis son numerosas. Aplicado a estudios de MA puede reducir drásticamente la cantidad de marcadores positivos, incluso llegar a no detectar ninguna asociación significativa. Una corrección alternativa es la propuesta por Benjamini y Hochberg (Benjamini y Hochberg, 1995) para controlar la proporción esperada de hipótesis mal rechazadas respecto a todas aquellas rechazadas (Miller *et al.*, 2001; Sabatti *et al.*, 2003; Schwartzman *et al.*, 2008; Tusher *et al.*, 2001).

El umbral de significación nominal α es inapropiado para reportar resultados del mapeo asociativo, no sólo por la multiplicidad de pruebas de hipótesis que se realizan sino también por la correlación esperable entre las pruebas debido a la correlación entre marcadores moleculares (pruebas no independientes). En 2001, se propuso un ajuste de valores-p para pruebas correlacionadas que se basa en la determinación del número efectivo (M_{eff}) de pruebas independientes (Cheverud, 2001). Li y Ji (2005) propusieron una estimación más exacta del M_{eff} basada en la descomposición por valor singular de una matriz de correlaciones entre marcadores y diseñaron un procedimiento (LJ) basado en este nuevo M_{eff} para controlar el error tipo I. El método LJ ha sido usado con éxito en el contexto del análisis de QTL clásico donde, aunque los marcadores pueden estar correlacionados, los casos son independientes porque provienen de una población

de mapeo sin estructura de correlación genética entre los individuos.

En este trabajo se propone una corrección por multiplicidad que contempla la estructura genética de la población de mapeo cuando esta existe. La propuesta está basada en el número de pruebas efectivas o independientes (similar a LJ). La modificación propuesta utiliza los ejes derivados de la descomposición aplicada sobre la matriz de estadísticos de Mantel y Haenszel (MH) (1959) incorporando la información conocida de la estructura genética poblacional. El método propuesto fue comparado con otros métodos de corrección por multiplicidad usando datos simulados. Los modelos de MA ajustados previamente a la corrección por multiplicidad fueron tres, uno que no contempla ninguna corrección por estructura genética poblacional (EGP) (Modelo naive) y otros con distintos tipos de ajuste de la EGP. La comparación se realizó usando bases de datos de marcadores moleculares simulados bajo distintos escenarios biológicos de EGP. Los ajustes de valores-p se realizaron luego de escoger las mejores estrategias de modelación para MA para los escenarios simulados. Para cada combinación modelo de MA-método de ajuste de valor-p, se obtuvieron tasas de falsos positivos y potencia (Φ), bajo dos escenarios con diferente nivel de EGP (bajo y alto F_{ST}). El objetivo de este trabajo es evaluar el desempeño de diferentes métodos de corrección de valor-p por multiplicidad cuando ellos son aplicados luego de ajustar modelos de mapeo asociativo, que contemplan o no la EGP subyacente, bajo distintos escenarios biológicos en lo que concierne a tamaño de la población de mapeo, cantidad de marcadores moleculares y nivel de divergencia genética entre subpoblaciones de la población de mapeo.

MATERIALES Y MÉTODOS

Datos

Los datos de marcadores moleculares usados en este trabajo fueron simulados a través de QMSim (Sargolzaei y Schenkel, 2009) involucrando escenarios con cantidad de genotipos que imitan datos usuales en mejoramiento genético vegetal. Se simuló un genoma con 300 marcadores multilocus-bialélicos, con diseño de cruzamientos y selección aleatorios para una EGP conformada por cinco poblaciones. Se crearon cuatro escenarios biológicos correspondientes a, dos niveles de divergencia genética entre poblaciones (bajo y alto F_{ST}), y dos tamaños distintos de

poblaciones de mapeo ($n \approx 150$ y $n \approx 300$), equivalente a 30 y 60 líneas por población simulada. Los datos simulados fueron creados a partir de una población histórica con un tamaño poblacional de 200 individuos y el sistema de cruzamiento basado en la unión al azar de gametos (cruzamientos aleatorios). La coancestría promedio fue baja como sucede en numerosas poblaciones usadas para MA en vegetales. Variando el número de generaciones desde la población fundadora, se crearon diferentes niveles de divergencia genética poblacional. Los datos simulados fueron codificados como 0 y 1 para cada marcador. El promedio del estadístico F_{ST} (Wright, 1951) provisto por el análisis molecular de la varianza (AMOVA) (Excoffier *et al.*, 2009) fue usado para cuantificar el grado de diferenciación genética entre poblaciones en cada escenario (Tabla 1).

Dada la matriz de marcadores moleculares simulados se escogieron aleatoriamente 20 marcadores y con ellos se realizó una combinación lineal con efectos que siguen una distribución gamma con media 2 y varianza 5 [$\Gamma(2,5)$] para simular el efecto de los *loci* ligados a un QTL. Adicionalmente, se anexó a cada perfil molecular la realización de una variable aleatoria con distribución normal de media 100 (representa la media del carácter que depende del efecto poligénico de *background*) y varianza 25 (representa la variabilidad experimental, *i.e.* desvío estándar 5, no superior al 5% de la media del carácter fenotípico). A esta variable simulada se le adicionaron los efectos de los marcadores ligados extraídos de la distribución gamma. Los valores resultantes fueron usados como variable fenotípica para los modelos de MA. La ubicación de cada uno de los 20 QTL simulados sobre los marcadores seleccionados aleatoriamente fue usada para determinar el carácter de verdad de la hipótesis nula.

Modelo de Mapeo Asociativo

Se estimaron ocho modelos de mapeo asociativo para evaluar el efecto del marcador sobre el fenotipo (Tabla 2). El modelo básico a partir del cual derivaron los modelos de MA comparados fue:

$$y = X\mathbf{b} + EGP\mathbf{v} + Z\mathbf{u} + e$$

donde y es el vector de valores fenotípicos (conteniendo un dato fenotípico por genotipo), X es la matriz de datos de los marcadores moleculares (tantas columnas como marcadores usados), \mathbf{b} es un vector desconocido de efectos de los alelos de cada marcador que debe ser estimado para identificar aquellos marcadores asociados con el fenotipo,

EGP es la matriz de estructura genética (construida alternativamente como la matriz Q de la salida del software *structure* o la matriz P de componentes principales estadísticamente significativas seleccionadas por el estadístico de Tracy-Widom (1994), ambos realizados previamente sobre los datos moleculares), \mathbf{v} es el vector de efectos de la estructura poblacional (en algunas aproximaciones considerado como vector de efectos fijos y en otras como vector de efectos aleatorios), Z es la matriz de incidencia que conecta el vector aleatorio \mathbf{u} de efectos de poligen con los datos fenotípicos (matriz identidad de dimensión igual al número de genotipos que componen la población de mapeo) y e es un vector de términos de error aleatorio, que se supone normalmente distribuido con media cero y varianza constante σ_e^2 . Se supone que el vector \mathbf{u} se distribuye independientemente del vector e y con matriz de varianzas y covarianzas dada por $\sigma_e^2 \Sigma_K$, siendo K la matriz de similitud entre todos los pares de perfiles moleculares derivadas del software EMMA (Kang *et al.*, 2008) y que es usada como indicador del parentesco o la filogenética existente entre los genotipos de la población de mapeo.

Criterios de comparación

Todos los modelos fueron ajustados usando *Info-Gen* (Balzarini y Di Rienzo, 2004) y su interfaz con R (Team, 2013). El desempeño de los ocho modelos se evaluó usando las curvas de distribución acumulada de valores- p . Para construir las curvas de distribución de valores- p , se usó la opción función de distribución empírica del software *Info-Gen* (Balzarini y Di Rienzo, 2004) usando como variable de análisis el valor- p asociado a cada una de las pruebas de hipótesis realizadas en un escenario. En cada escenario hay tantas pruebas de hipótesis de asociación como marcadores. Es importante resaltar que en una distribución acumulada de valores- p se espera que, si la modelación ha sido buena, la distribución se aproxime a una línea recta de 45 grados, ya que la distribución de los valores- p debiera ser simétrica. Una distribución asimétrica hacia valores- p pequeños indica mayor significancia de la esperada, lo que sugiere un posible incremento de falsos positivos, es decir presencia de asociaciones espurias.

Luego de aplicarse las correcciones por multiplicidad en los modelos seleccionados, se usó como criterio de evaluación la tasa de falsos descubrimientos o FDR (del inglés, *False Discovery Rate*) (Benjamini y Hochberg, 1995)

Tabla 1. Tamaño poblacional y diversidad genética poblacional que caracteriza la estructura genética subyacente en poblaciones de mapeo simuladas con 300 marcadores multilocus-bialélicos como dato genómico.

| Escenario | Diversidad genética | | Tamaño poblacional promedio |
|-----------|----------------------|-------|-----------------------------|
| | Estadístico F_{ST} | Nivel | |
| I | 0.03 | Bajo | 150 |
| II | 0.03 | Bajo | 300 |
| III | 0.20 | Alto | 150 |
| IV | 0.21 | Alto | 300 |

Tabla 2. Ocho modelos de mapeo asociativo para evaluar el efecto del marcador sobre el fenotipo en datos simulados.

| Matriz de Parentesco | Estructura Genética Poblacional | | | | |
|----------------------|---------------------------------|--------|----------|------------|---------------|
| | No | Q fijo | ACP fijo | Qaleatorio | ACP aleatorio |
| No | Naive | Q | P | QA | PA |
| Si | K | QK | PK | -- | -- |

Nota: Q es la matriz de probabilidades de pertenencia a los g grupos calculada por el software structure, P es la matriz de componentes principales retenidas mediante el estadístico de Tracy-Widom(1994) y K es la matriz de parentesco propuesta por Kang *et al.*, (2008).

y la potencia estadística. La tasa FDR se calculó en base a las proporciones de falsos positivos (FP) y verdaderos positivos (VP). Los FP son todos aquellos valores- p significativos vinculados a marcadores que no están asociados al fenotipo (no ligados a un QTL) y los VP son todos aquellos marcadores positivos que efectivamente están asociados al fenotipo (ligados a un QTL), de esta forma tenemos,

$$FDR = \frac{FP}{VP + FP}$$

La potencia estadística en la detección de marcadores asociados con el fenotipo está referida a una medida de eficacia de los modelos y es la probabilidad de que la hipótesis nula (H_0) sea rechazada cuando esta es falsa o dicho de otra manera cuando la hipótesis alternativa (H_a) es verdadera. La potencia estadística (Φ) puede interpretarse como la probabilidad de no cometer error del tipo II (error que producen los eventos conocidos como falsos negativos, FN). La potencia fue calculada como,

$$j = \frac{VP}{VP + FN}$$

Usando los datos de los 4 escenarios simulados consideramos el problema de contrastar simultáneamente m hipótesis nulas $H_0^j, j = 1, \dots, m$, con $m = 300$. Si R es la cantidad de hipótesis rechazadas, los resultados posibles luego del contraste de hipótesis pueden resumirse como en la Tabla 3. Los conjuntos de subíndices que corresponden a hipótesis nulas verdaderas y falsas $\Omega_0 = \{j: H_0^j \text{ es verdadera}\}$ y $\Omega_1 = \{j: H_0^j \text{ no es verdadera}\}$ son desconocidos y serán estimados mediante la simulación. El conjunto total de índices es $\Omega = \{1, 2, \dots, m\} = \Omega_0 \cup \Omega_1$. Las cantidades de hipótesis nulas verdaderas $m_0 = \#\Omega_0$ y falsas $m_1 = m - m_0 = \#\Omega_1$, fueron estimadas por conteo dentro de cada escenario. En cada escenario simulado se estimó la cantidad de hipótesis nulas rechazadas R y no rechazadas $m-R$ (variables aleatorias observables a través del conjunto de prueba de hipótesis).

Tabla 3. Situaciones posibles luego de realizar m pruebas de hipótesis.

| Realidad | Decisión | | Total de hipótesis |
|--------------------------|----------------------------|-------------------------|--------------------|
| | No rechazar hipótesis nula | Rechazar hipótesis nula | |
| Hipótesis nula verdadera | VN | FP(Falsos Positivos) | m_0 |
| Hipótesis nula falsa | FN(Falsos negativos) | VP | m_1 |
| Total | m-R | R | m |

Procedimientos de corrección por multiplicidad

Para cada escenario se implementaron 3 métodos de corrección por multiplicidad y con fines comparativos, también se observaron los resultados luego de contrastar las m hipótesis sin corrección por multiplicidad. Los métodos implementados para corregir valores- p del conjunto de pruebas por multiplicidad fueron tres: (1) BH, propuesto por Benjamini y Hochberg (1995); (2) LJ propuesto por Li y Ji (2005); y (3) un nuevo procedimiento propuesto en este trabajo que llamamos Li y Ji Modificado (MLJ). El método de Bonferroni (1935) tradicionalmente usado en problemas de multiplicidad, no fue usado por ser altamente conservador en situaciones como las que se producen en MA donde el número de contrastes de hipótesis asciende a cientos e incluso miles de pruebas.

Para implementar la corrección propuesta por Benjamini y Hochberg (1995) se realizó el siguiente procedimiento:

1. Los valores- p de las m pruebas de hipótesis se ordenaron de menor a mayor.
2. El valor- p mayor no fue ajustado.
3. Cada uno de los restantes valores- p se multiplicó por el número total de marcadores y se dividió por el valor que denota su orden en la lista de valores- p ordenados. Si el valor resultante era menor que 0,05, se rechazaba la hipótesis nula.

Para implementar el método de corrección de Li y Ji (2005), basado en la idea propuesta por Cheverud (2001) para ajustar pruebas de hipótesis correlacionadas, se realizaron los siguientes pasos:

1. Se calculó la matriz de correlación para todos los *loci*.
2. Se calculó el número efectivo (M_{eff}) de pruebas

independientes a través de la obtención de los valores propios de la matriz de correlación, donde M es el número de pruebas y $l_i (i = 1, \dots, M)$ son los valores propios:

$$M_{eff} = \sum_{i=1}^M f(|l_i|)$$

$$f(x) = I(x \geq 1) + (x - \lfloor x \rfloor), x \geq 0$$

donde $I(x \geq 1)$ es una función indicadora que vale 1 cuando $x \geq 1$ y 0 en otro caso, y $\lfloor x \rfloor$ es la función parte entera que devuelve el mayor entero posible menor o igual a x .

3. Se ajustó el nivel de significación de la prueba como si hubiera M_{eff} pruebas independientes usando la corrección de Sidak (1967): $\alpha_p = 1 - (1 - \alpha_e)^{1/M_{eff}}$
4. Se realizaron las m pruebas de hipótesis locus por locus y cuando el valor- p de alguna prueba era menor que α_p , la hipótesis de no asociación fue rechazada.

El método de corrección por multiplicidad propuesto en este trabajo está basado en la aproximación de Li y Ji (2005) con una modificación que contempla la posible EGP que subyace la población de mapeo. En caso de poblaciones estructuradas, la modificación analizará la correlación entre marcadores, controlando por la presencia de los grupos que definen la EGP, para derivar un M_{eff} . Con este fin, la matriz de correlación utilizada en el método LJ es reemplazada por una matriz de estadísticos C^2 de Mantel y Haenszel (1959). Los estadísticos C^2 fueron obtenidos a partir de tablas de contingencia construidas entre pares de marcadores, fijando la variable que indica el grupo al cual pertenecen los genotipos como variable de control.

