

Des systèmes d'information pour le développement et l'amélioration de plantes cultivées

Equipe Intégration des Données



30 Avril 2008
Stéphanie Sidibe Bocs

1. **Plateau technique de systèmes d'information (SI)**
2. **Recherche en SI**
3. **Besoins et souhaits de collaboration**

SI du plateau accessibles par internet

Gestion de données produites par l'UMR + partenaires extérieurs



Gestion des ressources génétiques et génomiques de plantes tropicales [Ruiz NAR 2004]



Gestion de ressources génétiques, phénotypiques et génomique du cacoyer



Exploration de la génétique inverse du riz [Droc NAR 2006]



Caractérisation phénotypique d'une collection de lignées d'insertion de T-DNA du riz [Larmande NAR 2008]

SI du plateau accessibles par intranet

Gestion de données produites par le département BIOS



Traçabilité des OGM



Gestion et annotation de la bibliographie

SI du plateau accessibles par internet

Analyse de données produites par l'UMR + partenaires extérieurs



Analyse de collection d'ESTs et
Identification de SNPs [Argout soumis]



Prédiction d'orthologues par
phylogénomique [Conte NAR 2008]



Identification de microsatellites dans des
banques enrichies [Dereeper BMC 2007]

GenDiversity

Analyse des données de diversité génétique

Recherches en SI

- ◉ Modèles de données, intégration et gestion de quantité massives de données complexes et hétérogènes
- ◉ Ontologies, vocabulaire contrôlé et Web service pour l'interopérabilité entre les SI
- ◉ Analyse statistique de données
- ◉ Reconstruction d'arbres phylogénétique et de graphes (réseaux)
- ◉ Combinaison de multiples sources d'évidences pour extraire des connaissances (senseurs)
- ◉ Représentation graphique et interfaces de navigation multi-échelle

SI en cours

Oryzon	Interopérabilité entre les systèmes OrygenesDB, Oryza Tag Line, Greenphyl et Tair
GnpAnnot	Annotation structurale, fonctionnelle et relationnelle de gènes et de TE dans les génomes eucaryotes
Haplophyle	Réseaux haplotypiques graphiques à la lumière de données éco-géographiques
MS-DMind	Gestion et exploration multi-échelle de connaissances sur le blé

Besoins : exemple Haplophyle

A Montpellier

- Etude de nombreuses plantes **cultivées**
- De plus en plus de données de génotypage (balayage du génome, gènes candidats...)
- Combinaison avec des données externes (phénotype, passeport, écologie, ethnographie, résultats d'expériences)

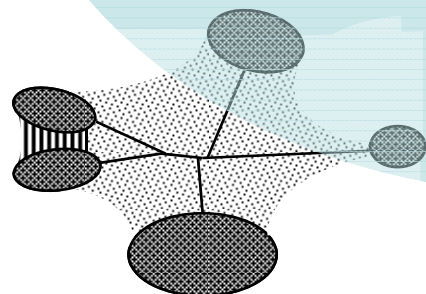
Objectifs

- Analyser la diversité nucléotidique de ces espèces au niveau de gènes candidats et intégrer cette diversité dans des scénarios
 - d'évolution de l'espèce avec la **domestication** et
 - de réponse des gènes à la **sélection**.

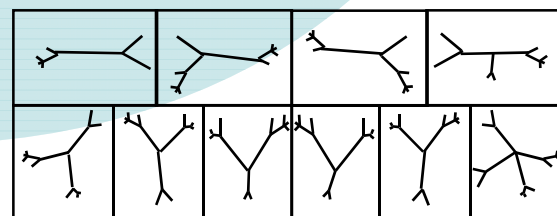
Utiliser différents types de données pour reconstituer l'histoire évolutive



