Construcción de un mapa genético tetraploide de papa e identificación de QTLs usando SNPs a partir de la población *Atlantic* x *Superior*

Cinthya Zorrilla*, Felix Navarro, Sandra Vega y Jiwan Palta
Departamento de Horticultura, Universidad de Wisconsin – Madison, Linden Drive 1575, Madison, USA
*cinzocis@gmail.com, fmnavarro@wisc.edu, sandra.vega-semorile@wisc.edu, jppalta@wisc.edu

RESUMEN

Uno de los principales intereses para los mejoradores de papa en los Estados Unidos de Norteamérica (EE.UU.) es mejorar la calidad para procesamiento ya que el consumo se da mayormente como papas fritas, hojuelas, deshidratada, y en menor proporción en fresco. En este trabajo se elaboró un mapa genético empleando una población biparental de 151 individuos producto del cruce entre dos variedades para hojuelas, *Atlantic* (parental femenino) y *Superior* (parental masculino). Los SNPs se generaron usando el *chip 8K Illumina Infinium Bead SNP* desarrollado por la iniciativa "*Solanaceae Coordinated Agricultural Project*" (SolCAP, por sus siglas en inglés). Los datos fenotípicos y genotípicos fueron analizados para obtener dos mapas genéticos que cubren 1254 cM para *Atlantic* y 939 cM para *Superior*. Asimismo, se identificaron loci para características cuantitativas (QTLs) de importancia comercial para las papas para hojuelas en los EE.UU; así como para el contenido de calcio en el tubérculo.

INTRODUCCION

La construcción de mapas genéticos es una herramienta importante para ubicar marcadores moleculares dentro de los cromosomas y permiten correlacionar la información genotípica con la información fenotípica de manera que se puedan identificar las regiones del genoma que controlan características cuantitativas (QTLs). Una de las aplicaciones del mapeo de QTLs es la identificación de genes mayores; así como la identificación de marcadores ligados que podrían ser de utilidad por ejemplo para el mejoramiento asistido por marcadores.
Las variedades *Atlantic* y *Superior* presentan características muy divergentes entre sí, *Atlantic* presenta alto rendimiento, alta gravedad específica pero altas incidencias de defectos internos y de sarna con lesiones profundas, y bajo contenido de calcio en el tubérculo. Por otro lado, *Superior* presenta características opuestas. En consecuencia, la población *Atlantic* x *Superior* presenta una amplia variación fenotípica (Zorrilla *et al*., 2014), lo que la hace muy útil para la identificación de regiones QTLs.

OBJETIVO

El objetivo del presente trabajo es construir un mapa genético tetraploide usando marcadores SNPs derivados del uso del *8K Illumina Infinium Bead SNP chip* y determinar las regiones del genoma de *Atlantic* y *Superior* que contienen loci que controlan rendimiento, gravedad específica, defectos internos, sarna con lesiones profundas, y contenido de calcio en el tubérculo.