La evaluación del impacto de los métodos de corrección se realizó considerando ambos niveles de F_{ST} usados en la simulación, ambos tamaños poblacionales y distintos modelos de MA para generar la lista de valores-p sin corregir. Los modelos ajustados seleccionados para evaluar la corrección por multiplicidad fueron: QK y K (Yu *et al.*, 2006) y el modelo de mapeo de regresión de efectos fijos que incluye los 300 marcadores como variables independientes y no incorpora de ninguna manera explícita el modelo de la EGP (modelo *naive*).

RESULTADOS

Las funciones de distribución acumulada para los 4 escenarios que involucraron datos genéticos de 300 marcadores moleculares multilocus-bialélicos,

consistentemente mostraron que los modelos con mejor ajuste fueron el modelo K y el modelo QK (Figura 1). Con bajo F_{ST} (Figura 1 arriba, escenarios I y II) se observó que el modelo de menor desempeño fue el modelo P. Los modelos se comportaron de manera parecida cuando fueron ajustados en un contexto de alto F_{ST} (Figura 1 abajo, escenarios III y IV), situación en la que se observó menor diferencias entre los ajustes, principalmente para el caso de mayor estructuración relativa en la población, correspondiente a un valor de F_{ST} de 0,20 y a una población de 150 individuos.

Las tasas FDR fueron menores para alto F_{ST} en los tres métodos de corrección por multiplicidad, *i.e.*, para los dos modelos de mapeo asociativo seleccionados por tener el mejor desempeño (modelo K y modelo QK) y para el modelo *naive*, seleccionado como modelo de referencia; tanto en poblaciones de tamaño 150 como de 300 se pudo

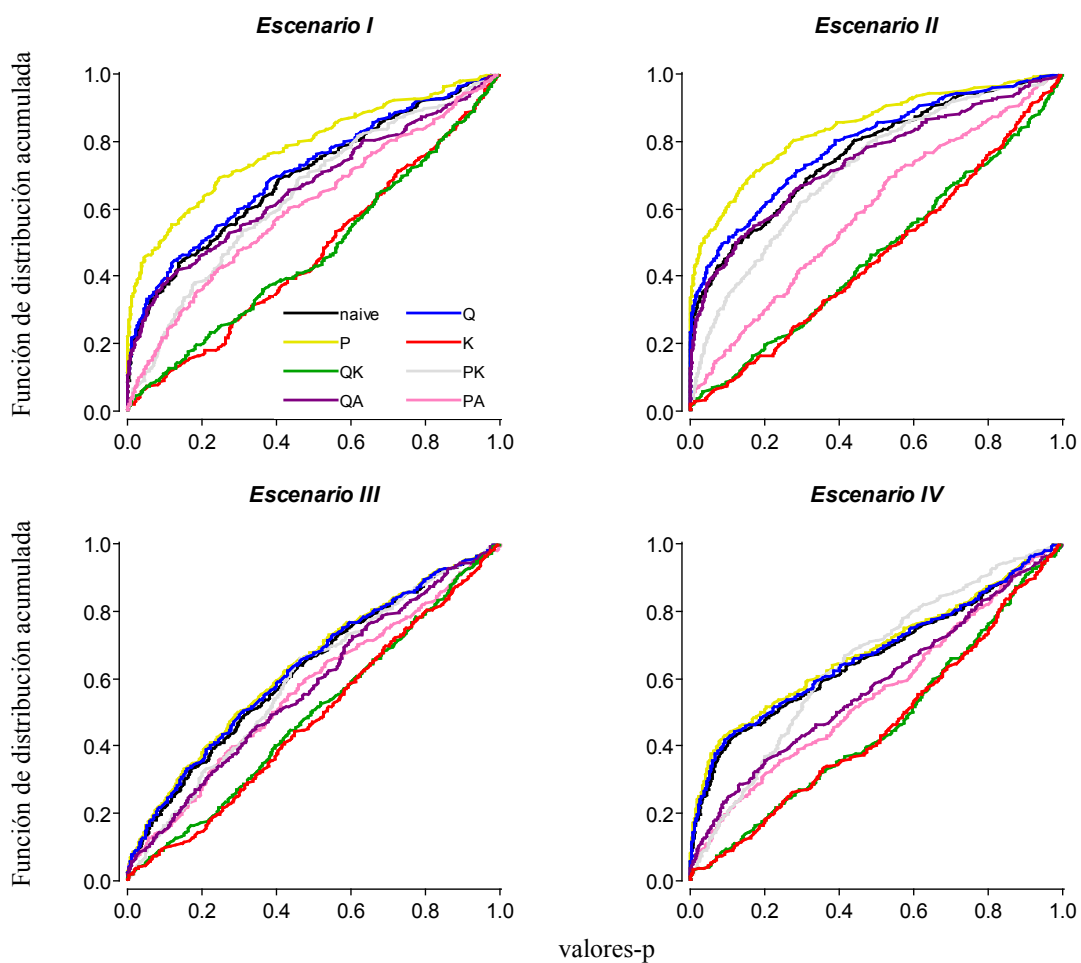


Figura 1. Función de distribución acumulada de los valores-p para cada uno de los ocho modelos evaluados en cuatro escenarios simulados que contienen 300 marcadores moleculares multilocus-bialélicos. En la columna de la izquierda escenarios con tamaño poblacional de 150 y en la columna derecha tamaño poblacional de 300. Arriba F_{ST} bajo y abajo F_{ST} alto.

observar que las tasas FDR fueron menores para alto FST. Cuando nos ubicamos en la situación de no corrección por estructura ni parentesco en el modelado, es decir cuando ajustamos un modelo *naive*, podemos observar que la tasa FDR disminuye con todas las correcciones respecto a sin corrección (SC), pero con alto FST baja en mayor medida corrigiendo con MLJ que con BH y LJ, esto se debe a que la estructura es grande y no fue corregida previamente en el modelado. Cuando la estructura es baja, es decir el nivel de convergencia entre poblaciones es alto, es más importante incluir la corrección por estructura en el modelado que en la corrección por multiplicidad, si bien las tasas FDR igualmente bajan con las correcciones por multiplicidad, dicha disminución no se produce de forma tan drástica como en la situación de alta estructura genética poblacional (Tabla 4 y Tabla 5).

Los resultados indicaron que aún para el caso de no corrección por EGP, es decir con el modelo *naive*, la aplicación de métodos de ajustes de valor-p por multiplicidad reduce la potencia significativamente. Potencias excesivamente bajas se observaron en escenarios correspondientes al menor tamaño poblacional (150 individuos) cuando las subpoblaciones tenían poca divergencia. Cuando se ajustaron los modelos de mapeo K o QK, la corrección de valores-p también produjo reducciones importantes de potencia. Estas reducciones fueron de mayor magnitud

que las producidas por el ajuste de un modelo de MA que controla EGP y sin corrección por multiplicidad. Es importante, mencionar que los métodos presentados para corrección por multiplicidad han sido diseñados para controlar el error de tipo I en una familia de pruebas y no para aumentar la probabilidad de detectar verdaderos positivos. Por la relación teórica existente entre los errores de tipo I y de tipo II en las pruebas de hipótesis es de esperar que la reducción significativa que estos métodos producen a nivel de FDR se encuentre asociada a pérdida de potencia. No obstante, la potencia con alto FST para el método MLJ fue igual o superior a la de los otros dos métodos de corrección de valores-p por multiplicidad (Tabla 6 y Tabla 7).

Con el mayor de los tamaños poblacionales (300), la aplicación del método MLJ directamente sobre los valores-p derivados del modelo más simple (de efectos fijos y sin corrección por EGP, *i.e.*, *naive*) produjo potencias similares a las obtenidas con el modelo de mapeo QK y sin ninguna corrección de valores-p. Las mayores pérdidas de potencia se obtuvieron con las dos estrategias usadas para controlar por EGP simultáneamente, es decir en el momento del modelado y al utilizar los valores-p para determinar significancia. MLJ produjo menor FDR que LJ en escenarios de alta estructura genética y no mostró mayores reducciones de potencia que LJ.

Tabla 4. Tasa de falsos descubrimientos (FDR) para tres modelos de mapeo asociativo, tres opciones de corrección de valores-p por inferencia simultánea bajo dos niveles de estructura genética poblacional, baja ($F_{ST}=0,03$) y alta ($F_{ST}=0,2$) divergencia genética, con un tamaño poblacional de 150.

| Modelo** | Bajo F_{ST} | | | | Alto F_{ST} | | | |
|--------------|---------------|------|------|------|---------------|------|------|------|
| | SC | BH | LJ | MLJ | SC | BH | LJ | MLJ |
| <i>naive</i> | 0.30 | 0.21 | 0.34 | 0.30 | 0.24 | 0.18 | 0.18 | 0.15 |
| K | 0.12 | 0.07 | 0.07 | 0.07 | 0.11 | 0 | 0.07 | 0 |
| QK | 0.13 | 0.07 | 0.07 | 0.07 | 0.11 | 0 | 0.03 | 0 |

*SC: Sin corrección por multiplicidad, BH: Benjamini y Hochberg, LJ: Li y Ji, MLJ: Li y Ji Modificado. ***naive*: sin corrección por estructura, K: con corrección por matriz de parentesco y QK: modelo mixto con Q, corrección mediante la matriz de probabilidades a posteriori obtenida con el software *Structure*, como factor de efectos fijos y K factor de efectos aleatorios.

Tabla 5. Tasa de falsos descubrimientos (FDR) para tres modelos de mapeo asociativo, tres opciones de corrección de valores-p por inferencia simultánea bajo dos niveles de estructura genética poblacional, baja ($F_{ST}= 0,03$) y alta ($F_{ST}= 0,2$) divergencia genética, con un tamaño poblacional de 300.

| Modelo** | Bajo F_{ST} | | | | Alto F_{ST} | | | |
|--------------|---------------|------|------|------|---------------|------|------|------|
| | Correcciones* | | | | | | | |
| | SC | BH | LJ | MLJ | SC | BH | LJ | MLJ |
| <i>naive</i> | 0.31 | 0.28 | 0.24 | 0.24 | 0.24 | 0.18 | 0.09 | 0.05 |
| K | 0.05 | 0 | 0.03 | 0 | 0.11 | 0 | 0 | 0 |
| QK | 0.07 | 0 | 0 | 0 | 0.11 | 0 | 0 | 0 |

*SC: Sin corrección por multiplicidad, BH: Benjamini y Hochberg, LJ: Li y Ji, MLJ: Li y Ji Modificado. ***naive*: sin corrección por estructura, K: con corrección por matriz de parentesco y QK: modelo mixto con Q, corrección mediante la matriz de probabilidades a posteriori obtenida con el software *Structure*, como factor de efectos fijos y K factor de efectos aleatorios.

Tabla 6. Tasa de falsos descubrimientos (FDR) para tres modelos de mapeo asociativo, tres opciones de corrección de valores-p por inferencia simultánea bajo dos niveles de estructura genética poblacional, baja ($F_{ST}= 0,03$) y alta ($F_{ST}= 0,2$) divergencia genética, con un tamaño poblacional de 300.

| Modelo** | Bajo F_{ST} | | | | Alto F_{ST} | | | |
|--------------|---------------|------|------|------|---------------|------|------|------|
| | Correcciones* | | | | | | | |
| | SC | BH | LJ | MLJ | SC | BH | LJ | MLJ |
| <i>naive</i> | 0.60 | 0.15 | 0.15 | 0.15 | 0.55 | 0.05 | 0.10 | 0.10 |
| K | 0.35 | 0 | 0.05 | 0.05 | 0.25 | 0.05 | 0.05 | 0.05 |
| QK | 0.35 | 0 | 0.05 | 0.05 | 0.30 | 0.05 | 0.10 | 0.05 |

*SC: Sin corrección por multiplicidad, BH: Benjamini y Hochberg, LJ: Li y Ji, MLJ: Li y Ji Modificado. ***naive*: sin corrección por estructura, K: con corrección por matriz de parentesco y QK: modelo mixto con Q, corrección mediante la matriz de probabilidades a posteriori obtenida con el software *Structure*, como factor de efectos fijos y K factor de efectos aleatorios.

Tabla 7. Potencia estadística para tres modelos de mapeo asociativo, tres opciones de corrección de valores-p por inferencia simultánea bajo dos niveles de estructura genética poblacional, baja ($F_{ST}= 0,03$) y Alto ($F_{ST}= 0,2$) divergencia genética, con un tamaño poblacional de 300.

| Modelo** | Bajo F_{ST} | | | | Alto F_{ST} | | | |
|--------------|---------------|------|------|------|---------------|------|------|------|
| | Correcciones* | | | | | | | |
| | SC | BH | LJ | MLJ | SC | BH | LJ | MLJ |
| <i>naive</i> | 0.75 | 0.35 | 0.55 | 0.45 | 0.75 | 0.25 | 0.35 | 0.45 |
| K | 0.40 | 0.10 | 0.20 | 0.10 | 0.45 | 0.05 | 0.20 | 0.20 |
| QK | 0.45 | 0.05 | 0.15 | 0.15 | 0.45 | 0.05 | 0.20 | 0.20 |

*SC: Sin corrección por multiplicidad, BH: Benjamini y Hochberg, LJ: Li y Ji, MLJ: Li y Ji Modificado. ***naive*: sin corrección por estructura, K: con corrección por matriz de parentesco y QK: modelo mixto con Q, corrección mediante la matriz de probabilidades a posteriori obtenida con el software *Structure*, como factor de efectos fijos y K factor de efectos aleatorios.

DISCUSIÓN

La corrección por multiplicidad con el método MLJ, propuesto en este trabajo, fue más potente que los métodos de corrección LJ y BH con los que fue comparado bajo distintos escenarios en relación al tamaño poblacional y al nivel de diferenciación genética entre las subpoblaciones de la población de mapeo. MLJ disminuyó la FDR en mayor proporción que BH y LJ en escenarios con alta divergencia genética poblacional aun cuando la EGP no había sido incluida en el ajuste del modelo, *i.e.* modelo *naive*. Para estudios de asociación genómica donde los marcadores pueden estar asociados, Li y Ji (2005) propusieron determinar el número efectivo de pruebas independientes (M_{eff}) para usarlo posteriormente en la corrección de los valores-p de las pruebas de hipótesis realizadas para descubrir asociaciones entre el estado de cada marcador y el fenotipo. Li y Ji *op. cit.*, usaron la descomposición espectral de la matriz de correlación entre marcadores para determinar el número de pruebas independientes. Esta aproximación es la base del método de corrección MLJ, aunque a diferencia de LJ, MLJ incorpora la estructura genética poblacional (EGP) en el cálculo de M_{eff} . La estrategia metodológica para contemplar la falta de independencia entre las pruebas de hipótesis que es ocasionada por la presencia de EGP, es ajustar la asociación entre cualquier par de marcadores por la estructura de grupo que determina la EGP. Así, la matriz de correlación usada por LJ es reemplazada por la matriz de medidas de asociación calculadas mediante el estadístico χ^2 de Mantel y Haenszel (1959) fijando como variable de control el grupo al que pertenece cada individuo de la población de mapeo.

