

Definición del núcleo optimizado de la colección de conservación del manzano español

C. Miranda¹, P. Errea², J. Urrestarazu¹, A. Pina², S. Pereira-Lorenzo³, E. Dapena⁴, V. Urbina⁵, M.A. Moreno⁶, M.B. Díaz-Hernandez³, A.M. Ramos-Cabrer³, J. Ascasibar-Errasti⁷, I. Iglesias⁸, M. Espiau², G. Reig⁶, J. Dalmases⁵, L.G. Santesteban¹, M.J. Laquidain¹ y J.B. Royo¹

¹Universidad Pública de Navarra, Departamento de Agronomía, Biotecnología y Alimentación, Campus de Arrosadía, 31006 Pamplona

²Unidad de Hortofruticultura, CITA de Aragón, E-50059, Zaragoza

³Escola Politécnica Superior, Departamento de Producción Vexetal, Universidad de Santiago de Compostela, Campus de Lugo, 27002 Lugo

⁴Servicio Regional de Investigación y Desarrollo de Asturias, 33300 Villaviciosa (Asturias)

⁵Departament d'Hortofruticultura, Botànica i Jardineria, Universitat de Lleida, 25198, Lleida

⁶Departamento de Pomología, Estación Experimental de Aula Dei (CSIC), Apartado 13034, 50080 Zaragoza

⁷Centro de Investigaciones Agrarias de Mabegondo. Xunta de Galicia, 15318 Abegondo (A Coruña)

⁸Institut de Recerca i Tecnologia Agroalimentàries, Parc Científic i Tecnològic de Lleida, 25003 Lleida

Palabras clave: *Malus x domestica* Borkh., colección nuclear, marcadores moleculares, microsatélites, optimización.

Resumen

Este trabajo tiene como objetivo determinar la estrategia más adecuada para la selección del conjunto mínimo de accesiones (núcleo optimizado) que represente eficientemente la variación genética del manzano conservado en las colecciones españolas. Este núcleo optimizado mediante criterios genéticos constituirá la base de la colección nuclear de conservación, que podrá ser complementado con accesiones seleccionadas por otros criterios (morfo-fisiológicos, agronómicos, valor histórico, etc.). Se ha evaluado la eficiencia de estrategias de selección por búsqueda local estocástica avanzada (ASLS) que diferían tanto por el tamaño final del núcleo como por la combinación (y peso relativo) de las medidas de distancia genética y riqueza alélica a optimizar. Las estrategias empleadas han proporcionado núcleos optimizados con grandes diferencias en la diversidad conservada, así como en el nivel de representación de la estructura genética general. Teniendo en cuenta el uso principal de la colección nuclear, la estrategia que ofrece un mejor equilibrio entre representatividad y adecuación al uso es la que combina la optimización de la distancia media entre cada accesión de la colección y la entrada en el núcleo más cercana con el índice de Shannon y la recuperación de alelos.

INTRODUCCIÓN

La Red de Colecciones de manzano del Programa Nacional de Recursos Fitogenéticos integra, entre otras, a las colecciones situadas en la Universidad Pública de Navarra (UPNA), el Centro de Investigaciones Agrarias de Mabegondo (CIAM), Cabildos

(Tenerife, La Palma y Gran Canaria), la Universidad de Lleida (UdL), el Servicio Regional de Investigación y Desarrollo de Asturias (SERIDA), la Estación Experimental de Aula Dei-CSIC (EEAD) y el Centro de Investigación y Tecnología Agroalimentaria de Aragón (CITA). Estas colecciones comprenden mayoritariamente cultivares locales prospectados en sus respectivas regiones y, en conjunto, representan la práctica totalidad de áreas productoras de manzano del país. En el Proyecto Coordinado RF2011-00017-C05-00, integrado por las colecciones citadas anteriormente, se estudiaron 1453 accesiones de manzano, empleando 13 SSR y una metodología armonizada. Este trabajo permitió detectar 737 genotipos únicos, así como la existencia de una población diferenciada del material más comercial y el originario del resto de Europa (Pereira-Lorenzo *et al.*, 2017). Los resultados de este trabajo de armonización permiten ahora desarrollar e implementar una estrategia de conservación eficiente del germoplasma de manzano, en la que se identifiquen las accesiones con mayor interés de conservación., siendo éste uno de los objetivos del Proyecto Coordinado RTA2015-00052-C02-00 que ha dado continuidad al anterior.

