The Potato Genome Sequencing Initiative

The Potato Genome Sequencing Consortium

Introduction

Potato is the world’s most important vegetable crop, the 3rd largest global food crop and a unique biological system belonging to the Solanaceae. In order to decipher the structure and function of its genome, the 4.5 Mb genome of potato (Solanum tuberosum L.), containing 12 chromosomes is currently being sequenced by the Potato Genome Sequencing Consortium (PGSC). The PGSC was initiated through Wageningen University and Research Centre and currently comprises member institutions from 15 different countries.

The Potato Genome Sequencing Initiative is an unparalleled investment, both financially and logistically, in the genetic and physical infrastructure of the potato genome. Sequence data from the sequencers will be released to the international scientific community for free. The project is also being accompanied by a comprehensive annotation and data dissemination plan.

Sequencing Strategy

Initial Strategy

The PGSC started in 2005 by taking a heterozygous diploid potato clone (PhF99-033-16) and adopting a chromosome by chromosome approach. The project was divided into three major stages:

1. **Genetic Linkage Map Construction**: A genetic linkage map was developed using BAC-end typing. This map was used to anchor the genomic scaffolds to specific chromosomes.

2. **Genome Assembly**: The genome was assembled using a combination of Illumina and 454 technologies. The assembly was then refined using long reads (Solexa) and paired-end Illumina data.

3. **Anchored BACs**: Ancillary BACs were produced to anchor the genome to specific chromosomes.

Additional use of highly homozygous genotypes (Figure 1b and 2) to get around heterozygosity and assembly problems of RH lines will provide a major boost to gaining a better understanding of potato trait biology and will underpin future breeding efforts.

Genome Assembly and Annotation

Genome Assembly

- Initial draft assembly of DM based on Illumina short reads and Sanger sequenced BAC ends and Promoter ends (Table 1) has already been generated using the short reads assembly software: SOAPsoutuex (version 1.104) developed by BGI (Figure 4 and 5, Table 2).
- Assembly of RH lines was done using NGS/WGS and Sanger data (Table 6).
- Integration of the two genome assemblies will generate three virtual molecules corresponding to the three haplotypes (Figure 6).

Structural and Functional Annotation

- Three gene prediction methods (Figure 7) parted to annotated gene models.
- More than 100% of the genes mapping were identified with the three methods.
- Integration of the two genome assemblies will generate three virtual molecules corresponding to the three haplotypes (Figure 6).

General Goals

- Sequence the complete genome of potato by early 2010.
- Deliver high-quality sequence data to a database infrastructure.
- Provide a major boost to gaining a better understanding of potato trait biology and will underpin future breeding efforts.

Problems With Initial Sequencing Strategy

- Significant resource and capacity development for potato genome sequencing but has also followed the development of NGS technologies.

- NGS methods have led to an increase in the number of reads and to a decrease in the number of reads per read length.

- The use of NGS technologies in addition to BAC sequencing to generate VGS sequence (Figure 9).

- Integration of data according to capability and available resource rather than a chromosome by chromosome approach.

Data Dissemination

- Enhance our ability to identify the desirable allelic variants of genes underlying traits.
- Increase scaffold size and generate hybrid assembly using Solexa, Roche 454 and Illumina technologies.
- Data is available to the international scientific community for free.
- Integration of the two genome assemblies will generate three virtual molecules corresponding to the three haplotypes (Figure 6).

Ongoing and Future Steps

- Complete potato genome by early 2010.
- Enhance scaffold size and generate hybrid assembly using Solexa, Roche 454 and Illumina technologies.
- Quality assessment of the DM assembly by Sanger sequenced DM BACs.
- Anchor genome assembly to a genetic map.
- Develop informatics tools to integrate resources (physical map, genome sequence, marker gene).
- Complete potato genome sequence by early 2010.

Acknowledgements

Scottish Crop Research Institute, United Kingdom (Glenn Bryan, David Marshall, Robbie Waugh and Sanjeev Kumar Sharma).

Benefits

- Facilitate detection of disease and breed resistance.
- Provide a major boost to gaining a better understanding of potato trait biology and will underpin future breeding efforts.

- Complete potato genome sequence by early 2010.