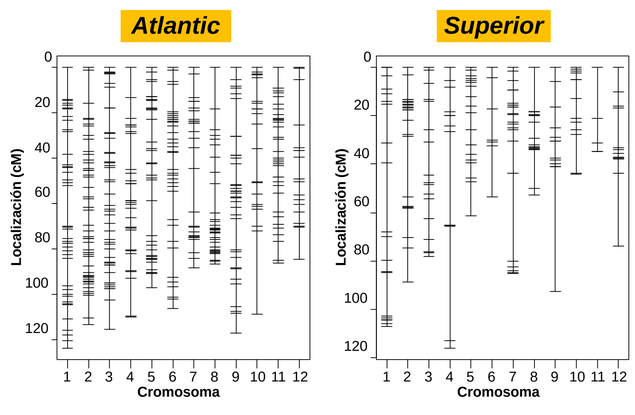
MATERIALES Y METODOS

Material genético: 151 individuos de la población *Atlantic* x *Superior*
Caracterización fenotípica: La evaluación se realizó en la Estación Experimental de Hancock en Wisconsin, EE.UU. Las características evaluadas fueron rendimiento, gravedad específica, color de las hojuelas, e incidencias de defectos internos e incidencia, sarna con lesiones profundas y contenido de calcio en el tubérculo. Los resultados de la caracterización fenotípica fueron presentados por Zorrilla *et al.* (2014).
Caracterización molecular: Se realizó empleando el *8K Illumina Infinium Bead SNP chip* desarrollado por SolCAP. Estos SNPs fueron identificados a partir del secuenciamiento de ESTs de las variedades *Bintje*, *Kennebec* y *Shepody*; así como del secuenciamiento de ARN de las variedades *Atlantic*, *Premier Russet* y *Snowden*; y para determinar el dosaje alélico se realizó mediante un análisis clúster para cada locus SNP usando el programa *Illumina GenomeStudio* (Hamilton *et al.,* 2011). El genotipado se realizó en el Instituto de Biotecnología de la Universidad de California Davis, EE.UU.
Construcción del mapa tetraploide e identificación de QTLs: Se empleó el programa *TetraploidMap* desarrollado por la Universidad de Wageningen (Hackett *et al*., 2007) y el paquete qtl (Broman *et al*., 2003) del programa R (2019). Para ello se filtraron los SNPs usando diversos criterios para la limpieza de datos, tales como contar con una secuencia homologa en el genoma de referencia, porcentaje de datos faltantes entre otras usando el programa R (**Tabla 1**). Seguidamente, se realizó el agrupamiento de marcadores dentro de cada grupo de ligamiento (cromosoma) y su ordenamiento. Luego, se determinó el tipo de fase, repulsión o acoplamiento. Posteriormente, se asignaron los marcadores a cada cromosoma homólogo y finalmente se construyó el mapa genético de cada parental. Luego de la construcción de los mapas genéticos se procedió a la identificación de los QTLs usando la metodología de mapeo por intervalos en TetraploidMap. El límite inferior del valor del LOD se determinó usando un análisis de permutaciones con 100 replicaciones.

Tabla 1. Procedimiento de limpieza de datos para mapeo genético

Pasos para limpieza de datos	Número de SNPs	Número de Individuos
Número inicial de SNPs	8303	184
Remover SNPs de baja calidad	637	
SNPs que no fueron localizados en el genoma de referencia	649	
SNPs con mas de un hit en el genoma de referencia	86	
SNPs sin datos	2326	
SNPs localizados en dos posiciones del genoma	107	
SNPs ausentes en ambos padres	414	
Eliminar genotipos mezclados y duplicados		7
Eliminar los cruces diferentes a nulliplex x simplex y viceversa	3117	
Remover individuos con muy alta doble reducción	6	
Eliminar marcadores duplicados	285	
Eliminar marcadores con alta distorsión de la segregación (<0.05/#markers)	49	
Eliminar marcadores con <100 ind genotipados	9	
Eliminar genotipos con <580 marcadores genotipados		11
Eliminar genotipos duplicados	1	15
Eliminar marcadores debido a la distorsión de segregación <0.001	5	
Eliminar debido al ligamiento a multiples homólogos	14	
Número final de SNPs usados en el mapeo	600	151

Figura 1. Mapas genéticos de *Atlantic* (414 SNPs) y *Superior* (186 SNPs) obtenidos usando loci del tipo *simplex* (ABBB) x *nulliplex* (BBBB) y *nulliplex* (BBBB) x *simplex* (ABBB).



RESULTADOS Y DISCUSIÓN

Producto del análisis genotípico se obtuvieron 8303 SNPs, luego del proceso de limpieza de datos y la selección de los marcadores *simplex* (ABBB) x *nulliplex* (BBBB) y *nulliplex* (BBBB) x *simplex* (ABBB), se obtuvieron 600 marcadores SNPs que se emplearon para la elaboración del mapa genético tetraploide usando la población *Atlantic* x *Superior*. Se mapearon 414 marcadores en *Atlantic* y 186 marcadores en Superior (**Figura 1**). Se puede concluir que el *8K Illumina Infinium Bead SNP chip* es útil para la generación de mapas tetraploides de papa debido a que es posible determinar el dosage de los alelos en cada loci en comparación con el uso de marcadores SSR y AFLP.
El mapeo genético a nivel tetraploide usando la población *Atlantic* x *Superior* fue usado exitosamente para detectar QTLs para características como rendimiento, gravedad específica, color de las hojuelas e incidencias de defectos internos, sarna con lesiones profundas y contenido de calcio en el tubérculo.
Por ejemplo, en el cromosoma 12 se encontraron 4 QTL: *i*) 1 para incidencia de sarna con lesiones profundas, *ii*) 2 para gravedad especifica y *iii*) 1 para contenido de calcio en el tubérculo (**Figura 2**). Asimismo, se puede observar que estos QTL explican alrededor del 10% de la varianza fenotípica, lo que indicaría que se trata de genes menores (**Tabla 2**).
Actualmente, existen mapas tetraploides con alta densidad de marcadores tales como el mapa basado en el análisis de la población *Stirling* x *12601ab1* que cuenta con 3839 marcadores SNPs mapeados (Hackett *et al*. 2013), el mapa genético biparental tetraploide más denso hasta la fecha. Sin embargo, la ventaja de nuestro estudio es la diversidad de características fenotipícas evaluadas y el haber considerado una característica que no había sido abordada en estudios previos, como es el contenido de calcio en el tubérculo.

