



THE PALOMA* COMPONENT OF THE MICROSCOPE PLATFORM

Claudine Médigue
CNRS & CEA – Institut de Génomique
LABGeM



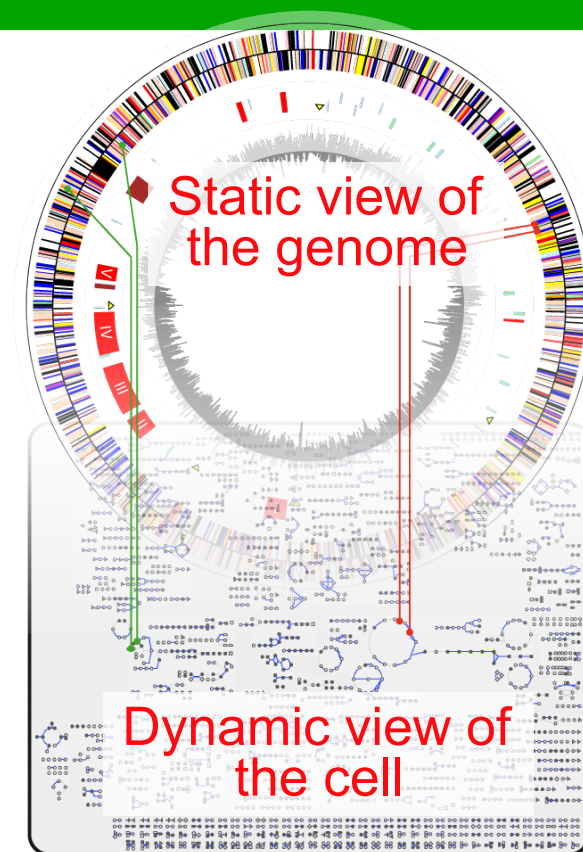
***Polymorphism Analyses in Light Of MAssive
DNA sequencing**



MICROSCOPE -OMICS

An integrated environment for:

- Microbial genome (re)annotation
- Comparative analyses
- Function and biological process predictions
- Expression and evolutionary studies by NGS data integration



Nucleic Acids Research, 2006, Vol. 34, No. 1 53–65
doi:10.1093/nar/gkj406

MaGe: a microbial genome annotation system supported by synteny results

David Vallenet*, Laurent Labarre, Zoé Rouy, Valérie Barbe¹, Stéphanie Bocs, Stéphanie Cruveiller, Aurélie Lajus, G raldine Pascal, Claude Scarpelli¹ and Claudine M digue

Atelier de G n mique Comparative, CNRS-UMR8030 and ¹Genoscope, 2 rue Gaston Cr mieux, 91057 Evry, Cedex, France



Database, Vol. 2009, Article ID bap021, doi:10.1093/database/bap021

Original article

MicroScope: a platform for microbial genome annotation and comparative genomics

D. Vallenet*, S. Engelen, D. Mornico, S. Cruveiller, L. Fleury, A. Lajus, Z. Rouy, D. Roche, G. Salvignol, C. Scarpelli and C. M digue

D636–D647 Nucleic Acids Research, 2013, Vol. 41, Database issue
doi:10.1093/nar/gks1194

Published online 27 November 2012

MicroScope—an integrated microbial resource for the curation and comparative analysis of genomic and metabolic data

David Vallenet^{1,2,3,*}, Eugeni Belda^{1,2,3}, Alexandra Calteau^{1,2,3}, St phanie Cruveiller^{1,2,3}, Stefan Engelen¹, Aur lie Lajus^{1,2,3}, Fran ois Le F vre^{1,2,3}, Cyrille Longin^{1,2,3}, Damien Mornico^{1,2,3}, David Roche^{1,2,3}, Zo  Rouy^{1,2,3}, Gregory Salvignol^{1,2,3}, Claude Scarpelli¹, Adam Alexander Thill Smith^{1,2,3}, Marion Weiman^{1,2,3} and Claudine M digue^{1,2,3,*}