Li *et al.* (2012) también trabajaron con la matriz de correlación entre marcadores, pero propusieron particionar dicha matriz en bloques de grupos de ligamiento para acelerar los tiempos computacionales relativos a la descomposición espectral. La estratificación en grupos de ligamiento permitió obtener una tasa de error de tipo I con valores cercanos al correcto de 0,05. Mientras mayor era la cantidad de estratos, más se acercaba al nivel de significación deseado (Li *et al.*, 2012). La propuesta fue comparada con otros métodos de corrección por multiplicidad, incluyendo el método de Li y Ji (2005) y el método de Moskvina y Schmidt (2008) que también estima el número de pruebas independientes a partir de la matriz de correlación entre marcadores, pero agrega el grado de independencia estadística entre las pruebas de hipótesis de un

marcador respecto a los marcadores que lo preceden (K_{eff}). LJ resultó, en esta comparación, más liberal que el método de Moskvina y Schmidt (2008) el cual fue menos conservador a medida que aumentaba la cantidad de bloques.

BH (Benjamini y Hochberg, 1995), es otro de los desarrollos metodológicos difundidos para la corrección de valores-p en casos de múltiples pruebas de hipótesis y de extenso uso en mapeo asociativo (Gutiérrez *et al.*, 2011; Muñoz-Amatriaín *et al.*, 2014; Olukolu *et al.*, 2014; Wang *et al.*, 2012). BH fue concebido como un método para corrección de la tasa de falsos descubrimientos. Para todos los métodos de control de multiplicidad, la potencia o capacidad de detectar asociaciones verdaderas disminuye a medida que aumenta el número de pruebas (Benjamini y Hochberg, 1995). En nuestro estudio, la pérdida de potencia fue mayor en BH y LJ que en MLJ y fue más abrupta para BH en el escenario con menor tamaño poblacional y bajo nivel de divergencia genética.

Wang *et al.* (2012) estudiaron el efecto del tamaño poblacional en la habilidad para detectar QTL usando un modelo que corregía por estructura sobre líneas endocriadas de cebada con EGP. A partir de las líneas disponibles generaron poblaciones de tamaños diferentes: 96, 192, 288, 384, 480, 576 y 672 individuos, encontrando que reducir el tamaño poblacional por debajo de 384 individuos produce una tasa alta de falsos descubrimientos. Cuando el tamaño de la población de mapeo disminuyó de 480 a 288 individuos, la FDR aumentó un 18%. Nuestros hallazgos también mostraron incremento de la FDR (7%), cuando el tamaño poblacional disminuyó de 300 a 150 individuos.

Beavis (1998) postuló que la construcción de un resultado estadístico en un análisis de asociación donde se pretende identificar QTL, puede caracterizarse según los valores de error de tipo I asociado a la tasa de falsos descubrimientos y según los valores de potencia para las pruebas de asociación. Beavis *op. cit.* estudió el efecto de la potencia simulando tres tamaños poblacionales (100, 500 y 1000 individuos para una progenie F2) para detectar 10 y 40 QTL, bajo tres niveles de heredabilidad expresado como variabilidad fenotípica explicada por los QTL (30, 63, 95%). Los valores de potencia estimados fueron menores a 6% con tamaños poblacionales de 100 individuos aun con QTL de alta heredabilidad. A medida que aumentaba el tamaño poblacional, los valores de potencia estimados aumentaron. Con 500 individuos las potencias fueron mayores al 50% sólo con 10 QTL, pero con 40 QTL se necesitaron 1000 individuos para alcanzar potencias

mayores al 50% en la detección de QTL de mediana a alta heredabilidad. Sin embargo, los QTL de menor efecto no fueron bien detectados aun con 1000 individuos en la población. Bradbury *et al.* (2011) usando datos genotípicos del programa de cebada (BarleyCAP), simularon datos con efectos fenotípicos para diferente cantidad de QTL con tres niveles de heredabilidad. En cada escenario ellos calcularon la potencia y el FDR para tamaños muestrales de 100 y 300 individuos. Bajo el modelo K para mapeo asociativo, las simulaciones con 100 líneas se desempeñaron pobremente para la detección de QTL, pero simulaciones con 300 líneas se desempeñaron adecuadamente. Las simulaciones con 300 líneas resultaron suficientemente potentes para detectar QTL cuando el carácter fenotípico fue controlado por pocos QTL con tamaño de efecto grande, aunque la potencia para detectar QTL de tamaño de efecto pequeño fue deficiente. En nuestro trabajo, las potencias fueron bajas debido a los tamaños poblaciones. La corrección por multiplicidad disminuye la FDR y provoca una pérdida de potencia en cualquiera de los métodos a los que se ha hecho referencia. Sin embargo, la corrección por multiplicidad con MLJ, usada en modelos donde no se ha descontado el efecto de la EGP previamente, fue la opción que condujo a la menor pérdida de potencia en poblaciones con alta divergencia genética y mayor tamaño poblacional. En poblaciones de mapeo de interés agronómico es frecuente la presencia de ancestros en común y por lo tanto la existencia de EGP, por lo que la práctica de usar un modelo que descuenta su efecto sobre la dependencia en la pruebas de hipótesis es recomendable para disminuir la tasa de falsos descubrimientos y trabajar con poblaciones de mapeos de más de 300 individuos para evitar que la tasa de no detección de QTL sea alta, sobre todo en contextos donde se esperan varios QTL de moderado o bajo efecto.

BIBLIOGRAFÍA

Balzarini M., Di Rienzo J. (2004) Info-Gen. Universidad Nacional de Cordoba, Córdoba.

Balzarini M.G., Gonzalez L., Tablada M., Casanoves F., Di Rienzo J.A., Robledo C.W. (2008) Infostat. Manual del Usuario, Córdoba, Argentina.

Beavis W.B. (1998) QTL analyses: power, precision, and ac-

curacy. In: Patterson A.H. (Ed.) Molecular dissection of complex traits. CRC Press, Boca Raton.

Benjamini Y., Hochberg Y. (1995) Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B* 57: 289-300.

Bernardo R. (2013) Genome wide markers as cofactors for precision mapping of quantitative trait loci. *Theor. Appl. Genet.* 126: 999-1009. doi:10.1007/s00122-012-2032-2.

Bonferroni C.E. (1935) Il calcolo delle assicurazioni su gruppi di teste. *Studi in Onore del Professore Salvatore Ortu Carboni*, pp. 13-60.

Bradbury P., Parker T., Hamblin M.T., Jannink J.L. (2011) Assessment of power and false discovery in genome-wide association studies using the BarleyCAP germplasm. *Crop Sci.* 51: 52-59.

Breseghele F., Sorrells M.E. (2006). Association Mapping of Kernel Size and Milling Quality in Wheat (*Triticum aestivum* L.) Cultivars. *Genetics*, 172(2), 1165-1177. <http://doi.org/10.1534/genetics.105.044586>.

Cappa E.P., El-Kassaby Y.A., Garcia M.N., Acuña C., Borralho N.M.G., Grattapaglia D., Marcucci Poltri S. (2013) Impacts of population structure and analytical models in genome-wide association studies of complex traits in forest trees: a case study in *Eucalyptus globulus*. *PLoS ONE* 8: e81267. doi:10.1371/journal.pone.0081267.

Cheverud J.M. (2001) A simple correction for multiple comparisons in interval mapping genome scans. *Heredity* 87: 52-58.

Draper N.R., Smith H. (1998) Applied regression analysis, 3rd Edition Wiley, New York.

Excoffier L., Hofer T., Foll M. (2009) Detecting loci under selection in a hierarchically structured population. *Heredity* 103: 285-298. doi:<http://www.nature.com/hdy/journal/v103/n4/supinfo/hdy200974s1.html>.

Gutiérrez L., Cuesta-Marcos A., Castro A.J., von Zitze-

- witz J., Schmitt M., Hayes P.M. (2011) Association mapping of malting quality Quantitative Trait Loci in winter barley: positive signals from small germplasm arrays. *Plant Gen.* 4: 256-272. doi:10.3835/plantgenome2011.07.0020.
- Gutiérrez L., Germán S., Pereyra S., Hayes P., Pérez C., Capettini F., Locatelli A., Berberian N.M., Falconi E.E., Estrada R., Fros D., Gonza V., Altamirano H., Huerta-Espino J., Neyra E., Orjeda G., Sandoval-Islas S., Singh R., Turkington K., Castro A.J. (2015) Multi-environment multi-QTL association mapping identifies disease resistance QTL in barley germplasm from Latin America. *Theor. Appl. Genet.* 128 (3): 501-516. doi:10.1007/s00122-014-2448-y.
- Kang H.M., Zaitlen N.A., Wade C.M., Kirby A., Heckerman D., Daly M.J., Eskin E. (2008) Efficient control of population structure in model organism association mapping. *Genetics* 178: 1709-1723. doi:10.1534/genetics.107.080101.
- Li J., Ji L. (2005) Adjusting multiple testing in multilocus analyses using the eigenvalues of a correlation matrix. *Heredity* 95: 221-227.
- Li M.X., Yeung J.M.Y., Cherny S.S., Sham P. (2012) Evaluating the effective numbers of independent tests and significant p-value thresholds in commercial genotyping arrays and public imputation reference datasets. *Hum. Genet.* 131(5): 747. <https://doi.org/10.1007/s00439-011-1118-2>.
- Locatelli A., Cuesta-Marcos A., Gutiérrez L., Hayes P., Smith K., Castro A. (2013) Genome-wide association mapping of agronomic traits in relevant barley germplasm in Uruguay. *Mol. Breeding* 31: 631-654. doi:10.1007/s11032-012-9820-x.
- Malosetti M, van der Linden CG, Vosman B, van Eeuwijk FA (2007) A mixed-model approach to association mapping using pedigree information with an illustration of resistance to *Phytophthora infestans* in potato. *Genetics* 175:879-889.
- Mantel N., Haenszel W. (1959) Statistical aspects of the analysis of data from retrospective studies of disease. *Journal of the National Cancer Institute* 22: 719-748.
- Miller C.J., Genovese C., Nichol R.C., Wasserman L., Connolly A., Reichart D., Hopkins A., Schneider J., Moore A. (2001) Controlling the false-discovery rate in astrophysical data analysis. *The Astronomical Journal* 122 (6): 3492-3505.
- Moskvina V., Schmidt K.M. (2008) On multiple-testing correction in genome-wide association studies. *Genet. Epidemiol.* 32: 567-573.
- Muñoz-Amatriáin M., Cuesta-Marcos A., Endelman J.B., Comadran J., Bonman J.M., Bockelman H.E., Chao S., Russell J., Waugh R., Hayes P.M., Muehlbauer G.J. (2014) The USDA Barley Core Collection: Genetic diversity, population structure, and potential for genome-wide association studies. *PLoS ONE* 9 (4): e94688. doi:10.1371/journal.pone.0094688.
- Olukolu B.A., Wang G.F., Vontimitta V., Venkata B.P., Marla S., Ji J., Gachomo E., Chu K., Negeri A., Benson J., Nelson R., Bradbury P., Nielsen D., Holland J.B., Balint-Kurti P., Gurmukh J. (2014) A Genome-Wide Association Study of the maize hypersensitive defense response identifies genes that cluster in related pathways. *PLoS Genet* 10: e1004562. doi:10.1371/journal.pgen.1004562.
- Parisseaux B., Bernardo R. (2004) *In silico* mapping of quantitative trait loci in maize. *Theor. Appl. Genet.* 109:508-514.
- Peña Malavera A. (2015) Aproximaciones estadísticas para el mapeo asociativo en estudios genéticos. Universidad Nacional de Córdoba, Córdoba.
- Pers T. H., Karjalainen J. M., Chan Y., Westra H. J., Wood A. R., Yang J., Luij. C., Vedantam S., Gustafsson S., Esko T., Frayling T., Speliotes E.K. Genetic Investigation of ANthropometric Traits (GIANT) Consortium, Boehnke M., Raychaudhuri S., Fehrmann R., Hirschhorn J., Franke L. (2015) Biological interpretation of genome-wide association studies using predicted gene functions. *Nature communications* 6: 5890.
- Sabatti C., Service S., Freimer N. (2003) False discovery

- rate in linkage and association genome screens for complex disorders. *Genetics* 164: 829-833.
- Sargolzaei M., Schenkel F. (2009) QMSim: a large-scale genome simulator for livestock. *Bioinformatics* 25: 680-681.
- Schwartzman A., Dougherty R.F., Taylor J.E. (2008) False discovery rate analysis of brain diffusion direction maps. *The Annals of Applied Statistics* 153-175. doi:10.1214/07-aoas133.
- Sidak Z. (1967) Rectangular confidence regions for the means of multivariate normal distributions. *Journal of the American Statistical Association* 62: 626-633. doi:10.2307/2283989.
- Spindel J., Begum H., Akdemir D., Virk P., Collard B., Redoña E., Atlin G., Jannink J.L., McCouch S. (2015) Genomic selection and association mapping in rice (*Oryza sativa*): effect of trait genetic architecture, training population composition, marker number and statistical model on accuracy of rice genomic selection in elite, tropical rice breeding lines. *PLoS Genet.* 11: e1004982.
- Tadesse W., Ogbonnaya F.C., Jighly A., Sanchez-Garcia M., Sohail Q., Rajaram S., Baum M. (2015) Genome-Wide Association Mapping of Yield and Grain Quality Traits in Winter Wheat Genotypes. *Plos One* 10(10): e0141339. <https://doi.org/10.1371/journal.pone.0141339>.
- Team R.D.C. (2013) R: A language and environment for statistical computing, Vienna, Austria.
- Tracy C.A., Widom H. (1994) Level-spacing distributions and the Airy kernel. *Comm. Math. Phys.* 159: 23.
- Tusher V.G., Tibshirani R., Chu G. (2001) Significance analysis of microarrays applied to the ionizing radiation response. *Proceedings of the National Academy of Sciences* 98: 5116-5121.
- Wang H., Smith K., Combs E., Blake T., Horsley R., Muehlbauer G. (2012) Effect of population size and unbalanced data sets on QTL detection using genome-wide association mapping in barley breeding germplasm. *Theor. Appl. Genet.* 124: 111-124. doi:10.1007/s00122-011-1691-8.
- Wright S. (1951) The genetical structure of populations. *Ann. Eugen.* 15: 31.
- Xiao J., Zhu W., Guo J. (2013) Large-scale multiple testing in genome-wide association studies via region-specific hidden Markov models. *BMC Bioinformatics* 14: 282.
- Yan J., Warburton M., Crouch J. (2011) Association mapping for enhancing maize (*Zea mays* L.) genetic improvement. *Crop Sci* 51: 433-449.
- Yu J., Pressoir G., Briggs W., Bi I., Yamasaki M., Doebley J., McMullen M.D., Gaut B.S., Nielsen D.N., Holland J.B., Kresovich S., Buckler E.S. (2006) A unified mixed model method for association mapping that accounts for multiple levels of relatedness. *Nat. Genet.* 2: 203-208.
- Zhou G. F., Broughton S., Zhang X. Q., Zhou M. X., Li C. D. (2016) Genome-wide association mapping of acid soil resistance in barley (*Hordeum vulgare* L.). *Front. Plant Sci.* 7:406 [10.3389/fpls.2016.00406](https://doi.org/10.3389/fpls.2016.00406).