données écogéographiques

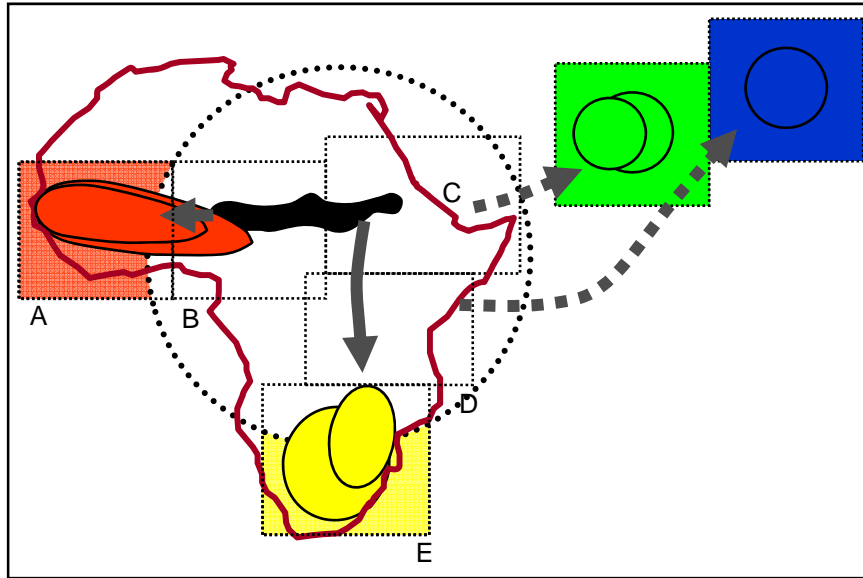


marqueurs neutres → structure globale



gènes candidats → phylogénies d'haplotypes

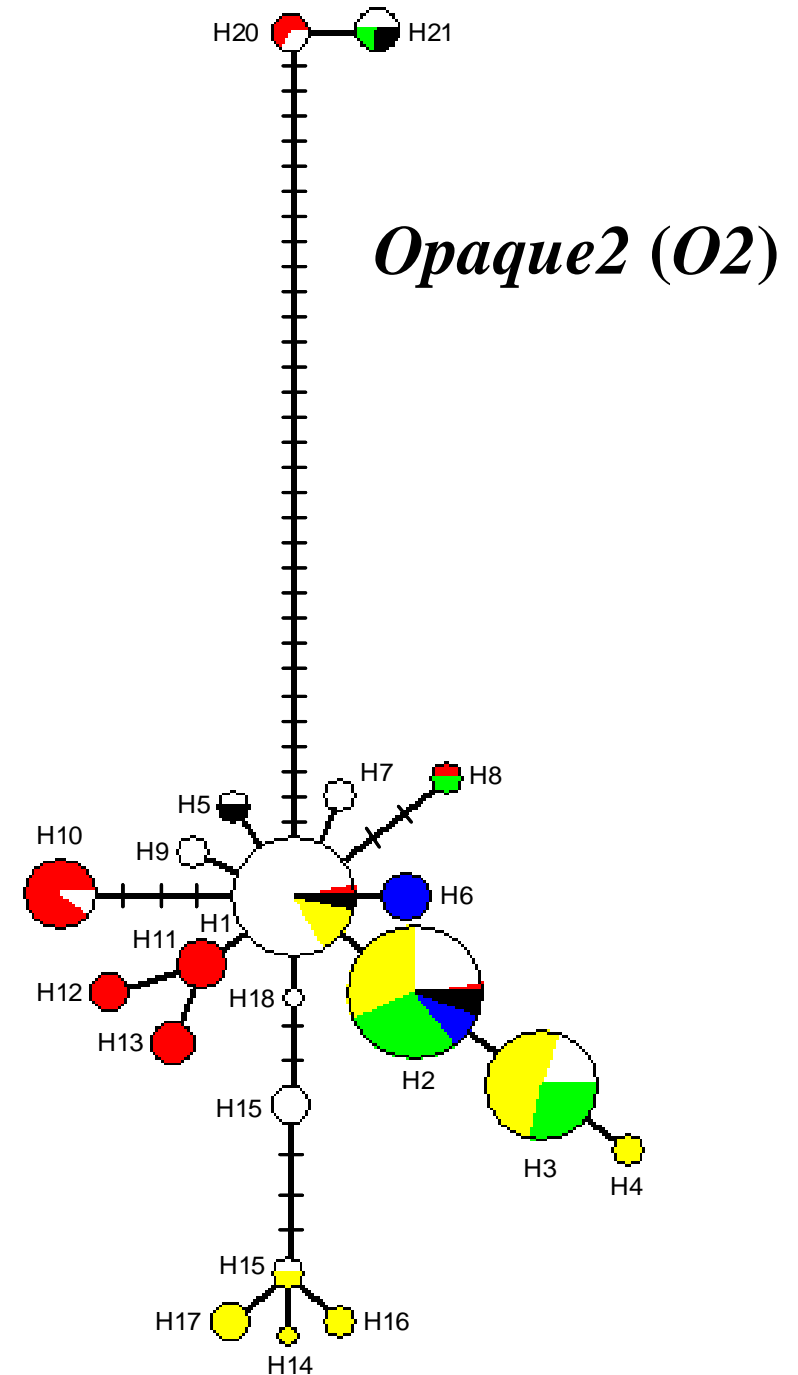
Evidence phylo-géographique de néo-diversité chez le sorgho cultivé