First publication in 2006

=> Synteny maps & curation of gene function

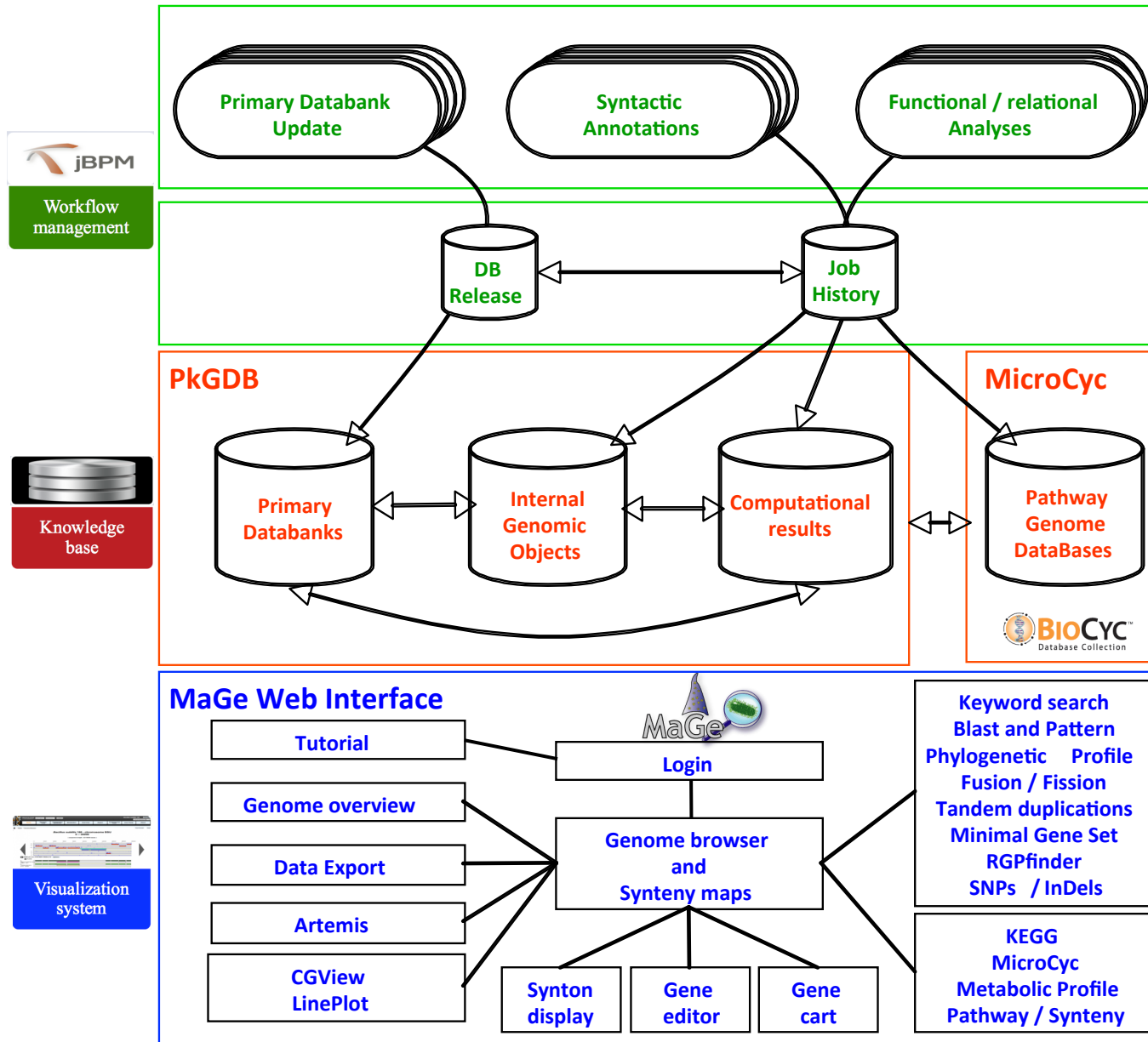
Release 1 in 2009

=> Comparative genomics tools & gene carts

Release 2 in 2013

=> Curation and analysis of metabolic data

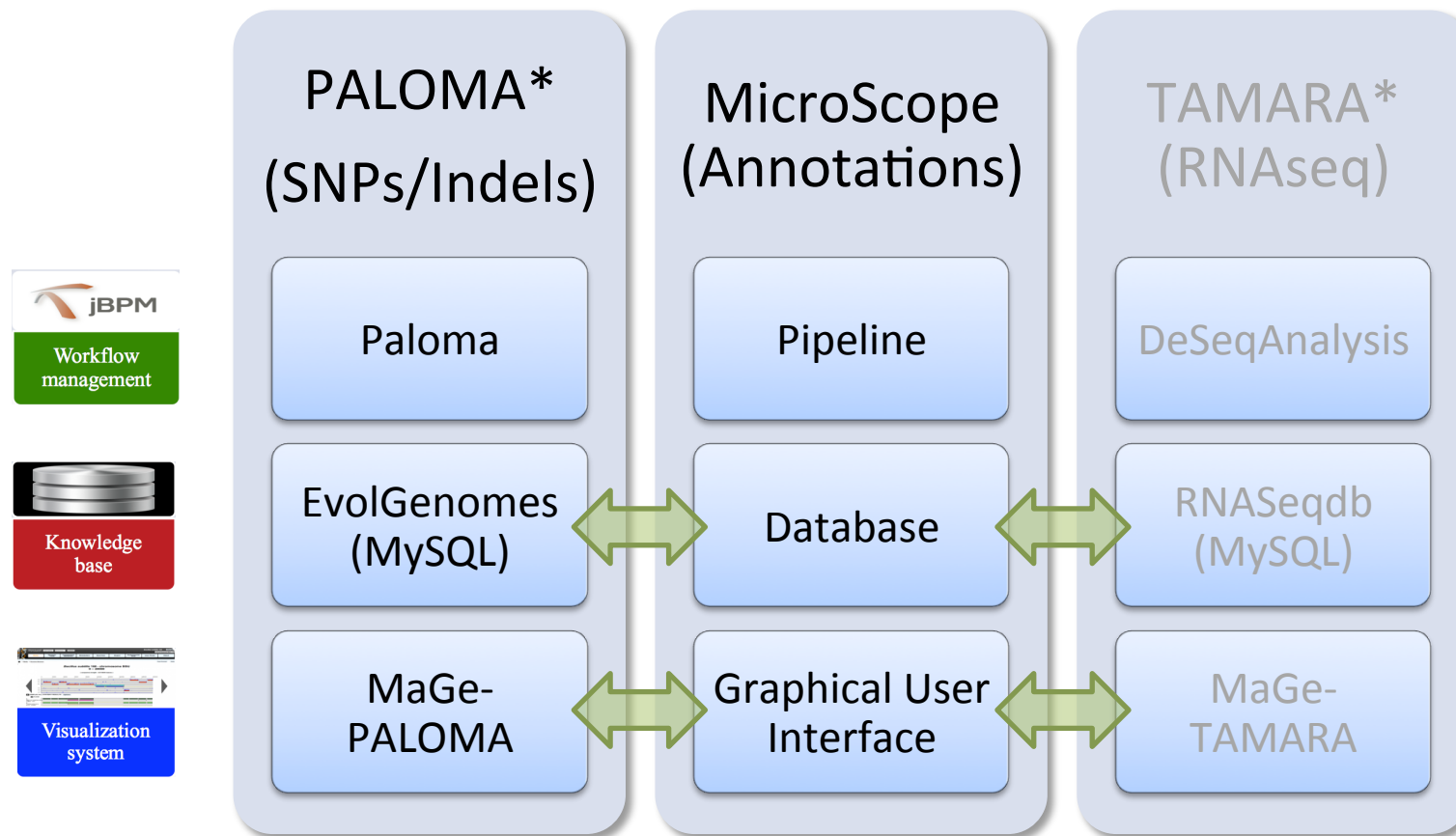
MICROSCOPE: 3 COMPONENTS



- ❖ **> 30 bioinformatics methods**
- ❖ **Continuous integration of genomic data**
- ❖ **Analysis of 3 genome a day** in 2013