Figura 2. SNPs y QTLs localizados en el cromosoma 12 de *Atlantic* y *Superior*. Las barras indican el intervalo de confianza ± 1 LOD para cada QTL. Los números identifican los QTLs.

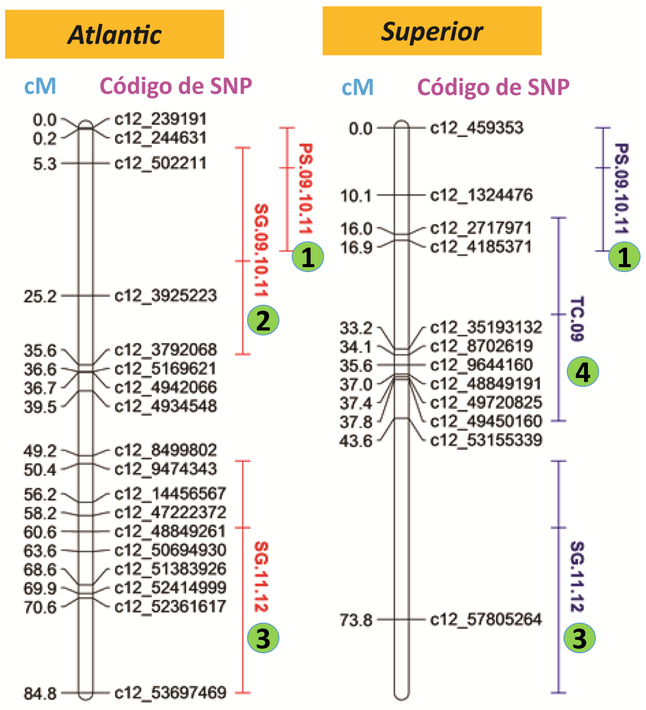


Tabla 2. Descripción de los QTLs localizados en el cromosoma 12 de *Atlantic* y *Superior*. **cM**: indica la localizacion dentro del mapa en centimorgan. **IC**: indica el intervalo de confianza en centimorgan. **LOD (%Varianza)**: Indica el valor minimo del LOD estimado usando 100 permutaciones y entre parentesis la estimacion del porcentaje de varianza explicado por el QTL.

	Caracter	Año	cM	IC	LOD (%Varianza)
1	Incidencia de sarna con lesiones profundas	2009 2010 2011	6	0-18.5	3.8 (10.2%)
2	Gravedad específica	2009 2010 2011	20	3-34	3.6 (11.8%)
3	Gravedad específica	2011 2012	60	50-84.8	2 (7.7%)
4	Contenido de calcio en el tubérculo	2009	28	13.5-44	3.9 (16.4%)

BIBLIOGRAFÍA

- Broman, K. W., Wu, H., Sen, Ś., & Churchill, G. A. (2003). R/qtl: QTL mapping in experimental crosses. Bioinformatics, 19(7), 889-890.
- Hackett, C. A., Milne, I., Bradshaw, J. E., & Luo, Z. (2007). TetraploidMap for Windows: linkage map construction and QTL mapping in autotetraploid species. Journal of heredity, 98(7), 727-729.
- Hackett, C. A., McLean, K., & Bryan, G. J. (2013). Linkage analysis and QTL mapping using SNP dosage data in a tetraploid potato mapping population. PloS one, 8(5), e63939.
- Hamilton, J. P., Hansey, C. N., Whitty, B. R., Stoffel, K., Massa, A. N., Van Deynze, A., ... & Buell, C. R. (2011). Single nucleotide polymorphism discovery in elite North American potato germplasm. BMC genomics, 12(1), 302.
- R Core Team (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.
- Zorrilla, C., Navarro, F., Vega, S., Bamberg, J., & Palta, J. (2014). Identification and selection for tuber calcium, internal quality and pitted scab in segregating 'Atlantic' x 'Superior' reciprocal tetraploid populations. American journal of potato research, 91(6), 673-687.

AGRADECIMIENTOS

Un agradecimiento especial al proyecto 2009-55605-05219 del Instituto Nacional de Alimentación y Agricultura (NIFA) y al Hatch grant de la Universidad de Wisconsin - Madison por financiar parcialmente este trabajo. Asimismo, gracias al proyecto SolCAP por realizar el genotipado de esta población .