AGENCE NATIONALE DE LA RECHERCHE

ANR

Presentation of the funded projects in 2010 for the « Genomics, Plant biotechnology » Programme

AFAQ ISO 9001

« Genomics, plant biotechnology » programme

YEAR 2010

Project title	**BRASSINAM** - Development of a Nested Association Mapping population for complex traits dissection in Brassica napus
Abstract	Rapeseed is the oilseed crop most widely cultivated in Europe.The cultivated surface kept growing since 2003 as a consequence of the development of non-food uses of rapeseed oil while the demand for human feeding remained high. Recently the demand for arable products and vegetable oil in particular, has increased dramatically. This is mainly due to the growing world population and the related demand for food and to the increasing demand for bioenergy. It is also becoming clear that to satisfy demand, productivity of arable crops, and more precisely oil crop, will need to be improved. At the same time inputs need to be reduced for both economic and ecological reasons. Genetic progress is of major interest in the context of low input-systems/sustainable agriculture. During the last years a technological breakthrough was observed in the domain of plant genetic improvement with the development of high throughput sequencing and genotyping technologies. Very promising results were also recently obtained using genetic association approaches to analyse traits. Association of new genotyping technologies and new methodologies for genetic data analysis, allow successful analysis of complex traits. Recently, the field of plant association genetics pioneered the use of a new type of association populations, designed to incorporate advantages of both linkage based and linkage disequilibrium based quantitative trait dissection approaches, in association studies. This population named NAM for Nested Association Mapping, was successfully developed in maize and is being developed in wheat. This kind of population is of great interest to validate quickly and at a large scale, candidate genes and to deeply characterize QTL in plants The aim of the OSR_NAM project is to develop a NAM population in oilseed rape which will give to scientific and breeders involved in oilseed rape improvement, a major tool

	for high resolution genetic mapping. This resource will be developed by two laboratories leader in genetics, genomics and genotyping. The project will gather public and private research. This secures the use of the resource at once for basic and applied research, to put in the market new varieties with improved agronomical traits.
Partners	Biogemma UMR_APBV
Coordinator	BRUNO GREZES-BESSET - Biogemma bruno.grezes-besset@biogemma.com
ANR funding	477 k€
Starting date and duration	January 2011 - 48 months
Reference	ANR-10-GENM-001
Cluster label	AGRIMIP INNOVATION

Project title	**CARTOSEQ - Large-scale identification of the causative genetic variants influencing traits of economic interest in French dairy breeds**

Abstract

In the past years, ANR funded several cattle-related research projects, including the CartoFine project.
The aim of this project was the fine-mapping of QTLs for production traits in the three main dairy breeds used in France, Holstein, Normande and Montbéliarde. Each of the 3,200 artificial insemination bulls selected for this project was characterized by the performances of 100 daughters for 25 production and adaptation traits (e.g. milk production and composition, fertility, mastitis resistance). All the bulls were genotyped for 54,001 SNPs using the Illumina Bovine SNP50K BeadChip and whole-genome association studies combining linkage disequilibrium and linkage analysis were done. Numerous QTLs have been detected and fine-mapped. A total of 305 significant QTL regions were found with in most cases, small confidence intervals. Interestingly, a large proportion of QTLs is not shared across breeds.
Genomic selection methods are already allowing better genetic improvement; however, as most of the SNPs present on the whole-genome cattle SNP genotyping microarray commonly used, are not in genes and also because of the extent of linkage disequilibrium, SNPs associated with economically important traits, will most likely, not be involved directly in these traits. The identification of the causative genetic variants involved in the phenotypes of interest, remain a difficult task. It is therefore, crucial to develop strategies to pinpoint more rapidly causative genetic variants underlying phenotypes of interest.
Until now, the identification of these causative genetic variants, also known as quantitative trait nucleotides (QTNs) involved the mapping of QTLs, the discovery of novel genetic markers in the QTL regions, the fine-mapping of QTLs and then the sequencing of candidate genes. This iterative process until recently was very time-consuming, but thanks to the availability of a large number of SNPs and of relatively low-cost whole-genome genotyping methodologies, the fine-mapping of QTL regions has now been expedited.
The advent of novel sequencing technologies offer now new opportunity for the identification of QTNs, with the ability to

re-sequence partially or completely mammalian genomes, in a relatively cost-effective manner. For example, Eck et al. (2009) have recently produced a whole-genome sequence at low coverage of a single Fleckvieh bull. They generated 24 Gb of sequence, using 36-bases paired-end reads, resulting in an average 7.4-fold sequence depth. This coverage was sufficient to identify 2.44 million SNPs, 82% of which were previously unknown, and 115,000 small insertion/deletions. A comparison with the genotypes of the same animal, generated using the Illumina SNP50K BeadChip, revealed a detection rate of 74% and 30% for homozygous and heterozygous SNPs, respectively.

The aim of the proposed research project is to develop a next-generation sequencing-based large-scale approach to identify QTNs influencing milk production in the three main French dairy breeds, Holstein, Normande and Montbéliarde. The proposed project will be the follow-up of the previously ANR-funded CartoFine project.

To our knowledge, this project will constitute the first large-scale whole-genome association study using re-sequencing in cattle. The methodology and tools, resulting from this project, could be used in future association studies to pinpoint more rapidly causative genetic variants and could be easily adapted to the identification of QTNs for other traits or other organisms of interest, provided their genome sequence are available. The proposed project will have therefore be In the past years, ANR funded several cattle-related research projects, including the CartoFine project.

The aim of this project was the fine-mapping of QTLs for production traits in the three main dairy breeds used in France, Holstein, Normande and Montbéliarde. Each of the 3,200 artificial insemination bulls selected for this project was characterized by the performances of 100 daughters for 25 production and adaptation traits (e.g. milk production and composition, fertility, mastitis resistance). All the bulls were genotyped for 54,001 SNPs using the Illumina Bovine SNP50K BeadChip and whole-genome association studies combining linkage disequilibrium and linkage analysis were done. Numerous QTLs have been detected and fine-mapped. A total of 305 significant QTL regions were found with in most cases, small confidence intervals. Interestingly, a large proportion of QTLs is not shared across breeds.

