
Report on Task 24 – Development of MOBY Applications

**Richard Bruskiwich¹, Martin Senger¹, Natalia Martins²,
Mathieu Rouard³, Reinhard Simon⁴, Masaru Takeya⁵**

¹IRRI, ²EMBRAPA, ³IPGRI/INIBAP, ⁴CIP and ⁵NIAS

Report Overview

- Overview of 2005 Task Activities
- IRRI
 - Core MOBY technology development
- MOBY use cases from other Task partners
 - IPGRI/INIBAP (see also their extensive poster)
 - CIP (see also their HPC poster)
- Proposal for next year's task work

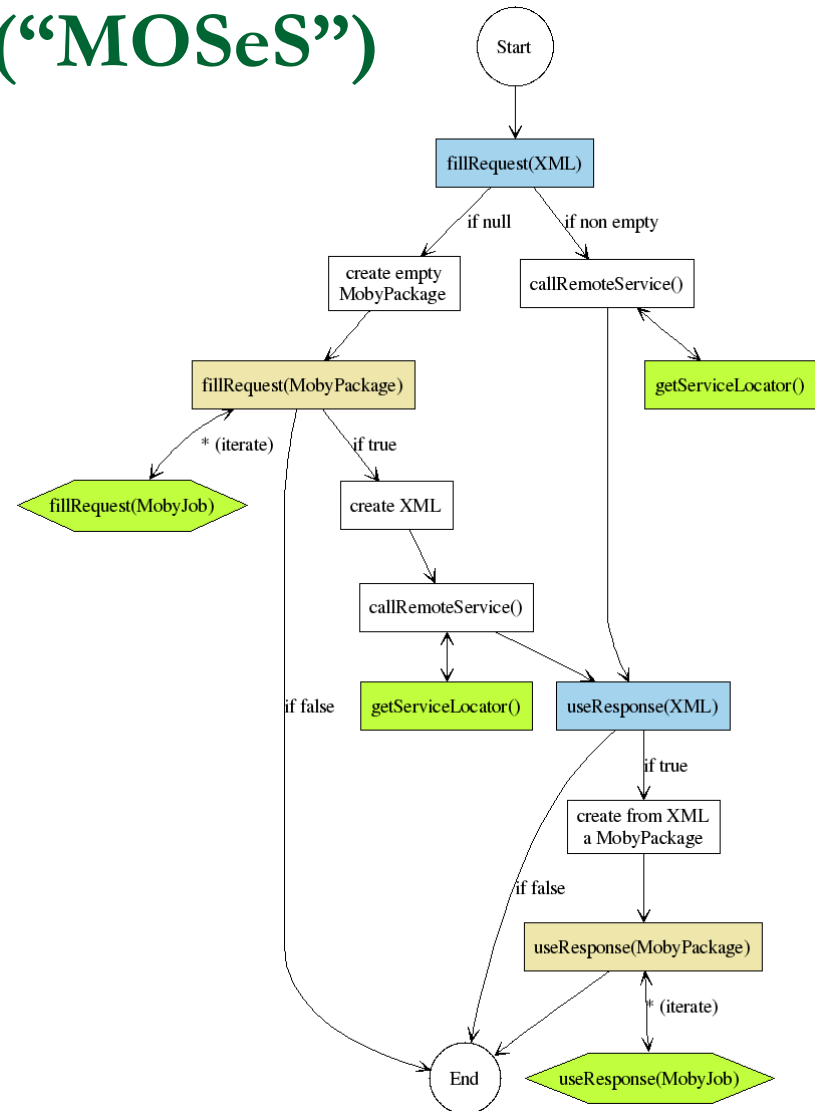
| Use Case Theme | Team | Contact(s) | Activity |
|---|---------|---|--|
| Germplasm, Phenotype, Genotype and Genomic data | IRRI | Richard Bruskiwich, Alexander Cosico | Specify& implement Java web services API to germplasm, phenotype, (including mutant), genotype and basic genomic (e.g. ortholog gene catalog) data |
| Perl-based web services API | INIBAP | Mathieu Rouard, Milko Skofic | Develop additional Perl-hosted web services in collaboration with CIAT and CIRAD |
| High Performance Computing System | CIP | Reinhard Simon, Luis Avila | Development and deployment of a web services API to the research applications on the GenerationCP High Performance Computing facilities. |
| Proteomics and protein structure | EMBRAPA | Natalia Martins | Embed web services client functionality in the EMBRAPA <i>Genoma</i> software, in collaboration with IRRI |
| Gene Expression Data | NIAS | Masaru Takeya, Shoshi Kikuchi, Koji Doi | Develop a web services interface to comparative gene expression data in collaboration with IRRI |