- ❖ **Data organization and persistence** in MySQL
- ❖ **One KB instance** for all projects
- ❖ **>3000 genomes** in PkGDB and PGDB in MicroCyc
- ❖ **Knowledge exploration and human expertise**
- ❖ **Analysis tools** for comparative genomics and metabolism

- Microscope Tools To Handle Nextgen Sequencing Data -

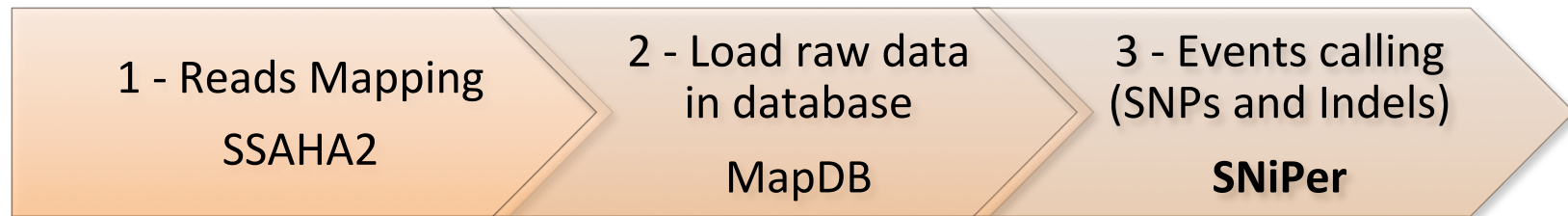
A shared structure for integrated tools...



