

## Caractérisation d'interactions protéine-protéine chez *Arabidopsis thaliana*

<http://www.evry.inra.fr/public/projects/bioinfo/ACI.html>

### OBJECTIFS

Proposer un nombre significatif d'interactions protéine-protéine expérimentalement validées permettant de mieux comprendre la dynamique des génomes végétaux et la co-évolution des gènes fonctionnellement liés.

**Approches utilisées :**

- Recherche automatique de gènes composites et leurs composants.
- Expertise bioinformatique des situations trouvées.
- Vérification expérimentale des interactions proposées.

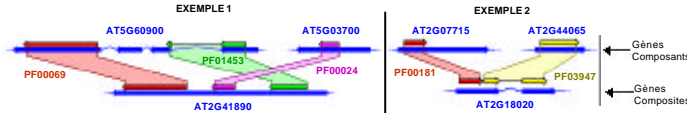
Cette combinaison d'approches bioinformatiques et moléculaires permettra d'aborder efficacement l'interactome d'une plante modèle sans *a priori* sur la fonction des gènes.

### METHODES

#### RECHERCHE DE FUSIONS DE GENES

Recherche de situations composite/composants chez *Arabidopsis thaliana*, le composite étant un gène dont le produit de l'expression contient des domaines protéiques également codés par deux gènes distincts (gènes composants).

Recherche basée sur les domaines PFAM annotés dans FLAGdb<sup>++</sup>.



> **514 situations identifiées** (de 2 à 5 domaines PFAM impliqués) représentant des gènes candidats dont les produits sont susceptibles d'interagir directement ou de collaborer dans une même fonction biologique.

**Outils utilisés :** HMMer, PFAM, FLAGdb<sup>++</sup>, scripts PERL.

#### EXPERTISE DES ANNOTATIONS

- Composites mal annotés. Ex:
- Composites non fonctionnels (pseudogènes, éléments transposables).
- Domaines PFAM douteux (petite taille, score faible).
- Sont également mis de côté, les situations pour lesquelles le nombre de combinaisons à considérer est incompatible avec une expertise manuelle et une validation expérimentale.

Ex : 94 gènes composites  
 582 gènes composants PF00646  
 20 gènes composants PF01344 } 11640 couples possibles

> **205 groupes analysables**

**Outils utilisés :** FLAGdb<sup>++</sup> (<http://urgv.evry.inra.fr/FLAGdb>), scripts PERL.

#### RECHERCHE D'INFORMATIONS SUR LES GENES COMPOSANTS

On recherche des informations pouvant aller dans le sens d'une interaction physique entre ces gènes afin d'estimer la qualité de la prédiction des interactions et de classer les groupes sélectionnés.

- Co-transcription : Nombre d'EST et de cDNA, d'étiquettes MPSS et de données transcriptome (62 expériences différentes utilisant les puces CATMA).
- Co-localisation : Prédiction de l'adressage des protéines composantes dans un même compartiment cellulaire (réticulum, chloroplaste, mitochondrie ou noyau).
- Appartenance à une même voie métabolique ou de signalisation : exploitation de l'annotation fonctionnelle et de la Gene Ontology (Molecular Function, Biological Process, Cellular Component).
- Interactions observées entre domaines PFAM au sein des entrées de la banque de structures 3D PDB.
- Présence d'une situation composite/composants similaire dans le génome complet du riz.

**Outils utilisés :** FLAGdb<sup>++</sup>, PFAM, MPSS database, CATdb, iPFAM.

### LABORATOIRE IMPLIQUE

URGV (Unité de Recherche en Génétique Végétale)

UMR INRA 1165 – CNRS 8114 – UEVE

2, rue Gaston Crémieux, CP5708, 91057 Evry cedex

<http://www.evry.inra.fr> ([aubourg@evry.inra.fr](mailto:aubourg@evry.inra.fr))

N. Bailly, C. Pommier, Y. De Oliveira, M.-L. Martin-Magniette, V. Brunaud, F. Samson, A. Avon, I. Small, A. Lechamy, C. Lurin & S. Aubourg



### MOTS CLES

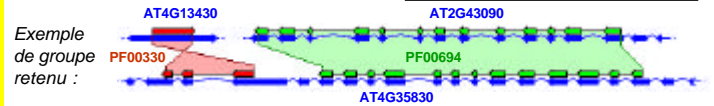
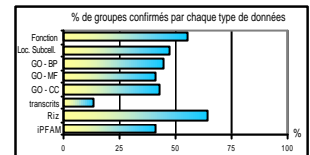
Interactions, Rosetta Stone, motifs PFAM, évolution, relations structure-fonction, co-transcription

### RESULTATS BIOINFORMATIQUES

> Réalisation d'une **base de données** relationnelle (Oracle 8i) incluant toutes les situations, leurs expertises et les informations fonctionnelles associées.

> **68 groupes** où au moins un des critères considérés va dans le sens d'une interaction protéine-protéine.

**Les gènes composants de ces groupes sont donc des gènes dont les produits sont susceptibles d'interagir ensemble.**

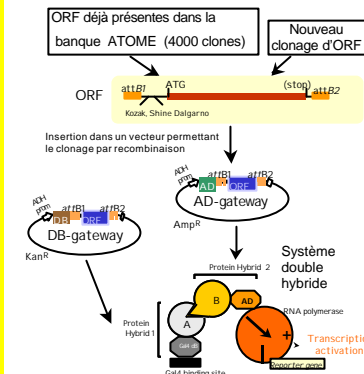


Composants	PFAM	Fonction	Localisation Subcellulaire	Gene Ontology			Nb d'EST et cDNA	Présence chez le riz
				Biological Process	Molecular Function	Cellular Component		
AT4G13430	PF00330	Aconitase family protein	Plastid	Metabolism	Lyase activity	3-isopropylmalate dehydratase complex	94	OS02G03260
AT2G43090	PF00694	Aconitase C-term domain containing protein	Plastid	Metabolism	Hydro-lyase activity	3-isopropylmalate dehydratase complex	45	OS02G43830

### VERIFICATION EXPERIMENTALE

Ne sont testés que les groupes pour lesquels le nombre de couples possibles (interactions) est inférieur à 6.

> On teste **86 couples de composants** (représentant 51 des 68 groupes retenus).



Exploitation de la collection d'ORF d'*Arabidopsis thaliana* 'ATOME' (Projet Génoplante): Pour l'ensemble des interactions proposées par l'approche bioinformatique, les partenaires potentiels sont recherchés dans cette collection. De nouveaux clonages d'ORF absents des collections sont également réalisés. Actuellement, des études d'interactions binaires sont réalisées en testant les protéines 2 à 2 par des approches double-hybrides chez la levure.

Les interactions positives seront vérifiées définitivement en utilisant un nouveau système permettant la détection d'interactions *in planta*. Il s'agit de reconstituer *in vivo* des interactions entre deux protéines de fusion portant d'une part les deux protéines X et Y pour lesquelles une interaction est recherchée et d'autre part deux demi enzyme DHFR (dihydrofolate reductase) qui une fois rapproché (grâce à l'interaction entre X et Y) reforment une activité détectable par fluorescence.

### PERSPECTIVES

- Estimer l'importance des paramètres informatiques pris en compte grâce aux résultats expérimentaux tout en fournissant des **données d'interactions validées**.
- Automatiser les différentes étapes pour permettre un débit plus important de la procédure et son application à d'autres génomes eucaryotes.
- Publication de la méthode et des résultats.