IRRI MOBY Activities

- Resourcing: IRRI Task 24 funds used for:
 - Two University of the Philippines graduate students undertaking MOBY Java software engineering thesis projects (May 2005)
 - One NRS web services programmer, Ms. Mylah Anacleto (July 2005)
 - 6 month consultancy contract to Martin Senger (August 2005)
- Task leader (RMB) attended MOBY (Vancouver) workshop in May.
- **MOBY Services Support (“MOSeS”)** Java software toolkit developed by Martin Senger and is being embedded directly into the GCP platform.
- BioCASE wrapped deployed on top of the IRRI IRIS online web server for GCP passport data publication (in collaboration with Task 23 and Task 25 data template task)
- MOSES applied to ICIS 5 code base to publish MOBY web services (in addition to last year’s WSDL defined web services).
- SoapLab deployed on IRRI HPC and being applied to publish EMBOSS, (Paracel) BLAST and (R statistics) microarray analysis services

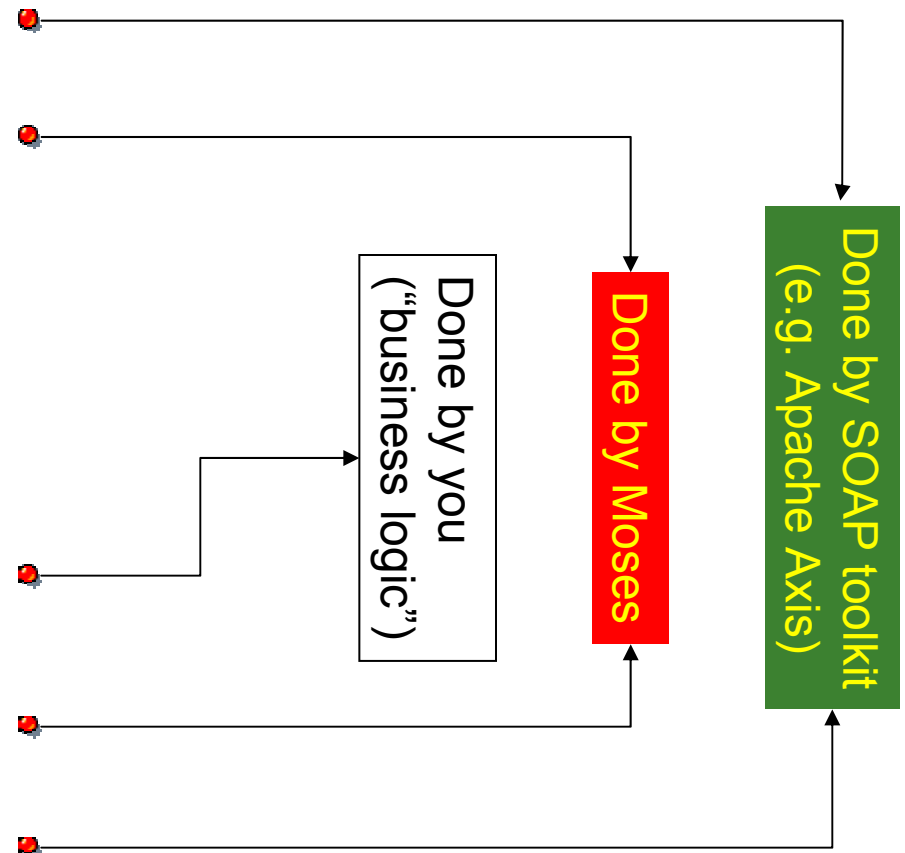
MOBY Services Support (“MOSeS”)

