SNPer, un outil web maison de visualisation des SNPs et de leurs effets le long du génome.

Journées PEPI IBIS – 6-7 juin 2019

Frédérique Bitton

06/062019

_01

L’unité GAFL et bioinformatique

GAFL

- **Accompagnement PGD et données FAIR.**
- **Accompagnement bioinformatique analyse de données**
- **Bonnes pratiques: reproductibilité, traçabilité. Utilisation de workflow manager (Snakemake).**
- **Portabilité des outils et pipelines (Containers)**

DADI

Gestion génotypages, phénotypages, NGS

I2B

Joël CHADOEUF,
Jacques LAGNEL,
Emmanuel LE-CALONNEC

ReDD

Gestion de données NGS et développement d'une base "résistance"

CRB

- Accompagnement SI de gestion des ressources CRB
- Système embarqué terrain

_02
L’application

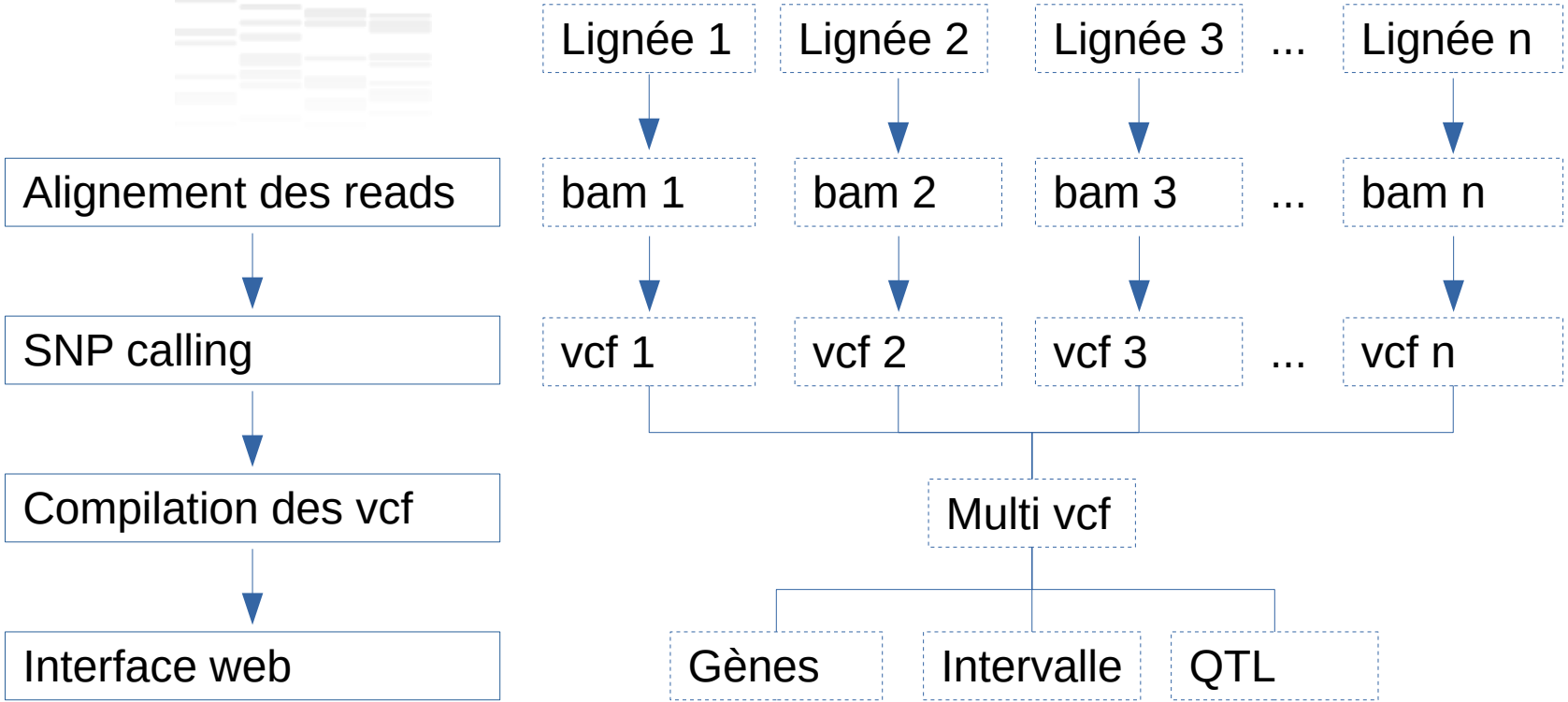
L’équipe DADI

- Reséquençage de 8 lignées parentales utilisées au labo,
- Mapping, SNP calling … et après ?
