

# Integración de herramientas informáticas para el diseño de marcadores moleculares ligados a un gen de interés a partir de un *microarray* de genotipado de trigo pan

Costa Tártara Sabrina<sup>1,2,3</sup>, Crescente Juan Manuel<sup>1,4</sup>, Vanzetti Leonardo<sup>1,4</sup>,  
Tranquilli Gabriela<sup>2</sup>, and Bonafede Marcos<sup>2</sup>

<sup>1</sup> Consejo Nacional de Investigaciones Científicas y Técnicas.

<sup>2</sup> Instituto Nacional de Tecnología Agropecuaria (INTA); Centro de Investigación en Recursos Naturales; Instituto de Recursos Biológicos.  
`costatartara.sabrina@inta.gob.ar`

<sup>3</sup> Universidad Nacional de Luján, Departamento de Tecnología.

<sup>4</sup> INTA; Estación Experimental Agropecuaria Marcos Juárez; Grupo Biotecnología y Recursos Genéticos.

**Resumen** El acceso a nuevas tecnologías moleculares para su aplicación en el mejoramiento de los cultivos demanda la utilización de herramientas informáticas de manera integrada para el tratamiento de los datos. Se presenta la confección de un mapa genético integrando 3232 SNP (en inglés, Single Nucleotide Polymorphism) obtenidos a partir de un *microarray* de SNP diseñado para trigo pan (*Triticum aestivum*) y 212 SSR (en inglés, Simple Sequence Repeat), utilizados previamente en el mapeo de una región cromosómica asociada a la resistencia de la Fusariosis de la Espiga de Trigo (FET), utilizando MapMerger y la librería RQTL. Se seleccionaron 15 SNP candidatos sobre los cuales se diseñaron cebadores *in silico*, para convertirlos en marcadores KASP (en inglés, Kompetitive allele-specific PCR), utilizando PolyMarker. Se espera contribuir con el desarrollo de un nuevo marcador asociado a un gen de resistencia a FET que pueda transferirse a la comunidad de mejoradores.

**Keywords:** trigo pan, SNP, SSR, Polymarker, RQTL, MapMerger

## 1. Introducción

El actual desarrollo de la tecnología para la secuenciación a gran escala (NGS), en conjunto con la caída exponencial del costo económico por dato y la simplificación en la preparación de librerías genómicas, hizo posible la disponibilidad de nuevas herramientas moleculares para su aplicación en el mejoramiento de los cultivos [1]. El avance de estas tecnologías y procesos permitió el desarrollo de plataformas de genotipado de alto rendimiento basadas en *microarrays* de SNPs (en inglés, Single Nucleotide Polimorphism), posicionandolos como los marcadores moleculares elegidos para estudios de diversidad, mapeo de genes y

## II

selección genómica, sucediendo en parte a los SSR (en inglés, Simple Sequence Repeat), ampliamente utilizados hasta el momento por ser especie-específicos, altamente polimórficos y abundantes en número.

Desde siempre los desafíos que presenta el mejoramiento genético de trigo pan (*Triticum aestivum*) se atribuyen a la naturaleza poliploide de su genoma, lo que hace que sea de gran tamaño, presente un gran número de secuencias repetitivas y una estrecha base genética dada por el reciente evento de poliploidización [2]. Para esta especie Allen et al. [3] desarrollaron un *microarray* dirigido a la comunidad de mejoradores (Breeders' 35K Affymetrix Axiom™array), a partir de un subset de SNPs contenidos en un *microarray* de ultra alta densidad (820K) [4]. El *microarray* de 35000 SNP fue diseñado teniendo en cuenta la ubicación genética consenso de cada SNP y el polimorfismo observado en germoplasma de trigo hexaploide existente a nivel mundial. Es un sistema adecuado, económico, de utilidad para investigación y mejoramiento, como por ejemplo generación de mapas genéticos de alta densidad.

En conjunción a la expansión de la tecnología de secuenciación se incrementaron los repositorios de recursos genómicos y bases de datos integradas de SNPs, que contienen la información de los diferentes *microarrays* de genotipado [2]. Además recientemente se cuenta con una nueva secuencia de referencia de trigo (cv Chinese Spring IWGSC Ref Seq v1.0) [5].

El acceso a las tecnologías mencionadas demanda la utilización de herramientas informáticas de manera integrada para el tratamiento de los datos. Teniendo como meta el desarrollo de germoplasma de trigo como estrategia para ampliar la diversidad genética del cultivo, el objetivo es exponer un flujo posible de trabajo para desarrollar marcadores moleculares que faciliten el proceso de selección de bases genéticas de interés, a partir de los datos de genotipado obtenidos con el *microarrays* de 35K, tomando como caso de estudio una población doble haploide (DH) previamente mapeada con marcadores SSR en la que se reporta una región cromosómica asociada a la resistencia de la Fusariosis de la Espiga de Trigo (FET) [6].