De Alencar Figueiredo et al, Genetics 2008, sous presse

- un projet financé par le Challenge Programme Generation (www.generationcp.org)

- une base pour un projet RTRA *Computational Plants and Ecosystems* (DAP, DiaPC, ...)?



Passport data

Line: AIV B06	
Cocultured callus	Other lines generated from the same callus
Construct	p4978
T2 seed stock	available
Delivered	No
FST	CU324322 CL520431

Mutant: AIV B06/0

Referenced or designated mutants	semi-dwarf (abbr. Sd1)
Phenotypic class	morphology
Trait	Size
Organ	all organs
Developmental stage	tillering to heading
Phenotype description	Semi-dwarf. Plant height is about 2/3 of normal height; normal to late flowering; normal to panicles incomp. exerted; small panicles and spikelets; with or without sterility; normal to narrow and small leaves (shorter than average size leaves); rolled to semi-rolled leaves; normal to twisted leaves (wavy, crinkly leaf); with or without erect leaves; normal to dark green leaves; normal to tillering; with or without thin tillers; normal to compact plant.
Overall observations	Semi-dwarf (2/3 of normal height) in proportion; late flowering; semi-rolled leaves; upper spikelets with hull (5 plants). (Photos 1, 2).
Observed segregation	5 mutant(s) over 22 plants => %



OryzaTagLine

OryGenesDB

Gene TIGR
Os10g36924.1
aquaporin NIP5.1, putative, expressed

Gene IRGSP
Os10g36924.1

Rice Gene Index
TC302368
TC320171

Flanking Sequence Tag
R07511
R07512
R07513
R07514
R07515
R07516
R07517
R07518
R07519
R07520
R07521
R07522
R07523
R07524
R07525
R07526
R07527
R07528
R07529
R07530
R07531
R07532
R07533
R07534
R07535
R07536
R07537
R07538
R07539
R07540
R07541
R07542
R07543
R07544
R07545
R07546
R07547
R07548
R07549
R07550
R07551
R07552
R07553
R07554
R07555
R07556
R07557
R07558
R07559
R07560
R07561
R07562
R07563
R07564
R07565
R07566
R07567
R07568
R07569
R07570
R07571
R07572
R07573
R07574
R07575
R07576
R07577
R07578
R07579
R07580
R07581
R07582
R07583
R07584
R07585
R07586
R07587
R07588
R07589
R07590
R07591
R07592
R07593
R07594
R07595
R07596
R07597
R07598
R07599
R07600

Rice Expression Database
R173469

BBM1 Similar Arabidopsis

Interpro Domain
IPR000425

FSTs

Accession Number: CL520431
Line ID: AIVB06
Phenotype: Yes
Location: 10:19151852-19451938
Mutagen: T-DNA
Source: Genoplante - OTL
Construct: Enhancer GUS or GFP
Enzym: SspI
Border: LB

Gene Arabidopsis
Name: AT4G10380.1
Location: 4:6431235-6434818
Description: water channel
GreenPhyl Information:
- Gene: gopst1
TAIR Information:
- Locus information
- Gene information
- Protein information
Publication:
- 12492832 (inferred from Reviewed Computational Analysis)
Swissprot Information:
- Swissprot entry: Q95V91
- Comment: Probable aquaporin NIP5.1

