Research on use of the EU-data GRID environment in high energy physics and ***bio-informatics***

(CKSC)

A high throughput analyses of glucosyltransferase genes and proteins by using EU-data GRID environment and hunting of novel genes

Doman Kim

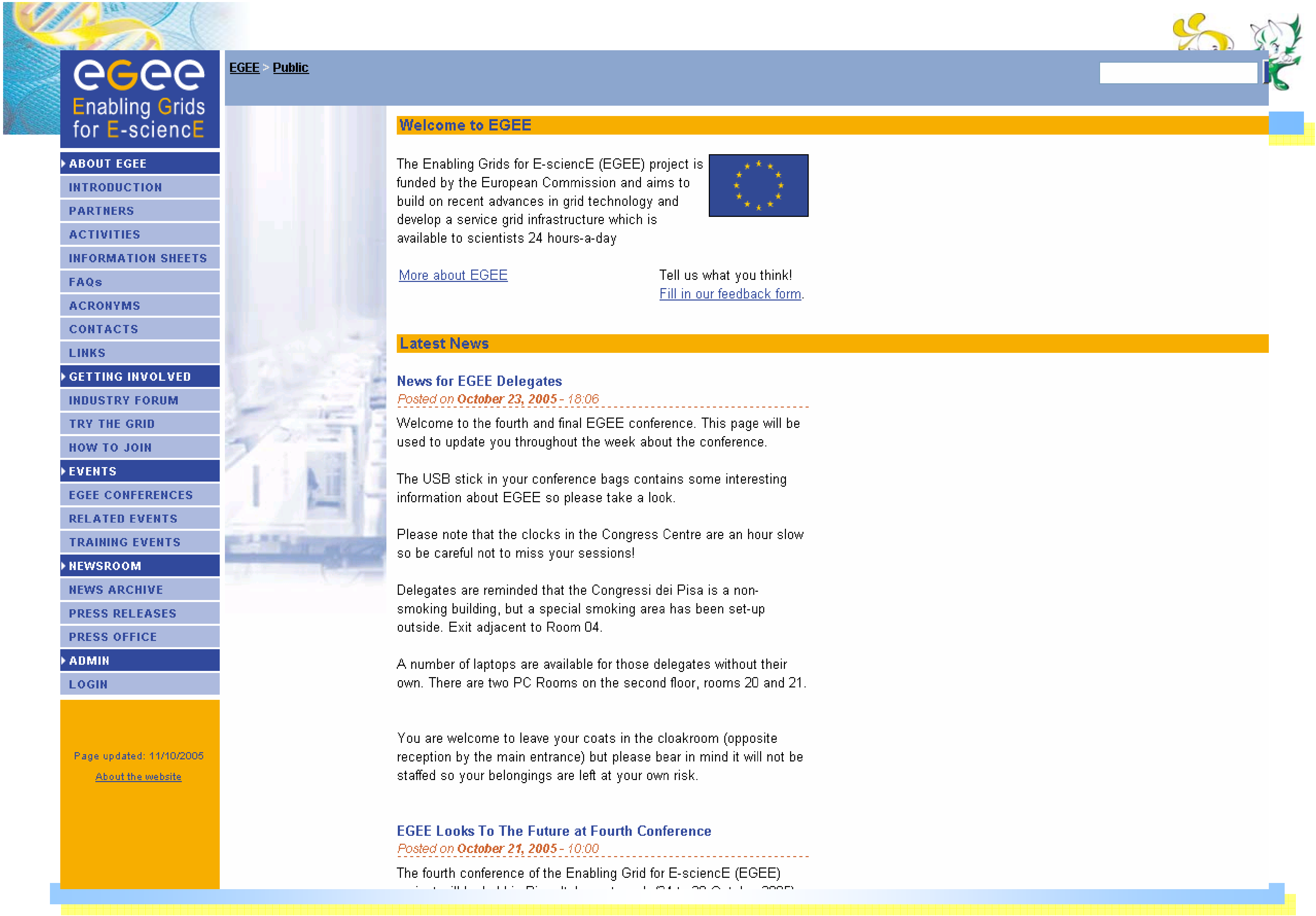
School of Biological Sciences and Technology
Chonnam National University

Functional Carbohydrate Enzymes and Microbial genomics *(http://altair.chonnam.ac.kr/~carboenz)*

전남대학교
CHONNAM NATIONAL UNIVERSITY

세종대학교
Sejong University

KANGNUNG NATIONAL UNIVERSITY
강릉대학교



Welcome to EGEE

The Enabling Grids for E-sciencE (EGEE) project is funded by the European Commission and aims to build on recent advances in grid technology and develop a service grid infrastructure which is available to scientists 24 hours-a-day

More about EGEE

Tell us what you think!
Fill in our feedback form.

Latest News

News for EGEE Delegates

*Posted on **October 23, 2005** - 18:06*

Welcome to the fourth and final EGEE conference. This page will be used to update you throughout the week about the conference.

The USB stick in your conference bags contains some interesting information about EGEE so please take a look.

Please note that the clocks in the Congress Centre are an hour slow so be careful not to miss your sessions!

Delegates are reminded that the Congressi dei Pisa is a non-smoking building, but a special smoking area has been set-up outside. Exit adjacent to Room 04.

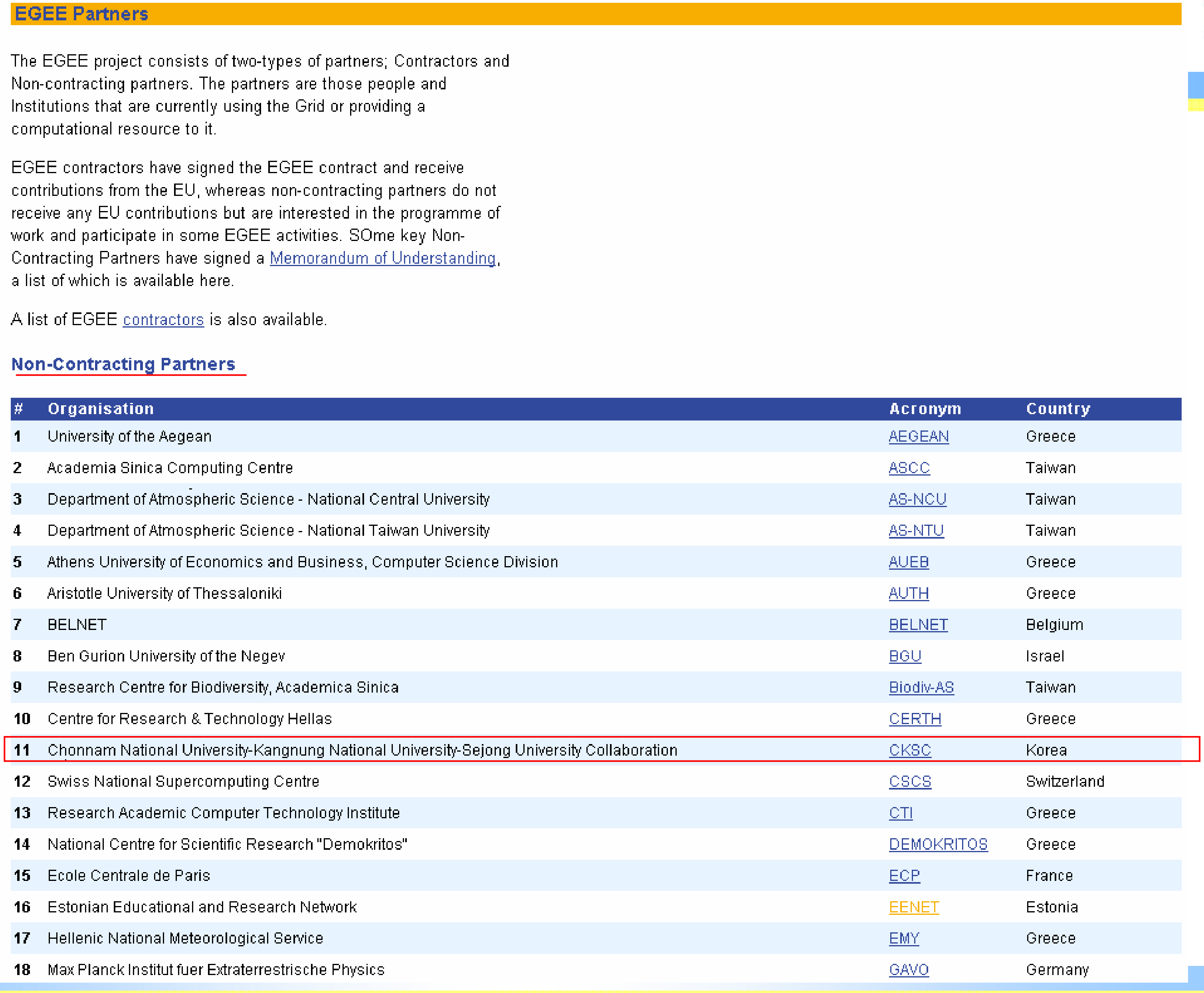
A number of laptops are available for those delegates without their own. There are two PC Rooms on the second floor, rooms 20 and 21.