PALOMA - **P**olymorphism **A**nalyses in **L**ight **O**f **M**assive DNA sequencing

TAMARA - **T**ranscriptome **A**nalyses based on **M**assive sequencing of **R**n**A**s

- The PALOMA Pipeline Detailed -



➤ **SNIper implementation :**

- Trim low quality 3' ends of reads
- Take into account strand bias that occurs during sequencing stage
- Can be rerun in seconds (if mapping info are already stored)
- All in one portable software & database "downloadable"

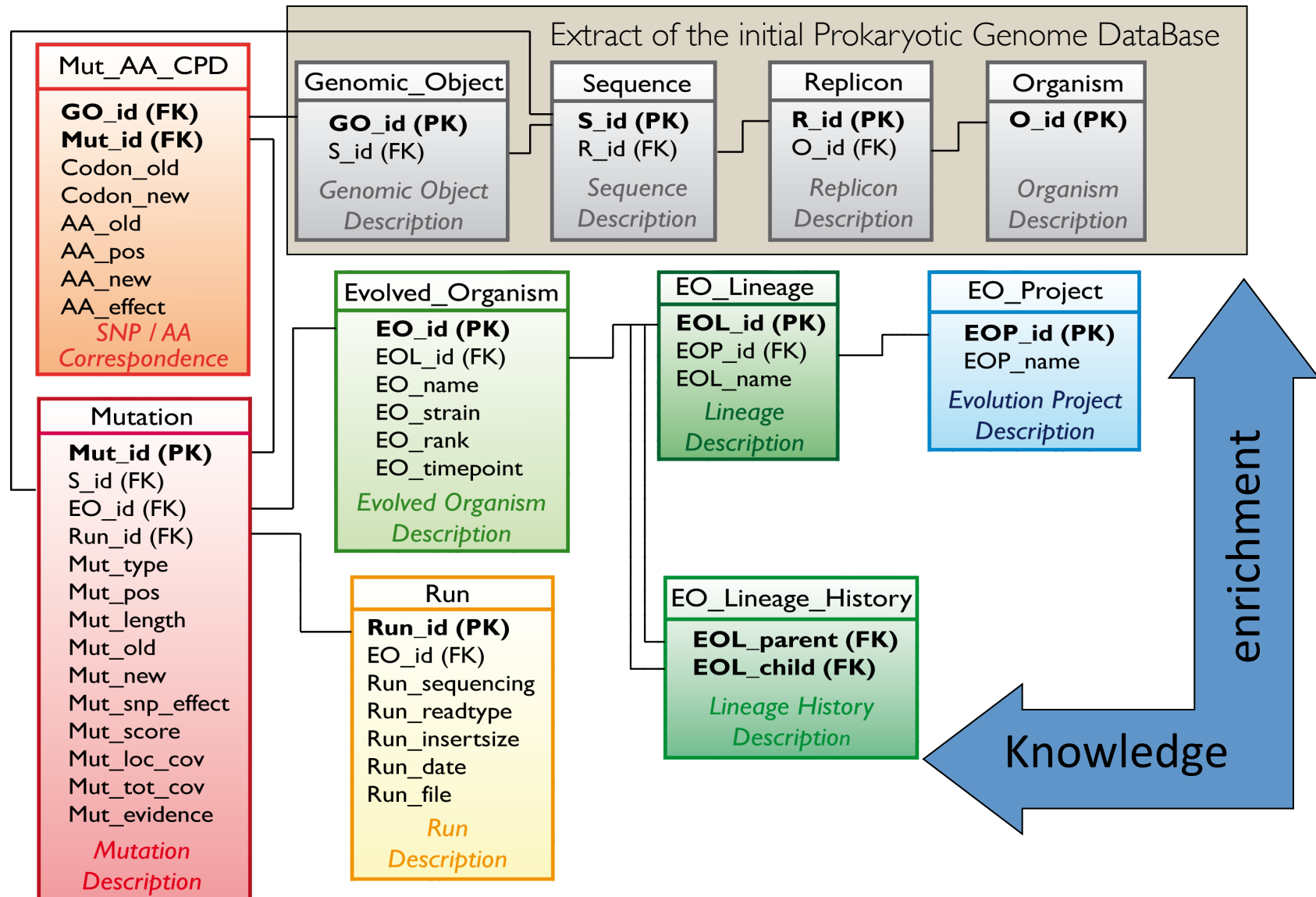
➤ **SNIper efficiency** (Collab. INRA-LIPM C. Masson & D. Capela)