- MOSES is a Java library and Ant build file that streamlines the integration of BioMOBY into a Java code base.
- MOSES retrieves MOBY data type descriptions directly from the MOBY central registry and generates skeleton Java code for said MOBY data types
- MOSES also generates skeleton Java code for registered MOBY services that can then be used for a full coding of business logic
- MOSES also auto-generates useful MOBY data type and service JavaDoc documentation



To write a Biomoby service, one needs:

- To extract data from a SOAP envelope
- To expect incoming data in different encoding (data can be a String or a byte array)
- To extract data from a Biomoby XML envelope
- To separate data into individual “jobs” (a request can consist of many of them)
- [To get installation parameters from the surrounding servlet engine]
- To do something meaningful with data (to create results)
- To convert results back into response “jobs”
- To wrap results into a Biomoby XML
- To send data back in a SOAP envelope



An example: a full Biomoby service

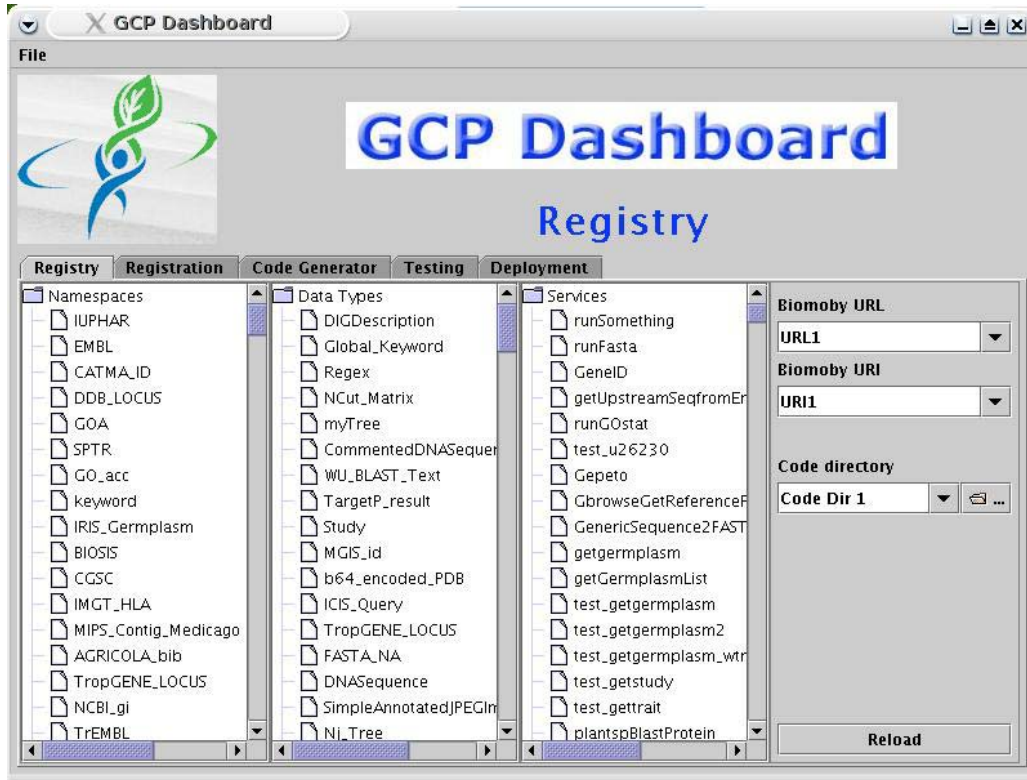
“HelloBiomobyWorld”

```
package org.jmoby.tutorial.service;

import net.jmoby.samples>HelloBiomobyWorldSkel;
import org.biomoby.shared.MobyException;
import org.biomoby.shared.parser.MobyPackage;
import org.biomoby.shared.parser.MobyJob;
import org.biomoby.shared.datatypes.*;

public class HelloBiomobyWorldImpl
    extends HelloBiomobyWorldSkel {
    public void processIt (MobyJob request, MobyJob response,
                          MobyPackage outputContext)
        throws MobyException {
        set_greeting (response, new MobyString ("Hello, World!"));
    }
}
```