You are welcome to leave your coats in the cloakroom (opposite reception by the main entrance) but please bear in mind it will not be staffed so your belongings are left at your own risk.

EGEE Looks To The Future at Fourth Conference

*Posted on **October 21, 2005** - 10:00*

The fourth conference of the Enabling Grid for E-sciencE (EGEE)



EGEE Partners

The EGEE project consists of two-types of partners; Contractors and Non-contracting partners. The partners are those people and Institutions that are currently using the Grid or providing a computational resource to it.

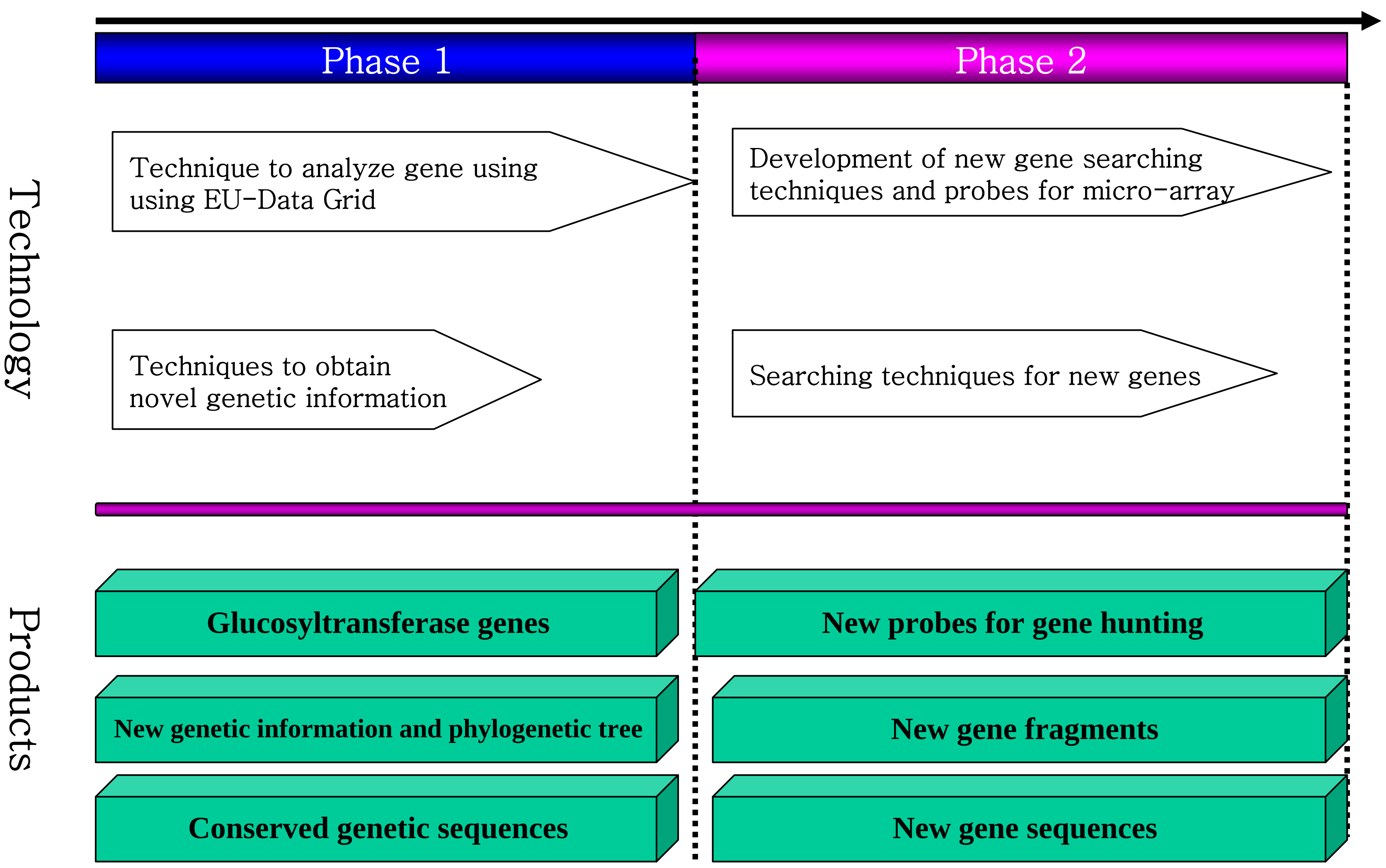
EGEE contractors have signed the EGEE contract and receive contributions from the EU, whereas non-contracting partners do not receive any EU contributions but are interested in the programme of work and participate in some EGEE activities. SOme key Non-Contracting Partners have signed a Memorandum of Understanding, a list of which is available here.

A list of EGEE contractors is also available.

Non-Contracting Partners

#	Organisation	Acronym	Country
1	University of the Aegean	AEGEAN	Greece
2	Academia Sinica Computing Centre	ASCC	Taiwan
3	Department of Atmospheric Science - National Central University	AS-NCU	Taiwan
4	Department of Atmospheric Science - National Taiwan University	AS-NTU	Taiwan
5	Athens University of Economics and Business, Computer Science Division	AUEB	Greece
6	Aristotle University of Thessaloniki	AUTH	Greece
7	BELNET	BELNET	Belgium
8	Ben Gurion University of the Negev	BGU	Israel
9	Research Centre for Biodiversity, Academica Sinica	Biodiv-AS	Taiwan
10	Centre for Research & Technology Hellas	CERTH	Greece
11	Chonnam National University-Kangnung National University-Sejong University Collaboration	CKSC	Korea
12	Swiss National Supercomputing Centre	CSCS	Switzerland
13	Research Academic Computer Technology Institute	CTI	Greece
14	National Centre for Scientific Research "Demokritos"	DEMOKRITOS	Greece
15	Ecole Centrale de Paris	ECP	France
16	Estonian Educational and Research Network	EENET	Estonia
17	Hellenic National Meteorological Service	EMY	Greece
18	Max Planck Institut fuer Extraterrestrische Physics	GAVO	Germany