9 evolved clones => about 490 mutations detected

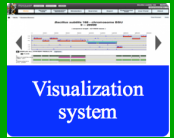
151 have been tested (PCR) => **147 (97%) were validated**

➤ **SNIper results directly linked to genome annotations**

- Evolution Data Storage Into The Relational Database PKGDB -



Exploration of PALOMA results



MicroScope Logged as Claudine MEDIGUE [Logout](#) <https://www.genoscope.cns.fr/agc/microscope>

MaGe Genomic Tools Comparative Genomics Metabolism Search/Export Transcriptomics **Variant Discovery** User Panel About Admin

PALOMA - Evolution

» Variant Discovery » PALOMA - Evolution

PALOMA - Polymorphism Analyses in Light Of MASSive DNA sequencing Evolution Project - SNPs/Indels

- 19 projects
- 477 evolved clones

↓ ● SYMPA

↓ ● EvoGeno

↓ ● Ferme_coli_CLU

↓ ● Ferme_coli_HOB

↓ ● MycoplasmEvolScope

↓ ● LegionellEvolScope

↓ ● MUTA

● **i**

- Lineage Name: 206
- Evolved Organisms Number: 18
- Run Number: 18
- Evolved organism sequence(s):
 - Ralstonia solanacearum GMI1000 : chromosome NC_003295, plasmid NC_003296
 - Cupriavidus taiwanensis LMG19424 : plasmid p2RT

● **i**

- Lineage Name: 349
- Evolved Organisms Number: 5
- Run Number: 5
- Evolved organism sequence(s):
 - Ralstonia solanacearum GMI1000 : chromosome NC_003295, plasmid NC_003296
 - Cupriavidus taiwanensis LMG19424 : plasmid p2RT

- A Dedicated Query Interface : « SNPs / InDels » -



Identify genetic polymorphisms shared by several clones in different lineages

1. Choose an Analysis Type

Analysis: ☒ Comparative ☐ Parallelism ☐ Graphical

Focus on:

- Clones grouped by lineage
- Clones grouped by timepoint
- Lineages

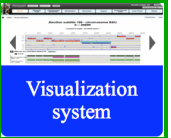
Find a set of mutations present in some organisms and absent from others

Plot the mutations on a circular representation of the reference genome

2. Manage your Analysis Query

Reference sequence:	<div>Cupriavidus taiwanensis LMG19424 chromosome RALTA_A RALTA_Av2 411 Cupriavidus taiwanensis LMG19424 chromosome RALTA_B RALTA_B 280 Cupriavidus taiwanensis LMG19424 plasmid pRALTA_pRALTA 560 Ralstonia solanacearum GMI1000 chromosome RSc NC_003295 55</div>		
Find mutational events:	<div><div>Present in:</div><div>Absent from:</div><div>(Select at least one)</div><div><div>Lineage CBM124</div><div>PoolJ5 Tp 1</div><div>Lineage CBM206 (=CBM124GenR)</div><div>CBM206 Tp -1</div><div>PoolJ1 Tp 1</div><div>PoolJ2 Tp 1</div><div>PoolJ4 Tp 1</div><div>PoolJM1 Tp 1</div><div><input type="checkbox"/> ALL selected clones</div></div><div><div>Lineage CBM124</div><div>PoolJ5 Tp 1</div><div>Lineage CBM206 (=CBM124GenR)</div><div>CBM206 Tp -1</div><div>PoolJ1 Tp 1</div><div>PoolJ2 Tp 1</div><div>PoolJ4 Tp 1</div><div>PoolJM1 Tp 1</div></div></div>		
Score parameters:	<div>Score <input type="text" value="≥"/> <input type="text" value="0.5"/></div> <div>HQ Reads <input type="text" value="≥"/> <input type="text" value="5"/></div> <div>Score <input type="text" value="≥"/> <input type="text" value="0"/></div> <div>HQ Reads <input type="text" value="≥"/> <input type="text" value="0"/></div>		
With these restrictions:	Events Type :	Where ?	Technology :
	<input type="text" value="SNPs/InDels"/>	<input type="text" value="Everywhere"/>	<input type="text" value="Solexa/454"/>
	<input type="checkbox"/> Display only the downstream genes of intergenic events		
	Genome Position from <input type="text" value="1"/> to <input type="text" value=""/> bp		Mut Length ≥ <input type="text" value="1"/> nt
Displayed characteristics:	<div>Nucleotide change + <div>Mutation Type Nuc. Change Effect Codon Change</div></div>		