- 3 questions distinctes :
  - Quels sont les SNP inclus dans mes gènes préférés ?
  - Quels sont les SNP dans tous les gènes d’un intervalle ?
  - Quels sont les SNPs qui vont dans le même sens que mon QTL (recherche de gènes candidats) ?
- Et quels sont leurs effets sur les gènes ?

Le projet SNPer

- Jusqu'en 2015 : j'étais biologiste moléculaire + bioanalyse
- 2015-2016 : Master bioinformatique
- UE Python : Projet à coder et présenter :
  - Python 3
  - cgi pour l'interface web (pas de framework)
  - Devait utiliser une liste de fonctionnalités et structures vues en cours (Dictionnaires, lecture/écriture de fichiers ...)
-

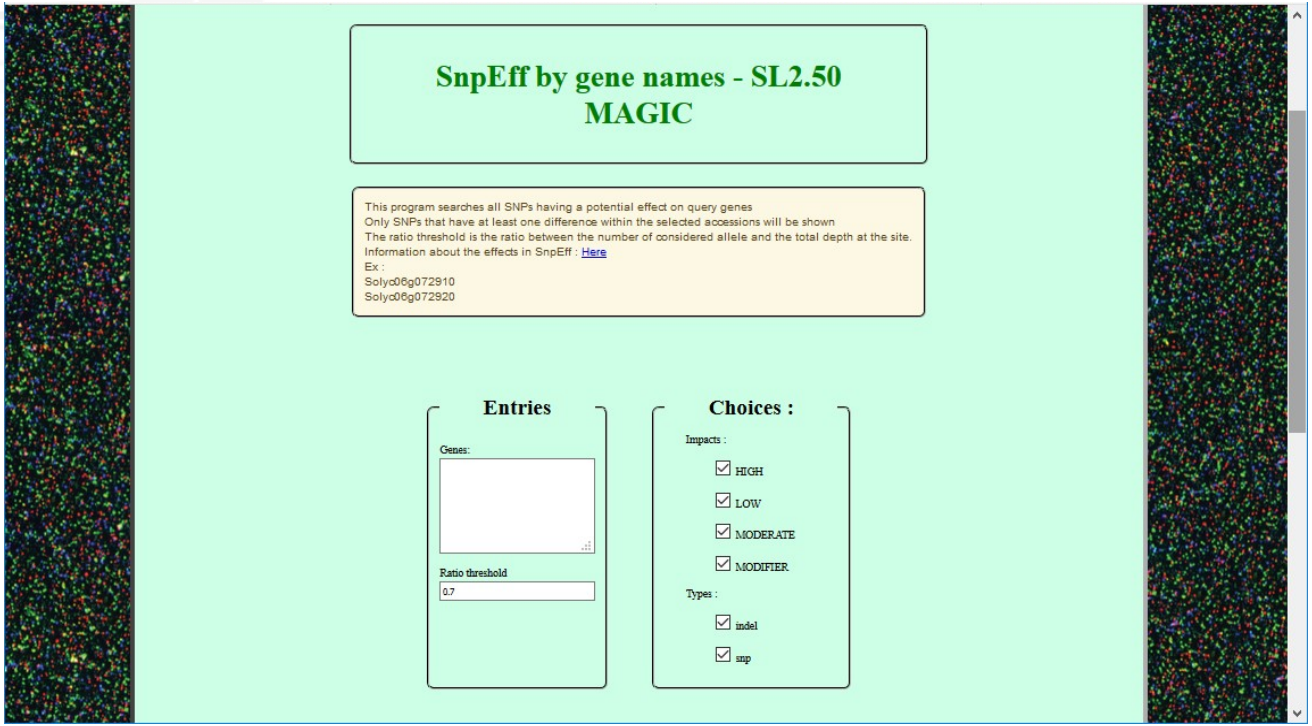
Workflow


Lignée 1
Lignée 2
Lignée 3
...
Lignée n
Alignement des reads
bam 1
bam 2
bam 3
...
bam n
SNP calling
vcf 1
vcf 2
vcf 3
...
vcf n
Compilation des vcf
Multi vcf
Interface web
Gènes
Intervalle
QTL