Genomic selection methods are already allowing better genetic improvement; however, as most of the SNPs present on the whole-genome cattle SNP genotyping microarray commonly used, are not in genes and also because of the extent of linkage disequilibrium, SNPs associated with economically important traits, will most likely, not be involved directly in these traits. The identification of the causative genetic variants involved in the phenotypes of interest,

remain a difficult task. It is therefore, crucial to develop strategies to pinpoint more rapidly causative genetic variants underlying phenotypes of interest.
Until now, the identification of these causative genetic variants, also known as quantitative trait nucleotides (QTNs) involved the mapping of QTLs, the discovery of novel genetic markers in the QTL regions, the fine-mapping of QTLs and then the sequencing of candidate genes. This iterative process until recently was very time-consuming, but thanks to the availability of a large number of SNPs and of relatively low-cost whole-genome genotyping methodologies, the fine-mapping of QTL regions has now been expedited.
The advent of novel sequencing technologies offer now new opportunity for the identification of QTNs, with the ability to re-sequence partially or completely mammalian genomes, in a relatively cost-effective manner. For example, Eck et al. (2009) have recently produced a whole-genome sequence at low coverage of a single Fleckvieh bull. They generated 24 Gb of sequence, using 36-bases paired-end reads, resulting in an average 7.4-fold sequence depth. This coverage was sufficient to identify 2.44 million SNPs, 82% of which were previously unknown, and 115,000 small insertion/deletions. A comparison with the genotypes of the same animal, generated using the Illumina SNP50K BeadChip, revealed a detection rate of 74% and 30% for homozygous and heterozygous SNPs, respectively.
The aim of the proposed research project is to develop a next-generation sequencing-based large-scale approach to identify QTNs influencing milk production in the three main French dairy breeds, Holstein, Normande and Montbéliarde. The proposed project will be the follow-up of the previously ANR-funded CartoFine project.
To our knowledge, this project will constitute the first large-scale whole-genome association study using re-sequencing in cattle. The methodology and tools, resulting from this project, could be used in future association studies to pinpoint more rapidly causative genetic variants and could be easily adapted to the identification of QTNs for other traits or other organisms of interest, provided their genome sequence are available. The proposed project will have therefore benefits not only to the dairy and bovine genome research community but also potentially to other scientists.nefits not only to the dairy and bovine genome research community but also potentially to other scientists.

Partners

G2B
UNCEIA
PlaGe
SIGENAE

Coordinator	Dominique Rocha - G2B dominique.rocha@unilim.fr
ANR funding	412 k€
Starting date and duration	January 2011 - 36 months
Reference	ANR-10-GENM-018
Cluster label	

« Genomics, plant biotechnology » programme

YEAR 2010

Project title	**CHLOROTYPES** - Chloroplast adaptation to abiotic stresses: use of proteomics to reveal molecular phenotypes
Abstract	The goal of the present project is to enlarge our knowledge of the regulation of chloroplast metabolism to adaptation phenomena, i.e. to relate the dynamic of chloroplast biogenesis to environmental changes. To this, we will first decipher the molecular mechanisms underlying the dynamic of the chloroplast proteome and then relate these mechanisms to the plant response to environmental cues. During a previous ANR project (ANR Génoplante "GlycoChloroplast" 2007-2010) associating LPCV and EDyP, we produced important tools to investigate the dynamics of the chloroplast proteome at the scale of a full organelle. One important and still unique tool is the first AMT database dedicated to chloroplasts, the chloroplast AMT database AT-CHLORO. This proteomic database is conceived in such a way that it can be used for further quantitative studies (e.g. comparisons of mutants or impact of variable growth conditions) based on the accurate mass and time tags (AMT) strategy. Another finding from previous Genoplante projects involving both LPCV and EDyP laboratories is the identification of a new regulatory mechanism for protein trafficking between the cytosol and the chloroplast. In this pathway, chloroplast proteins are first sequestered into the cytosol through interaction with a protein partner, and then released to the cytosol by a specific signal before being targeted to the chloroplast. Preliminary data indicate that this regulatory mechanism should control essential metabolic functions of the chloroplast and might be under the control of environmental stimuli. The aim of this project is therefore to i) to identify abiotic stresses that impact on the targeting of these specific chloroplast proteins, ii) to decipher the stress-induced regulatory mechanisms that control their subcellular localization, and iii) to understand the physiological consequences of these regulatory mechanisms.
Partners	CNRS-LPCV INSERM-LEDyP

Coordinator	Norbert ROLLAND - CNRS - LPCV norbert.rolland@cea.fr
ANR funding	371 k€
Starting date and duration	January 2011 - 48 months
Reference	ANR-10-GENM-002
Cluster label	

« Genomics, plant biotechnology » programme

YEAR 2010

Project title	**CNV-MAIZE** - Genome-wide association study between structural variation, agronomic traits variation and heterosis in maize
Abstract	Structural variation (SV) – that encompasses sequence deletions, duplications, insertions, inversions and translocations – has long been considered as rare and meaningless for phenotypic evolution. From recent studies, it is now clear that SV is common to many species including maize, and involves a much greater proportion of the genome than previously thought. Understanding the nature and content of SV in maize will allow to elucidate the structure, evolution and variability of the maize genome. Because maize is a major crop, discovery of particular SVs such as Copy Number Variants (CNVs) or Presence Absence Variants (PAVs) is important regarding their potential contribution to phenotypes, especially as they may account for part of the genetic variation of complex traits and heterosis. The maize germplasm contains contrasted lines that are likely to show large SV. Therefore, to investigate the potential implications of SV in important agronomic traits and heterosis, it is essential to characterize SV for several representative lines. In the CNV-Maize project, we propose to address these issues using an original approach that combines next generation sequencing and array-based Comparative Genomic Hybridization (aCGH) to reveal maize SVs at the whole genome level. This project will be conducted by 3 French laboratories which are leaders in maize genetics and genomics and statistics. Their partnership gathers complementary skills in molecular biology, bioinformatics and statistics. This project will also take advantage of strong interactions with a U.S. laboratory pioneer in this domain which has been deeply involved in maize genome sequencing, as well as support of a biotechnology company that develops arrays, and of the french National Center for Sequencing (CNS). In the first part of our project will be to capture SVs among a core collection of genetically distant European and American maize lines, using aCGH array. The maize genome has been sequenced by an American consortium using the B73 U.S. line, and a CGH array for this sequence is available. To capture as many SVs as possible from our core collection, we

will complement this array with sequences that are specific of the French F2 line, which is widely used in French breeding programs. Hybridization of the DNAs of our core collection on this pangenomic CGH array will allow to characterize the extent, organization and nature of SV in maize, and will generate a comprehensive dataset of maize SVs.
In the second part of the project, the most relevant SVs will be selected to develop an innovative CGH array dedicated to the genotyping of a large association panel. To decipher whether the use of SVs as genotyping markers will provide information complementary to that of conventional markers such as SNPs for genetic association studies, we will characterize the nature and prevalence of each relevant SV, as well as its link with other polymorphisms. SV data will then be used to investigate the genotype-phenotype association for agronomic traits using an SV-adapted Linkage Disequilibrium mapping approach, with a particular interest in understanding how SVs can be predictive to the heterotic response observed when crossing lines from different genetic groups. We will also develop methodologies to include this new type of information within breeding programs. Finally, high-throughput molecular markers will be derived from the most associated SVs, leading to the construction of a «CNV markers toolkit » that will be useful for the whole maize community, including breeders and researchers. These CNV markers will likely be particularly powerful for further prospects on recombination, heterosis and chromosome plasticity, which are key processes to be elucidated for genetics and breeding programs. Inclusion of markers that are specific to European lines will lead to major improvement for European breeding and is likely to pave the way for future maize research in Europe.

