

BIOMART: Integración de datos genotípicos, fenotípicos y climáticos de papa: experiencias en el Centro Internacional de la Papa (CIP)

Cordova, Raúl¹; Tejada, Sofia¹; Salas, Elisa¹; Flores, Mirella¹; Simon, Reinhard¹; Bonierbale, Merideth¹ y Hualla, Vilma¹.

¹ Centro Internacional de la Papa (CIP), Perú. E-mail: r.cordova@cgiar.org

Introducción

La integración de datos implica el proceso de combinar datos heterogéneos provenientes de diferentes fuentes, facilitando una vista unificada de los mismos a los investigadores. En el CIP se tienen registros con información pasaporte, genealógica, genómica, fenotípica (evaluación y caracterización) además de datos geográficos y climáticos. Ésta combinación de diversos tipos de datos sumados al gran número de variables registradas, medidas y/o calculadas proponen un desafío en la gestión de datos donde es necesario considerar tanto la flexibilidad y la complejidad. En la comunidad genómica el problema general ha sido abordado por BioMart, impulsado para proporcionar un acceso unificado a las bases de datos biológicos, para simplificar las búsquedas entre bases de datos que de otra manera podrían requerir varios pasos complicados [1]. Utilizando BioMart los datos genotípicos y fenotípicos se pueden combinar de forma transparente para su posterior análisis en el proceso de decisión para la selección de nuevos genotipos.

Materiales y Métodos

Siguiendo las mejores prácticas para organizar y estructurar los sistemas de búsqueda de información, la interfaz de BioMart se estructuró utilizando la ontología de papa, ya que dichos conceptos se utilizan para limpiar las bases de datos agronómicos y describir los datos. Los términos de ontología sirven para describir las variables genotípicas, fenotípicas y climáticas, para sistemas de bases de datos como BioMart, es importante ya que proporciona la descripción armonizada de los datos y, por tanto, facilita la recuperación de la información [2].

La ontología de la papa continúa en fase de desarrollo y también se utiliza para la documentación de datos. Así mismo BioMart proporciona a los usuarios un acceso autorizado diferencial.

Resultados

En BioMart se almacena información de 326 experimentos de 9 países y 62 localidades. El número de ensayos de campo es: evaluaciones fisiologías (4), evaluaciones bioquímicas (73), evaluaciones

de rendimiento bajo exposición intencional de tizón tardío (131) y evaluaciones de rendimiento (118). Además, tiene información histórica de datos: pasaporte (47,531 registros), caracterización morfológica (60,419 registros), distribución (93,054 registros), climáticos diarios (23 localidades 1974 a 2012 - 41,036 registros), climáticos horario (8 localidades de 2004 a 2012 - 422,735 registros) y de suelo (127 localidades del 1989 a 2012 - 1,753 registros). Se cuenta con 52 pares de iniciadores SSR que cubren los 12 cromosomas de papa donde se considera la información de la población, pedigrí, parentesco, marcadores moleculares y otros. También se tienen registrada información de 22 primers para 80 clones avanzados y 52 primers para 1564 genotipos, siendo en total 431,840 registros. En total BioMart almacena 676,729 registros.

Conclusiones

BioMart es una herramienta libre y gratuita que facilita la búsqueda, combinación y recuperación de información por parte del investigador. Es importante para todo sistema de información cumplir con las jerarquías y estandarización establecidas en una ontología previamente definida.

Agradecimientos

Mejoradores del CIP: Amoros, W.; Gastelo, M., Mihovilovich, M.A el programa "Generation Challenge Program".

Referencias bibliográficas

- [1] BioMart Central Portal: an open database network for the biological community
- [2] International Potato Center (CIP). 2012. Procedures for Standard Evaluation and Data Management of Advanced Potato Clones.
- [3] Integrating and sharing accession-level and omics-size genotype, phenotype and environmental data: Experiences at the International Potato Center (CIP)