Este trabajo tiene como objetivo determinar la estrategia más adecuada para la selección de un conjunto mínimo de accesiones que optimice la representatividad de la variación genética conservada en las colecciones. Este núcleo optimizado constituirá la base de la colección nuclear de conservación, que podrá ser complementado con accesiones seleccionadas por otros criterios.

MATERIAL Y MÉTODOS

Material vegetal

En el estudio se han utilizado 457 genotipos únicos, caracterizados con 13 SSR (Pereira-Lorenzo *et al.*, 2017), correspondientes a 763 de las 1340 accesiones de manzano conservadas en siete colecciones españolas. Los genotipos empleados fueron todos los que en el análisis de estructura genética presentado en Pereira-Lorenzo *et al.* (2017) fueron asignados a la población reconstituida asociada exclusivamente con cultivares españoles, a los que se sumaron los que, en el mismo estudio, quedaron asignados en admixis ($qI < 80\%$) a los grupos asociados a cultivares internacionales.

Definición de núcleos optimizados

Para la formación de la colección nuclear de conservación del manzano español se ha considerado un enfoque mixto, en el que se selecciona primero un núcleo de genotipos que maximice la representatividad de la diversidad presente en las colecciones (núcleo optimizado) y posteriormente se complementa con accesiones seleccionadas por otros criterios, tales como la relevancia histórica o la presencia de características morfo-fisiológicas o agronómicas únicas o de interés.

Se han evaluado cuatro estrategias de selección de entradas candidatas para el núcleo optimizado, mediante selección por búsqueda local estocástica avanzada (ASLS), empleando el programa *CoreHunter* v3.2.1 (Thachuck *et al.*, 2009):

- *AEC_{PURA}*: optimización de la distancia media de cada genotipo presente en la colección con la entrada más cercana (*A-EC*), definida por Odong *et al.* (2013).
- *AEC_{MIXTA}*: optimización simultánea de la distancia *A-EC*, el índice de diversidad de Shannon (S_H) y la proporción de alelos recuperados (cobertura alélica, C_V). En el proceso se dio a la optimización de *A-EC* un peso relativo del 50%, mientras que para S_H y C_V el peso asignado fue del 25%.

- EEC_{PURA} : optimización de la distancia media entre cada entrada seleccionada y la entrada seleccionada más cercana ($E-EC$), también definida por Odong *et al.* (2013).
- EEC_{MIXTA} : como la estrategia $A-EC$ mixta, pero empleando la distancia $E-EC$.

Para cada estrategia evaluada se obtuvieron núcleos optimizados que incluían el 5%, 10% y 15% de los genotipos, es decir, que contenían respectivamente 23, 46 y 69 genotipos. Se generaron 50 réplicas de cada estrategia y tamaño, y los núcleos optimizados finales se formaron con los genotipos seleccionados con mayor frecuencia.

Los núcleos optimizados obtenidos se evaluaron en función de los valores de distancia $A-EC$ y $E-EC$, los índices S_H y C_V , y la distribución de los genotipos en las poblaciones reconstituidas.

RESULTADOS Y DISCUSIÓN

Los valores de diversidad y distancia obtenidos para las estrategias de optimización se resumen en la Tabla 1. Como era de esperar, cada estrategia optimizaba mejor la medida de distancia incluida en el análisis. Por ello, las estrategias mixtas optimizaron mejor el índice S_H y la recuperación de alelos que las puras. La estrategia AEC_{PURA} obtuvo valores de distancia e índices de diversidad más distintos de los demás, optimizó $A-EC$ un 5-15% mejor, pero a cambio la optimización de $E-EC$ y, sobre todo, la de la recuperación de alelos fueron más bajas (hasta un 40% menores), especialmente para el tamaño de núcleo más pequeño. Las diferencias de distancias entre el resto de las estrategias fueron menores al 10%, y en la recuperación de alelos menores al 20%.

La medida de distancia empleada en la optimización del núcleo influyó notablemente sobre los genotipos seleccionados. Las estrategias puras compartieron menos del 6% de los genotipos, mientras que las mixtas compartieron en torno al 50%. AEC_{PURA} compartió entre un 15-30% con AEC_{MIXTA} , mientras que las estrategias con más genotipos comunes (50-60%) fueron las obtenidas al utilizar $E-EC$ como distancia a optimizar. Las diferencias fueron también grandes en lo referido a los grupos genéticos a los que pertenecían los genotipos y a su nivel de ploidía: las estrategias AEC_{PURA} y AEC_{MIXTA} seleccionaban accesiones de cada población reconstituida de una forma más proporcional al tamaño de cada grupo, sobrerrepresentando los genotipos triploides. Sin embargo, las otras dos estrategias tenían una proporción de genotipos diploides más similar a la del conjunto, pero sobrerrepresentaban a las poblaciones “internacionales”.