¿FUNCIONA LA LEY DE EQUILIBRIO DE HARDY-WEINBERG EN AUTOPOLIPOIDES IGUAL QUE EN DIPLOIDES?

DOES THE HARDY-WEINBERG LAW OF EQUILIBRIUM WORK IN AUTOPOLYPLOIDS AS IT DOES IN DIPLOIDS?

Núñez-Colín C.A.

Programa de Biotecnología, Universidad de Guanajuato, Mutualismo 303,
Col. La Suiza, Celaya, Guanajuato, 38060, México.

email: carlos.nunez@ugto.mx

ABSTRACT

The law of equilibrium Hardy-Weinberg is the cornerstone of the population genetics and of the quantitative genetics; however, for its calculation in autopolyploids it is necessary take in account that the allelic and gametic frequencies are different, contrary to the diploids where they are the same. This causes that the calculations must be done with the allelic frequencies or gametic based on allelic frequencies. Otherwise the equilibrium is broken in the population and the bias that this entails in the calculation of other genetical test like Wright's F statistics, the Nei's GST or Bayesian models that are based on the disequilibrium that populations show. That is why in this work they developed models of one locus with two alleles in autotetraploid and autooctoploid genotypes to make a generalization of the law of equilibrium in autopolyploid populations.

Key words: Population Genetics, Quantitative Genetics, Genotype frequencies calculation

RESUMEN

La ley del equilibrio Hardy-Weinberg es la piedra angular de la genética de poblaciones y la genética cuantitativa; sin embargo, para su cálculo en autopoliploides hay que tener en cuenta que las frecuencias alélicas y gaméticas son diferentes, caso contrario que los diploides donde son iguales. Esto ocasiona que el cálculo de esta fórmula deba hacerse con las frecuencias alélicas o las gaméticas basadas en las alélicas, de otra forma se puede romper el equilibrio de la población y el sesgo que esto conlleva en cálculos de otras pruebas como los estadísticos F de Wright, la G_{ST} de Nei o modelos bayesianos que se basan en los desequilibrios que presentan las poblaciones. Por eso este ensayo desarrolla los modelos de un *locus* con dos alelos en genotipos autotetraploides y autooctoploides para poder realizar una generalización de la ley del equilibrio en poblaciones autopoliploides.

Palabras clave: Genética de poblaciones, Genética Cuantitativa, Cálculo de frecuencias genotípicas

Fecha de recepción: 18/01/2017
Fecha de aceptación de versión final: 17/05/2018

INTRODUCCIÓN

La genética es la ciencia del siglo XX, que se origina con el redescubrimiento de las leyes de Mendel (De Vries, 1900; 1902; Correns, 1900), originalmente publicadas en 1866 (Mendel, 1866) y que a partir de este punto empieza su estudio, pasando por la Teoría cromosómica de Thomas Morgan (Morgan *et al.*, 1915) que dio el sustento biológico a las mismas. Sin embargo, la genética de poblaciones nace con un conocimiento muy especial al inicio de la genética como ciencia: la ley del equilibrio (Crow, 1988; 1999).

Hardy (1908) y Weinberg (1908) de manera autónoma, lograron descifrar que el cuadrado de la suma de las frecuencias alélicas daba como resultado la suma de las frecuencias genotípicas, y que las frecuencias genotípicas segregaban exactamente las mismas frecuencias alélicas. No obstante, en un modelo diploide donde $2n=2x$ y los gametos son $n=x$, las frecuencias alélicas y gaméticas son idénticas debido a que hay un solo alelo por gameto (Haldane, 1930). Esto hace que exista confusión sobre lo que se está calculando en estudios con marcadores moleculares si es que se está calculando las frecuencias alélicas o gaméticas. En los autopoliplóides estas son frecuencias diferentes, debido a que en el caso de los autotetraploides tenemos genotipos $2n=4x$ que segregan gametos $n=2x$, los autohexaploides genotipos $2n=6x$ que segregan gametos $n=3x$ y los autooctoploides que presentan genotipos $2n=8x$ que segregan gametos $n=4x$ y con sólo dos alelos por *locus* generan un mayor número de gametos y genotipos (Tabla 1).

Hoy en día la mayoría de los programas de análisis genético tienen programado la obtención de frecuencias alélicas basadas en el modelo del equilibrio HW con dos alelos, principalmente para marcadores denominados dominantes como los ISSR, RAPD, AFLP, entre otros (Yeh *et al.*, 1999; Peakall y Smouse, 2012). No obstante, Dufresne *et al.* (2014) hacen una revisión del estado del arte sobre análisis de poliplóides y uno de los principales conflictos es que el cálculo de las frecuencias alélicas es la adaptación del modelo diploide y no obtienen realmente frecuencias alélicas sino gaméticas, razón por la cual este ensayo tuvo como objetivo calcular el equilibrio Hardy-Weinberg (HW) en autopoliplóides a partir de frecuencias alélicas y de frecuencias gaméticas y poder conocer los posibles errores que se cometen con los programas que calculan con unas u otras frecuencias; para hacerlo se utilizó el modelo de un *locus* con dos alelos con genotipos autotetraploides (genotipos $2n=4x$ y gametos $n=2x$) y en genotipos autooctoploides (genotipos $2n=8x$ y gametos $n=4x$). Hay que mencionar que estos cálculos son aplicables sólo para autopoliplóides y deben omitirse aquellas ploidías con polisomías.

Tabla 1. Número de gametos y genotipos en modelos de dos alelos por *locus* con diferentes niveles de ploidía.

| Nivel de ploidía | Número de gametos | Número de genotipos |
|------------------|-------------------|---------------------|
| $2n=2x$ | 2 | 3 |
| $2n=4x$ | 3 | 5 |
| $2n=6x$ | 4 | 7 |
| $2n=8x$ | 5 | 9 |

EQUILIBRIO HW EN UN MODELO DIPLOIDE

Numéricamente, el equilibrio HW se explica así en un modelo de un *locus* con dos alelos (*A/a*) en diploides, el alelo *A* tiene una frecuencia de *p* (*pA*) y el alelo *a* una frecuencia de *q* (*qa*) sabiendo entonces que $p+q=1$ entonces:

$$[pA + qa]^2 = p^2AA + 2pqAa + q^2aa$$

Es decir, la suma de estas frecuencias gaméticas al cuadrado es igual a la suma de las frecuencias genotípicas.

Al mismo tiempo sabemos que el genotipo *AA* sólo segrega gametos *A* al igual que el genotipo *aa* sólo segrega gametos *a*; mientras que el genotipo *Aa* segrega ½ de gametos con el alelo *A* y ½ con el alelo *a*, por tanto, para volver a obtener la frecuencia del gameto con el alelo *A* se procede de la siguiente manera:

$$\left(p^2 + \frac{1}{2}(2pq)\right)A = (p^2 + pq)A = [p(p+q)]A = [p(1)]A = pA$$

De manera análoga se hace para el alelo *a*

$$\left(\frac{1}{2}(2pq) + q^2\right)a = (pq + q^2)a = [q(p+q)]a = [q(1)]a = qa$$

Lo que demuestra que en apareamiento aleatorio y en ausencia de mutación, migración o selección, las frecuencias gaméticas, y por obvias razones las genotípicas, se mantienen de generación en generación y que sólo es necesario una generación en apareamiento aleatorio para obtener el equilibrio.

DESARROLLO DEL EQUILIBRIO HW EN UN MODELO AUTOTETRAPLOIDE

En un modelo autotetraploide de un *locus* con dos alelos (*A/a*) se pueden tener 5 genotipos diferentes: *AAAA*, *AAAa*, *AAaa*, *Aaaa* y *aaaa* y estos pueden segregar 3 gametos: *AA*, *Aa* y *aa*; en este sentido, se sigue la fórmula de la ley del equilibrio Hardy-Weinberg (equilibrio HW) suponiendo frecuencias gaméticas de p_1, p_2 y p_3 , respectivamente, se obtiene lo siguiente:

$$(p_1AA + p_2Aa + p_3aa)^2 = p_1^2AAAA + p_2^2AAaa + p_3^2aaaa + 2p_1p_2AAAA + 2p_1p_3AAaa + 2p_2p_3Aaaa$$

Como puede notarse existen dos combinaciones gaméticas que expresan el mismo genotipo: *AA* x *aa* y *Aa* x *Aa* dan como resultado el genotipo *AAaa* por lo que la expresión se reduce a

$$(p_1AA + p_2Aa + p_3aa)^2 = p_1^2AAAA + 2p_1p_2AAAA + (p_2^2 + 2p_1p_3)AAaa + 2p_2p_3Aaaa + p_3^2aaaa$$

Ahora, los genotipos *AAAA* y *aaaa* sólo segregan gametos *AA* y *aa*, respectivamente; mientras que los otros tres genotipos pueden generar dos o los tres gametos por lo que se segregarán de acuerdo a sus combinaciones de 4 en 2,

$$\binom{4}{2} = \frac{4!}{2! \cdot (4-2)!} = \frac{4 \cdot 3 \cdot 2!}{2! \cdot 2!} = \frac{12}{2} = 6$$

Es decir, existen seis combinaciones, y si a cada genotipo se numera de acuerdo a sus combinaciones dará como resultado la frecuencia con la que segrega cada uno de los gametos (Tabla 2).

Lo que se traduce que el genotipo *AAAA* segrega sólo gametos *AA*, mismo caso de *aaaa* que sólo segrega gametos *aa*, mientras que el genotipo *AAAa* segrega ½ de gametos *AA* y ½ de gametos *Aa* y de manera análoga el genotipo *Aaaa* segrega ½ de gametos *Aa* y ½ de gametos *aa*. Sin embargo, el genotipo *AAaa* segrega □ de gametos *AA*, ⅔ de gametos *Aa* y □ de gametos *aa*.

Los gametos parecen estar en equilibrio como podemos constatarlo en la Tabla 2; sin embargo, al comprobar la frecuencia del gameto *AA*, ésta estaría calculada de la siguiente forma:

$$\begin{aligned} [p_1^2 + \frac{1}{2}(2p_1p_2) + \frac{1}{6}(p_2^2 + 2p_1p_3)]AA &= [p_1^2 + p_1p_2 + \frac{1}{6}p_2^2 + \frac{1}{3}p_1p_3]AA \\ &= [p_1(p_1 + p_2 + \frac{1}{3}p_3)] + \frac{1}{6}p_2^2 AA = [p_1(1 - \frac{2}{3}p_3)] + \frac{1}{6}p_2^2 AA \\ &= [p_1 - \frac{2}{3}p_1p_3 + \frac{1}{6}p_2^2]AA = [p_1 + [\frac{1}{6}(p_2^2 - 4p_1p_3)]]AA \end{aligned}$$

Tabla 2. Combinaciones gaméticas de genotipos tetraploides.

| Combinación | Genotipos | | | | |
|-------------|-----------|------|------|------|------|
| | AAAA | AAAa | AAaa | Aaaa | aaaa |
| 1,2 | AA | AA | AA | Aa | aa |
| 1,3 | AA | AA | Aa | Aa | aa |
| 1,4 | AA | Aa | Aa | Aa | aa |
| 2,3 | AA | AA | Aa | aa | aa |
| 2,4 | AA | Aa | Aa | aa | aa |
| 3,4 | AA | Aa | aa | aa | aa |

De manera análoga para el gameto Aa resulta

$$\begin{aligned} \left[\frac{1}{2} (2p_1 p_2) + \frac{1}{2} (2p_1 p_3) + \frac{2}{3} (p_2^2 + 2p_1 p_3) \right] Aa &= [p_1 p_2 + p_1 p_3 + \frac{2}{3} p_2^2 + \frac{4}{3} p_1 p_3] Aa \\ &= \left[[p_2 (p_1 + \frac{2}{3} p_2 + p_3)] + \frac{4}{3} p_1 p_3 \right] Aa = \left[[p_2 (1 - \frac{1}{3} p_2)] + \frac{4}{3} p_1 p_3 \right] Aa \\ &= [p_2 - \frac{1}{3} p_2^2 + \frac{4}{3} p_1 p_3] Aa = [p_2 - \frac{1}{3} (p_2^2 - 4p_1 p_3)] Aa \end{aligned}$$

Y finalmente para el gameto aa resulta

$$\begin{aligned} \left[\frac{1}{6} (p_2^2 + 2p_1 p_3) + \frac{1}{2} (2p_2 p_3) + p_3^2 \right] aa &= \left[\frac{1}{6} p_2^2 + \frac{1}{3} p_1 p_3 + p_2 p_3 + p_3^2 \right] aa \\ &= \left[[p_3 (\frac{1}{3} p_1 + p_2 + p_3)] + \frac{1}{6} p_2^2 \right] aa = \left[[p_3 (1 - \frac{2}{3} p_1)] + \frac{1}{6} p_2^2 \right] aa \\ &= [p_3 - \frac{2}{3} p_1 p_3 + \frac{1}{6} p_2^2] aa = [p_3 + \frac{1}{6} (p_2^2 - 4p_1 p_3)] aa \end{aligned}$$

Por lo que existe una desviación del equilibrio de $\frac{1}{3} (p_2^2 - 4p_1 p_3)$ del cual se resta del gameto heteroalélico y se suma la mitad de este factor a cada uno de los gametos homoalélicos.