Se presenta la confección de un mapa genético integrando SNPs y SSR, y la elección de SNPs candidatos para el desarrollo de marcadores moleculares alelo específicos tipo KASP (Kompetitive allele-specific PCR).

## 2. Materiales y Métodos

### 2.1. Datos genotípicos

Se realizó el genotipado de la población original DH Milan x Catbird (MxC) con el *microarray* de 35K Axiom™Wheat Breeder's [3] en una plataforma GeneTitan™de Affymetrix con un Robot de Beckma. Se obtuvieron 35143 SNP para 91 líneas recombinantes (en inglés, RIL: Recombinant Inbred Lines) y los dos genotipos parentales. Los SNPs cuya variante fue igual a Catbird (parental resistente a FET) se codificaron como "A" y aquellos cuya variante fue igual a Milan (parental susceptible a FET) como "B". Los SNP cuyo patrón fue heterocigota

(variante de ambos padres) se codificaron como dato faltante ya que al ser una población DH se consideraron como error de genotipado. Los SNPs monomórficos, aquellos con alta frecuencia de datos faltantes ( $\geq 10\%$ ) y aquellos cuya segregación no se ajustó a 30:70 se excluyeron del análisis. Luego se excluyeron individuos con alta frecuencia de datos faltantes ( $\geq 40\%$ ). Se seleccionaron 212 marcadores SSR de Cativelli et al. [6] para integrar al nuevo mapa de MxC.

## 2.2. Fusión de los marcadores co-segregantes y construcción de un mapa de ligamiento

Para fusionar los marcadores co-segregantes se utilizó MapMerger, un script en python <sup>1</sup> que fusiona los marcadores que co-segregan, guardando un solo dato de ubicación cromosómica y física, siendo esta última la menor de todas las posibles de acuerdo a los marcadores que está fusionando. El nombre del marcador aparecerá entre llaves, corchetes, paréntesis ó sin símbolos, de acuerdo a que la posición dentro de los grupos de marcadores sea o no la única posible, y/o presente datos faltantes. Previo a utilizar MapMarger, los marcadores se anclaron físicamente al genoma de referencia [5] utilizando BLAST y se anclaron genéticamente a los cromosomas teniendo en cuenta la posición consenso resultado de otras poblaciones de mapeo. Para la construcción del mapa de ligamiento se utilizó la librería RQTL [7]. Mediante la función *formLinkageGroups* se agruparon los marcadores en grupos de ligamiento (GL) utilizando una frecuencia de recombinación máxima de 0,35 entre pares de marcadores y un LOD mínimo de 6. Los marcadores de cada GL se ordenaron mediante la función *orderMarkers* utilizando la función de Haldane para estimar las distancias genéticas entre marcadores, asumiendo una probabilidad de error en el genotipado de 0,005.

## 2.3. Diseño de marcadores

Se seleccionaron SNP candidatos para el diseño *in silico* de marcadores alelos específicos tipo KASP utilizando PolyMarker [8], un pipeline desarrollado para genomas poliploides. PolyMarker convierte la información de la secuencia de entrada (secuencia adyacente al SNP) en un archivo FASTA; utiliza el Exonerate [9] para alinear la secuencia que contiene el SNP con la secuencia de referencia de trigo (cv Chinese Spring IWGSC Ref Seq v1.0); selecciona el mejor hit de cada cromosoma y realiza un alineamiento local con MAFFT [10] en el cual se enmascaran las variaciones entre secuencias homeólogas. En el output se caracterizan los polimorfismos en categorías, de acuerdo a la homología de la variación detectada a través de los cromosomas. Finalmente, utiliza el Primer3 [11] para diseñar los cebadores que se utilizarán para amplificar el SNP. Para los KASP se diseñan tres cebadores, dos específicos (uno para cada alelo del SNP objetivo) y uno común que puede ser específico, semi-específico o inespecífico, respecto a los genomas A, B o D del trigo hexaploide.

<sup>1</sup> <https://github.com/juancrescente/mapmerger>

IV

### 3. Resultados

De la matriz de genotipado se obtuvieron 3232 SNPs útiles para la confección del mapa y se integraron con 212 marcadores SSR. Luego de la fusión con Map-Merger se obtuvieron 505 Loci de Segregación Unica (LSU) para 88 RILs, cada uno de los cuales contiene uno o más marcadores. Se formaron 77 GL a partir de los 505 LSU, que luego se agruparon de acuerdo a la ubicación cromosómica, conformando los 21 cromosomas que posee el trigo pan. Se procedió al ordenamiento de los LSU en cada cromosoma, en función de la distancia genética entre ellos.