Research Plans



2004 2006 2008

Phase 1

Phase 2

Technology

Technique to analyze gene using using EU-Data Grid

Techniques to obtain novel genetic information

Development of new gene searching techniques and probes for micro-array

Searching techniques for new genes

Products

Glucosyltransferase genes

New genetic information and phylogenetic tree

Conserved genetic sequences

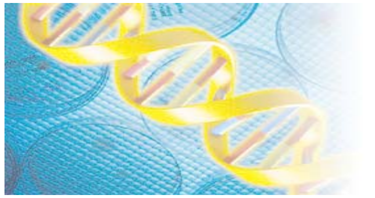
New probes for gene hunting

New gene fragments

New gene sequences

Glycosyltransferases

- One of five carbohydrate related enzymes such as glycosyltransferase, glycosidases, transglycosidases, polysaccharide lysases, and carbohydrate esterase.
- Glycosyltransferases (EC 2.4.x.y) enzyme which catalyses the transfer of sugar moieties from activated donor molecules to specific acceptor molecules, forming glycosidic bonds.
- Glycosyltransferases involve in biosynthesis of disaccharides, oligosaccharides and polysaccharides, and misglycosylation leads to many kinds of diseases such as diabetes, cancer, rheumatoid arthritis, asthma, coagulation, inflammation.
- It forms glycoprotein or glycolipid by attaching onto their side chains and exists on cell surface in order to assist cell immune system as recognizers from intrusion of harmful organisms outside.



Why this study is important?

1. Glycobiology shows many important functions in many biological organism including human.

2. "Bioinformatics" becomes stressed for its importance to analyze or reveal systemically huge genomic information from various organisms including human.

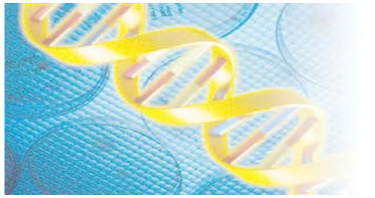
3. Study focus is hunting the novel genes with new genetic information through DNA and protein sequence analysis of glucosyltransferases and furthermore inventing useful biomaterials using those enzymes in real world.

Why this study is important?

4. Study outcomes make it produce the distinct phylogenetic tree of glucosyltransferase among biological organisms and confirm the relationship between enzyme characteristics and gene sequences.

5. It will contribute on potential applications in the food, cosmetic, and medical industries, understanding on disease mechanisms and therapeutic pathways.

6. This can give functionality of glycosides in human proteins and/or enzymes. The unique characteristic genes of glucosyltransferase will be designed and evolved further for synthesis of valuable functional carbohydrates and proteins.



What EGEE makes it possible?

- Glycobiology-relevant genes and proteins information will be collected for analysis with high computing speed from differently formatted database .
- Huge information related to glycosyltransferase will be analyzed and interpreted with EGEE but will not with other systems.
- EGEE could draw the precise correlation from information resources for a short time period.
- EGEE could lead to develop and upgrade tools for fast comparison/ analysis of vast genetic information.

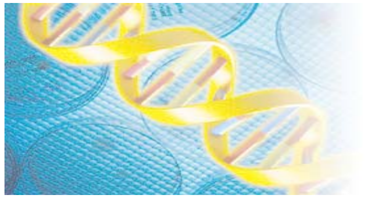
Overall Goals

➢ Scientifically…

- Analysis of glucosyltransferase genes and proteins through EU-data GRID environment and construction of phylogenetic tree based on output results.
- Development of oligonucleotide probes for microarray to obtain new genetic information by comparing properties of glucosyltransferase and novel genes.
- Identification of new characteristic glucosyltransferase and Synthesis of useful functional biomaterials.

➢ Practically …

- Development of novel characteristic glucosyltransferase gene
- Functional biomaterials: foods, cosmetics, and pharmaceuticals-medical industries, understanding on disease mechanisms and therapeutic pathways
- Development of new functional materials and construction of production systems



Research Approach

1. Information related to glycosyltransferase genes
2. Construction of the phylogenetic tree for glycosyltransferases
3. Development of DNA microarray probe and genes
4. Gene sequence determination, property studies & gene evolution
5. Gene sequence comparison, functionality detection, and artificial gene construction

1. Information related to glycosyltransferase genes:

A. Analysis of Glycosyltransferase Genomes

Glucosyl-transferring enzyme groups

Current enzyme classification : 347 families (http://afmb.cnrs-mrs.fr/CAZY/CBM_intro.html)

nucleotides (6997), proteins (17610), genomes (247), structures (214), 3D domain (980) are registered in NCBI (National Center For biotechnology Information ;http://www.ncbi.nlm.nih.gov)

Glycosidase & Transglycosidase

(100 families)

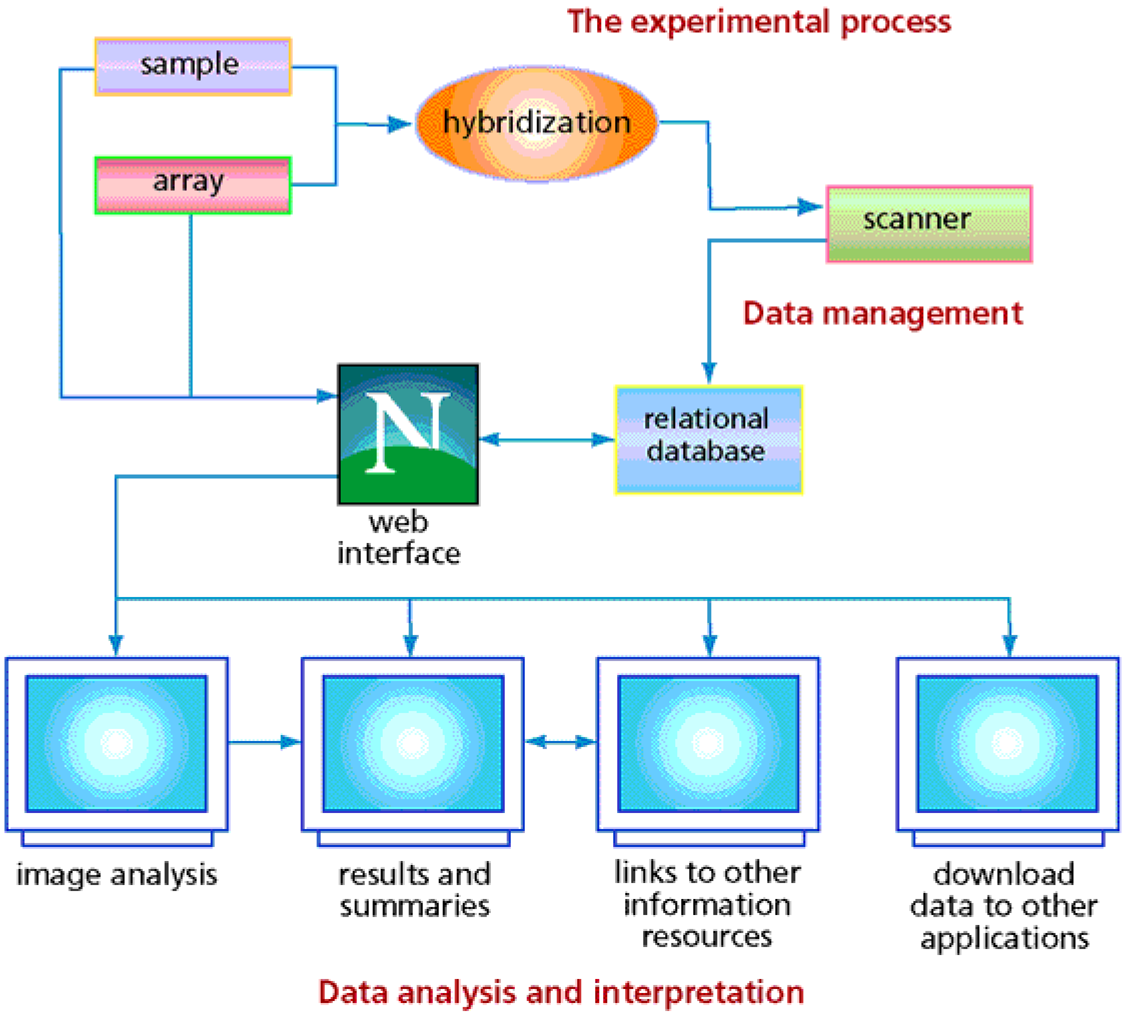
Glycosyltransferase (78), Polysaccharide Lyases (18), Carbohydrate esterase (14), Carbohydrate-binding module (43)