Si vemos a este sistema desde la perspectiva de la frecuencia alélica, es decir, si vemos la frecuencia de pA y qa en el modelo autotetraploide (genotipos $2n=4x$ y gametos $n=2x$) tenemos lo siguiente:

$$\begin{aligned} \left[\left(\frac{2!}{2! \cdot 0!} p^2 \right) AA + \left(\frac{2!}{1! \cdot 1!} pq \right) Aa + \left(\frac{2!}{0! \cdot 2!} q^2 \right) aa \right]^2 &= [p^2 AA + 2pqAa + q^2 aa]^2 \\ &= p^4 AAAA + 4p^3 q AAAa + 6p^2 q^2 AAaa + 4pq^3 Aaaa + q^4 aaaa \end{aligned}$$

Si seguimos el mismo principio para la asignación de las frecuencias gaméticas en base a las frecuencias alélicas de la siguiente generación, se tiene para el caso del gameto AA lo siguiente:

$$\begin{aligned} [p^4 + \frac{1}{2} (4p^3 q) + \frac{1}{6} (6p^2 q^2)] AA &= [p^4 + 2p^3 q + p^2 q^2] AA = [p^2 (p^2 + 2pq + q^2)] AA \\ &= [p^2 (p + q)^2] AA = [p^2 (1^2)] AA = p^2 AA \end{aligned}$$

Para el gameto Aa se deriva de la siguiente manera:

$$\begin{aligned} [\frac{1}{2} (4p^3 q) + \frac{1}{2} (4pq^2) + \frac{2}{3} (6p^2 q^2)] Aa &= [2p^3 q + 2pq^2 + 4p^2 q^2] Aa = [2pq (p^2 + 2pq + q^2)] Aa \\ &= [2pq (p + q)^2] Aa = [2pq (1^2)] Aa = 2pq Aa \end{aligned}$$

Finalmente, para el gameto aa resulta lo siguiente:

$$\begin{aligned} [\frac{1}{6} (6p^2 q^2) + \frac{1}{2} (4pq^2) + q^4] aa &= [p^2 q^2 + 2pq^2 + q^4] aa = [q^2 (p^2 + 2pq + q^2)] aa = [q^2 (p + q)^2] aa \\ &= [q^2 (1^2)] aa = q^2 AA \end{aligned}$$

Es decir, se mantiene el equilibrio siempre y cuando se utilicen las frecuencias gaméticas basadas en las frecuencias alélicas y también las frecuencias alélicas se mantienen en equilibrio considerando que el genotipo $AAAA$ tiene sólo alelos A y $aaaa$ tiene sólo alelos a , mientras que el genotipo $AAAAa$ tiene $\frac{3}{4}$ del alelo A y $\frac{1}{4}$ del alelo a y $Aaaa$ tiene $\frac{1}{4}$ del alelo A y $\frac{3}{4}$ del alelo a y finalmente el genotipo $AAaa$ tiene $\frac{1}{2}$ de cada uno de los alelos.

Por tanto, el cálculo de la frecuencia alélica para el alelo A es el siguiente:

$$[p^4 + \frac{3}{4}(4p^3q) + \frac{1}{2}(6p^2q^2) + \frac{1}{4}(4pq^3)]A = [p(p^3 + 3p^2q + 3pq^2 + q^3)]A = [p(p+q)^3]A = [p(1^3)]A = pA$$

De manera análoga para el cálculo de la frecuencia del alelo a es la siguiente:

$$[\frac{1}{4}(4p^3q) + \frac{1}{2}(6p^2q^2) + \frac{3}{4}(4pq^3) + q^4]a = [q(p^3 + 3p^2q + 3pq^2 + q^3)]a = [q(p+q)^3]a = [q(1^3)]a = qa$$

Es decir, el equilibrio alélico, gamético y genotípico se mantiene siempre y cuando se estén utilizando las frecuencias gaméticas basadas en las frecuencias alélicas. De otra forma se estaría extrapolando un modelo diploide de un *locus* con alelos múltiples y con su obvio desequilibrio HW que eso conlleva.

DESARROLLO DEL EQUILIBRIO HW EN UN MODELO AUTOOCTOPLOIDE

En un modelo autooctoploide de un *locus* con dos alelos (A/a) se pueden tener 9 genotipos diferentes: $AAAAAAAA$, $AAAAAAAa$, $AAAAAAaa$, $AAAAAaaa$, $AAAAaaaa$, $AAAaaaaa$, $AAaaaaaa$, $Aaaaaaaa$ y $aaaaaaaa$ y estos pueden segregar 5 gametos: $AAAA$, $AAAa$, $AAaa$, $Aaaa$ y $aaaa$; en este sentido, y al igual que en modelo autotetraploide; se sigue la fórmula de la ley del equilibrio HW suponiendo frecuencias gaméticas p_1, p_2, p_3, p_4 y p_5 , respectivamente, se obtiene lo siguiente:

$$(p_1AAAA + p_2AAAa + p_3AAaa + p_4Aaaa + p_5aaaa)^2 = p_1^2AAAAAAAA + 2p_1p_2AAAAAAa + (p_2^2 + 2p_1p_3)AAAAAAaa + (2p_1p_4 + 2p_2p_3)AAAAAaaa + (p_3^2 + 2p_1p_5 + 2p_2p_4)AAAAaaaa + (2p_2p_5 + 2p_3p_4)AAAaaaaa + (p_4^2 + 2p_3p_5)AAaaaaaa + 2p_4p_5Aaaaaaaa + p_5^2aaaaaaaa$$

Como puede notarse en este modelo existen varias combinaciones gaméticas que expresan el mismo genotipo: $AAAa \times AAAa$ y $AAAA \times AAaa$ dan como resultado $AAAAAAaa$; $AAAA \times Aaaa$ y $AAAa \times AAaa$ dan como resultado $AAAAAaaa$; $AAAA \times aaaa$, $AAAa \times Aaaa$ y $AAaa \times AAaa$ dan como resultado $AAAAaaaa$; $AAAa \times aaaa$ y $AAaa \times Aaaa$ dan como resultado $AAAaaaaa$, y finalmente, $AAaa \times aaaa$ y $Aaaa \times Aaaa$ dan como resultado $AAaaaaaa$.

Al igual que en el caso del modelo autotetraploide, los genotipos $AAAAAAAA$ y $aaaaaaaa$ sólo segregar gametos $AAAA$ y $aaaa$, respectivamente, los otros genotipos pueden segregar dos o más de los cinco gametos posibles por lo que en este modelo los gametos de cada genotipo se pueden obtener con las combinaciones posibles.

$$\binom{8}{4} = \frac{8!}{4! \cdot (8-4)!} = \frac{8 \cdot 7 \cdot 6 \cdot 5 \cdot 4!}{4! \cdot 4!} = \frac{8 \cdot 7 \cdot 6 \cdot 5}{4 \cdot 3 \cdot 2 \cdot 1} = \frac{1680}{24} = 70$$

Por lo que se muestra en la Tabla 3 las frecuencias con la que se manifiesta cada gameto en cada genotipo en este modelo autooctoploide.

Tabla 3. Proporción de la segregación de gametos a partir de genotipos octoploides en un modelo de un *locus* con dos alelos.

| Genotipo | Gametos | | | | |
|-----------|---------|------|-------|------|------|
| | AAAA | AAAa | AAaa | Aaaa | aaaa |
| AAAAAAAA | 1 | 0 | 0 | 0 | 0 |
| AAAAAAAAa | 1/2 | 1/2 | 0 | 0 | 0 |
| AAAAAAaa | 3/14 | 4/7 | 3/14 | 0 | 0 |
| AAAAAaaa | 1/14 | 3/7 | 3/7 | 1/14 | 0 |
| AAAAaaaa | 1/70 | 8/35 | 18/35 | 8/35 | 1/70 |
| AAAaaaaa | 0 | 1/14 | 3/7 | 3/7 | 1/14 |
| AAaaaaaa | 0 | 0 | 3/14 | 4/7 | 3/14 |
| Aaaaaaaa | 0 | 0 | 0 | 1/2 | 1/2 |
| aaaaaaaa | 0 | 0 | 0 | 0 | 1 |

Por tanto, para obtener la frecuencia gamética del gameto AAAA se obtiene de la siguiente manera:

$$\begin{aligned}
 & [p_1^2 + 1/2 (2p_1 p_2) + 3/14 (p_2^2 + 2p_1 p_3) + 1/14 (2p_1 p_4 + 2p_2 p_3) + 1/70 (p_3^2 + 2p_1 p_5 + 2p_2 p_4)]AAAA \\
 & = [p_1^2 + p_1 p_2 + 3/14 p_2^2 + 3/7 p_1 p_3 + 1/7 p_1 p_4 + 1/7 p_2 p_3 + 1/70 p_3^2 + 1/35 p_1 p_5 \\
 & + 1/35 p_2 p_4]AAAA \\
 & = [p_1(p_1 + p_2 + 3/7 p_3 + 1/7 p_4 + 1/35 p_5) + 3/14 p_2^2 + 1/7 p_2 p_3 + 1/35 p_2 p_4 \\
 & + 1/70 p_3^2]AAAA \\
 & = [p_1(1 - 4/7 p_3 - 6/7 p_4 - 34/35 p_5) + 3/14 p_2^2 + 1/7 p_2 p_3 + 1/35 p_2 p_4 + 1/70 p_3^2]AAAA \\
 & = [p_1 - 4/7 p_1 p_3 - 6/7 p_1 p_4 - 34/35 p_1 p_5 + 3/14 p_2^2 + 1/7 p_2 p_3 + 1/35 p_2 p_4 \\
 & + 1/70 p_3^2]AAAA
 \end{aligned}$$

De manera análoga para el gameto AAAa se deriva de la siguiente manera:

$$\begin{aligned}
 & [1/2 (2p_1 p_2) + 4/7 (p_2^2 + 2p_1 p_3) + 3/7 (2p_1 p_4 + 2p_2 p_3) + 8/35 (p_3^2 + 2p_1 p_5 + 2p_2 p_4) \\
 & + 1/14 (2p_2 p_5 + 2p_3 p_4)]AAAa \\
 & = [p_1 p_2 + 4/7 p_2^2 + 8/7 p_1 p_3 + 6/7 p_1 p_4 + 6/7 p_2 p_3 + 8/35 p_3^2 + 16/35 p_1 p_5 + 16/35 p_2 p_4 \\
 & + 1/7 p_2 p_5 + 1/7 p_3 p_4]AAAa \\
 & = [p_2(p_1 + 4/7 p_2 + 6/7 p_3 + 16/35 p_4 + 1/7 p_5) + 8/7 p_1 p_3 + 6/7 p_1 p_4 + 8/35 p_3^2 \\
 & + 16/35 p_1 p_5 + 1/7 p_3 p_4]AAAa \\
 & = [p_2(1 - 3/7 p_3 - 1/7 p_4 - 19/35 p_5 - 6/7 p_3) + 8/7 p_1 p_3 + 6/7 p_1 p_4 + 8/35 p_3^2 \\
 & + 16/35 p_1 p_5 + 1/7 p_3 p_4]AAAa \\
 & = [p_2 + 8/7 p_1 p_3 + 6/7 p_1 p_4 + 16/35 p_1 p_5 - 3/7 p_2^2 - 1/7 p_2 p_3 - 19/35 p_2 p_4 - 6/7 p_2 p_5 \\
 & + 8/35 p_3^2 + 1/7 p_3 p_4]AAAa
 \end{aligned}$$

Para obtener la frecuencia del gameto $AAaa$ resulta lo siguiente:

$$\begin{aligned}
 & \left[\frac{3}{14} (p_2^2 + 2p_1p_3) + \frac{3}{7} (2p_1p_4 + 2p_2p_3) + \frac{18}{35} (p_3^2 + 2p_1p_5 + 2p_2p_4) + \frac{3}{7} (2p_2p_5 + 2p_3p_4) \right. \\
 & \quad \left. + \frac{3}{14} (p_4^2 + 2p_3p_5) \right] AAaa \\
 & = \left[\frac{3}{14} p_2^2 + \frac{3}{7} p_1p_3 + \frac{6}{7} p_1p_4 + \frac{6}{7} p_2p_3 + \frac{18}{35} p_3^2 + \frac{36}{35} p_1p_5 + \frac{36}{35} p_2p_4 \right. \\
 & \quad \left. + \frac{6}{7} p_2p_5 + \frac{6}{7} p_3p_4 + \frac{3}{14} p_4^2 + \frac{3}{7} p_3p_5 \right] AAaa \\
 & = \left[p_2 \left(\frac{3}{7} p_1 + \frac{6}{7} p_2 + \frac{18}{35} p_3 + \frac{6}{7} p_4 + \frac{3}{7} p_5 \right) + \frac{3}{14} p_2^2 + \frac{6}{7} p_1p_4 + \frac{36}{35} p_1p_5 \right. \\
 & \quad \left. + \frac{36}{35} p_2p_4 + \frac{6}{7} p_2p_5 + \frac{3}{14} p_4^2 \right] AAaa \\
 & = \left[p_2 \left(1 - \frac{4}{7} p_1 - \frac{1}{7} p_2 - \frac{17}{35} p_3 - \frac{1}{7} p_4 - \frac{4}{7} p_5 \right) + \frac{3}{14} p_2^2 + \frac{6}{7} p_1p_4 \right. \\
 & \quad \left. + \frac{36}{35} p_1p_5 + \frac{36}{35} p_2p_4 + \frac{6}{7} p_2p_5 + \frac{3}{14} p_4^2 \right] AAaa \\
 & = \left[p_2 - \frac{4}{7} p_1p_3 + \frac{6}{7} p_1p_4 + \frac{36}{35} p_1p_5 + \frac{3}{14} p_2^2 - \frac{1}{7} p_2p_3 + \frac{36}{35} p_2p_4 + \frac{6}{7} p_2p_5 \right. \\
 & \quad \left. - \frac{17}{35} p_3^2 - \frac{1}{7} p_3p_4 - \frac{4}{7} p_3p_5 + \frac{3}{14} p_4^2 \right] AAaa
 \end{aligned}$$

Mientras que para obtener la frecuencia del gameto $Aaaa$ se obtiene de la siguiente manera:

$$\begin{aligned}
 & \left[\frac{1}{14} (2p_1p_4 + 2p_2p_3) + \frac{8}{35} (p_2^2 + 2p_1p_5 + 2p_2p_4) + \frac{3}{7} (2p_2p_5 + 2p_3p_4) + \frac{4}{7} (p_4^2 + 2p_3p_5) \right. \\
 & \quad \left. + \frac{1}{2} (2p_4p_5) \right] Aaaa \\
 & = \left[\frac{1}{7} p_1p_4 + \frac{1}{7} p_2p_3 + \frac{8}{35} p_2^2 + \frac{16}{35} p_1p_5 + \frac{16}{35} p_2p_4 + \frac{6}{7} p_2p_5 + \frac{6}{7} p_3p_4 \right. \\
 & \quad \left. + \frac{4}{7} p_4^2 + \frac{8}{7} p_3p_5 + p_4p_5 \right] Aaaa \\
 & = \left[p_4 \left(\frac{1}{7} p_1 + \frac{16}{35} p_2 + \frac{6}{7} p_3 + \frac{4}{7} p_4 + p_5 \right) + \frac{1}{7} p_2p_3 + \frac{8}{35} p_2^2 + \frac{16}{35} p_1p_5 \right. \\
 & \quad \left. + \frac{6}{7} p_2p_5 + \frac{8}{7} p_3p_5 \right] Aaaa \\
 & = \left[p_4 \left(1 - \frac{6}{7} p_1 - \frac{19}{35} p_2 - \frac{1}{7} p_3 - \frac{3}{7} p_4 \right) + \frac{1}{7} p_2p_3 + \frac{8}{35} p_2^2 + \frac{16}{35} p_1p_5 \right. \\
 & \quad \left. + \frac{6}{7} p_2p_5 + \frac{8}{7} p_3p_5 \right] Aaaa \\
 & = \left[p_4 - \frac{6}{7} p_1p_4 + \frac{16}{35} p_1p_5 + \frac{1}{7} p_2p_3 - \frac{19}{35} p_2p_4 + \frac{6}{7} p_2p_5 + \frac{8}{35} p_2^2 - \frac{1}{7} p_3p_4 \right. \\
 & \quad \left. + \frac{8}{7} p_3p_5 - \frac{3}{7} p_4^2 \right] Aaaa
 \end{aligned}$$