De acuerdo a Cattivelli et al. [6] la región cromosómica que mostró mayor asociación con la expresión de resistencia a la FET se localizó en el cromosoma 7D, denominándose "Qfhs.inta-7D". La región se encuentra estrechamente ligada al locus SSR *efd14*, dentro de un intervalo de 1,5 centiMorgan (cM) de distancia genética, flanqueado por los loci SSR *barc128* y *wmc702*. En el nuevo mapa se obtuvieron 21 LSU para el 7D, compuestos por 27 SSR y 31 SNPs, que determinaron una longitud de cromosoma de 125,35 cM. Los SSR ligados al *Qfhs.inta-7D* colapsaron en un mismo LSU, probablemente porque el número de RILs que se utilizaron en el genotipado con el *microarray* fue menor que la población original. No obstante, delimitando un mayor intervalo genético que contiene el *Qfhs.inta-7D* (entre los loci SSR *wms44* y *wmc94*), se observan dos LSU que contienen 7 SSR y 16 SNPs candidatos para el desarrollo de los nuevos marcadores. En términos físicos la región presenta una longitud de 434 Megabases (Mb).

Las secuencias de los 16 SNPs se analizaron con PolyMarker [8]. Todas se alinearon con la secuencia de referencia del cromosoma 7D, aunque solamente dos resultaron específicos para este genoma. El diseño *in silico* de cebadores fue exitoso para 15 SNP, 7 presentaron polimorfismo homólogo, mientras que 8 SNP resultaron no homólogos. El cebador común diseñado para amplificar cada SNP resultó ser específico para 3 SNP, semi-específico para 8 SNP y específico para 4 SNP.

### 4. Conclusiones y Trabajos futuros

La utilización de un *microarray* de genotipado permitió contar con 3232 marcadores nuevos para utilizar en la población DH MxC, aunque el rendimiento de la tecnología fuera del 9%. La versatilidad de la librería RQTL para la confección de mapas genéticos permitió integrar la información con marcadores previamente reportados en la población, potenciando la tecnología del *microarray*. La existencia de una herramienta cuyo flujo de trabajo es específico para el desarrollo de marcadores KASP en trigo pan simplifica el diseño de cebadores para el desarrollo de nuevos marcadores, frente a la complejidad que plantea el genoma de la especie en estudio. El trabajo futuro es la validación de los marcadores diseñados *in silico* en genotipos recombinantes para la región *Qfhs.inta-7D* y la determinación de un nuevo marcador que pueda transferirse a la comunidad de mejoradores para su utilización en programas de mejoramiento.

## 5. Agradecimientos

Los autores agradecen al Centro de Investigación, Docencia y Extensión en TIC de la Universidad Nacional de Luján (CIDETIC<sup>2</sup>) por proveer recursos humanos y computacionales necesarios para este trabajo.

## Referencias

1. Rasheed A., Hao Y., Xia X.C., et al.: Crop breeding chips and genotyping platforms: progress, challenges and perspectives. *Mol Plant* 10:1047-1064 (2017)
2. Rasheed A., Xia X.: From markers to genome-based breeding in wheat. *Theor Appl Genet* 132(3):767-784 (2019)
3. Allen A.M., Winfield M.O., BurrIDGE A.J., et al.: Characterization of a Wheat Breeders' Array suitable for high-throughput SNP genotyping of global accessions of hexaploid bread wheat (*Triticum aestivum*). *Plant Biotechnology J* 15:390-401 (2017)
4. Winfield M.O., Allen A.M., BurrIDGE A.J., et al.: High-density SNP genotyping array for hexaploid wheat and its secondary and tertiary gene pool. *Plant Biotechnol J* 14:1195-1206 (2016)
5. Appels et al. Shifting the limits in wheat research and breeding using a fully annotated reference genome. *Science*. Vol. 361, Issue 6403 (2018)
6. Cativelli M., Lewis S., Appendino L.: A Fusarium Head Blight Resistance Quantitative Trait Locus on Chromosome 7D of the Spring Wheat Cultivar Catbird. *Crop Sci.* 53:1464–1471 (2013)
7. Broman K., Wu H., Sen S., Churchill G.: 2003. R/qtl: QTL mapping in experimental crosses. *Bioinformatics* 19(7): 889-890.
8. Ramirez-Gonzalez R., Uauy C., Caccamo M.: PolyMarker: A fast polyploid design pipeline. *Bioinformatics*, 31(12), 2038-2039 (2015)
9. Slater G.S., Birney E.: Automated generation of heuristics for biological sequence comparison. *BMC Bioinformatics* 31(12): 2038–2039 (2005)
10. Katoh K., Standley D.M.: MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* 30: 772–780 (2013)
11. Rozen S., Skaletsky H.: Primer3 on the WWW for general users and for biologist programmers. *Methods Mol. Biol.* 132: 365–386 (2000)

---

<sup>2</sup> <http://cidetic.unlu.edu.ar/>