

Accession rareness, allele specificity and core set selection with an information theory approach: application to wheat marker data

M. Humberto Reyes-Valdés

Universidad Autónoma Agraria Antonio Narro. Área: Biometría

Agosto 31, 2017

Resumen

● Motivación

Los bancos de germoplasma son valiosos reservorios de información genética para enfrentar los retos poblacionales y climáticos. Su manejo requiere de herramientas analíticas para la caracterización informacional de las colectas.

● Resultados

Proponemos una visión de los bancos de germoplasma basada en teoría de la información, así como parámetros para medir la especificidad de alelos y la rareza de colectas. Probamos con éxito este paradigma en su aplicación para la obtención de colecciones núcleo.

El proyecto

• Objetivo general

Desarrollar una visión basada en teoría de la información, y herramientas para estudiar y manejar colecciones de germoplasma a través de marcadores genéricos, con razas criollas de trigo como caso de estudio.

• Objetivos específicos

- Proveer un estimador de especificidad de alelos, en el sentido de su exclusividad en un conjunto de colectas.
- Proveer un estimador de rareza de colectas, en función de la especificidad de alelos.
- Desarrollar una estrategia para la formación de conjuntos núcleo, con base en los parámetros desarrollados, y comparar con otros métodos.
- Aplicar los métodos desarrollados en la formación de conjuntos núcleo de la colección de trigos del CIMMYT, con base en información de marcadores SNP.

Conceptos desarrollados en esta investigación

- Mutual information between genetic markers and accessions

$$I(X; M)$$

- Allele specificity

$$S_i$$

- Accession rarity

$$R_j$$

- Accession divergence

Rarity of accessions

Rarity Degree in which an accession is different to the rest of the collection, measured as the weighted average of the specificity of its alleles.

$$R_j = \sum_{i=1}^k p_{ij} S_i,$$

where p_{ij} is the frequency of the i -th allele in the j -th accession



Divergence of accessions

Divergence Degree in which the allele frequency distribution of an accession diverges from the pooled allele frequency distribution in the collection, measured by the Kullback-Leibler divergence.

$$D_j = \sum_{i=1}^k p_{ij} \log_2 \left(\frac{p_{ij}}{p_i} \right),$$

where p_{ij} is the frequency of the i -th allele in the j -th accession and p_i is the average frequency of the i -th allele in the collection.



Resultados

- La especificidad alélica, rareza y divergencia resultaron exitosos para la caracterización de alelos y colecciones.
- Los datos de trigo de MasAgro fueron clave en la evaluación de los métodos propuestos.
- Se implementó un método para formar conjuntos núcleo

	Method	Divergence	MR	SH	AR	LA
1	HCore	0.442	0.438	7.954	0.985	59
2	MixRep	0.435	0.435	7.950	0.986	53
3	REMC	0.408	0.419	7.932	0.975	98
5	MSTRAT	0.406	0.417	7.931	0.983	67
6	Random	0.402	0.416	7.929	0.976	95

Experiencias

- Resultados satisfactorios
- Expectativas alcanzadas con las herramientas de MasAgro Biodiversidad y la valiosa colaboración de Juan Burgueño
- Los datos de genotipificación de trigo de MAB fueron fundamentales para la prueba de concepto
- MasAgro Biodiversidad puede tener más impacto a través del apoyo con becas a estudiantes de postgrado