Partners	UMR320 Biogemma UMR518
Coordinator	Stéphane Nicolas - UMR320 stephane.nicolas@versailles.inra.fr
ANR funding	232 k€
Starting date and duration	November 2011 - 36 months
Reference	ANR-10-GENM-003
Cluster label	Céréales Vallée

Project title	**DURAVITIS** - Developmental, molecular and genetic bases of the grapevine adaptation to thermal stress
Abstract	The grapevine, like other important crop species, is faced to the effects of ongoing and predicted climate change, for instance of temperature increase. Temperature is well-known factor involved in grapevine vegetative and reproductive developments with important consequences for wines yields and quality. Temperature increase has already been shown to alter yield components (eg: flowers initiation, anthesis, flower and berry abortions) and wine quality (taste (acid balance), aspect (anthocyanidin pigmentation), flavors (aroma precursors) and nutritional properties (antioxidant properties, ethanol content). Plant breeding combined with molecular plant biotechnology have the potential to deliver stable yields and quality under warmer climate conditions, but this requires the identification of key traits of tolerance to heat and their incorporation into new cultivars. In order to success in such tasks, the mechanisms of grapevine response to thermal stress need to be better understood and traits of adaptation need to be identified. The central goal of the DURAVITIS project is to re-evaluate the impact of temperature on grapevine vegetative and reproductive developments, through various angles of analysis including whole plant and microenvironment levels. A developmental analysis framework will be established to investigate the complex processes involved in microvine adaptation to temperature (C flow, berry respiration and metabolisms). Another goal of DURAVITIS is to identify genetic traits of temperature tolerance in grapevine screening of Vitis germplasm. Given the various objectives of this project and the complexity of analyzing grapevine response to temperature in terms of vegetative and reproductive developments, this project is designed in a multidisciplinary research format involving ecophysiologists, molecular physiologists and geneticists. We believe that the release of high-throughput grapevine genomic tools and the development of innovative experimental approaches (fruiting cuttings, microvine) will help to dissect ecophysiological and molecular mechanisms underlying grapevine adaptative responses to high

temperature, and facilitate the identification of genetic traits of heat tolerance. Several outcomes are expected to be produced in DURAVITIS, including:
i) Scientific advances - DURAVITIS is expected to release original information about these aspects in perennial fruit crops leading to a major breakthrough in the comprehension of fruiting plant responses to environmental constraints.
ii) Technical outcomes - The DURAVITIS project proposes a new methodological format based on the use of fruiting cuttings and microvines to accelerate and improve both ecophysiology and genetic investigations. By-passing several experimental drawbacks associated with traditional grapevines, this project offers the possibility to establish a phenotypic pipeline that may ultimately be used to address a range of questions related to grapevine adaptation to environmental stresses.
iii) Economical opportunities - The selection of new cultivars capable of withstanding climate warming while retaining quality appears to be most efficient and sustainable approach in a long-term perspective. The studies proposed in the DURAVITIS project are a pre-requisite to establish the principles for phenotypic analysis and for the selection of tolerant genotypes to temperature. Such tools are critical to initiate or design breeding programs for cultivar selection to temperature tolerance.

Partners	UMR DIAPC UMR LEPSE UMR SPO UMR EGFV
Coordinator	Laurent Torregrosa - UMR DIAPC laurent.torregrosa@supagro.inra.fr
ANR funding	547 k€
Starting date and duratio	January 2011 - 48 months
Reference	ANR-10-GENM-004
Cluster label	

Project title	**EFG-MIG** - Evolutionary and Functional Genomics of Modified Indole Glucosinolates
Abstract	The glucosinolate-myrosinase system is an activated defense system in the model plant Arabidopsis thaliana and related species from the order Brassicales. This system protects plants effectively from most herbivorous insects and other enemies. It relies on the generation of toxic effector molecules from biologically inactive precursors upon enemy attack. While the basic functional principle of this activated defense is simple, the system itself is nonetheless extremely complex and displays an enormous amount of structural and regulatory variation within and among species. Previous research has largely focused on methionine-derived glucosinolates, the most abundant and structurally diversified class of glucosinolates in Arabidopsis. The enormous importance of indole glucosinolates for ecological interactions between plants and their natural enemies has been realized only very recently. The genetic architecture that controls variation in indole glucosinolate structures is complex, and we have cloned only the first of several QTL a short while ago. The fundamental research that is described in this proposal is designed to dissect the genetic architecture that controls structural and quantitative variation in this important class of secondary metabolites in the model plant Arabidopsis. Moreover, it aims at providing insight into the factors that determine variability in indole glucosinolate biosynthesis and control across different taxonomic levels, by a comparative investigation of key genes within the species Arabidopsis thaliana, among close and distant Arabidopsis relatives, and between the Brassicaceae, Cleomaceae, Capparaceae and Resedaceae families of the order Brassicales.
Partners	UPS/ESE IBP
Coordinator	Juergen Kroymann - UPS/ESE juergen.kroymann@u-psud.fr
ANR funding	481 k€
Starting date and duration	January 2011 - 48 months

Reference	ANR-10-GENM-005
Cluster label	

« Genomics, plant biotechnology » programme

YEAR 2010

Project title	**GEMBAL** - Multi-breed Genomics of Beef and Dairy Cattle
Abstract	Adapting selection tools and objectives to efficiently manage French cattle meat and milk productions is a major challenge of the next years. Genomic selection provides a fantastic opportunity to reorient bovine selection towards a more sustainable breeding. The Gembal project aims at developing a multi-breed genomic selection to extend its use to all beef and dairy breeds, including the small ones. Special attention will be paid for functional traits and maternal traits: calving ease, fertility and longevity of cows in both beef and dairy breeds. At national level, this project should be a common foundation for all breeding schemes, thus avoiding a multiplication of too small and inefficient initiatives. The core of the project is the making-up of the technical basis for the development of multi-breed genomic selection in beef and dairy cattle. The basic idea is that a sample - so-called imputation population - will be genotyped with a high density chip in each breed, whereas most other individuals will be genotyped at a lower cost for a medium density chip. The condition required to build the imputation populations is an extensive use of a new molecular tool, a high density chip with 800,000 SNP developed by Illumina with a consortium including INRA and UNCEIA. Task 2 is dedicated to this technical part of the project. In Task 3, the large multi-breed resource cattle population generated in Task 2 will be the basis for academic researches aimed at characterizing the genetic diversity across breeds and the history of each population submitted to its own context, i.e. drift and selection. This task will also be useful to detect the conserved chromosomal segments across breeds that can be used in multi-breed genomic selection as it will be envisioned in Task 5. Task 4 corresponds to imputation, i.e the statistical procedure to infer missing genotypes in most individuals from the complete genotype information in a limited imputation sample. We will study the quality of the imputation according to breed effective and imputation sample size. We will also develop more computationally efficient algorithms, as imputation will be very demanding with the fast development of genomic selection.