Finalmente, la frecuencia del gameto $aaaa$ se deriva de la siguiente manera:

$$\begin{aligned}
 & \left[\frac{1}{70} (p_3^2 + 2p_1p_5 + 2p_2p_4) + \frac{1}{14} (2p_2p_5 + 2p_3p_4) + \frac{3}{14} (p_4^2 + 2p_3p_5) + \frac{1}{2} (2p_4p_5) + p_5^2 \right] aaaa \\
 & = \left[\frac{1}{70} p_3^2 + \frac{1}{35} p_1p_5 + \frac{1}{35} p_2p_4 + \frac{1}{7} p_2p_5 + \frac{1}{7} p_3p_4 + \frac{3}{14} p_4^2 + \frac{3}{7} p_3p_5 + p_4p_5 \right. \\
 & \quad \left. + p_5^2 \right] aaaa \\
 & = \left[p_5 \left(\frac{1}{35} p_1 + \frac{1}{7} p_2 + \frac{3}{7} p_3 + p_4 + p_5 \right) + \frac{1}{70} p_3^2 + \frac{1}{35} p_2p_4 + \frac{1}{7} p_3p_4 \right. \\
 & \quad \left. + \frac{3}{14} p_4^2 \right] aaaa \\
 & = \left[p_5 \left(1 - \frac{34}{35} p_1 - \frac{6}{7} p_2 - \frac{4}{7} p_3 \right) + \frac{1}{70} p_3^2 + \frac{1}{35} p_2p_4 + \frac{1}{7} p_3p_4 \right. \\
 & \quad \left. + \frac{3}{14} p_4^2 \right] aaaa \\
 & = \left[p_5 - \frac{34}{35} p_1p_5 + \frac{1}{35} p_2p_4 - \frac{6}{7} p_2p_5 + \frac{1}{70} p_3^2 + \frac{1}{7} p_3p_4 - \frac{4}{7} p_3p_5 \right. \\
 & \quad \left. + \frac{3}{14} p_4^2 \right] aaaa
 \end{aligned}$$

Por lo que puede notarse el desbalance es bastante fuerte y complicado en los gametos e involucran todos los valores del equilibrio, principalmente los términos de los dobles productos y al menos dos términos al cuadrado; no obstante, cuando se suman todos los gametos esos desbalances se cancelan y al igual que en el caso del modelo autotetraploide hay desequilibrios que un gameto gana y es la suma de lo que pierden uno, dos o más gametos o viceversa uno pierde y los otros gametos ganan una proporción de esos desequilibrios (Tabla 4).

Tabla 4. Factores de desequilibrio de los cinco gametos en un modelo octoploide con dos alelos por *locus*.

| Gameto | Valor de equilibrio | Términos de desequilibrio | | | | | | | | | | |
|--------|---------------------|---------------------------|--------------|----------------|--------------|--------------|----------------|--------------|---------------|--------------|--------------|--------------|
| AAAA | p_1 | $-4/7p_1p_3$ | $-6/7p_1p_4$ | $-34/35p_1p_5$ | $+3/14p_2^2$ | $+1/7p_2p_3$ | $+1/35p_2p_4$ | | $+1/70p_2^2$ | | | |
| AAAa | p_2 | $+8/7p_1p_3$ | $+6/7p_1p_4$ | $+16/35p_1p_5$ | $-3/7p_2^2$ | $-1/7p_2p_3$ | $-19/35p_2p_4$ | $-6/7p_2p_5$ | $+8/35p_2^2$ | $+1/7p_2p_4$ | | |
| AaaA | p_3 | $-4/7p_1p_3$ | $+6/7p_1p_4$ | $+36/35p_1p_5$ | $+3/14p_2^2$ | $-1/7p_2p_3$ | $+36/35p_2p_4$ | $+6/7p_2p_5$ | $-17/35p_2^2$ | $-1/7p_2p_4$ | $-4/7p_2p_5$ | $+3/14p_2^2$ |
| Aaaa | p_4 | | $-6/7p_1p_4$ | $+16/35p_1p_5$ | | $+1/7p_2p_3$ | $-19/35p_2p_4$ | $+6/7p_2p_5$ | $+8/35p_2^2$ | $-1/7p_2p_4$ | $+8/7p_2p_5$ | $-3/7p_2^2$ |
| aaaa | p_5 | | | $-34/35p_1p_5$ | | | $+1/35p_2p_4$ | $-6/7p_2p_5$ | $+1/70p_2^2$ | $+1/7p_2p_4$ | $-4/7p_2p_5$ | $+3/14p_2^2$ |

Asimismo, se puede observar que el gameto heteroalélico *AAaa* es el que tiene mayor desbalance, seguido por los otros dos gametos heteroalélicos y en menor medida los dos gametos homoalélicos (Tabla 4).

Ahora, siguiendo el mismo principio que para el modelo autotetraploide se procede al cálculo del equilibrio basado en las frecuencias gaméticas basadas en las frecuencias alélicas, es decir, si vemos la frecuencia de pA y qa en el modelo autooctoploide (genotipos $2n=8x$ y gametos $n=4x$) tenemos lo siguiente:

$$\begin{aligned} & \left(\frac{4!}{4!0!} p^4 AAAA + \frac{4!}{3!1!} p^3 q AAAa + \frac{4!}{2!2!} p^2 q^2 AAaa + \frac{4!}{1!3!} p q^3 Aaaa + \frac{4!}{0!4!} q^4 aaaa \right)^2 \\ &= (p^4 AAAA + 4p^3 q AAAa + 6p^2 q^2 AAaa + 4p q^3 Aaaa + q^4 aaaa)^2 \\ &= p^8 AAAAAAAAAA + 8p^7 q AAAAAAAAAa + 28p^6 q^2 AAAAAAAAAaa + 56p^5 q^3 AAAAAAAAAaaa \\ &+ 70p^4 q^4 AAAAAaaaa + 56p^3 q^5 AAAAaaaaa + 28p^2 q^6 AAAAAaaaaa + 8p q^7 Aaaaaaaaaa \\ &+ q^8 aaaaaaaaaa \end{aligned}$$

Si seguimos el mismo principio para la asignación de las frecuencias gaméticas en base a las frecuencias alélicas de la siguiente generación se tiene para el caso del gameto *AAAA* lo siguiente:

$$\begin{aligned} & [p^8 + 1/2 (8p^7 q) + 3/14 (28p^6 q^2) + 1/14 (56p^5 q^3) + 1/70 (70p^4 q^4)] AAAA \\ &= [p^8 + 4p^7 q + 6p^6 q^2 + 4p^5 q^3 + p^4 q^4] AAAA \\ &= [p^4 (p^4 + 4p^3 q + 6p^2 q^2 + 4p q^3 + q^4)] AAAA = [p^4 (p + q)^4] AAAA = [p^4 (1^4)] AAAA \\ &= p^4 AAAA \end{aligned}$$

Para el gameto *AAAa* se deriva de la siguiente manera:

$$\begin{aligned} & [1/2 (8p^7 q) + 4/7 (28p^6 q^2) + 3/7 (56p^5 q^3) + 8/35 (70p^4 q^4) + 1/14 (56p^3 q^5)] AAAa \\ &= [4p^7 q + 16p^6 q^2 + 24p^5 q^3 + 16p^4 q^4 + 4p^3 q^5] AAAa \\ &= [4p^3 q (p^4 + 4p^3 q + 6p^2 q^2 + 4p q^3 + q^4)] AAAa = [4p^3 q (p + q)^4] AAAa \\ &= [4p^3 q (1^4)] AAAa = 4p^3 q AAAa \end{aligned}$$

Ahora para el gameto $AAaa$ se deriva de la siguiente manera:

$$\begin{aligned} & \left[\frac{3}{14} (28p^6q^2) + \frac{3}{7} (56p^5q^3) + \frac{18}{35} (70p^4q^4) + \frac{3}{7} (56p^3q^5) + \frac{3}{14} (28p^2q^6) \right] AAaa \\ & = [6p^6q^2 + 24p^5q^3 + 36p^4q^4 + 24p^3q^5 + 6p^2q^6] AAaa \\ & = [6p^2q^2(p^4 + 4p^3q + 6p^2q^2 + 4pq^3 + q^4)] AAaa = [6p^2q^2(p+q)^4] AAaa \\ & = [6p^2q^2(1^4)] AAaa = 6p^2q^2 AAaa \end{aligned}$$

Siguiendo con el gameto $Aaaa$ se obtiene lo siguiente:

$$\begin{aligned} & \left[\frac{1}{14} (56p^5q^3) + \frac{8}{35} (70p^4q^4) + \frac{3}{7} (56p^3q^5) + \frac{4}{7} (28p^2q^6) + \frac{1}{2} (8pq^7) \right] Aaaa \\ & = [4p^5q^3 + 16p^4q^4 + 24p^3q^5 + 16p^2q^6 + 4pq^7] Aaaa \\ & = [4pq^3(p^4 + 4p^3q + 6p^2q^2 + 4pq^3 + q^4)] Aaaa = [4pq^3(p+q)^4] Aaaa \\ & = [4pq^3(1^4)] Aaaa = 4pq^3 Aaaa \end{aligned}$$

Finalmente, para el gameto $aaaa$ se tiene lo siguiente:

$$\begin{aligned} & \left[\frac{1}{70} (70p^4q^4) + \frac{1}{14} (56p^3q^5) + \frac{3}{14} (28p^2q^6) + \frac{1}{2} (8pq^7) + q^8 \right] aaaa \\ & = [p^4q^4 + 4p^3q^5 + 6p^2q^6 + 4pq^7 + q^8] aaaa \\ & = [q^4(p^4 + 4p^3q + 6p^2q^2 + 4pq^3 + q^4)] aaaa = [q^4(p+q)^4] aaaa = [q^4(1^4)] Aaaa \\ & = q^4 aaaa \end{aligned}$$

Es decir, se mantiene el equilibrio siempre y cuando se utilicen las frecuencias gaméticas basadas en las frecuencias alélicas y también las frecuencias alélicas se mantienen en equilibrio considerando que el genotipo $AAAAAAA$ tiene sólo alelos A y $aaaaaaaa$ tiene sólo alelos a , mientras que el genotipo $AAAAAa$ tiene $\frac{7}{8}$ del alelo A y $\frac{1}{8}$ del alelo a y $Aaaaaaaa$ tiene $\frac{1}{8}$ del alelo A y $\frac{7}{8}$ del alelo a , el genotipo $AAAAAaa$ $\frac{3}{4}$ del alelo A y $\frac{1}{4}$ del alelo a y $AAaaaaaa$ tiene $\frac{1}{4}$ del alelo A y $\frac{3}{4}$ del alelo a , el genotipo $AAAAaaaa$ tiene $\frac{5}{8}$ del alelo A y $\frac{3}{8}$ del alelo a y el genotipo $AAaaaaaa$ tiene $\frac{3}{8}$ del alelo A y $\frac{5}{8}$ del alelo a finalmente el genotipo $AAAaaaaa$ tiene $\frac{1}{2}$ de cada uno de los alelos.

Por tanto, el cálculo de la frecuencia alélica para el alelo A es el siguiente:

$$\begin{aligned} & [p^8 + \frac{7}{8} (8p^7q) + \frac{3}{4} (28p^6q^2) + \frac{5}{8} (56p^5q^3) + \frac{1}{2} (70p^4q^4) + \frac{3}{8} (56p^3q^5) + \frac{1}{4} (28p^2q^6) \\ & + \frac{1}{8} (8pq^7)] A = [p^8 + 7p^7q + 21p^6q^2 + 35p^5q^3 + 35p^4q^4 + 21p^3q^5 + 7p^2q^6 + pq^7] A \\ & = [p(p^7 + 7p^6q + 21p^5q^2 + 35p^4q^3 + 35p^3q^4 + 21p^2q^5 + 7pq^6 + q^7)] A = [p(p+q)^7] A \\ & = [p(1^7)] A = pA \end{aligned}$$

De manera análoga para el cálculo de la frecuencia del alelo a es la siguiente:

$$\begin{aligned} & \left[\frac{1}{8} (8p^7q) + \frac{1}{4} (28p^6q^2) + \frac{3}{8} (56p^5q^3) + \frac{1}{2} (70p^4q^4) + \frac{5}{8} (56p^3q^5) + \frac{3}{4} (28p^2q^6) + \frac{7}{8} (8pq^7) \right. \\ & \left. + q^8 \right] a = [p^7q + 7p^6q^2 + 21p^5q^3 + 35p^4q^4 + 35p^3q^5 + 21p^2q^6 + 7pq^7 + q^8] a \\ & = [q(p^7 + 7p^6q + 21p^5q^2 + 35p^4q^3 + 35p^3q^4 + 21p^2q^5 + 7pq^6 + q^7)] a = [q(p+q)^7] a \\ & = [q(1^7)] a = qa \end{aligned}$$

Al igual que en el caso anterior se mantiene el equilibrio HW, es decir, el equilibrio alélico, gamético y genotípico se mantiene siempre y cuando se estén utilizando las frecuencias gaméticas basadas en las frecuencias alélicas. De otra forma se estaría extrapolando un modelo diploide de un *locus* con alelos múltiples y con su obvio desequilibrio HW que eso conlleva.

DESARROLLO DE LA GENERALIZACIÓN DEL EQUILIBRIO HW

Como se habrá notado en los desarrollos anteriores el equilibrio HW se mantiene utilizando las frecuencias alélicas. Además, se pudo notar que al duplicar el nivel de ploidía, las frecuencias gaméticas de los autotetraploides son iguales a las frecuencias genotípicas de los diploides y las frecuencias gaméticas de los autooctoploides son iguales a las frecuencias genotípicas de los autotetraploides por lo que se puede resumir el cálculo de las frecuencias genotípicas del modelo de dos alelos por *locus* con las formulas expuestas en la Tabla 5.

Esto significa que la suma de las frecuencias alélicas debe ser elevada al exponente del nivel de ploidía que se esté analizando y con esto se obtienen las frecuencias genotípicas y como se demostró anteriormente se alcanza el equilibrio HW.

Y se puede tomar la generalización de la fórmula del equilibrio HW desarrollada por Haldane (1930) y esta puede ser utilizada sin importar su nivel de ploidía ni el número de alelos mediante la siguiente expresión:

$$\left(\sum_{i=1}^k p_i A_i \right)^C \quad \text{Donde } k = \text{número de alelos por } \textit{locus} \text{ y } C = \text{Nivel de ploidía.}$$

Esto es válido siempre y cuando se trate de autopoliplóides de reproducción sexual.

Sin embargo, la dificultad es que para modelos multialélicos es difícil encontrar una única solución como si lo puede haber para diploides mediante la fórmula:

$$\left(\sum_{i=1}^n p_i A_i \right)^2 = \sum_{i=1}^n p_i^2 A_i A_i + 2 \sum_{\substack{i,j=1 \\ \forall i \neq j}}^n p_i p_j A_i A_j$$

Debido a que los coeficientes son diferentes dependiendo de cuantos alelos tengan repetidos en el genotipo y cuantos sean diferentes, pudiendo tener uno, dos, tres o cuatro alelos diferentes, un autotetraploide y entre uno y ocho alelos diferentes un autooctoploide y hay que hacer el cálculo específico para cada uno de ellos.