(247 families)

They are classified based on substrate or/and acceptors specificity.

- Detail structural analysis among individual enzymes were studied within or among groups

A. Analysis of Glycosyltransferase Genomes



The experimental process

sample

array

hybridization

scanner

Data management

web interface

relational database

image analysis

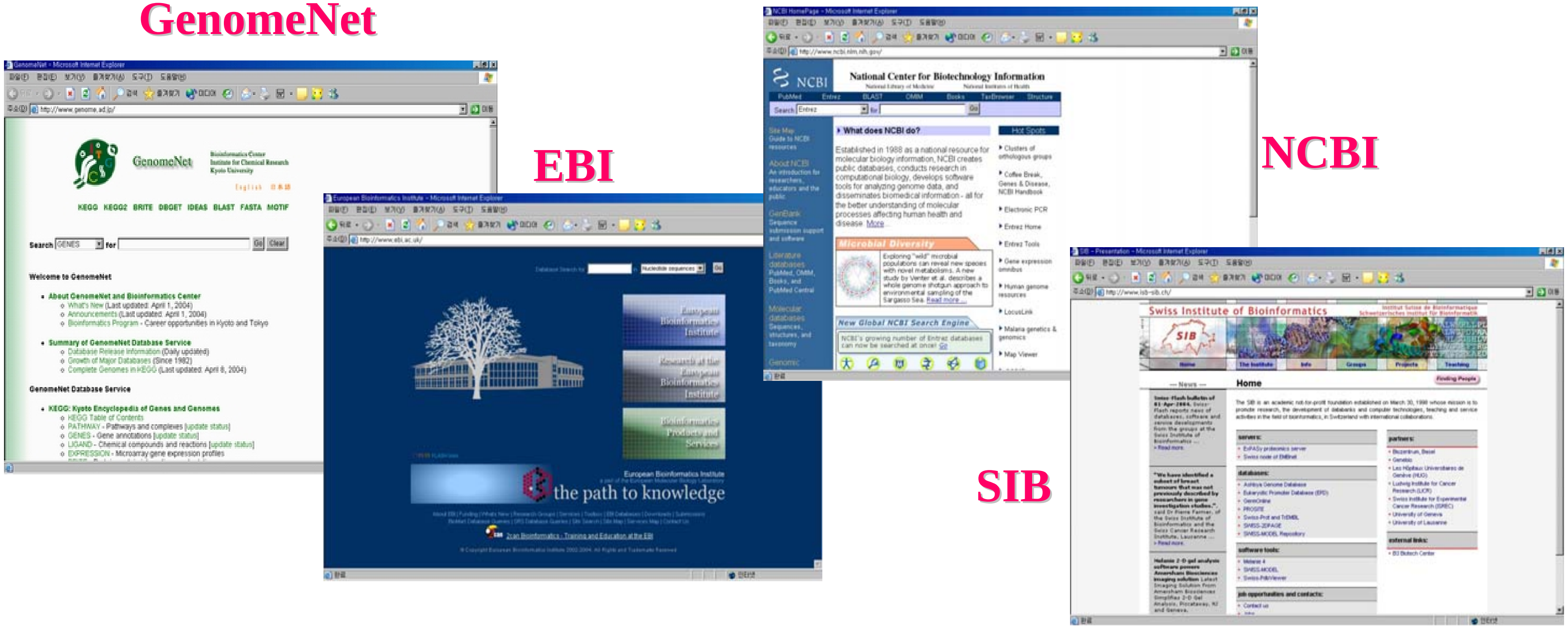
results and summaries

links to other information resources

download data to other applications

Data analysis and interpretation

B. Current Bioinformatics System



GenomeNet

http://www.genome.ad.jp/

GenomeNet
Bioinformatics Center
Institute for Chemical Research
Kyoto University

KEGG KEGG2 BRITE DBGET IDEAS BLAST FASTA MOTIF

Welcome to GenomeNet

- About GenomeNet and Bioinformatics Center
  - What's New (Last updated: April 1, 2004)
  - Announcements (Last updated: April 1, 2004)
  - Bioinformatics Program - Career opportunities in Kyoto and Tokyo
- Summary of GenomeNet Database Service
  - Database Release Information (Daily updated)
  - Growth of Major Databases (Since 1982)
  - Complete Genomes in KEGG (Last updated: April 8, 2004)

GenomeNet Database Service

- KEGG: Kyoto Encyclopedia of Genes and Genomes
  - KEGG Table of Contents
  - PATHWAY - Pathways and complexes [update status]
  - GENES - Gene annotations [update status]
  - LIGAND - Chemical compounds and reactions [update status]
  - EXPRESSION - Microarray gene expression profiles

EBI

http://www.ebi.ac.uk/

European Bioinformatics Institute

the path to knowledge

NCBI

http://www.ncbi.nlm.nih.gov/

National Center for Biotechnology Information

What does NCBI do?

Established in 1988 as a national resource for molecular biology information, NCBI creates public databases, conducts research in computational biology, develops software tools for analyzing genome data, and disseminates biomedical information - all for the better understanding of molecular processes affecting human health and disease. More...

Microbial Diversity

Exploring "wild" microbial populations can reveal new species with novel metabolisms. A new study by Venter et al. describes a whole genome shotgun approach to environmental sampling of the Sargasso Sea. Read more ...