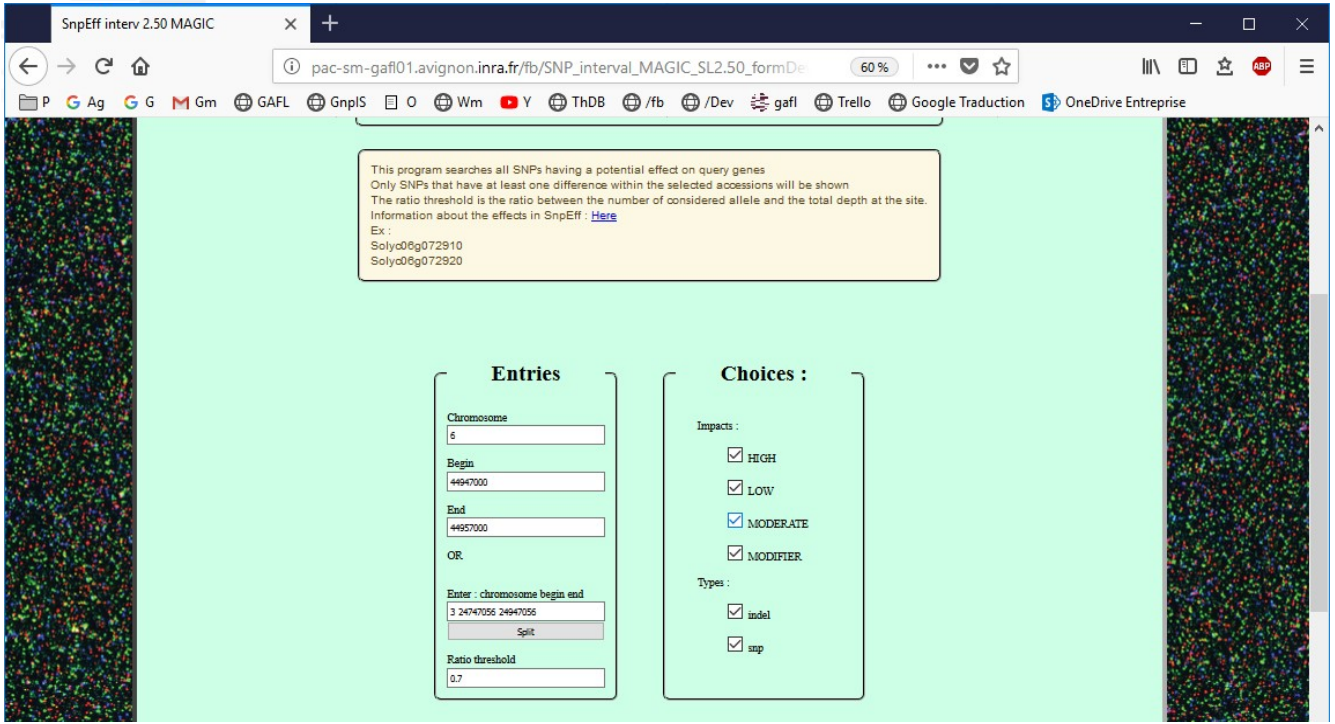
Accueil


SNPer sur gafl01
News
Merci à Hussein pour l'idée de nom !
Genes
Location of genes on different genome versions with Microtom
Displays the location of the genes on all the versions of the genome : ITAG 2.3, 2.5, 3.0, 3.1 and microtom, + functions and GO when known
NEW : Location of SNPs on the cDNA and protein
Displays the location of the SNP in the genes when in the coding sequence
MAGIC - Microtom
SnpEff on several genes - Microtom
Asks for gene IDs and displays information on SNP effect on Microtom
SnpEff by interval on a chromosome - Microtom
Asks for chromosome number, begin and end of an interval and displays information on SNP effect on Microtom
QTL-analyser : SNPs compatibles with a QTL - Microtom
Asks for chromosome number, begin and end of an interval and information the QTL effect on each MAGIC accession and displays information on SNP effect on Microtom
MAGIC - SL2.50
SnpEff on several genes - SL2.50
Asks for gene IDs and displays information on SNP effect on SL2.50 (ITAG 2.4)
SnpEff by interval on a chromosome - SL2.50
Asks for chromosome number, begin and end of an interval and displays information on SNP effect on SL2.50 (ITAG 2.4)
QTL-analyser SL2.50 : SNPs compatibles with a QTL - MAGIC SL2.50 V2 : NEW : with "almost good genes"
Asks for chromosome number, begin and end of an interval and information the QTL effect on each MAGIC accession and displays information on SNP effect on SL2.50 (ITAG 2.4)
SnpEff by interval on a chromosome - SL2.50 + expression + annotation
Asks for chromosome number, begin and end of an interval and displays information on SNP effect on SL2.50 (ITAG 2.4)
SnpEff by genes - SL2.50 + expression + annotation
Asks for gene IDs and displays information on SNP effect on SL2.50 (ITAG 2.4)