MOSES “Dashboard” Interface



- MOSES deployed web services currently specified by a command line programs.
- The “Dashboard” being developed by Martin Senger is a GUI being developed to integrate and facilitate MOBY web services registration, code generation, testing and deployment (to Tomcat/Axis)

Web-based workflow: evaluate the genetic diversity of germplasm collections

- A web services workflow implemented: starting from phenotypic criteria, users obtain a list of *Musa* accessions with information on markers and alleles, that are sent to MOBY services wrapping the R statistical software to calculate a distance matrix and to generate a phylogenetic tree.
- 10 Moby services implemented
- Data retrieval across two database (MGIS and TropGENE DB) on two web service provider hosts
- Connection with a software to perform analyses and graphical outputs
- Moby Web client developed (<http://biomoby.inibap.org>)

Using BioMOBY to evaluate the genetic diversity of *Musa* germplasm collections

Germplasm selection form on phenotypic criteria

List of accessions

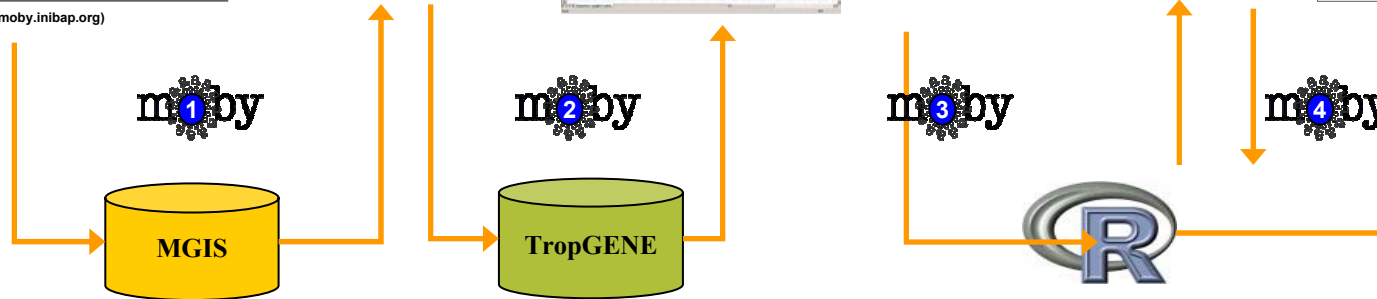
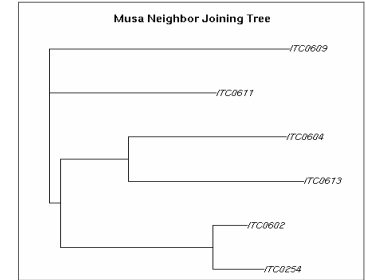
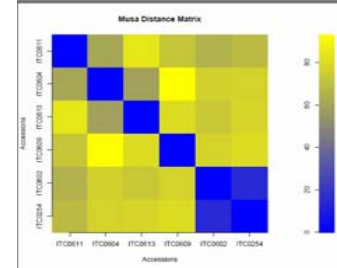
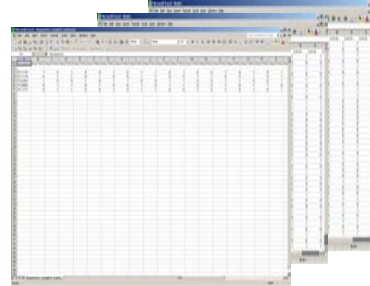
Allelic or disjunctive complete matrix

Distance matrix

Tree



(<http://biomoby.inibap.org>)



Step1: The web query form displays the criteria to extract, from MGIS, a list of *Musa* accessions with traits of interest.