Then, a genomic prediction model, using linkage disequilibrium information across breeds, will be developed in Task 5. The methodological challenges are the development of powerful and robust statistical approaches as well as and computing tools for the prediction in a multi-breed context, especially for functional traits with correlated direct and maternal genetic effects. The applications regarding functional traits will be carried out in Task 6 and Task 7 for dairy and beef breeds, respectively.
In Task 6, the existence of three breeds in France for which reference populations of reasonable to very large size are available and for which genomic selection programs are already implemented will allow us to undertake reliable comparisons of within vs multi-breed genomic evaluations, hopefully revealing what are the underlying conditions for a successful implementation of multi-breed evaluation. An alternative strategy will consist in checking whether the conserved genome fragments corresponding to favourable haplotypes of QTL detected in any large breed are also segregating in the smaller breeds. Then a genomic evaluation based on these haplotypes could be implemented for the smaller breeds.
In Task 7, the multi-beef breed reference population will be composed of the 2,300 bulls that also constitute the beef imputation populations. If a sufficient number of QTL are commonly detected across beef and dairy breeds, a QTL detection and a computation of prediction equations from the beef and dairy pooled reference populations will be undertaken for maternal functional traits.

Partners	Inra - GABI Institut de l'Elevage UNCEIA Races de France
Coordinator	Florence PHOCAS - GABI Florence.phocas@jouy.inra.fr
ANR funding	662 k€
Starting date and duration	January 2011 - 36 months
Reference	ANR-10-GENM-014
Cluster label	

Project title	**GENAMIBE** - Understating the Pathogenicity of Entamoeba Using Comparative Transcriptomics and Phylogenomics
Abstract	Entamoeba histolytica is a protozoan parasite and an amitochondriate pathogenic amoeba, which may cause dysentery and liver abscess in humans (i.e. amoebiasis). Disease (transmitted by cyst-contaminated water) develops in approximately 10% of infected individuals, resulting in 50 million clinical cases and 100,000 deaths annually. Differential pathogenicity is observed among E. histolytica strains: strain HM1:IMSS is virulent while strain Rahman is naturally attenuated. A closely related species, E. dispar, is non-pathogenic and produces asymptomatic infections. Another Entamoeba species, E. moshkovskii, is primarily free-living and rarely infects humans. Despite the fact that these Entamoeba species are morphologically indistinguishable and phylogenetically closely related, their clinical outcomes are dramatically different. Their phenotypic differences form an excellent theoretical basis for genome-wide comparative analyses to search for factors relevant to pathogenicity and adaptation to humans. Our aim is to understand the phenotypic differences between Entamoeba species/strains using comparative phylogenomic and transcriptomic approaches, which is supported by two primary objectives; 1) To catalogue the relevant and high-resolution transcriptomic and phylogenomic differences between Entamoeba species/strains and 2) To functionally annotate the Entamoeba protein families and discover the relevant gene sets relevant for the phenotypic differences between Entamoeba species/strains. First, taking advantage of the next-generation sequencing technologies, we plan to characterize the transcriptional landscape of the Entamoeba species/strains at an unprecedented scale and resolution, including the generation of genome wide maps for coding and non-coding transcripts, small RNAs and anti-sense RNAs, as well as the expression profiles of these transcripts in three culture conditions (axenic culture, nitric oxide treatment and ex vivo colon culture). By discovering the diversity and expression profiles of small RNAs and anti-sense RNAs we expect to provide valuable insights into potential roles of non-coding RNA across the

Entamoeba transcriptomes.
Second, from an evolutionary perspective, the phenotypic differences between pathogenic and non-pathogenic Entamoeba species result from natural selection of loci mutations (i.e. adaptive selection). Identification of these loci might shed light on the genomic basis of their phenotypic differences. Therefore, using integrated phylogenomic methods, we plan to perform a genome-wide scanning of the interesting mutations from an evolutionary standpoint, including adaptively evolving sites and point mutations that are likely to have functional impacts. By co-analyzing these coding differences with the transcriptomes described above, this study is expected to provide a comprehensive picture of the genotypic differences of Entamoeba species.
Finally, the relatively poor annotations of Entamoeba genomes represent the bottleneck for the analysis of high-throughput data. We plan to re-annotate the coding regions with traceable functional annotations and present them with a desktop application, enabling researchers to easily integrate and analyze their data on a pathway/network basis rather than the laborious "from-gene-to-gene" basis.
This project represents the first and the most comprehensive study on protists in terms of : 1) the types of transcripts we are able to capture; 2) the quality of the coding transcript map (i.e. map of transcription start site, splicing junctions, alternative splicing pattern and poly-adenylation sites); 3) the scope of comparisons (i.e. inter-species, inter-strain and inter-culture-condition comparisons).
We expect this study to have a significant impact on the fundamental understanding of transcription in lower eukaryotes and set an example for high-quality standard for studies in similar kinds of protists.

Partners	Institut Pasteur
Coordinator	NANCY GUILLEN – Institut Pasteur nguillen@pasteur.fr
ANR funding	642 k€
Starting date and duration	January 2011 - 36 months
Reference	ANR-10-GENM-011
Cluster label	