TAIR

Home Help Contact About Us Login

Search Browse Tools Stocks Portals Download Submit News

Gene Search

Locus: AT4G10380

Date last modified: 2003-05-02
TAIR Accession: Locus 2122829
Representative Gene Model: AT4G10380.1

Other names: F7L13.6, NIP5.1, NLM6, NLM8, NOD26-LIKE INTRINSIC PROTEIN 5.1

Description: Essential for efficient Boron uptake and plant development under boron limitation

Annotations

Category	Relationship Type	Keyword
GO Biological Process	involved in	transport, boron transport, response to boron
GO Cellular Component	required for	boron transport
GO Cellular Component	located in	membrane, plasma membrane
GO Molecular Function	has	water channel activity, boron transporter activity
Plant structure	expressed in	root, flower, leaf
	expressed only in	root

Annotation Detail

Genoplasm	Name/Insg	Polymorphisms	Background	Stock Name	Select
Showing 5 of 2 entries	SALK_122287	SALK_122287.48.10.x		SALK_122287	not available
	SALK_122287C	SALK_122287.48.10.x		SALK_122287C	<input type="checkbox"/>

Phenotypes

- Inhibition root, shoot, flowers and siliques growth when Boron was supplied at concentrations of <10 uM. Takano, et al. (2006)

Publication

title	source	associated gene models	date
The Arabidopsis Major Intrinsic Protein NIP5.1 Is Essential for Efficient Boron Uptake and Plant Development under Boron Limitation.	THE PLANT CELL	AT4G10380.1	2006
Whole gene family expression and drought stress regulation of aquaporins.	PLANT MOLECULAR BIOLOGY	AT4G10380.1	2005
The complete set of genes encoding major intrinsic proteins in Arabidopsis provides a framework for a new nomenclature for major intrinsic proteins in plants.	PLANT PHYSIOLOGY	AT4G10380.1	2001

AT4G10380.1

major intrinsic family protein / MIP family protein, contains Pfam profile: MIP PF0023

OryGenesDB entry: AT4G10380.1 UniProt entry: Q95V91 TIGR entry: AT4G10380.1

Annotation page

Locus alias: NIP51

No GO information from IPR motif

Sequence classification:
126 (MCL_11.2): Major intrinsic family (MIP)
90 (MCL_12):
332 (MCL_13)
313 (MCL_15): NOD26-like intrinsic protein (NIP) family

Similarity evidence (BBM1 and INPARANOID)

* Most similar BBM1: Os10g36924.1

* Query belong to Inparanoid group N° 5048

Species	Confidence value %	Confidence value %
Arabidopsis	100	
Oryza		100
Os10g36924.1		100

GREENPHY, Phylogenetic predictions

Orthology (o) Subtree-neighbor (n) SuperOrthologs (s) Distance (D)