Por ejemplo, para el caso de un autotetraploide tendríamos que la generalización sería

$$\begin{aligned} \left(\sum_{i=1}^n p_i A_i \right)^4 = & \sum_{i=1}^n p_i^4 A_i A_i A_i A_i + 4 \sum_{\substack{i,j=1 \\ \forall i \neq j}}^n p_i^3 p_j A_i A_i A_i A_j + 6 \sum_{\substack{i,j=1 \\ \forall i \neq j}}^n p_i^2 p_j^2 A_i A_i A_j A_j + 12 \sum_{\substack{i,j,k=1 \\ \forall i \neq j \neq k}}^n p_i^2 p_j p_k A_i A_i A_j A_k \\ & + 24 \sum_{\substack{i,j,k,l=1 \\ \forall i \neq j \neq k \neq l}}^n p_i p_j p_k p_l A_i A_j A_k A_l \end{aligned}$$

Y para el caso de un modelo autooctoploide se derivaría lo siguiente

$$\begin{aligned}
 \left(\sum_{i=1}^n p_i A_i\right)^8 = & \sum_{i=1}^n p_i^8 A_i A_i A_i A_i A_i A_i A_i A_i + 8 \sum_{\substack{i,j=1 \\ \forall i \neq j}}^n p_i^7 p_j A_i A_i A_i A_i A_i A_i A_i A_j + 28 \sum_{\substack{i,j=1 \\ \forall i \neq j}}^n p_i^6 p_j^2 A_i A_i A_i A_i A_i A_i A_j A_j \\
 & + 56 \sum_{\substack{i,j,k=1 \\ \forall i \neq j \neq k}}^n p_i^6 p_j p_k A_i A_i A_i A_i A_i A_i A_j A_k + 56 \sum_{\substack{i,j=1 \\ \forall i \neq j}}^n p_i^5 p_j^3 A_i A_i A_i A_i A_i A_j A_j \\
 & + 168 \sum_{\substack{i,j,k=1 \\ \forall i \neq j \neq k}}^n p_i^5 p_j^2 p_k A_i A_i A_i A_i A_i A_j A_j A_k + 336 \sum_{\substack{i,j,k,l=1 \\ \forall i \neq j \neq k \neq l}}^n p_i^5 p_j p_k p_l A_i A_i A_i A_i A_j A_k A_l \\
 & + 70 \sum_{\substack{i,j=1 \\ \forall i \neq j}}^n p_i^4 p_j^4 A_i A_i A_i A_i A_j A_j A_j A_j + 280 \sum_{\substack{i,j,k=1 \\ \forall i \neq j \neq k}}^n p_i^4 p_j^3 p_k A_i A_i A_i A_i A_j A_j A_k \\
 & + 420 \sum_{\substack{i,j,k=1 \\ \forall i \neq j \neq k}}^n p_i^4 p_j^2 p_k^2 A_i A_i A_i A_i A_j A_j A_k A_k + 840 \sum_{\substack{i,j,k,l=1 \\ \forall i \neq j \neq k \neq l}}^n p_i^4 p_j^2 p_k p_l A_i A_i A_i A_i A_j A_k A_l \\
 & + 420 \sum_{\substack{i,j,k=1 \\ \forall i \neq j \neq k}}^n p_i^4 p_j^2 p_k^2 A_i A_i A_i A_i A_j A_j A_k A_k + 840 \sum_{\substack{i,j,k,l,t=1 \\ \forall i \neq j \neq k \neq l \neq t}}^n p_i^4 p_j p_k p_l p_t A_i A_i A_i A_i A_j A_k A_l A_t \\
 & + 560 \sum_{\substack{i,j,k=1 \\ \forall i \neq j \neq k}}^n p_i^3 p_j^3 p_k^2 A_i A_i A_i A_j A_j A_j A_k A_k + 1120 \sum_{\substack{i,j,k,l=1 \\ \forall i \neq j \neq k \neq l}}^n p_i^3 p_j^3 p_k p_l A_i A_i A_i A_j A_j A_k A_l \\
 & + 1680 \sum_{\substack{i,j,k,l=1 \\ \forall i \neq j \neq k \neq l}}^n p_i^3 p_j^2 p_k^2 p_l A_i A_i A_i A_j A_j A_k A_k A_l + 3360 \sum_{\substack{i,j,k,l,t=1 \\ \forall i \neq j \neq k \neq l \neq t}}^n p_i^3 p_j^2 p_k p_l p_t A_i A_i A_i A_j A_j A_k A_l A_t \\
 & + 6720 \sum_{\substack{i,j,k,l,t,u=1 \\ \forall i \neq j \neq k \neq l \neq t \neq u}}^n p_i^3 p_j p_k p_l p_t p_u A_i A_i A_i A_j A_k A_l A_t A_u \\
 & + 2520 \sum_{\substack{i,j,k,l=1 \\ \forall i \neq j \neq k \neq l}}^n p_i^2 p_j^2 p_k^2 p_l^2 A_i A_i A_j A_j A_k A_k A_l A_l + 5040 \sum_{\substack{i,j,k,l,t=1 \\ \forall i \neq j \neq k \neq l \neq t}}^n p_i^2 p_j^2 p_k^2 p_l p_t A_i A_i A_j A_j A_k A_k A_l A_t \\
 & + 10080 \sum_{\substack{i,j,k,l,t,u=1 \\ \forall i \neq j \neq k \neq l \neq t \neq u}}^n p_i^2 p_j^2 p_k p_l p_t p_u A_i A_i A_j A_j A_k A_l A_t A_u \\
 & + 20160 \sum_{\substack{i,j,k,l,t,u,v=1 \\ \forall i \neq j \neq k \neq l \neq t \neq u \neq v}}^n p_i^2 p_j p_k p_l p_t p_u A_i A_i A_j A_k A_l A_t A_u A_v \\
 & + 40320 \sum_{\substack{i,j,k,l,t,u,v,w=1 \\ \forall i \neq j \neq k \neq l \neq t \neq u \neq v \neq w}}^n p_i p_j p_k p_l p_t p_u p_w A_i A_j A_k A_l A_t A_u A_v A_w
 \end{aligned}$$

Tabla 5. Fórmulas para el cálculo de frecuencias genotípicas a partir de las frecuencias alélicas modelo de dos alelos por locus en diferentes niveles de ploidía.

| Nivel de ploidía | Fórmula para calcular las frecuencias genotípicas |
|------------------|--|
| $2n=2x$ | $(pA + qa)^2 = p^2 + 2pq + q^2$ |
| $2n=4x$ | $(pA + qa)^4 = p^4AAAA + 4p^3qAAAa + 6p^2q^2AAaa + 4pq^3Aaaa + q^4aaaa$ |
| $2n=6x$ | $(pA + qa)^6 = p^6AAAAAA + 6p^5qAAAAAa + 15p^4q^2AAAAaa + 20p^3q^3AAAAaaa + 15p^2q^4AAaaaa + 6pq^5Aaaaaa + q^6aaaaaa$ |
| $2n=8x$ | $(pA + qa)^8 = p^8AAAAAAA + 8p^7qAAAAAAAa + 28p^6q^2AAAAAaa + 56p^5q^3AAAAaaa + 70p^4q^4AAAAaaaa + 56p^3q^5AAAAaaaa + 28p^2q^6AAAAaaaaa + 8pq^7Aaaaaaaa + q^8aaaaaaaa$ |

DISCUSIÓN

Como se pudo observar el equilibrio HW en autopoliplóides debe ser calculado por métodos apropiados para no tener un desequilibrio por generalizar la estructura de un modelo diploide. Esto es importante porque muchos análisis genéticos se basan en el equilibrio HW, desde el cálculo de frecuencias genotípicas y heterocigocidad así como las pruebas F (F_{ST} , F_{IT} y F_{IS}) de Wright (Wright, 1949; 1965), el estadístico G_{ST} de Nei (1973) y los métodos bayesianos (Pritchard *et al.*, 2000; Falush *et al.*, 2003).

Haldane (1930) desarrolló las bases para la generalización de la ley del equilibrio HW para poliploides, las cuales también fueron revisadas y ampliadas por Wright (1938), Fisher y Mather (1942), Fisher (1941; 1943) y Little (1945); más recientemente Moody *et al.* (1993) desarrollaron la teoría completa sobre el cálculo de varios parámetros genéticos en autotetraploides, desarrollos similares fueron obtenidos por Vargas (2000) también para tetraploides. No obstante, el asunto sigue siendo de interés por no generar aún los algoritmos para cualquier tipo de ploidía en los programas actuales de análisis genético (Dufresne *et al.*, 2014).

La ley del equilibrio HW es una buena herramienta para estimar las frecuencias genotípicas, la heterocigocidad y ser el soporte de las otras pruebas genéticas antes mencionadas, si se están calculando las frecuencias genotípicas de poliploides de manera incorrecta con el modelo diploide, entonces todos los cálculos de parámetros genéticos están sesgados, lo que puede traducirse en una mala interpretación

de la genética de poblaciones propia para estas poblaciones poliploides, es decir, si su cálculo se hace con el modelo de ploidía adecuado entonces se puede mantener el equilibrio, si no se hace así y se generaliza el modelo diploide, lleva a tener un desequilibrio HW por hacer mal uso del modelo apropiado y no por desequilibrios que pueden darse de manera natural por mutación, selección o deriva genética (Stern, 1943).

BIBLIOGRAFÍA

- Correns C. (1900) G. Mendels Regel Über das Verhalten der Nachkommenschaft der Rassenbastarde. *Berichte der Deutschen Botanischen Gesellschaft* 18: 158-168. <http://www.esp.org/foundations/genetics/classical/holdings/c/cc-00.pdf>.
- Crow J.F. (1988) Eighty Years Ago: The Beginnings of Population Genetics. *Genetics* 119 (3): 473-476. <http://www.genetics.org/content/119/3/473>.
- Crow J.F. (1999) Hardy, Weinberg and Language Impediments. *Genetics* 152 (3): 821-825. <http://www.genetics.org/content/152/3/821>.
- De Vries H. (1900) Sur la loi de disjonction des hybrides. *Comptes Rendus de l'Academie des Sciences (Paris)*, 130: 845-847. <http://www.esp.org/foundations/genetics/classical/holdings/v/hdv-00.pdf>.

- De Vries H. (1902) The Origin of Species by Mutation. *Science* 15 (384): 721-729. <http://www.jstor.org/stable/1629679>.
- Dufresne F., Stift M., Vergilino R., Mable B.K. (2014) Recent Progress and Challenges in Population Genetics of Polyploids Organisms: An Overview of Current State-of-the-Art Molecular and Statistical Tools. *Molecular Ecology* 23: 40-69. DOI: 10.1111/mec.12581.
- Falush D., Stephens M., Pritchard J.K. (2003) Inference of Population Structure Using Multilocus Genotype Data: Linked Loci and Correlated Allele Frequencies. *Genetics* 164 (4): 1567-1587. <http://www.genetics.org/content/164/4/1567.full>.
- Fisher R.A. (1941) The Theoretical Consequences of Polyploid inheritance for the mid style form of *Lythrum salicaria*. *Annals of Eugenics* 11 (1): 31-38. DOI: 10.1111/j.1469-1809.1941.tb02268.x.
- Fisher R.A. (1943) Allowance for Double Reduction in the Calculation of Genotype Frequencies with Polysomic Inheritance. *Annals of Eugenics* 12 (1): 169-171. DOI: 10.1111/j.1469-1809.1943.tb02320.x.
- Fisher R.A., Mather K. (1942) Polyploid inheritance in *Lythrum salicaria*. *Nature* 150: 143. DOI: 10.1038/150430a0.
- Haldane J.B.S. (1930) Theoretical genetics of autopolyploids. *Journal of Genetics* 22 (3): 359-372. DOI: 10.1007/BF02984197.
- Hardy G.H. (1908) Mendelian Proportion in a Mixed Populations. *Science* 28 (706): 49-50. <http://www.jstor.org/stable/1636004>.
- Little T.M. (1945) Gene Segregation in Autotetraploids. *Botanical Review* 11 (1): 60-85. <http://www.jstor.org/stable/4353314>.
- Mendel G. (1866) Versuche über pflanzen-hybriden. *Verhandlungen des naturforschenden Vereines in Brünn* 4: 3-47. <http://www.esp.org/foundations/genetics/classical/gm-65.pdf>.
- Moody M.E., Muller L.D., Soltis D.E. (1993) Genetic Variation and Random Drift in Autotetraploid Populations. *Genetics* 134 (2): 649-657. <http://www.genetics.org/content/134/2/649>.
- Morgan T.H., Sturtevant A.H., Muller H.J., Bridges C.B. (1915) The Mechanism of Mendelian Heredity. New York, Henry Holt and Company.
- Nei M. (1973) Analysis of Gene Diversity in Subdivided Populations. *Proceedings of the National Academy of Science of U.S.* 70 (12): 3321-3323. DOI: 10.1073/pnas.70.12.3321.
- Peakall R., Smouse P.E. (2012) GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research—an update. *Bioinformatics* 28: 2537-2539. DOI: 10.1093/bioinformatics/bts460.
- Pritchard J.K., Stephens M., Donnelly P. (2000) Inference of Population Structure Using Multilocus Genotype Data. *Genetics* 155 (2): 945-959. <http://www.genetics.org/content/155/2/945.full>.
- Stern C. (1943) The Hardy-Weinberg Law. *Science* 97 (2510): 137-138. DOI: 10.1126/science.97.2510.137.
- Vargas J.A. (2000) Hardy-Weinberg Theory for Tetraploidy with Mixed Mating. *Advances in Applied Mathematics* 24: 369-383. DOI: 10.1006/aama.1999.0680.
- Wienberg W. (1908) Über Vererbungsgesetze beim Menschen. *Zeitschrift für Induktive Abstammungs- und Vererbungslehre* 1 (1): 440-460. DOI: 10.1007/BF01990626.
- Wright S. (1938) The Distribution of Gene Frequencies in Populations of Polyploids. *Proceedings of the National Academy of Sciences of the United States of America* 24 (9): 372-377. <http://www.jstor.org/stable/86892>.
- Wright S. (1949) The Genetical Structure of Populations. *Annual of Eugenics* 15 (1): 323-354. DOI: 10.1111/j.1469-1809.1949.tb02451.
- Wright S. (1965) The Interpretation of Population Structure by F-Statistics with Special Regard to

Systems of Mating. *Evolution* 19 (3): 395-420. <http://www.jstor.org/stable/2406450>.

Yeh F.C., Yang R., Boyle T. (1999) POPGENE version 1.31, Microsoft Window-based Freeware for Population Genetic Analysis, Quick User Guide. Edmonton: University of Alberta and Centre for International Forestry Research

AGRADECIMIENTOS

A la Universidad de Guanajuato por el financiamiento mediante el proyecto CIFOREA 012/2015. Asimismo, se agradece al PRODEP por el financiamiento mediante el proyecto DSA/103.5/15/7007.