Project title	**GnpAsso** - A generic tool for managing and exploiting genetic association studies results using high throughput genotyping data and high throughput phenotyping
Abstract	Over several years, large efforts have been made to produce crop genetic and genomic data at genome scale. Bioinformatics and data integration was more and more required to take full advantage of these data [1]. Several databases were developed at national and international levels, dedicated to one species in particular or dedicated to one type of data. Association genetic methods, based on the exploitation of linkage disequilibrium (LD), are valuable tools for the dissection of complex traits. In many plant species indeed, they have shown to be useful for the analysis of the genetic determinism of quantitative traits and to pinpoint the genes involved in the variation of traits of agronomic interest. With the increase of genome sequencing and resequencing plant projects and the rapid decrease of cost in molecular typing, with at the same time, the set up of high-throughput genotyping methods, it is now possible to genotype a lot of SNP polymorphisms to conduct exhaustive association mapping approaches at the whole-genome scale "Whole genome association mapping" or in regions in which QTL have been detected by linkage approaches, no longer focusing only on candidate genes putatively involved in the variation of traits of interest. This strategy is today possible for an increasing number of crops. By combining this method with traditional ones, as QTL mapping, it is now possible to i) fine map QTL, ii) identify faster new markers useful for selection, iii) to find interesting alleles in genetic resources collections and to use them in new material dedicated to selection. The rationale of GnpAsso project is to create a new bioinformatics database resource dedicated to the storage and the query of genome association mapping results in relation with highthroughput sequencing data (in particular third generation markers "Single Nucleotide Polymorphism, SNP data). This resource will also give access ii) in details to all the data linked to these association studies results (the genetic resources, genotypes, SNP, traits, alleles, QTL, gene

	annotation, gene expression…) via links to other existing resources in GnpIS, URGI information system.
Partners	INRA URGI Biogemma UMR Génétique végétale du Moulon UMR LEG UMR DIA PC UMR RPB
Coordinator	Delphine Steinbach - INRA URGI delphine.steinbach@versailles.inra.fr
ANR funding	382 k€
Starting date and duration	February 2011 - 36 months
Reference	ANR-10-GENM-006
Cluster label	Céréales Vallée

Project title	**Immunit-Ae** - Genetic diversity and mechanisms of resistance to Aphanomyces euteiches in legumes

Abstract

Aphanomyces euteiches, a root oomycete of legumes is the main limiting biotic factor for French and European pea production. To accelerate genetic improvement of resistance in pea to A. euteiches we propose to use genetic and genomic resources developed in the model legume M. truncatula. This includes notably the recent discovery of a major locus of resistance to A. euteiches in this plant. This locus was identified in two different sources of resistance with several isolates of this parasite. In the A17 sequenced line,this locus, named prAe1 conferred a partial recessive resistance and was narrowed to a 135 kb genomic region which does not contain any classical NBS-LRR resistance gene. In the DZA45.5 line, the AER1 gene, located in the same genomic region is responsible for a complete dominant resistance and act epistatically with other loci to confer resistance against several A. euteiches isolates.

From this results, the first objective of this proposal is to identify sequence(s) located in the prAe1/AER1 sequences that are involved in the resistance to A. euteiches. Following the sequencing of this locus in DZA45.5 and in the susceptible F83005.5 line, data comparison with A17 will lead to fine mapping studies to identify the gene(s) (or sequences) involved in resistance. These sequence(s) will then be functionally validated in complementation studies performed in transformed M. truncatula roots, in available mutants and/or in transgenic M. sativa plants. . To analyze prAe1-related molecular mechanisms, near-isogenic lines (NIL) at this locus will be analyzed following A. euteiches inoculation with Affymetrix chips in microarray experiments. The transcriptome of divided pericycle cells, which are only observed in the resistant NIL, wil also be analyzed by using RNAs from these cells that will be microdissected by laser capture microdissection.

The second objective of this proposal is to mine the diversity of loci involved in resistance to several strains of A. euteiches in different resistance sources of a M. truncatula collection. We will take advantage of a privileged access to data generated by an ongoing NSF Hapmap re-sequencing

genome project in M. truncatula. Two approaches will be used : i)- a genome-wide association mapping analysis in a core-collection of 192 M. truncatula lines, which genomes are currently under re-sequencing and, ii)- a QTL mapping approach in three available joint family populations, which could use whole genome sequence data of the founder lines for developing a nested association mapping study for increasing resolution of QTL detection. Results from the two approaches will be compared.

The third goal of this project is to assess conservation of genes, loci and mechanisms involved in resistance to A. euteiches between M. truncatula and main leguminous crops, especially pea. To reach it, we will take advantage of the high amount of knowledge and data previously acquired, particularly in pea (markers, sequences, QTL). First, synteny of Aphanomyces resistance loci identified in M. truncatula and pea will be studied by generating bridge markers in target regions between the two genomes. Second, orthologous genes to candidate genes at the main Aphanomyces resistance loci detected in M. truncatula will be searched in different leguminous crops, especially pea. Markers at these orthologous genes will be developed. In pea, association between resistance and allelic diversity at best candidate genes (i.e. in syntenic regions and/or colocalizing with QTL) will be validated in a pea core-collection. Functional role in resistance of a few validated candidate genes will be analyzed with TILLING mutants. Finally, cytological studies will be undertaken to compare resistance mechanisms observed in M. truncatula and several other crop legumes.

Partners	UPS - SCSV UMR INRA-Agrocampus Ouest-Université Rennes I APBV INRA DIAPC Montpellier URGV Evry
Coordinator	Christophe Jacquet - UPS - SCSV jacquet@scsv.ups-tlse.fr
ANR funding	686 k€
Starting date and duration	January 2011 - 36 months
Reference	ANR-10-GENM-007
Cluster label	

« Genomics, plant biotechnology » programme

YEAR 2010

Project title	**PHENOBLE** - Development and utilization of new generation phenotyping tools to analyse genetic determinants of nitrogen fertilisers use efficiency in bread wheat
Abstract	To match the increase in demand of cereals in the next 50 years, wheat yield has to be improved by 2% per year. This challenge will be met only if a new paradigm is built in wheat breeding and agronomic practices. Advances in crop productivity are related to the understanding of yield limiting factors and the development of strategies for future genetic improvement. This should be achieved through the advances made by searchers and breeders in two main areas: genotyping and phenotyping. Notable efforts have been applied in recent years to develop numerous cost-effective molecular markers to construct genetic materials suited to the objectives of research programmes and more recently to sequence entire genomes. In spite of these efforts, there has not been much progress in the area of phenotyping, especially in field conditions which is required for the main agronomical traits. Recent developments concentrated on plant phenomics as an emerging field that develops and provides tools such as technologies to: - Characterize plant performance and the dynamics of plant structures and functions - Design a high throughput evaluation of plants in response to desired environmental scenarios, including novel field techniques based on proxi-identification. The precise measurements of phenotypic characters is of paramount importance to discover genes that explain such complex traits. Right now, the main development has been made in controlled environments as greenhouses and not in the field. In this project named PHENOBLE, we propose to build a first platform adapted to fine and accurate phenotyping with a new set of tools. This platform will be dedicated to bread wheat and will aim to decipher the genetic factors involved in genotype x environment interactions regarding nitrogen uptake. PHENO BLE will aim to validate, adapt and improve available new phenotyping tools by evaluating a collection of wheat elite lines under different nitrogen fertilizer regimes.