New Global NCBI Search Engine

NCBI's growing number of Entrez databases can now be searched at once! Go

SIB

http://www.isb-sib.ch/

Swiss Institute of Bioinformatics

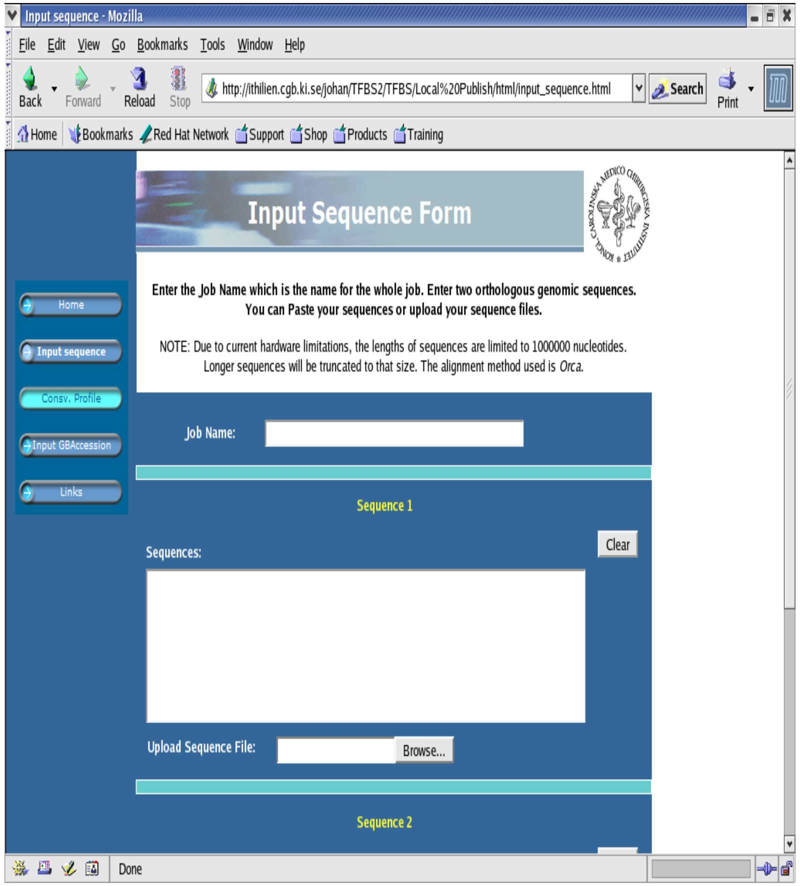
Home

The SIB is an academic not-for-profit foundation established on March 30, 1998 whose mission is to promote research, the development of databanks and computer technologies, teaching and service activities in the field of bioinformatics, in Switzerland with international collaborations.

Current information store for genes, genomes, proteins and their structures

- Limited size of CPU and poor storage capacity
- Local resources have limited permission for tool uses and analyzing data sizes

B. Current Bioinformatics System



Enter the Job Name which is the name for the whole job. Enter two orthologous genomic sequences. You can Paste your sequences or upload your sequence files.

NOTE: Due to current hardware limitations, the lengths of sequences are limited to 1000000 nucleotides. Longer sequences will be truncated to that size. The alignment method used is *Orca*.

Job Name:

Sequence 1

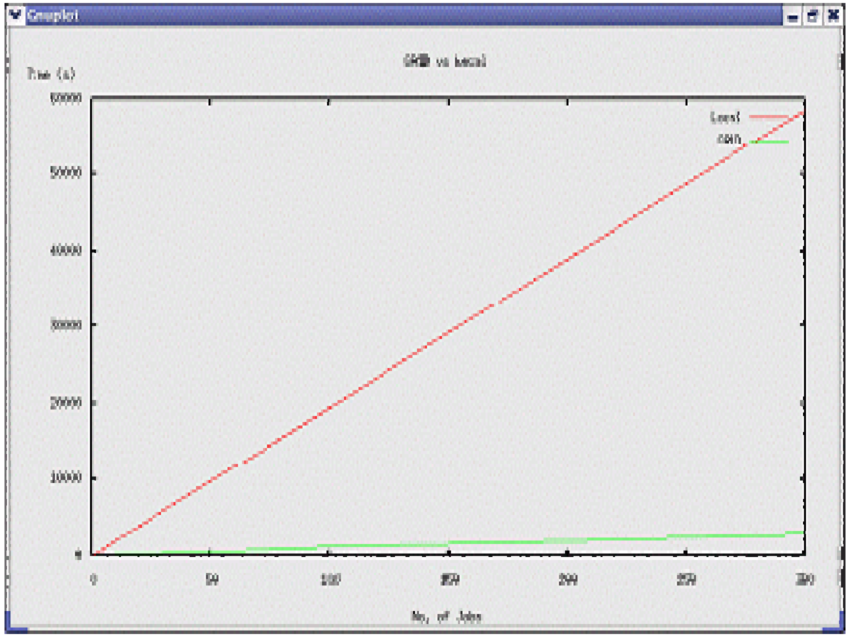
Sequences:

Upload Sequence File:

Sequence 2

- Web Interface

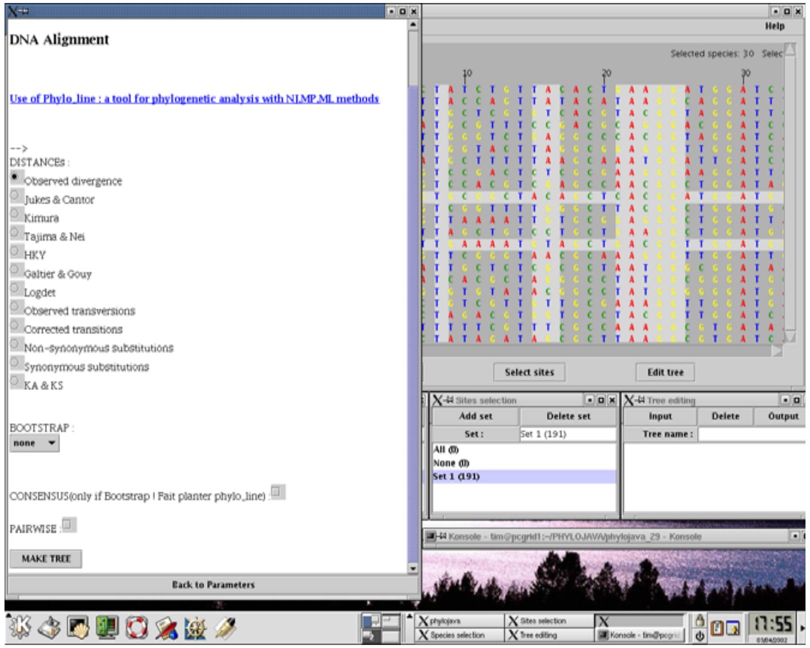
- Promoter/ Expression



- Comparison between Data Grid (green line) and conventional method (red line) in the respect of time required for analysis – Data Grid required 25 times less time

2. Construction of the phylogenetic tree for glycosyltransferases

- **EU Bioinformatics Group - NPS@(Network Protein Sequence Analysis, http://npsa-pbil.ibcp.fr/) Construction**
- **GPS@Grid Web - Supply analysis tools and storage through IST Data Grid project**


DNA Alignment

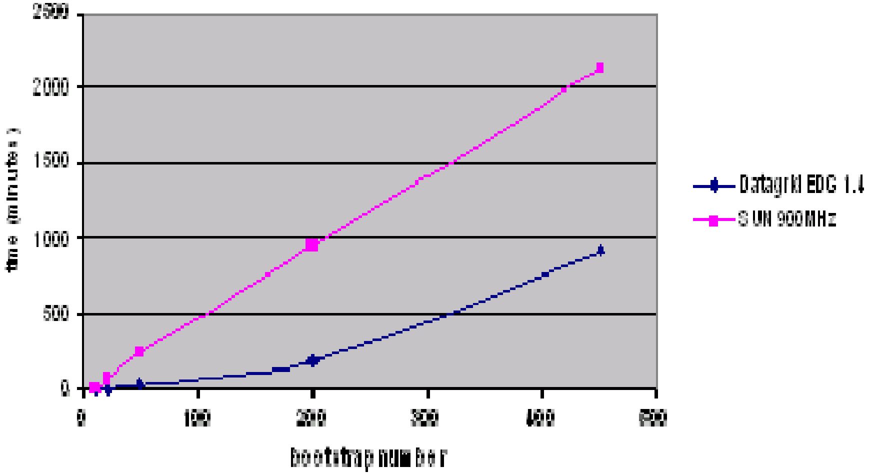
Use of Phylo_line : a tool for phylogenetic analysis with NJ,MP,ML methods