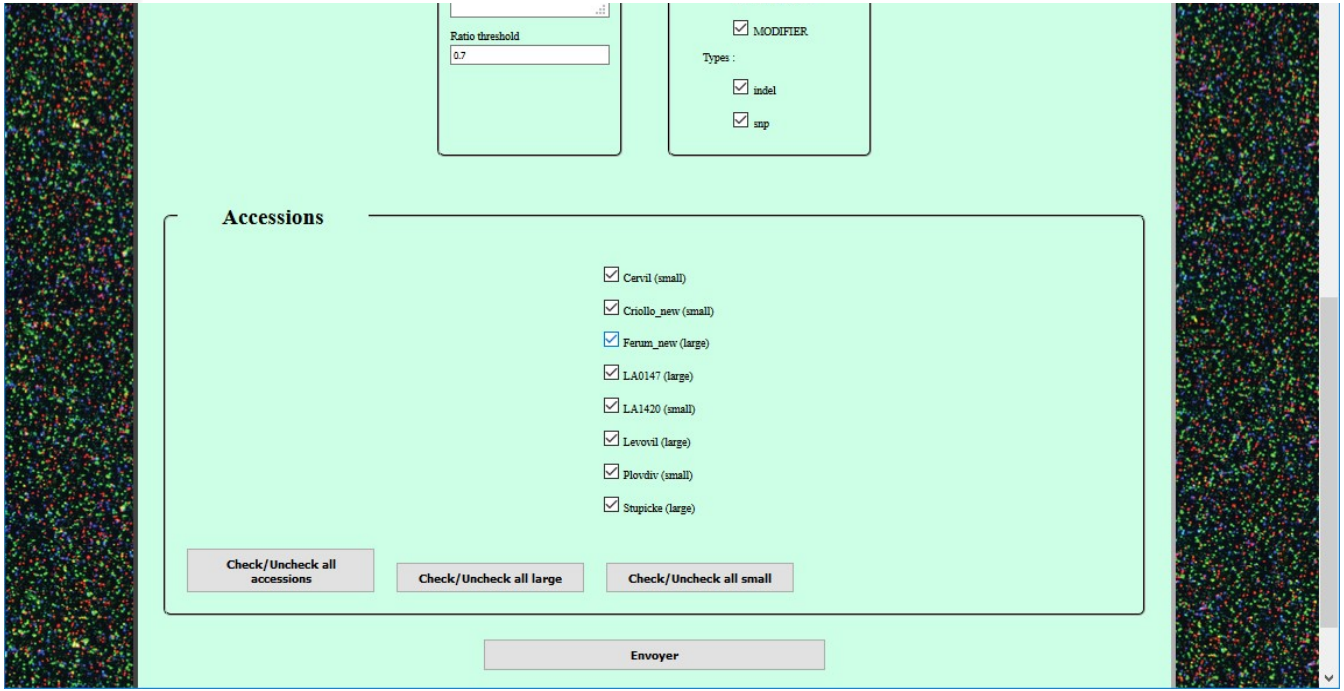
Formulaire : liste de gènes