	The project will focus on nitrogen because it represents a very important target for economic, environmental and agronomic reasons. Nitrogen influences characters as leaf area, senescence, chlorophyll content or metabolites content that are among the ones that appear as possibly measurable by new generation phenotyping tools. The main outputs of this project will be not only validated phenotyping tools or platforms usable in breeding or genetics programs but a set of heritable phenotyping traits and markers associated. Acquiring rapidly and dynamically large data sets will able to efficiently carry out association studies that should allow at the end of the project to identify the main genetic determinants of a better nitrogen use efficiency. Both markers and tools will benefit to the scientific and breeding community.
Partners	ARVALIS BIOGEMMA UMR GDEC INRA-Bordeaux INRA-EMMAH UMR DIAPC UMR AGIR
Coordinator	Katia BEAUCHENE - ARVALIS k.beauchene@arvalisinstitutduvegetal.fr
ANR funding	365 k€
Starting ate and duration	January 2011 - 48 months
Reference	ANR-10-GENM-008
Cluster label	Céréales Vallée

« Genomics, plant biotechnology » programme

YEAR 2010

Project title	**PhyloFish** - RNASeq-based PHYLOgenomic analysis of gene duplications in teleost FISHes

Abstract

Genome duplication is a recurring theme in vertebrate evolution and for instance, the human genome still contains numerous gene families that arose from the two rounds of genome duplication at the origin of vertebrates. Teleost fish are in that regards very interesting models due to an additional whole genome duplication (WGD) event that occurred at the base of their radiation, resulting in duplicated copies of many single-copy human genes. These 'extra' genes are, in principle, available for the evolution of new functions that could drive the origin of novelties and thus contribute to the diversification of life on Earth. Gene evolution after genome duplication is thus a crucial question to understand the mechanisms by which genomes evolved and drive the development and physiology of vertebrates. Unfortunately, we still do not yet have a sufficient understanding of vertebrate genomes to fully answer this question. The present project will make use of the additional WGDs of teleosts to address this question. By using next generation sequencing this project will first provide novel and genome-wide information on the transcript repertoires of different fish species chosen at key taxonomic positions with regards to teleost fish evolution. This evolutionary-relevant sequence dataset will then be used as a basis for the development of a high throughput analysis combining gene phylogenies, conserved syntenies and expression profiling. Results of this project should provide genome-wide answers on how often different gene copies are lost independently in different fish lineages and whether lineage-specific changes in duplicate gene content or in expression patterns of duplicated genes is important for the evolution of the remarkable diversity among teleosts. In addition and because these gene duplications also have a major impact on the quality of gene annotation in teleosts, this project will propose, supported by the results of our evolutionary-based analysis, the refinement of teleost gene nomenclature. This will link gene information across many vertebrate species, allowing to bridge functional information from model species to economically-relevant species.

Partners	INRA SCRIBE UOregon MGX INRA SIGENAE U Aix Mars 1
Coordinator	Julien BOBE - INRA SCRIBE Julien.Bobe@rennes.inra.fr
ANR funding	481 k€
Starting date and duration	February 2011 - 48 months
Reference	ANR-10-GENM-017
Cluster label	

Project title	**RepliColScope** - Sequencing and comparative genomics of Escherichia coli plasmids: a way to understand the evolutionary success of extended-spectrum ß-lactamases

Abstract

The spread of antibiotic-resistant pathogens is becoming an extremely serious clinical and public health problem worldwide. In recent year, we have observed a dramatic increase of the resistance to the third generation cephalosporins that are regularly used for empirical therapy of Enterobacteriaceae. Acquired resistance to third generation cephalosporins is mainly mediated by extended-spectrum beta-lactamases (ESBLs) that confer resistance to all beta-lactams except carbapenems and cephamycins. Currently, the high prevalence of all ESBL coding genes, mainly carried on plasmids, is caused by different mechanisms, not mutually exclusive: (i) horizontal transfer of ESBL genes between related or unrelated plasmids, (ii) horizontal transfer of plasmids among unrelated clones and (iii) spread of local or international epidemic clones. During the 1990s, TEM-ESBLs and SHV-ESBLs were dominant. Since the 2000s, CTX-M enzymes have become the most prevalent ESBLs worldwide. Escherichia coli, a versatile human and animal major opportunist pathogen which was up to now sensitive to third generation cephalosporins, has been recognise as the major source of ESBLs. ESBL-producing E. coli are commonly isolated in human community or hospital infection and from faecal carriage and are also increasingly detected in food-producing animals, companion and wild animals and in the environment.

The sudden emergence of E. coli producing ESBLs and the prominence of CTX-M enzymes cannot be explained simply as a result of selective pressure exerted by the use of third generation cephalosporins. Other factors may influence this evolution at ecological and/or molecular levels.

The objective of this project is to understand the evolutionary success of the ESBLs in
E. coli, and especially the CTX-M enzymes emergence as compared to the TEM and SHV enzymes.

To reach this objective, we propose (i) to sequence 80 E. coli plasmids, carrying the three major types of ESBLs (70 plasmids from human isolates and 10 plasmids from animal

isolates) and 20 comparative plasmids from periods before the use of third generation cephalosporins (17 plasmids from the ECOR collection strains and 3 plasmids from the E. coli strains of the Murray collection dating from the pre-antibiotic period), (ii) to perform comparative genomics with the generated and already available data and (iii) to test in vitro some hypotheses generated from the in silico data.
The plasmids will be sequenced using a strategy involving high-density pyrosequencing (A 454 Titanium) allowing reads of 400 bp. Novel strategies will be developed in term of automatic structural and functional annotation of the plasmid sequences and the results will be integrated into a special database to perform relevant expert annotations and to explore the data. Comparative genomic and phylogenetic analyses will be performed to reconstruct the evolutionary history of the plasmid genes by identifying those having the most similar gene content. The data will also be analysed in the light of the chromosomal genetic background, represented by the phylogenetic history of the strains and the CRISPRs. We will explore the possible cross talk between modules carried on plasmids and those found on the chromosomes of E. coli, which can play a key role in the stabilization of these extrachromosomal elements.
This project should provide us amounts of quantitative and qualitative data that will bring both fundamental and medical insights on plasmid evolution and on antibiotic resistance. Our project could also offer opportunity for advances in new strategies to prevent resistant plasmid dissemination in bacteria.
ReplicolScope is an association between teams that have a strong expertise on the evolutionary biology of E. coli (U722), the epidemiology of the ESBLs among human and animal strains (UR8 and UR1282), the bacterial plasticity (UPGB) and, a team that is skilled in gene sequencing and in bioinformatics (Genoscope).