Step2: A first service is launched to search for existing markers and alleles in TropGENE DB corresponding to the selected accessions and to display results in different formats (e.g. allelic matrix, disjunctive complete matrix).

Step3: The outputs of step 2 are processed by the second service using the statistical R software to obtain a distance matrix based on the method chosen by the user (genetic, euclidean, jaccard etc.).

Step4: Another service processes the distance matrix using R software and generates a phylogenetic tree by the Neighbour Joining (NJ) method.

HPC, R and BioMoby (CIP)

F. Mendiburu, L. Avila, R. Simon

R applications for parallel processing on HPC further refined and new one added (resampling for linear models and optimizing choice of molecular markers).

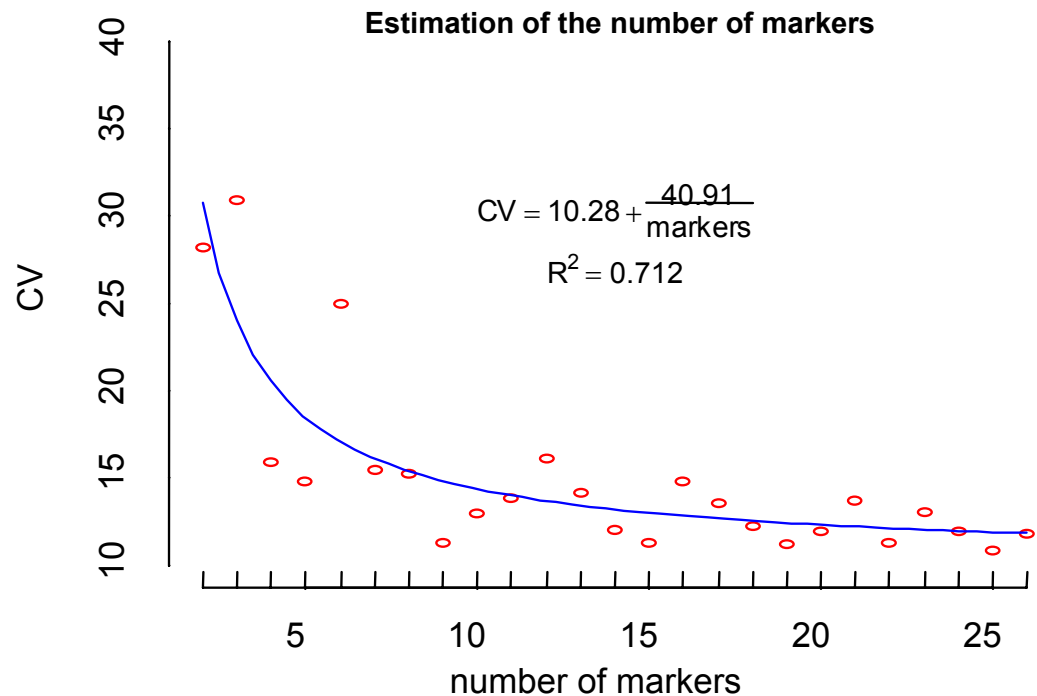
| Use case identification | R impl. (agricolae library) | R impl./HPC | Bio-Moby |
|--|-----------------------------------|----------------|----------|
| Find optimal number of molecular markers | ok | In process | TBD |
| Analyze linear models with non-normal distribution | ok | In process | TBD |

Resampling application I

F. Mendiburu, R. Simon

Find optimal number of molecular markers given a variability in set of germplasm

```
> study<-resampling.cv(  
  markers,  
  size=2,  
  npoints=25  
)
```