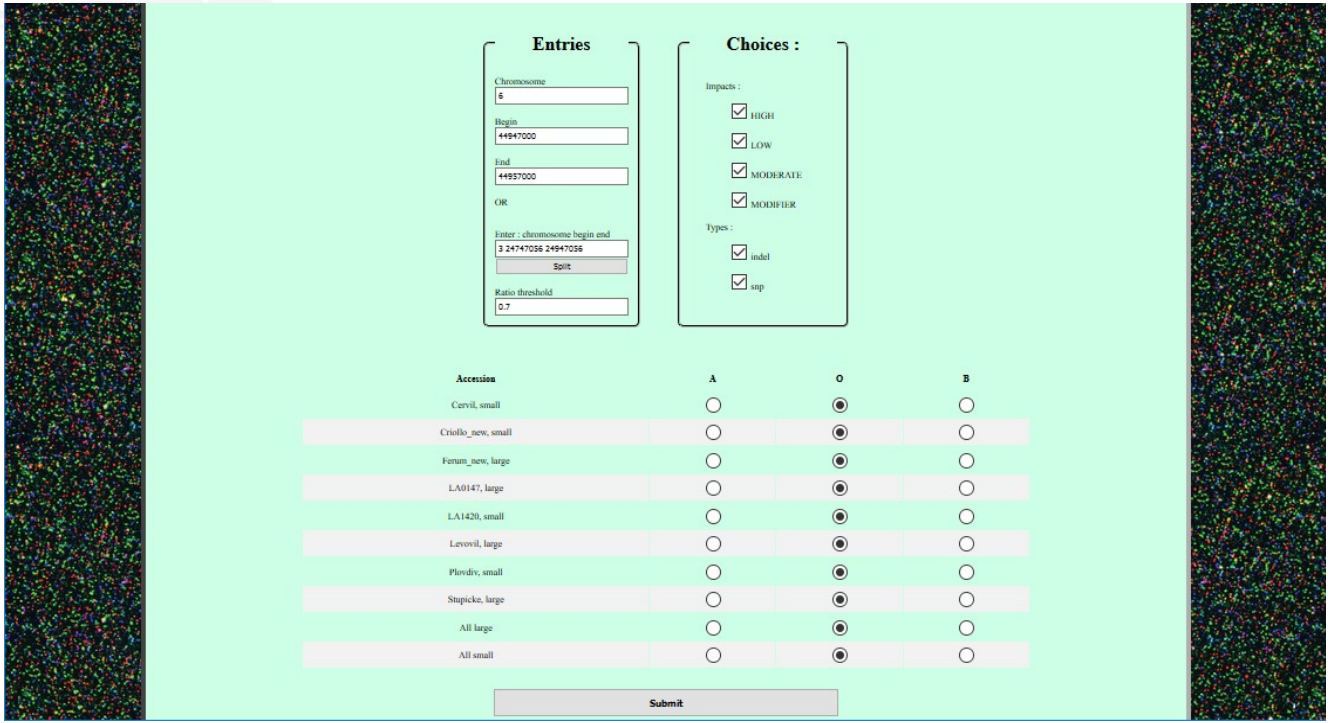
Formulaire : intervalle sur un chromosome