Partners	U722 Genoscope ER8 UR1282 UPGB
Coordinator	Catherine Branger - U722 catherine.branger@lmr.aphp.fr
ANR funding	380 k€
Starting date and duration	January 2011 - 36 months

Reference	ANR-10-GENM-012
Cluster label	

Project title	**SOLAR** - Developing and testing new biotechnological tools to increase seed oil yield: a key to the production of hydrocarbon chains in oleaginous crop species
Abstract	Project aims: Oil from seeds constitutes a key component of both human and livestock diets, which consumption is steeply increasing worldwide. Then, fatty acids composing triacylglycerols (TAGs) accumulated in plant seeds are structurally similar to long chain hydrocarbons and consequently represent logical and competitive alternatives to hydrocarbon-based products for the production of detergents, paints, plastics and lubricants (green chemistry). The increasing demand of plant oils for such industrial and nutritional applications highlights the urgent need to develop new methodologies to increase seed oil content when only little progress has been made by conventional breeding over the last decade. The successful engineering of domesticated high yielding oil crop species now requires a full elucidation of the mechanisms controlling the production of fatty acids and their assembly into TAGs. The TAG biosynthetic network comprises two blocks of reactions: block A is composed of plastidial enzymes involved in fatty acid synthesis, while block B of reactions is composed of acylating enzymes involved in TAG assembly in the ER. So far, most biotechnological approaches aimed at increasing seed oil content have put the focus on block B, thought to contain the few metabolic bottlenecks limiting seed oil production. However, metabolic control analysis experiments have recently shown that control of flux is exerted both by block A and block B. Within each of these blocks, metabolic control is then further shared between several enzymatic steps. Thus, it is essential to elucidate the regulation of block A to find out new biotechnological tools able to efficiently stimulate fatty acid synthesis. Recent data indicate that block A is highly regulated at the transcriptional level and that a coordinated activation of most genes encoding enzymes of block A is necessary to stimulate the rate of fatty acid production. This is the reason why this project is focused on the transcriptional regulation of block A. Work plan:

This project aims to elucidate the transcriptional regulation of fatty acid biosynthesis in the model plant Arabidopsis and to provide us with new ways to modify seed filling in the support of sustainable agriculture. Our project has three major objectives:
- The first is to isolate an original set of transcription factors involved in the control of fatty acid synthesis in Arabidopsis.
- The second is to provide a deep understanding of the regulatory complex controlling the transcription of lipogenic genes.
- The third is to exploit this knowledge to boost fatty acid production in seeds in the frame of a biotechnological approach.
Expected results:
- The main outcome of the project will consist in original knowledge about the regulation of fatty acid biosynthesis in plants.
- The expected results of the project are mainly fundamental, even though biotechnological tools will be developed in the model species Arabidopsis as a proof of concept. Tools and strategies tested in this project may then be applied to the improvement of oil yields on oilcrops such as Brassica napus and the generation of new varieties of bioeconomic interest.

Partners	INRA IJPB SDQ INRA IJPB PLHS
Coordinator	Sébastien BAUD - INRA IJPB SDQ sbaud@versailles.inra.fr
ANR funding	320 k€
Starting date and duration	February 2011 - 48 months
Reference	ANR-10-GENM-009
Cluster label	

Project title	**SUS-FLORA** - Gut microbiota and homeostasis of immune system in swine: a joint genetic and genomic study
Abstract	In the near future, a broader definition of breeding goals should balance productivity with improved functional traits such as health, in order to describe animals as an integrated component of sustainable production systems. In this context, there is a need to study the genetic control of general immune response and analyse whether health traits referred to as immunocompetence might be included in future selection schemes. In collaboration with pig breeders, we have started a large survey of innate and adaptive immune response in French Large White pigs (IMMOPIG project, ANR 2007-2009) and moderate to strong heritability estimates have been obtained for the majority of the traits investigated, showing that immune response is largely under genetic control, in agreement with reports from other laboratories. Animals of the IMMOPIG project are being genotyped with porcine 60K iSelect chips for genetic association studies. In addition, a divergent selection for an index of four immune parameters has been launched and animals of the G1 generation will be born in 2010. Commensal bacteria inhabit all body surfaces that are exposed to environment and the lower gastrointestinal tract may be considered as the best example of bacteria hostage with a possible coevolution of host and bacteria. The gut microbiota develops as a host-specific parameter that gets stabilized early during lifetime. There is a growing interest in studying the relationships between the mammalian immune system and the bacteria that are present in the mammalian gut. Analyzing microbiota as a potent relevant phenotype to integrate into a global approach on immunocompetence represents a prospective and original approach. The aim of the work will be to study gut microbiota in pigs that are scored for conventional immune response and consider the gut microbiota as a new parameter to include in genetic analysis. The project is divided into two main parts. On the one hand, global studies including microbiota characterisation, animal phenotyping and genotyping will be carried out. On the other hand, local studies targeting interactions between gut epithelium and microbiota will be investigated for a subgroup of piglets well

characterized for phenotypes and genotypes. Ninety mononuclear families will be produced and four piglets per family will be studied. Microbiota of 60 day-old piglets will be characterized together with the microbiota of the sows after piglet birth. Variation of microbiota during early lifetime will be studied for a subset of 30 piglets from birth to the growing period. Microbiota heritability and correlations between gut microbiota and immune response parameters will be calculated. The 360 piglets and the 90 boars will be genotyped using the genome-wide porcine 60K iSelect chips in order to accumulate preliminary data for further genome-wide association studies. Major Histocompatibility Complex haplotypes and Toll-like receptor polymorphisms will be characterised. Local interaction studies will be carried out on the subgroup of 30 piglets followed for microbiota variations. We shall have the possibility to also include animals produced during the ongoing divergent selection on an immunity index. The differential expression of the host genes at five distinct gut sites (duodenum, jejunum, ileum, colon and Peyer's patches) will be studied by a transcriptomic approach. We will target three key functions of local immunity: histocompatibility by analysing expression of non-classical class I genes (SLA-Ib, MIC-2, MR1, CD1), motif recognition at the cell surface (TLRs, NODs) and analysis of IgA immune response expansion. In addition, peripheric T lymphocytes will be compared to intestinal T lymphocytes.

