68750

RIKEN PSC

Conception for organizing information on cassava

Kenji Akiyama[1], Manabu Ishitani[2], Kazuo Shinozaki[1], Tetsuya Sakurai[1]
[1]RIKEN Plant Science Center, 1-7-22 Suehiro, Tsurumi, Yokohama, Kanagawa 230-0045, Japan
[2]Agrobiodiversity and Biotechnology Project, International Center for Tropical Agriculture (CIAT), A.A. 6713, Cali, Colombia

226874

Abstract

Cassava is a starch crop known for its ability to grow in diverse environments, ranging from dry to humid climates and acidic to alkaline soils. Further, this crop can grow in nutrient-poor soil and is found worldwide. Cassava is considered an energy source and is used in industries and as food. Therefore, it is thought that cassava is a relatively better energy source than other starch crops, such as corn. In fact, information on cassava, such as published journal articles and sequence registration in public databanks, is increasing, and this increase is associated with high expectations. However, unfortunately, the information is scattered and not organized. Therefore, at present, we are unable to refer to it effectively. Similar to the study of Arabidopsis, comprehensive researches such as a cDNA collection and microarray experiments will be performed for cassava. In fact, we have successfully obtained a cassava full-length cDNA collection. Because it is expected that the volume of information will be generated from such researches will be large, organization of the information will be important. The TAIR and GRAMENE portals for Arabidopsis and monocot crops, respectively, play important roles. Hence, I would like to explain our conception for organizing information on cassava.

Status of plant database

Until July 2008, the number of databases related to plants, particularly model plants, has increased. The numbers of databases related to different plants as published in a Nucleic Acids Research Database issue are as follows.

Arabidopsis:	26
Rice:	17
Other Plants:	18

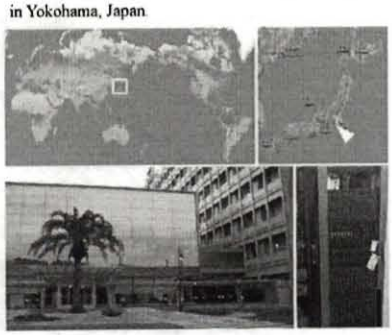
Database Summary Paper Categories
http://www.oxfordjournals.org/nar/database/c

Conclusion

We propose to construct a portal site for Cassava, which is one of the important crops. To construct a portal site, enrichment of information such as genome and expression of genes and gene markers is necessary. Therefore, we expect co-ordination among many research institutes. We hope that the use of information available on the portal site will result in the generation of a new breed; hence, immediate improvements in information management are required.

RIKEN Plant Science Center is

in Yokohama, Japan.



Stage 1: Creation of resource-based databases

Information such as genome, cDNAs, gene markers, and mutants of each species has been accumulated.

Major databases

[Arabidopsis cDNAs, Mutants, and Microarrays]
RARGE - RIKEN Arabidopsis Genome Encyclopedia
[Rice Genome]
Rice Annotation Project Database
[Cassava cDNAs]
Cassava Full-Length cDNA Database

Resource types

- Genome
- ESTs
- Microarrays
- Small RNA
- Gene Markers
- Mutants
- Traits
- .

Sakurai *et al.*, Nucleic Acids Res. 2005, 33(Database Issue):D647-D650
http://rarge.psc.riken.jp/
Sakurai *et al.*, BMC Plant Biol. 2007, 7:66
http://amber.gsc.riken.jp/cassava/

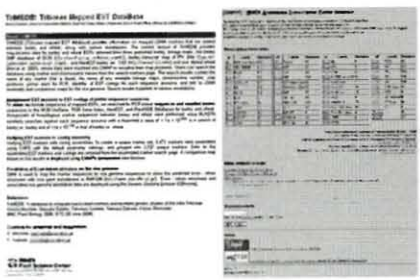
The available data on primary resources such as genome and ESTs of cassava are inadequate; therefore, data improvement is necessary.
RIKEN and CIAT have particularly focused on the creation of such resources. Hence, we have good experience in data handling and database development.

Stage 2: Integration and Comparison

Accumulation of information is increasing; therefore, comparison across various classes can be performed.

Major databases

[Triticeae]
TriMEDB - Triticeae Mapped EST DataBase
[Transcription Factor]
Plant Transcription Factor Database
RIKEN Arabidopsis Transcription Factor Database



The release of cassava genome sequence would result in the detection of possible links among different types of data. Cross-sectional or exhaustive analysis of the data gathered would be possible by using approaches such as comparison between species, gene classification, and alternative splicing events.
Obtaining new findings by combining several data would also be important.
RIKEN has already created such databases.

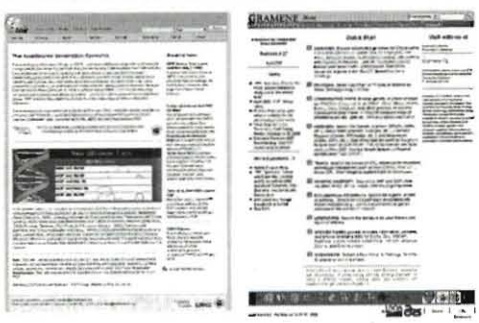
Mochida *et al.*, BMC Plant Biol. 2008, 8:72
http://trimedb.psc.riken.jp/cgi-bin/TriMEDB/index.pl
Iida *et al.*, DNA Res. 2005, 12 247-256
http://rarge.gsc.riken.jp/rantf/

Stage 3: Portalization

Since increasing amounts of data are being collected, a portal site that provides the available data for various analyses will be constructed.

Major databases

[Arabidopsis]
TAIR - The Arabidopsis Information Resource
[Rice]
GRAMENE



Types of Information provided by TAIR

- Genome Sequence
- Genome Annotation
- Gene markers
  - RFLP
  - SSR
- Mutants
- Microarray
- Pathway
- Chemical Structure
- .

Types of Information provided by GRAMENE

- Genome Sequence
- Gene markers
  - RFLP
  - SSR
- Protein
- Comparative Map
- Biochemical Pathway
- Traits
- Genetic Diversity
- .