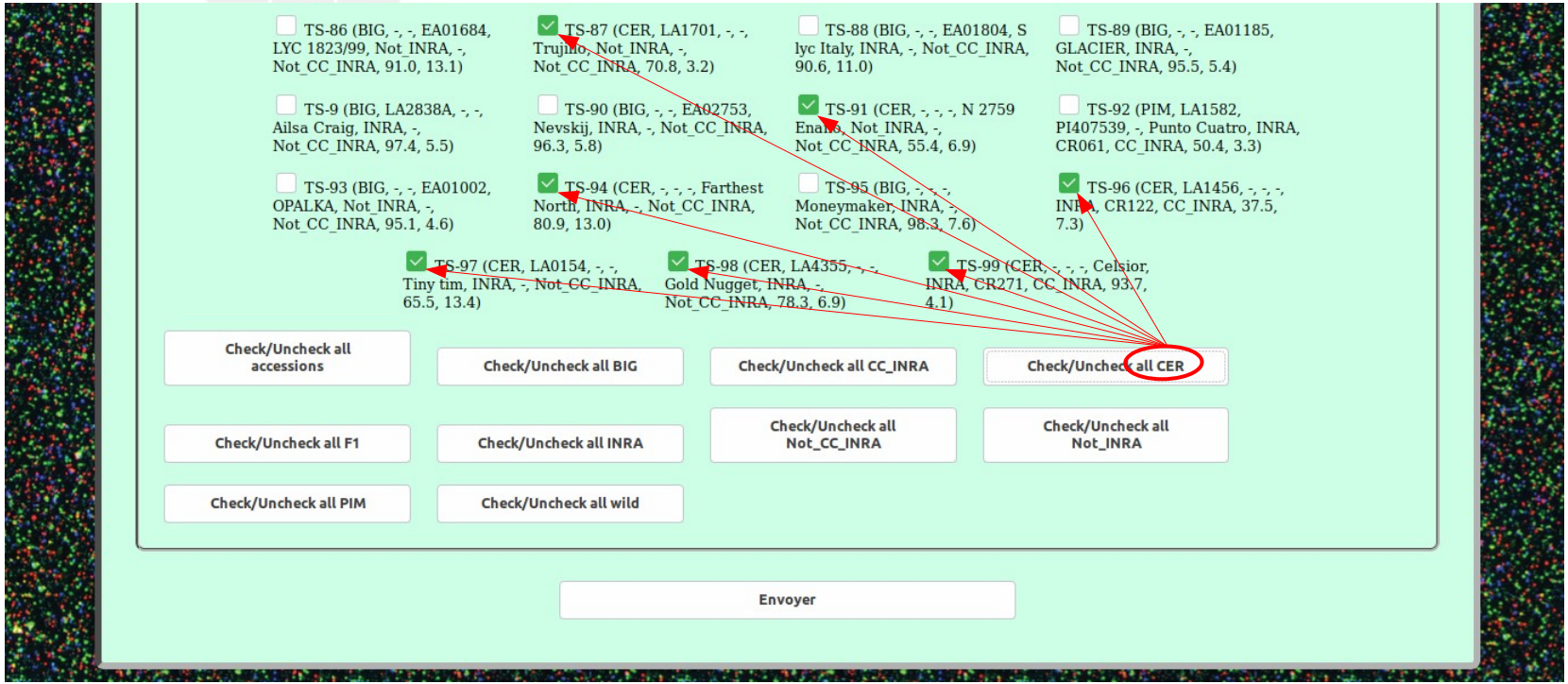
Formulaire : lignées



Formulaire : QTL analyser



Formulaire : grand nombre de lignées



Résultats : gènes / intervalles

Results SNP genes MAGIC SL2.50

Exportation fichier CSV

Chromosome	Location	Gene	Function	Ref_nt	Typ_Mut	Effect	Impact	Cervil	Criollo_new	Ferum_new	LA0147	LA1420	Levovil	Plovdiv	Stupicl
SL2.50ch06	44943005	Solyc06g072910.2	Aluminum-activated malate transporter-like	A	snp	upstream_gene_variant	MODIFIER	A (10/10)	A (6/6)	T (5/5)	A (17/17)	T (11/11)	A (15/15)	A (14/14)	A (9/9)
SL2.50ch06	44943008	Solyc06g072910.2	Aluminum-activated malate transporter-like	C	snp	upstream_gene_variant	MODIFIER	C (10/10)	C (6/6)	T (5/5)	C (16/16)	C (10/10)	C (15/15)	C (13/13)	C (9/9)
SL2.50ch06	44943247	Solyc06g072910.2	Aluminum-activated malate transporter-like	T	snp	upstream_gene_variant	MODIFIER	T (9/9)	T (9/9)	C (5/5)	T (13/13)	C (9/9)	T (19/19)	T (11/11)	T (9/9)
SL2.50ch06	44943282	Solyc06g072910.2	Aluminum-activated malate transporter-like	G	snp	upstream_gene_variant	MODIFIER	G (5/5)	G (5/5)	A (6/6)	G (14/14)	G (9/9)	G (17/17)	G (11/10)	G (8/8)
SL2.50ch06	44943659	Solyc06g072910.2	Aluminum-activated malate transporter-like	G	snp	upstream_gene_variant	MODIFIER	G (8/8)	G (4/4)	A (5/5)	G (4/4)	G (17/17)	G (21/21)	G (8/8)	G (8/8)
SL2.50ch06	44943684	Solyc06g072910.2	Aluminum-activated malate transporter-like	T	indel	upstream_gene_variant	MODIFIER	T (9/9)	X (x/3)	T (6/6)	T (7/7)	+A (13/11)	T (22/22)	T (8/8)	T (7/7)
SL2.50ch06	44943693	Solyc06g072910.2	Aluminum-activated malate transporter-	T	snp	upstream_gene_variant	MODIFIER	T (10/10)	X (x/3)	T (6/6)	T (8/8)	C (9/9)	T (19/19)	T (7/7)	T (7/7)