Los resultados obtenidos en este estudio muestran, como era esperable, que ninguna de las estrategias empleadas ha sido capaz de optimizar mejor que las demás todos los indicadores de calidad de los núcleos optimizados, ya que éstos están interrelacionados (El Bakkali *et al.*, 2013). Por ello, es preciso lograr un equilibrio entre la diversidad retenida en el núcleo, su representatividad y la adecuación al uso que se le quiera dar a la colección nuclear. Odong *et al.* (2013) recomienda que las evaluaciones incluyan criterios basados en distancias, definiendo a tal fin $A-EC$ y $E-EC$, y que se de mayor peso en la evaluación a los criterios más estrechamente relacionados con el objetivo final de la colección.

En este estudio se ha optado por crear una colección de tipo “generalista” o “multipropósito”, en la que se busca maximizar la representatividad de la diversidad genética de la colección. Es decir, cada accesión de la colección original debería estar representada por una entrada en el núcleo con la mayor similitud posible. Teniendo esto en cuenta, la medida de distancia más adecuada sería $A-EC$, y AEC_{PURA} la estrategia más

eficiente. No obstante, la estrategia AEC_{MIXTA} consigue un mejor equilibrio entre optimización de los parámetros de distancia, la recuperación de alelos, y la representatividad de la estructura genética de la colección.

AGRADECIMIENTOS

Este trabajo ha sido financiado por los proyectos INIA RF2011-00017-C05-00 y RTA2015-00052-C02-00.

Referencias

- El Bakkali, A., Haouane, H., Moukhli, A., Costes, E., van Damme, P., Khadari, B. 2013. Construction of core collections suitable for association mapping to optimize use of Mediterranean olive (*Olea europaea* L.) genetic resources. PLoS One 8(5): e61265
- Odong, T.L., Jansen, J., van Eeuwijk, F.A., van Hintum, T.J.L. 2013. Quality of core collections for effective utilisation of genetic resources review, discussion and interpretation. Theor. Appl. Genet. 126:289-305.
- Pereira-Lorenzo, S., Urrestarazu, J., Ramos-Cabrer, A.M., Miranda, C., Pina, A., Dapena, E., Moreno, M.A., Errea, P., Llamero, N., Díaz-Hernández, M.B., Santesteban, L.G., Laquidain, M.J., Gogorcena, Y., Urbina, V., Dalmases, J., Ascasibar-Errasti, J. and Royo, J.B.. 2017. Analysis of the genetic diversity and structure of the Spanish apple genetic resources suggests the existence of an Iberian genepool. Ann Appl Biol 171: 424–440.
- Thachuk, C., Crossa, J., Franco, J., Dreisigacker, S., Warburton, M., Davenport, G.F. 2009. Core Hunter: an algorithm for sampling genetic resources based on multiple genetic measures. BMC Bioinformatics 10:243.

Tabla 1. Comparación de los núcleos optimizados obtenidos según la estrategia de optimización y el tamaño de núcleo considerados.

Estrategia de optimización	Tamaño núcleo (%)	Distancia media				Proporción de genotipos (%)		
		A-EC	E-EC	S _H	C _V (%)	Gr. "español"	Diploides	Únicos ¹
AEC _{PURA}	5	0,44	0,45	4,0 3	0,49	73,9	52,2	65,2
	10	0,39	0,44	4,1 7	0,68	67,4	60,9	69,6
	15	0,35	0,44	4,1 7	0,74	63,8	71,0	72,5
AEC _{MIXTA}	5	0,49	0,56	4,4 3	0,82	56,5	43,5	87,0
	10	0,41	0,52	4,4 0	0,93	56,5	54,3	82,6
	15	0,37	0,50	4,3 8	0,99	50,7	63,8	79,7
EEC _{PURA}	5	0,52	0,62	4,4 0	0,73	26,1	91,3	82,6
	10	0,45	0,57	4,3 6	0,79	23,9	89,1	95,7
	15	0,40	0,54	4,3 5	0,82	31,9	89,9	94,2
EEC _{MIXTA}	5	0,52	0,58	4,4 4	0,83	34,8	87,0	87,0
	10	0,45	0,57	4,4 3	0,92	30,4	80,4	91,3
	15	0,41	0,54	4,4 1	0,94	29,0	78,3	89,9
Genotipos empleados en el análisis (n=457)						62,4	80,7	84,6

¹Genotipo único: aquellos presentes en sólo una de las colecciones de la Red.