Partners

INRA-GABI
CEA-DSV/IRCM/SREIT/LREG
INRA - MICALIS
INRA-LPT
INRA-VIM
INRA-GEPA
BIOPORC
INRA-IASP

Coordinator

Claire Rogel-Gaillard - INRA-GABI
claire.rogel-gaillard@jouy.inra.fr

ANR funding

766 k€

Starting date and duration

March 2011 - 36 months

Reference

ANR-10-GENM-016

Cluster label

Project title	**UTOPIGE** - Towards the Optimal Use of Genomic Information in pyramid schemes
Abstract	The aim of the project is to give the information necessary to the implementation genomic selection in pyramid schemes which concern : selection of purebred animals to produce crossbred animals; selection in high health environment but production under commercial conditions. Experimental approach will be carried out in pig and in chicken. Resource populations, purebred and crossbred, controlled in different environments, will be created and controlled for a large number of traits. Genotyping will be carried out thanks to high density SNP Chips (64K already available). Different methods of genomic evaluaation will be compared according to their ability to predict accurately the genetic values of the reproducers in these situations (genetic type, environment). Data sets will be also exploit to optimize the breeding schemes. Results will be presented to the breeders' community of concerned species.
Partners	GARen GABI LGC SAGA UE TP IFIP SYSAAF Novogen
Coordinator	Pascale LE ROY - GARen pascale.leroy@rennes.inra.fr
ANR funding	968 k€
Starting date and duration	January 2011 - 48 months
Reference	ANR-10-GENM-015
Cluster label	VALORIAL - l'Aliment de demain

Project title	**WALLARAY II** - Glycan microarrays of cell wall polysaccharides in functional genomics

Abstract

The project "WallArray" was submitted to the ANR call for proposals entitled "Genomics" in April, 2008. In this framework, the project was funded for a feasibility study over one year. The obtained results confirm the objectives initially proposed by the five partners involved in the project. The new project is based on the main results of the feasibility study, and develops new technical solutions and new propositions to achieve the objectives. The submission of the project WallArray 2 to the call for proposals "Genomics, Plant Biotechnologies" thus joins in the continuity of the previous project. It brings together the same consortium of research teams from the public institutes of CEA, CNRS and INRA, from the universities of Grenoble and Toulouse, and from the private company Genoptics. They develop researches in the fields of physico-chemistry, chemistry, structural biology, and plant biology.

Cell walls are natural composite structures, mostly made of high molecular weight polymers (polysaccharides and lignins) and proteins interacting with these polymers. Cell walls are dynamic structures involved in cell division, expansion, differentiation and adhesion governing plant morphology. They are also the sources of oligosaccharide signals for molecular recognition to control developmental processes and even the outcome of plant-microorganism interactions. In recent years, significant headway in genomics, proteomics, glycomics and bioinformatics increased our knowledge of cell wall-related genes and proteins, in particular those related to the modification and interaction with polysaccharides. The function of only few of them was experimentally proven. Therefore, structural and functional aspects concerning the interplay of the cell wall glycome (defined as a set of carbohydrate structures present in the cell wall) and the cell wall proteome (set of proteins) are of great interest to understand cell wall dynamics and signalling.

The aim of the project is to develop carbohydrate microarrays of plant cell wall polysaccharides as innovative tools for screening cell wall carbohydrate-binding proteins (CBPs) and assessing their specificity. The functional and structural data obtained by the consortium will be integrated into existing

database freely opened to the scientific community (glyco3d). Also, the data will be integrated to design softwares for the prediction of protein and carbohydrate structures. These studies should ultimately lead to more predictive capabilities in protein/carbohydrate interactions. The project brings together a diverse set of established novel technologies which were implemented during the feasibility study: from the generation of glycan libraries and their linking to a surface, to the production of carbohydrate binding proteins and the analysis of carbohydrate-protein binding by surface plasmon resonance imaging.
Understanding the carbohydrate-protein interactions at the plant cell surfaces complements data of the glycome and the proteome, and represents an essential step in the post-genomic area. On the other hand, plant cell walls represent the most abundant components of biomass. The carbohydrate-containing materials (cellulose, pectins) have been exploited by industries for food and non-food uses throughout centuries. With the recognition of the carbohydrates as carriers of biological properties, glyco-biotechnology companies have emerged. The carbohydrate microarrays will be a useful tool for discovering new polysaccharides acting enzymes, elaborating new materials, and engineering new plants.

Partners	UPS - SCSV INRA INAC/SPrAM GENOPTICS CNRS
Coordinator	Hervé CANUT - UPS - SCSV canut@scsv.ups-tlse.fr
ANR funding	568 k€
Starting date and duration	January 2011 - 36 months
Reference	ANR-10-GENM-010
Cluster label	AGRIMIP INNOVATION

Project title	**XANTHOMIX** - Comparative genomics and transcriptomics of plant-pathogenic xanthomonads
Abstract	The project XANTHOMIX aims at providing genomic and transcriptomic resources for a better understanding of the infection process, host and tissue specialization, evolutionary history and epidemiology of five Xanthomonas species. Xanthomonads are gram-negative bacteria which cause diseases of more than 400 different host plants belonging to both monocots and dicots, many of them having great economic importance (e.g. barley, rye, wheat, rice, beans, cabbage, cassava, citrus and cotton). Xanthomonads display a high degree of host and tissue specificity. The genus Xanthomonas is an excellent model to study the molecular, evolutionary and epidemiological aspects of host-bacteria interaction. We wish to sequence 22 strains of Xanthomonas which serve as models in our labs to study all phases of the infection process and/or which are of special importance as quarantine organisms that represent a significant threat for agriculture and global food security. All chosen strains belong to genetic lineages for which no genomic or transcriptomic data are available, among them two new species and 14 new pathovars. Yet, due to the close genetic relationship of carefully selected strains comparative analyses will allow to discover candidate components responsible for tissue and/or host specificity. We will use next-generation sequencing technology. Specifically, we will combine 454 sequencing of genomic DNA with Illumina sequencing of cDNA. This combination will allow to obtain high-quality draft genome sequences and at the same time to study the transcriptome under different conditions. cDNAs will be generated from bacteria grown under standard conditions and under conditions that mimic important steps of the plant infection process. Comparison of data will allow to formulate hypotheses about regulons which can easily be tested thanks to the available genome sequences. For cDNA synthesis, we will use specific protocols which will allow us to identify small, non-coding RNAs and anti-sense RNAs, in addition to the mRNA pool. Knowledge of the transcription start sites will be generated and help to identify

regulatory cis-acting promoter elements, riboswitches and translation initiation codons. The plethora of information will lead to excellent structural annotation of these and other, already available genome sequences of Xanthomonas. Expert functional annotation of genes involved in the infection process will generate an invaluable resource for the international scientific community.
Genome-/transcriptome-derived hypotheses about the role of new components of the infection process will be complemented by functional analyses. Our new genomic resources will also be used to develop high-resolution molecular typing tools which will serve in the epidemiological surveillance of these important pathogens.
The project involves four French partners from institutes being active in fundamental and applied research. Researchers involved in this project, mainly belonging to IRD, CIRAD and INRA, are all well recognized as experts in their field (microbiology, plant pathology, phylogeny, genetics, genomics, epidemiology) with a broad expertise in genetics, genomics and epidemiology of pathogenic bacteria with particular expertise on the Xanthomonas genus.

Partners	RPB LIPM PaVé PVBMT
Coordinator	Ralf KOEBNIK – RPB Ralf.Koebnik@ird.fr
ANR funding	887 k€
Starting date and duration	January 2011 - 30 months
Reference	ANR-10-GENM-013
Cluster label	