Résultats : résumés par gène

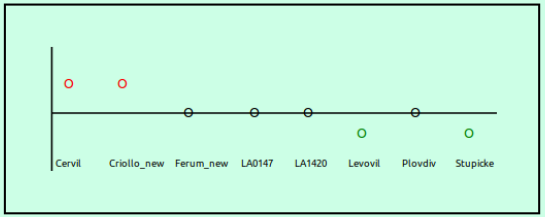
Gene Summary Table : effects

Gene	TOTAL	3_prime_UTR_variant-MODIFIER	5_prime_UTR_premature_start_codon_gain_variant-LOW	5_prime_UTR_variant-MODIFIER	downstream_gene_variant-MODIFIER	intron_variant-MODIFIER	missense_variant-MODERATE	splice_region_variant&intron_variant-LOW	splice_region_variant&synonymous_variant-LOW
solyc06g072890	**14**	0	0	0	14	0	0	0	0
solyc06g072900	**14**	0	0	0	14	0	0	0	0
solyc06g072910	**20**	0	1	2	10	0	1	1	0
solyc06g072920	**41**	0	0	0	23	0	1	0	0
solyc06g072930	**31**	2	0	0	0	5	2	0	1
solyc06g072940	**9**	0	0	0	0	0	0	0	0

Gene Summary Table : impacts

Gene	TOTAL	HIGH	LOW	MODERATE	MODIFIER
solyc06g072890	**14**	0	0	0	14
solyc06g072900	**14**	0	0	0	14
solyc06g072910	**20**	0	4	1	15
solyc06g072920	**41**	0	1	1	39
solyc06g072930	**31**	0	1	2	28
solyc06g072940	**9**	0	0	0	9

Résultats : QTL analyser



Number of genes : 8
Number of SNPs : 14 (+ 11)

Good Genes

Chromosome	Localization	gene	Function	Ref	Type_mut	Effect	Impact	Cervil	Criollo_new	Ferum_new	LA0147	LA1420	Levovil	Plovdiv	Stupicke
SL2.50ch03	23805909	Solyc03g058320.1-Solyc03g058330.2	Pentatricopeptide repeat-containing protein;Unknown Protein	C	snp	intergenic_region	MODIFIER	C 11 (C:11)	C 5 (C:5)	X 3 (G:3)	G 13 (G:13)	C 9 (C:9)	G 30 (G:30)	C 8 (C:8)	G 9 (G:9)
SL2.50ch03	24012334	Solyc03g058360.1-Solyc03g058370.1	Pentatricopeptide repeat-containing protein At4g39530;UDP-glucosyltransferase family 1 protein	G	snp	intergenic_region	MODIFIER	G 9 (G:9)	G 9 (G:9)	T 9 (T:9)	T 13 (T:13)	G 8 (G:8)	T 21 (T:21)	G 16 (G:16)	T 9 (T:9)
SL2.50ch03	24025409	Solyc03g058360.1-Solyc03g058370.1	Pentatricopeptide repeat-containing protein At4g39530;UDP-glucosyltransferase family 1 protein	G	snp	intergenic_region	MODIFIER	G 11 (G:11)	G 11 (G:11)	G 9 (G:9)	A 21 (A:21)	G 11 (G:11)	A 37 (A:37)	G 13 (G:13)	A 10 (A:10)
SL2.50ch03	24268646	Solyc03g058400.2-Solyc03g058410.1	Aspartyl protease family protein;Unknown Protein	C	snp	intergenic_region	MODIFIER	C 11 (C:11)	C 10 (C:10)	C 9 (C:9)	A 9 (A:9)	C 14 (C:14)	A 60 (A:60)	C 10 (C:10)	A 7 (A:7)
SL2.50ch03	24319526	Solyc03g058400.2-Solyc03g058410.1	Aspartyl protease family protein;Unknown Protein	G	snp	intergenic_region	MODIFIER	G 20 (G:20)	G 7 (G:7)	A 7 (A:7)	A 19 (A:19)	G 10 (G:10)	A 26 (A:26)	G 10 (G:10)	A 13 (A:13)

Développements en cours

- Harmonisation des différents codes,
- Amélioration de la phase de compilation des fichiers,
- Dépôt sur SourceSup.

Perspectives

- Passage en BdD plutôt que lecture de fichier,
- Migration sous Django.