-->
DISTANCEs :
● Observed divergence
○ Jukes & Cantor
○ Kimura
○ Tajima & Nei
○ HKY
○ Galtier & Gouy
○ Logdet
○ Observed transversions
○ Corrected transitions
○ Non-synonymous substitutions
○ Synonymous substitutions
○ KA & KS

BOOTSTRAP : none

CONSENSUS(only if Bootstrap ! Fait planter phylo_line) : ☐

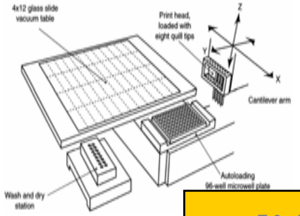
PAIRWISE : ☐



Comparison of CPU time

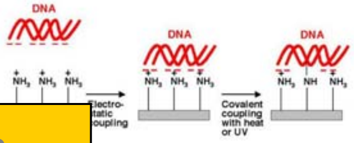
3. Development of DNA Microarray Probe/ Genes

Clone
1 ↓
Plasmid
2 ↓
PCR
3 →

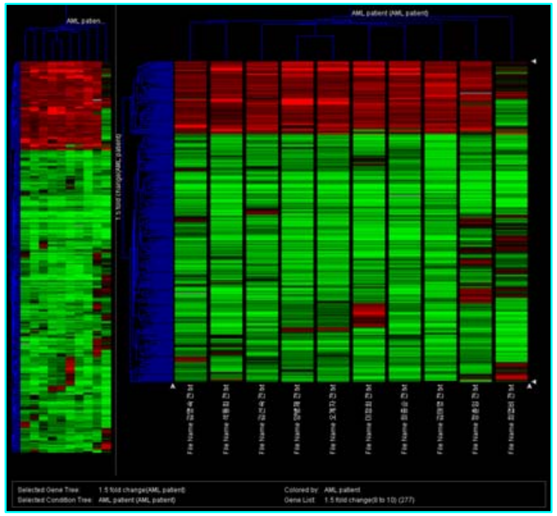


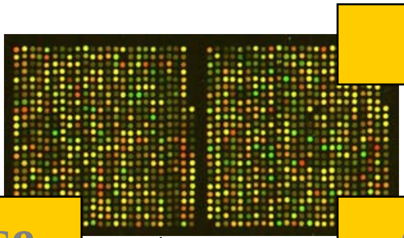
Slide glass
4 ↓
Array
5 ↓
Post-proc
→

Sample
↓
Total RNA
6 ↓
Probe

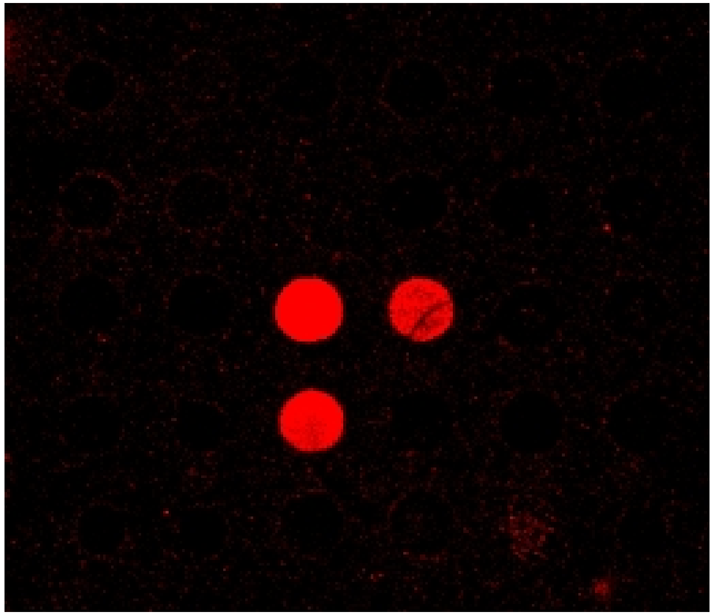


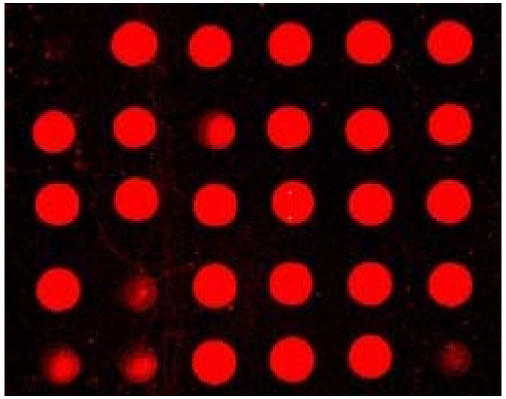
↓
Hybrid
7 ↓
Scan
8 ↓
Analysis
←
Database

