

# An Object-Protocol Model Multidatabase Query System for Rice and Maize

**Bal A. Antonio**<sup>1</sup>

antonio@staff.or.jp

**Hector Sanchez**<sup>2</sup>

hector@teosinte.agron.missouri.edu

**Su-Shing Chen**<sup>2</sup>

schen@ecn.missouri.edu

**Yoshiyuki Mukai**<sup>1</sup>

mukai@staff.or.jp

**Zhiwei Fang**<sup>2</sup>

zwwang@teosinte.agron.missouri.edu

**Mary Polacco**<sup>2</sup>

maryp@teosinte.agron.missouri.edu

**Katsumi Sakata**<sup>1</sup>

ksakata@abr.affrc.go.jp

**Isamu Ohta**<sup>1</sup>

ohta@staff.or.jp

**Ed Coe, Jr.**<sup>2</sup>

ecoe@teosinte.agron.missouri.edu

**Takuji Sasaki**<sup>1</sup>

tsasaki@abr.affrc.go.jp

<sup>1</sup> Rice Genome Research Program, National Institute of Agrobiological Resources / Institute of the Society for Techno-innovation of Agriculture, Forestry and Fisheries, Tsukuba, Ibaraki, Japan

<sup>2</sup> Missouri Maize Project, University of Missouri-Columbia, Columbia, Missouri, USA

## 1 Introduction

With the recent explosion in genome analysis of various organisms, access to different databases often involves browsing through repositories characterized by various degrees of heterogeneity. In most cases, each database is characterized by a specific view of the domain and implemented using a specific database management system.

The Rice Genome Research Program (RGP) and the Missouri Maize Project (MMP) have developed a scheme for cross querying of rice and maize databases using the OPM (Object-Protocol Model) tools of Gene Logic, Inc. This enables users to explore semantics and to assemble biological information from the two databases. As rice is considered a model for genome analysis of cereal crops, genome information of rice can be directly reflected in maize and vice-versa.

## 2 Rice Database INE and MaizeDB

The rice genome database called INE (INtegrated rice genome Explorer, <http://www.dna.affrc.go.jp:82/giot/INE.html>) has been constructed to integrate all genomic information in rice such as the genetic map, physical map and ESTs. It also serves as the repository of the rice genome sequence from the International Rice Genome Sequencing Project (<http://www.dna.affrc.go.jp:82/Seqcollab.html>). The MaizeDB (<http://www.agron.missouri.edu/>) covers all genomic information on maize that range from genetic maps, ESTs, QTLs, mutant traits, gene products as well as supporting data such as references and persons involved in maize research. The two databases combined represents the most comprehensive genome databases on major cereal crops. Therefore, a query system that combines and manipulates data from both databases will be useful in genome analysis of both rice and maize.

## 3 Multidatabase Access

The flowchart for querying and combining data from multiple databases is shown in Fig. 1. Queries expressed in OPM language are executed using a query processor which interprets the language,

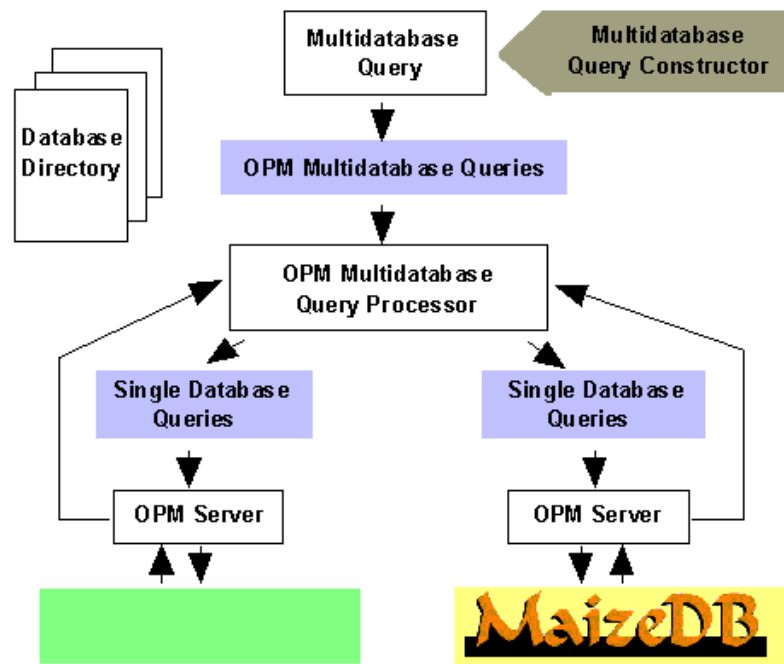


Figure 1: OPM multidatabase query tools for INE and MaizeDB (modified from Data Logic Technical Document, 1998).

generates the queries in each database and performs data manipulation. A database directory records the necessary access and information about interdatabase links [1]. MaizeDB has been retrofitted into an OPM/Sybase platform and we have also adapted database query in INE using the OPM tools. At present, genetic markers which have been commonly mapped in both rice and maize are used to establish multidatabase access links and overlaps between INE and MaizeDB at the level of OPM classes and attributes. Since the locus names are represented in different formats in the two databases, query can be made using accession numbers. This multidatabase query system is useful for comparative genetic/physical mapping, functional genomics, and analysis of genome sequence. A query system to multiple crop databases will maximize efficient application of genome information in various cereal crops.

## References

- [1] Markowitz, V.M., Chen, I.A., and Kosky, A., Exploring heterogeneous molecular biology databases in the context of the object-protocol model, *Theoretical and Computational Genome Research*, S. Suhai (ed.), Plenum, 393–418, 1996.