Species	o	n	s	D
Os10g36924.1				
UniProt	Alias	100	100	93
				0.26246

UltraParalogy (p) Distance

Species	Distance	D
UniProt	Alias	p

GreenPhylDB

Un exemple

Oryzon

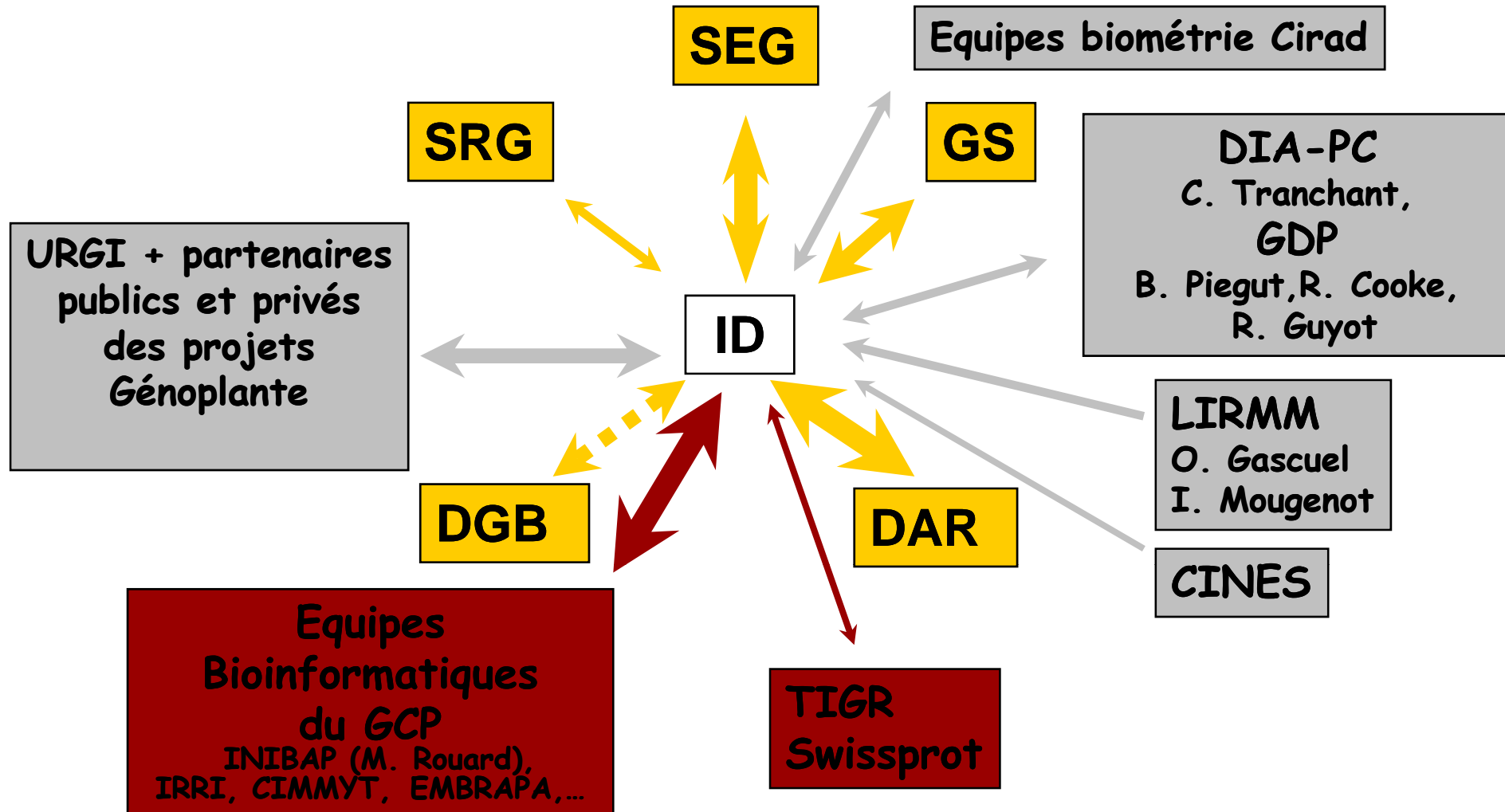
<http://orygenesdb.cirad.fr/login.html>

Ex : Os07g36170.1

Activités d'appui de l'équipe

- Développement d'outils spécifiques en réponse aux demandes des utilisateurs (comité utilisateurs en cours)
- Appui en biostatistiques
 - + Diversité génétique; cartographie génétique; analyse QTL; déséquilibre de liaison; génétique des populations...
- Formation à la bioinformatique des partenaires
 - + en interne (UMR; agents Cirad Guadeloupe)
 - + en externe (Indonésie, Madagascar)

Principales collaborations et partenariats de l'équipe Bioinformatique



L'équipe ID : Intégration des Données

Plateforme matérielle

- **Les serveurs = le nerf de la guerre**
 - 1 serveur de calcul
 - 2 serveurs pour la plateforme logicielle
 - 1 serveur web internet
 - 1 serveur de stockage
 - Administration des serveurs bioinformatiques par l'équipe
 - Appui ponctuel pour l'accès web et la maintenance des serveurs par la DSI Cirad
 - Recours au CINES pour calculs coûteux en temps (sur projets)
 - Connexion avec l'URGI : en cours de discussion

L'équipe ID : Intégration des Données

6 permanents

Xavier ARGOUT

Gaëtan DROC

Chantal HAMELIN

Pierre LARMANDE

Manuel RUIZ

Stéphanie SIDIBE-BOCS

CDD, étudiants : nombre fluctuant