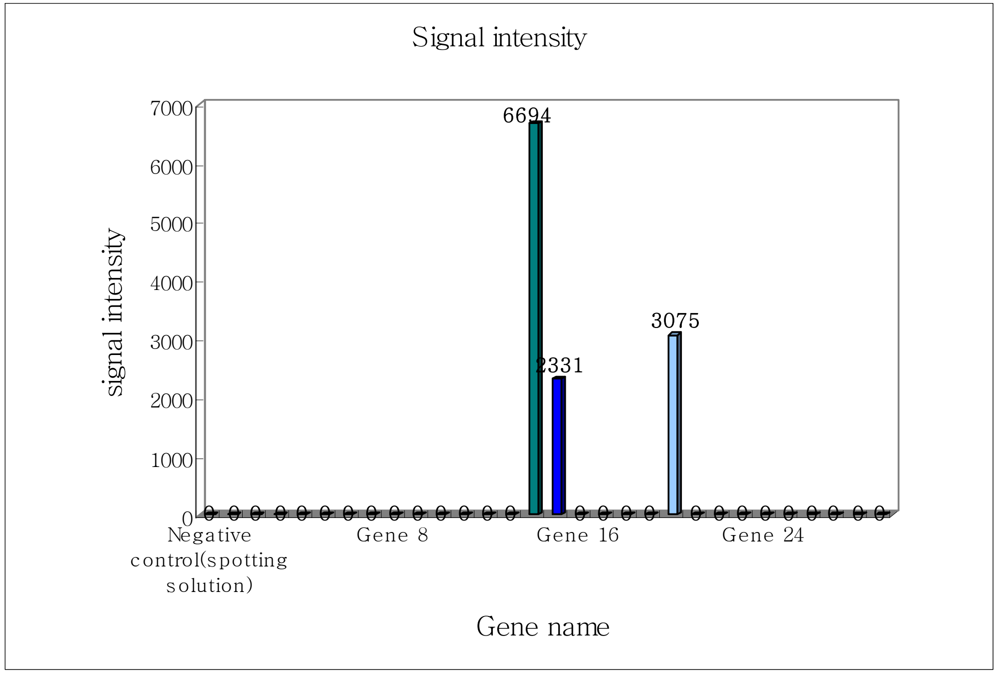




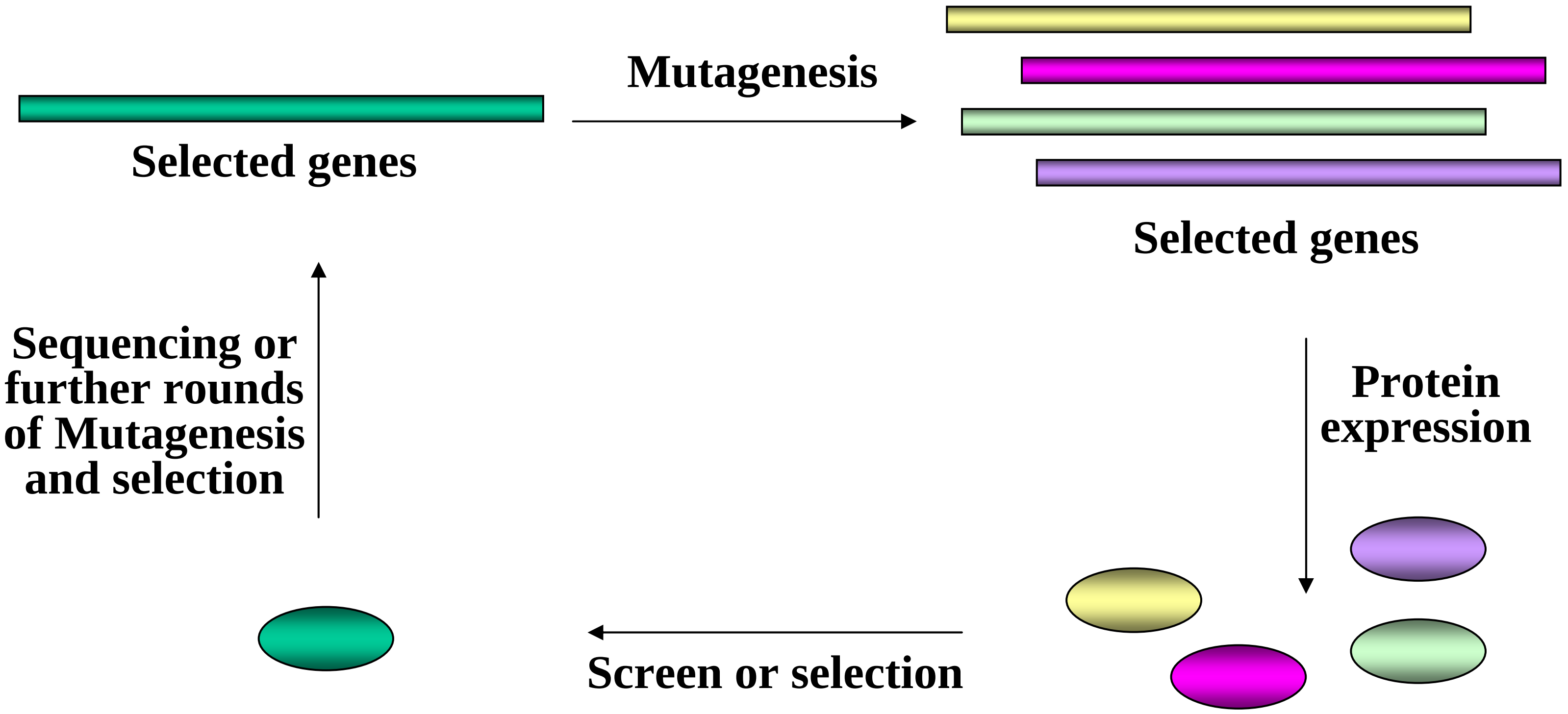
Example : Dextran hydrolyzing enzyme







4. Gene sequence determination, Property studies/ Gene evolution



Selected genes

Mutagenesis

Selected genes

Protein expression

Library of mutant proteins

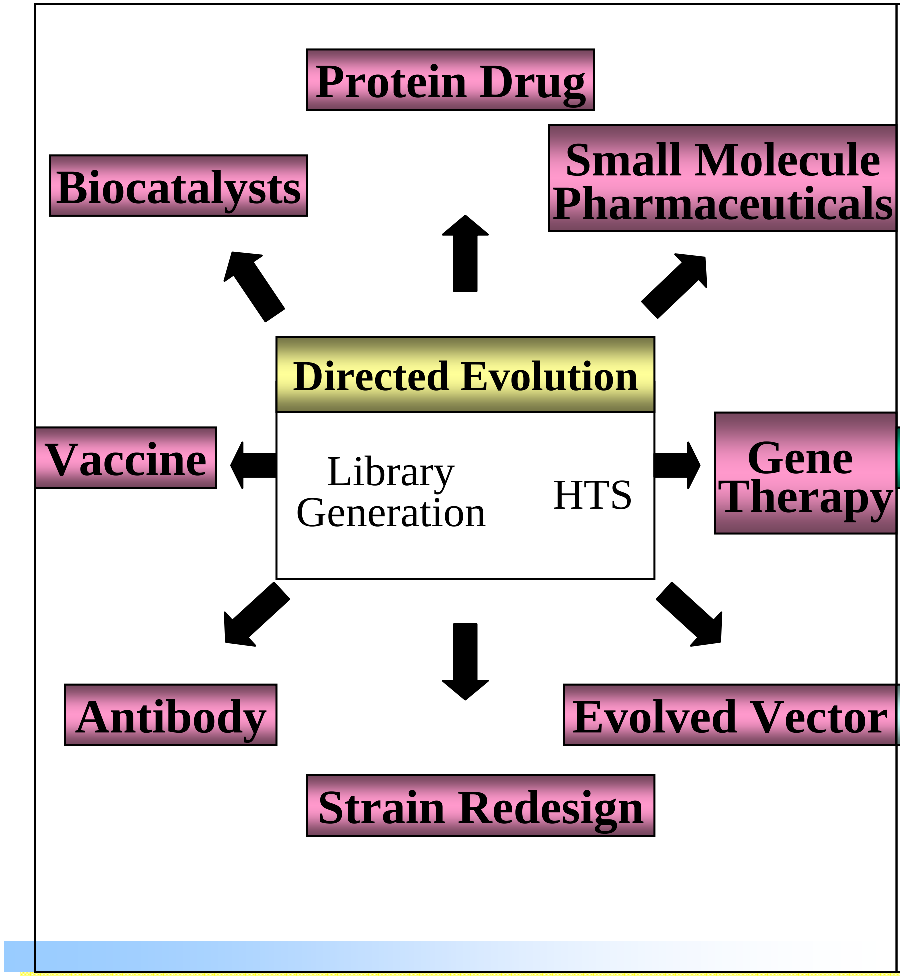
Screen or selection

Proteins with desired property

Sequencing or further rounds of Mutagenesis and selection

5. Gene Sequences/Functionality/Artificial Gene construction

Application of Directed Gene Evolution



Directed Evolution

Library Generation

HTS

Protein Drug

Biocatalysts

Small Molecule Pharmaceuticals

Vaccine

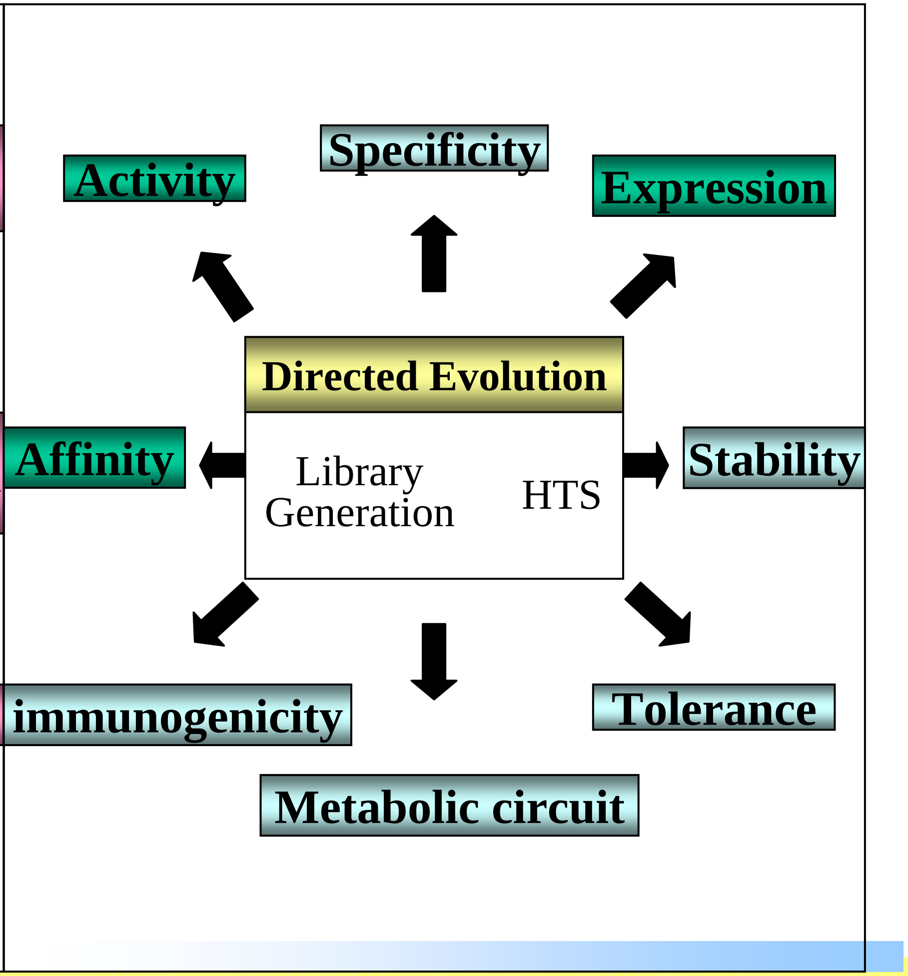
Gene Therapy

Antibody

Evolved Vector

Strain Redesign

Purpose of Directed Gene Evolution



Directed Evolution

Library Generation

HTS

Activity

Specificity

Expression

Affinity

Stability

immunogenicity

Tolerance

Metabolic circuit

Expecting Effects

- Developing and upgrading of tools for fast comparison/analysis of huge genetic information.
- Novel-designed genes with unique characteristics will be produced based on the analyzed data, and then biomaterials will be developed and synthesized by using the novel enzymes for practical applications.