CHILEAN PLANTS CYTOGENETIC DATABASE: COVERAGE, FEATURES AND USAGES

BASE DE DATOS CITOGÉNÉTICOS DE PLANTAS CHILENAS: COBERTURA, CARACTERÍSTICAS Y USOS

Jara-Seguel P.^{1,2}, Urrutia J.^{1*}

¹ Departamento de Ciencias Biológicas y Químicas.

² Núcleo de Estudios Ambientales (NEA), Facultad de Recursos Naturales,
Universidad Católica de Temuco, Rudecindo Ortega 2950,
Temuco-Chile.

* Corresponding author: pjara@uct.cl

ABSTRACT

Chilean Plants Cytogenetic Database (CPCD) is a resource available on line in electronic format, which provides a cytogenetics catalogue for continental and insular Chilean vascular plants. In this report, we made reference to the CPCD, discussing aspects on its coverage, features and usages. Currently, the database stores cytogenetic information for 247 species, 107 genera, and 55 families belonging to Pteridophyta, Pinophyta, and Magnoliophyta.

Key words: chromosome number, karyotype, Chilean Plants

RESUMEN

La Base de Datos Citogenéticos de Plantas Chilenas (CPCD) es un recurso disponible en formato electrónico en línea, que provee de un catálogo citogenético para plantas vasculares de Chile continental e insular. En esta comunicación hacemos referencia a la CPCD, discutiendo aspectos sobre su cobertura, rasgos y usos. Actualmente, la base de datos contiene información citogenética para 247 especies, 107 géneros y 55 familias pertenecientes a Pteridophyta, Pinophyta y Magnoliophyta.

Palabras clave: número cromosómico, cariotipo, plantas chilenas

Fecha de recepción: 23/01/2017
Fecha de aceptación de versión final: 10/02/2018

INTRODUCTION

The electronic databases have taken relevance in biological investigation from the beginning of the genomic revolution. The most used databases –the so-called GenBank– are a tool related to the molecular biology but with multiple applications in different fields of the biological sciences. These databases contain public available nucleotide sequences and protein translation for more than 260,000 species (Benson *et al.*, 2013). Other available databases most related to cytology store information on genome size (or DNA C-values) for a high number of Eukaryote species including plants, animals, fungi, and algae (more than 10,000 species according to Gregory *et al.*, 2006). In the specific case of plants, current reports estimate that the only existing electronic catalogue on genome sizes comprises data for more than 8,000 species of around the world (Bennett and Leitch, 2016). Thus, this database provides a truly comprehensive catalogue on plant genome size, and represents a much-needed resource for cytogeneticists, with innumerable applications that have been broadly discussed in the literature (Bennett and Leitch, 2005; Leitch *et al.*, 2005; Peruzzi *et al.*, 2009; Roa and Pires de Campos Telles, 2017).

Regarding on line resources for plant chromosomes, fourteen cytogenetic databases in electronic format are available around the world (Peruzzi and Bedini, 2014). For all of them, data have been obtained from different bibliographic sources, being chromosome number ($2n$, n) the main presented character. Chilean plants have a scarce representation in most databases, even though many species have been included during the last decades in the Index to Plant Chromosome Number (IPCN, Goldblatt and Johnson, 2006), together with species from other latitudes. This low representation of Chilean plants in databases makes it difficult to conduct literature reviews aimed to explain tendencies on chromosome evolution in a wide geographic scale, including species from continental and insular Chile. It is important to remark that continental Chile is a mosaic of environments along its latitudinal configuration, having more than 4,000 km in length from 18° S to 56° S, with marked climatic North to South gradients. In addition, the longitudinal gradient of Chile is marked by differences in altitudes (up to 6,000 m.a.s.l.) which originate a variation of microclimates and, thus, of environments, many of which are favorable for plant growth. Of this territory, 55% is covered by vegetation

with 45% of floristic endemism. In addition, insular Chile is also an interesting mosaic with 320 km² in surface, and an endemism that varies between 64% and 87% (Villagrán and Hinojosa, 2005; CONAF, 2013; Urbina-Casanova *et al.*, 2015). Therefore, cytogenetic studies on spatial scale on the Chilean flora are of relevance to analyze tendencies on genome variation along geographic gradients. The first review reporting the state of the art on cytogenetic of Chilean angiosperms emphasized on these matters as well as on the difficulties encountered to compile the information, specifically for data documented in the oldest literature (Jara-Seguel and Urrutia, 2012). At present, theoretical and/or empirical investigations on different fields of biology (*i.e.*, cytogeography, genomics, taxonomy, evolution, and/or palaeobotany) require the data to be readily available and compiled from diverse sources. Thus, at present, databases constitute a valuable resource in which the information can be easily accessible to the scientific community (Gregory *et al.*, 2006; Jara-Seguel and Urrutia, 2012; Roa and Pires de Campos Telles, 2017).

In this communication we make reference to the Chilean Plants Cytogenetic Database (CPCD, Jara-Seguel and Urrutia, 2016), and discuss aspects on its coverage, features and uses.

COVERAGE, FEATURES AND USAGES

The CPCD provides the first electronic catalogue exclusive for the Chilean flora, and was launched for the first time in September 2010, with 101 recorded species. Since then, two updates were made, the first in March 2011, which increased the number of species to 163 (release 2.0), and the second in June 2016, in which this number was increased to 247 (release 3.0). The latter number of species represents almost 4.8% of the Chilean Vascular Plants. The adscription of the species to families and/or genera was based on the classification given in the original source cited in the references. The CPCD assembles information obtained from 88 bibliographic sources covering 55 years of investigation on the topic. In total, CPCD contains information for 247 species representing 107 genera and 55 families belonging to Pteridophyta, Pinophyta, and Magnoliophyta. As remarkable contribution, the database includes a wide gamma of cytogenetic characters [chromosome number, karyotype morphology, banding patterns (C, Ag-NOR, CMA3 and DAPI), nuclear DNA

Table 1. Cytogenetic data for each plant division recorded in the Chilean Plants Cytogenetic Database (CPCD) and the current percentage of species representation in the database regarding to the total Chilean vascular flora.

| Division | Chromosome number (2n) | | Ploidy level | DNA C-values (pg) | Number of families in CPCD | Number of species in CPCD | Number of species in Chile | Representation of species in CPCD (%) |
|---------------|------------------------|-----|--------------|-------------------|----------------------------|---------------------------|----------------------------|---------------------------------------|
| | Min | Max | | | | | | |
| Pteridophyta | 44 | 328 | 2, 3, 4, 8 | - | 5 | 24 | 167 ⁽¹⁾ | 14.3 |
| Pinophyta | 22 | 44 | 2, 4 | 10.2 - 25.8 | 2 | 3 | 16 ⁽²⁾ | 18.7 |
| Magnoliophyta | 8 | 94 | 2, 4 | 1.46 – 39.4 | 48 | 220 | 4.946 ⁽²⁾ | 4.4 |

(1)Based in Rodríguez (1995).

(2)Based in Marticorena (1990).

content (C-value) and/or fluorescent *in situ* hybridization (rDNA 5S, 18S, 28S)], which is advantageous regarding other databases focused only on chromosome numbers or rDNA chromosome location. Moreover, all bibliographic references are appended to the cytogenetic data of each species. Several statistical details on the database coverage are summarized in Table 1.

The cytogenetic data comprised in the CPCD increases the potential applications of the accumulated knowledge on genome structure and diversity of Chilean native plants, focused on relevant cytogenetic characters. In general, it is remarkable that the higher number of species included in the database (196 continental species) are representative of discontinuous latitudinal strips localized between 18° S and 53° S, and most of them (192 species) are found within the Chilean Hotspots (Myers *et al.*, 2000). Nevertheless, cytogenetic data of species that inhabit in the Juan Fernández Archipelago and Isla de Pascua are also included (51 insular species).

The database is a comprehensive, ordered, easily accessible and useful catalogue for experimented and incipient cytogeneticist interested in cytogenetic aspects of Chilean plants or for comparative investigations with co-familial members that inhabit other geographic regions. The database is of free access, and the traffic has increased steadily since its launching in September 2010, with more than 7,400 visitors belonging to 46 countries of around the world. From the time of launching and on, the database has been cited in seven publications, some of them analyzing

tendencies on chromosome number variation of global or local floras (Jara-Seguel and Urrutia, 2012; Peruzzi *et al.*, 2012; Rice *et al.*, 2015), and others discussing the usefulness of databases assembling data on plant rDNA chromosome localization or genome size (García *et al.*, 2012; Dagher-Kharrat *et al.*, 2013). Recently, CPCD has been included in the largest community resource of plants chromosome numbers at the worldwide level (see CCDB, Rice *et al.*, 2015), which may increase its visibility and use. It is also listed as an on line resource for chromosome numbers among other 17 databases available around the world (Peruzzi and Bedini, 2014). In fact, the data on Chilean plants pooled in the CPCD not only have been used to analyze tendencies, but have also served as substrate to propose new quantitative methods to analyze large data sets of chromosome numbers of animals and plants (see Peruzzi *et al.*, 2014). Thus, the CPCD is a valuable tool for generating knowledge in different cytogenetic fields.

FUTURE PROSPECTS

The CPCD could be annually or biannually updated, as new cytogenetic information may published or obtained from old sources. However, for continuously being an important source of information for cytogeneticists and/or taxonomists, it is necessary to increase the number of cytogenetic studies in all taxonomic divisions that represent the Chilean plant biodiversity.

BIBLIOGRAPHY

- Bennett M., Leitch I. (2005) Plant genome size research: A field in focus. *Ann. Bot.* 95 (1): 1-6.
- Bennett M., Leitch I. (2016) Plant DNA C-values database. URL: <http://data.kew.org/cvalues/> (Viewed: January 08, 2018).
- Benson D., Cavanaugh M., Clark K., Karsch-Mizrachi I., Lipman D., Ostell J., Sayers E. (2013) GenBank. *Nucleic Acids Res.* 42: 36-42.
- Conaf (2013) Por un Chile forestal sustentable. Corporación Nacional Forestal, Santiago, Chile.
- Dagher-Kharrat M., Abdel-Samad N., Douaihy B., Bourge M., Firdlender A., Siljak-Yakovlev S., Brown S. (2013) Nuclear DNA C-values for biodiversity screening: Case of the Lebanese flora. *Plant Biosyst.* 147 (4): 1128-1237.
- García S., Garnatje T., Kovařík A. (2012) Plant rDNA database: ribosomal DNA loci information goes online. *Chromosoma* 121 (4): 389-394.
- Goldblatt P., Johnson D. (2006) Index to Plant Chromosome Numbers 2001-2003. *Syst. Bot.* 106: 1-242.
- Gregory R., Nicol J., Tamm H., Kullman B., Kullman K., Leitch I., Murray B., Kapraun D., Greilhuber J., Bennett M. (2006) Eukaryotic genome size databases. *Nucleic Acids Res.* 35: 332-338.
- Jara-Seguel P., Urrutia J. (2012) Cytogenetics of Chilean Angiosperms: Advances and prospects. *Rev. Chil. Hist. Nat.* 85: 1-12.
- Jara-Seguel P., Urrutia J. (2016) Chilean Plants Cytogenetic Database. URL: <http://www.chileanpcd.com/>. (Viewed: January 15, 2018).
- Leitch I., Soltis D., Soltis P., Bennett M. (2005) Evolution of DNA amounts across Land Plants (Embryophyta). *Ann. Bot.* 95: 207-217.
- Martícorena C. (1990) Contribución a la estadística de la flora vascular de Chile. *Gayana Bot.* 47 (3-4): 85-113.
- Myers N., Mittermeier R., Mittermeier C., Fonseca G., Kent J. (2000) Biodiversity hotspots for conservation priorities. *Nature* 403: 853-858.
- Peruzzi, L., Bedini, G. (2014) Online resources for chromosome number databases. *Caryologia* 67(4): 292-295.
- Peruzzi L., Caparelli K., Bedini G. (2014) A new index for the quantification of chromosome number variation: An application to selected animal and plant groups. *J. Theor. Biol.* 353: 55-60.
- Peruzzi L., Góralski G., Joachimiak A., Bedini G. (2012) Does actually mean chromosome number increase with latitude in vascular plants? An answer from the comparison of Italian, Slovak and Polish floras. *Comp. Cytogen.* 6 (4): 371-377.
- Peruzzi L., Leitch I., Caparelli K. (2009) Chromosome diversity and evolution in Liliaceae. *Ann. Bot.* 103: 459-475.
- Rice A., Glic L., Abadi S., Einhorn M., Kopelman N., Salman-Minkov A., Mayzel J., Chay O., Mayrose I. (2015) The Chromosome Counts Database (CCDB) – a community resource of plant chromosome numbers. *New Phytol.* 206: 19-26.
- Roa E., Pires de Campos Telles M. (2017) The Cerrado (Brazil) plant cytogenetics database. *Comp. Cytogen.* 11 (2): 285-297.
- Rodríguez R. (1995) Pteridophyta. In: Martícorena C., Rodríguez R. (Eds.) *Flora de Chile*. Ediciones Universidad de Concepción, Concepción, Chile, pp. 119-350.
- Urbina-Casanova R., Saldivia P., Scherson R. (2015) Consideraciones sobre la sistemática de las familias y los géneros de plantas vasculares endémicos de Chile. *Gayana Bot.* 72 (2): 272-295.

Villagrán C., Hinojosa L. (2005) Esquema biogeográfico de Chile. In: Llorente-Bousquets J., Morrone J. (Eds.) Regionalización biogeográfica en Iberoamérica y tópicos afines. Ediciones de la Universidad Nacional Autónoma de México, México, pp. 551-577.

ACKNOWLEDGEMENTS

To Núcleo de Estudios Ambientales (NEA), Universidad Católica de Temuco, for funding the current implementation of the Chilean Plants Cytogenetic Database (CPCD). Our gratefulness to the authors cited in the database for their valuable contribution to the knowledge of the Chilean flora.

OBITUARY

Dr. Eduardo E. Castilla

Buenos Aires, 09/10/1933 – 31/12/2017



Nació en Buenos Aires en 1933. Médico por la Universidad de Buenos Aires y doctor en Genética Médica por la Universidad Federal de São Paulo. Fue miembro fundador y director del Centro de Genética Médica de Buenos Aires, el cual hoy lleva su nombre, y del Estudio Colaborativo Latino Americano de Malformaciones Congénitas (ECLAMC) desde 1967. Investigador Superior del CONICET, profesor titular de Genética Médica en el Instituto Universitario CEMIC en Buenos Aires e investigador visitante del Consejo Nacional de Pesquisa (CNPq) en la Fundación Oswaldo Cruz (FIOCRUZ), Río de Janeiro. Miembro, desde 2008, de la Academia Brasileira de Ciencias y de la Academia Nacional de Medicina de Uruguay. Recibió el premio Konex de Platino 2003.



BAG
Journal of Basic & Applied Genetics