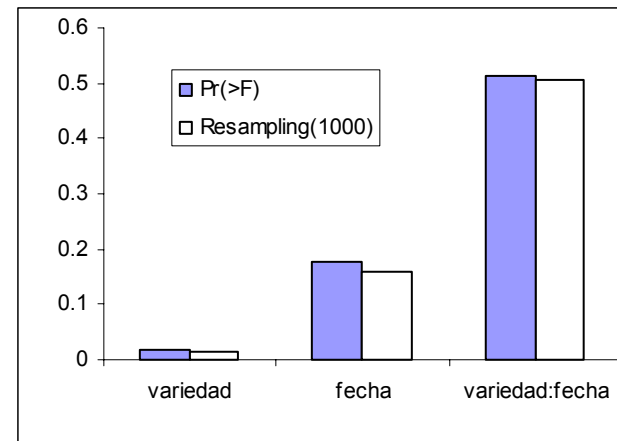
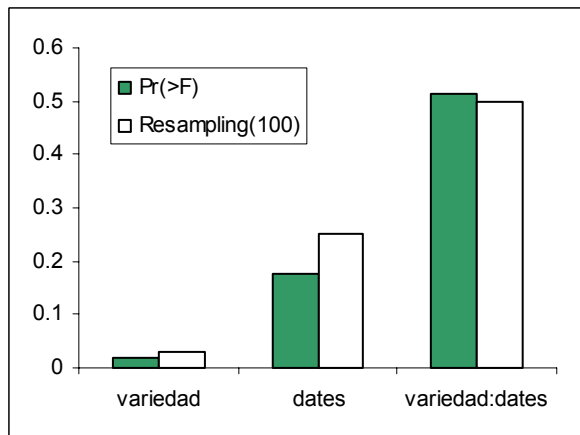


Resampling application II

F. Mendiburu, R. Simon

Use if you cannot assume a normal distribution of data and want to test for significance

```
> model<-"cutting~variety + date + variety:date"  
> resampling.model(100,data,model)
```



Effect of number of resamplings on test value using normal distributed data for reference

EMBRAPA

- EMBRAPA representatives (Marcos Costa and Georgios Pappas) participated in IRRI workshop in May and discussed Genoma project
- Genoma team working on deploying jMOBY “MOSES” technology into Genoma (ongoing)

NIAS

- Masaru attended MOBY (Vancouver, Canada) workshop in May.
- Visit by Martin Senger to NIAS (Task 23 funded). With Martin's guidance, the NIAS team successfully designed and deployed first generation of MOBY web services for the Task 32 gene expression repository using MOSES.

Looking Ahead?

Remainder of 2005

- IRRI:
 - Completion and deployment of MOSES Dashboard application
 - Full MOBY of ICIS use cases (a.k.a. GCP germplasm domain model reference implementation) and deployment of web service providers for testing by other GCP partner sites
- IPGRI/INIBAP
 - Target use case generalized to a few other crops and sites?
- CIP
 - R statistics MOBY web services provider completed and deployed to GCP HPC nodes
- EMBRAPA
 - MOBY enabled *Genoma* binary posted for download by GCP scientists
- NIAS:
 - Complete commissioning of first generation gene expression data access web services with distributed client side support in GCP platform?

Looking Further Ahead??

2006

- Q: what does “MOBY Application Development” *really* mean? Could MOBY data types and web services be auto-generated from the GCP domain model? Is a generic specification of core MOBY services for the GCP domain models possible? Would the MOBY job then be done?
- Impact of NCGR VPIN funded project on GCP objectives?
- Strive toward collaborative integration of MOBY with BioCase, GDPC and LSID development communities?
- MOBY software engineering training workshop and web service deployment hackathon for GCP developers and data source administrators (in collaboration with NCGR (Gary Schiltz) and UBC (Mark Wilkinson)).
- Could Martin Senger assume task leadership for 2006?

Proposed Workplan for 2006

- (\$40K) Contribution toward primary GCP MOBY software engineering and task leadership (Martin Senger, March-December, 2005)
 - Software engineering for tighter integration of MOBY data types and services with the GCP domain models and middleware
 - Collaborative efforts to integrate with GDPC and BioCASE
 - Training of GCP scientists in web service provider deployment using GCP (Java) platform technologies
- (\$20K) GCP web service provider (“bring your own datasource”) developers’ hackathon in early-to-mid 2006.
- (\$20K) novel web service applications/workflows commissioned via competitive partner proposals.