- Genetic information support for developments and production of new functional biomaterials as for the Growth Engines of Korea (and others).
- Industrial improvements of food, pharmaceutical-medical fields through genetic information gathering and applications.

Study Outcomes

- 18 yrs Research History for glycosyltransferases
    - Secure many glycosyltransferase genes, enzymes, strains

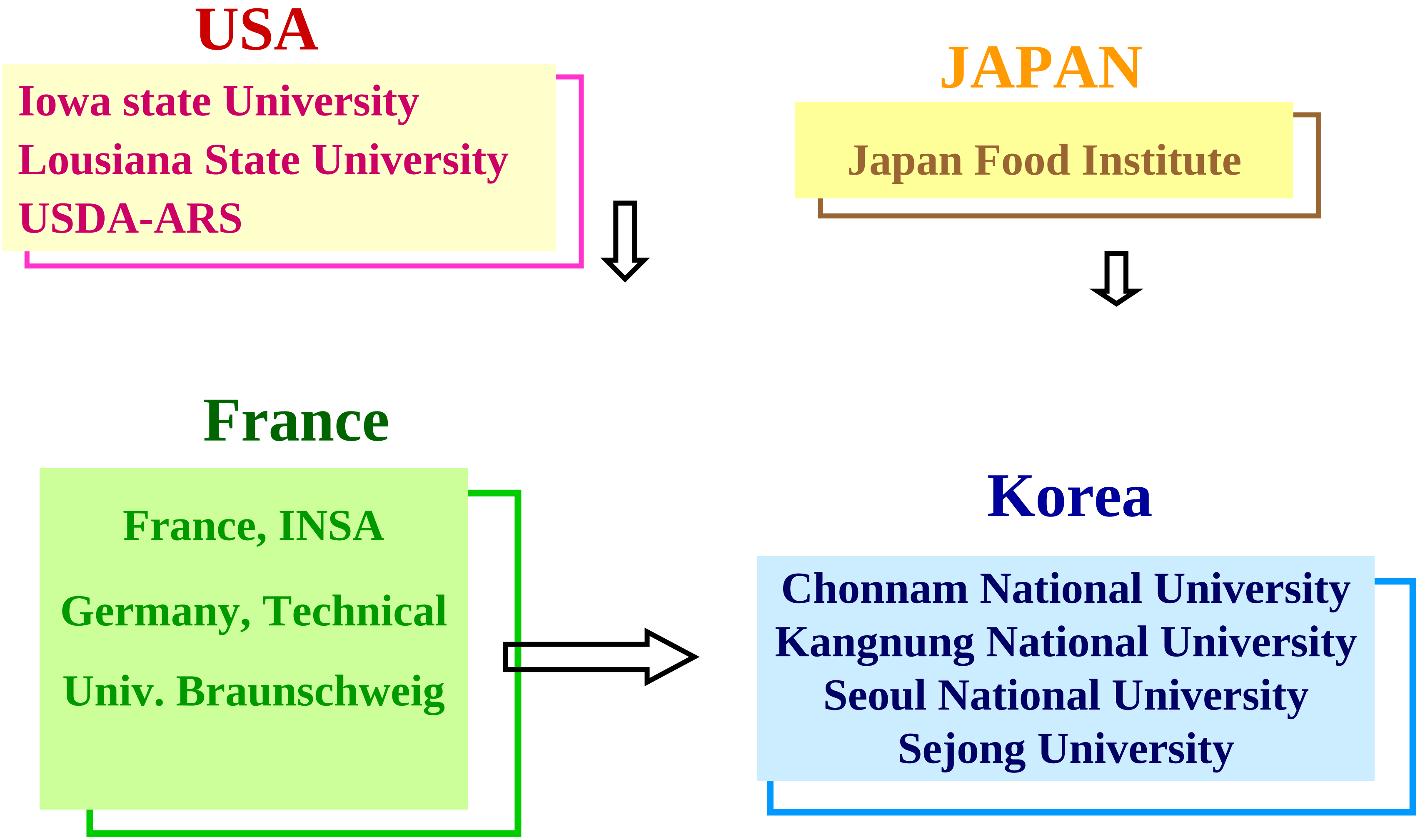
- Study focuses:
    - Synthesis of functional carbohydrates and glycosides using glycosyltransferase.

    - Understanding of catalysis mechanisms to develop novel functionalities.

- Publication:

    - 54 SCI International Journals
    - 26 Domestic Journals
    - 8 Patent Registrations including 8 Pending

Collaboration



USA

Iowa state University
Lousiana State University
USDA-ARS

JAPAN

Japan Food Institute

France

France, INSA

Germany, Technical

Univ. Braunschweig

Korea

Chonnam National University
Kangnung National University
Seoul National University
Sejong University

전남대학교 CHONNAM NATIONAL UNIVERSITY

세종대학교 Sejong University

KANGNUNG NATIONAL UNIVERSITY 강릉대학교

Thank you !!!