- A Dedicated Query Interface : « SNPs / InDels » -



↓ **Ralstonia solanacearum GMI1000 chromosome RSc NC_003295 55** ^[35] [Export to Gene Cart](#)

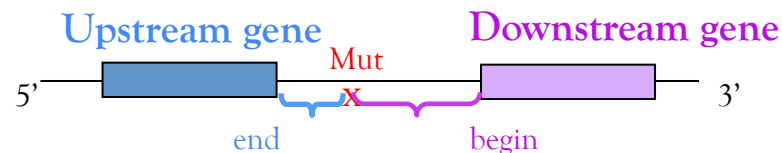
Showing 1 to 35 of 35 results Show **All** Results

x	Abs Position	Rel Position	GO Label	GO Description	Distance to the flanking GO	CBM124		CBM206 (=CBM124GenR)	
						PoolJ5 Tp 1	PoolJ4 Tp 1	PoolJ1 Tp 1	PoolJ2 Tp 1
<input checked="" type="checkbox"/>	24152	1575	RALSOc_0022	putative response regulator, LuxR family (modular protein) 23990 25726 -1		-	G/A	-	-
<input checked="" type="checkbox"/>	116046	1890	RSc0102	putative calcium binding hemolysin protein 114282 117935 -3		A/G	A/G	-	-
<input checked="" type="checkbox"/>	116049	1887	RSc0102	putative calcium binding hemolysin protein 114282 117935 -3		-	G/A	-	-
<input checked="" type="checkbox"/>	200268	915	RSc0177	glmU bifunctional: N-acetyl glucosamine-1-phosphate uridylyltransferase (N-terminal); glucosamine-1-phosphate acetyl transferase (C-terminal) 199354 200721 +1		-	C/T	-	-
<input checked="" type="checkbox"/>	558596		RSc0520 RSc0521	gloA glyoxalase I, nickel isomerase 557915 558322 -1 putative metal-dependent hydrolase 559056 559922 -3	274 460	-/T	-/T	-	-
<input checked="" type="checkbox"/>	558598		RSc0520 RSc0521	gloA glyoxalase I, nickel isomerase 557915 558322 -1 putative metal-dependent hydrolase 559056 559922 -3	276 458	A/G	A/G	-	-
<input checked="" type="checkbox"/>	717490	467	RSc0672	putative composite protein : Response regulator receiver of two component system system (sensor histidine kinase and response regulator)/transcription regulator, similiraty with pilL protein 717024 723170 +3		-	C/T	-	-

Genomic Objects

Evolved clones are grouped by lineage

- Mutations are replaced in a genomic and functional context :



- The dynamics of genomic changes can easily be drawn during the studied evolutionary time

- Some Variants Discovery Projects -

- **Experimental Evolution**

- EvoGeno: *Escherichia coli* B REL606 (12 lineages, 170 clones)
- **SYMPA**: Chimaeric *Ralstonia Solanacearum* GMI1000 including Nod plasmid from *Cupriavidus taiwanensis* LMG1924 (3 lineages, 80 clones)
- **SHAPE**: Continuity of SYMPA project (24 clones + 27 pools of 20 clones)
- PutidaEvolScope: *Pseudomonas putida* KT2440 (WT + 3 engineered mutants)

- **Natural Evolution**

- MyctuEvolscope: Comparative genomics including various mycobacteria (5 clones)
- FrancisellaScope: Evolution of *Francisella tularensis* clones (30 clones)

- **Strains Profiling**

- *Mycoplasma* sp. (10 clones)
- *Legionella* sp. (5 clones)

- **Resequencing Projects**

- *Pseudomonas putida* KT2440