

Implication des ARN non codant dans la virulence du phytopathogène Agrobacterium fabrum C58

Magali Dequivre

▶ To cite this version:

Magali Dequivre. Implication des ARN non codant dans la virulence du phytopathogène Agrobacterium fabrum C58. Microbiologie et Parasitologie. Université Claude Bernard - Lyon I, 2015. Français. NNT: 2015LYO10016. tel-01474897

HAL Id: tel-01474897 https://theses.hal.science/tel-01474897

Submitted on 23 Feb 2017 $\,$

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers. L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.





THÈSE DE L'UNIVERSITÉ DE LYON

Délivrée par

L'UNIVERSITÉ CLAUDE BERNARD LYON 1

DIPLÔME DE DOCTORAT

En Micro-organismes, Interactions, Infections

(Arrêté du 7 août 2006)

Implication des ARN non codant dans la virulence du phytopathogène *Agrobacterium fabrum* C58

Par

Mme Magali DEQUIVRE

Soutenue publiquement le 20 février 2015

Directeur de thèse : Florence HOMMAIS

Jury :

Yves DESSAUX	Directeur de recherches CNRS	Rapporteur
Olga SOUTOURINA	Maître de conférences Université Paris 7	Rapporteur
Wafa ACHOUAK	Directeur de recherche CNRS	Examinateur
Céline BROCHIER	Professeure Université Lyon 1	Examinateur
Guy CONDEMINE	Directeur de recherches CNRS	Examinateur
Florence HOMMAIS	Maître de conférences Université Lyon 1	Directeur de thèse





THÈSE DE L'UNIVERSITÉ DE LYON

Délivrée par

L'UNIVERSITÉ CLAUDE BERNARD LYON 1

DIPLÔME DE DOCTORAT

En Micro-organismes, Interactions, Infections

(Arrêté du 7 août 2006)

Implication des ARN non codant dans la virulence du phytopathogène *Agrobacterium fabrum* C58

Par

Mme Magali DEQUIVRE

Soutenue publiquement le 20 février 2015

Directeur de thèse : Florence HOMMAIS

Jury :

Yves DESSAUX	Directeur de recherche CNRS	Rapporteur
Olga SOUTOURINA	Maître de conférences Université Paris 7	Rapporteur
Wafa ACHOUAK	Directeur de recherches CNRS	Examinateur
Céline BROCHIER	Professeure Université Lyon 1	Examinateur
Guy CONDEMINE	Directeur de recherches CNRS	Examinateur
Florence HOMMAIS	Maître de conférences Université Lyon 1	Directeur de thèse

UNIVERSITE CLAUDE BERNARD - LYON 1

Président de l'Université

Vice-président du Conseil d'Administration Vice-président du Conseil des Etudes et de la Vie Universitaire Vice-président du Conseil Scientifique Directeur Général des Services

M. François-Noël GILLY

M. le Professeur Hamda BEN HADIDM. le Professeur Philippe LALLEM. le Professeur Germain GILLETM. Alain HELLEU

COMPOSANTES SANTE

Faculté de Médecine Lyon Est – Claude Bernard	Directeur : M. le Professeur J. ETIENNE	
Faculté de Médecine et de Maïeutique Lyon Sud – Charles Mérieux	Directeur : Mme la Professeure C. BURILLON	
Faculté d'Odontologie	Directeur : M. le Professeur D. BOURGEOIS	
Institut des Sciences Pharmaceutiques et Biologiques	Directeur : M. le Professeur Y. MATILLON	
Département de formation et Centre de Recherche en Biologie Humaine	Directeur : Mme. la Professeure A-M. SCHOTT	

COMPOSANTES ET DEPARTEMENTS DE SCIENCES ET TECHNOLOGIE

Faculté des Sciences et Technologies	Directeur : M. F. DE MARCHI
Département Biologie	Directeur : M. le Professeur F. FLEURY
Département Chimie Biochimie	Directeur : Mme Caroline FELIX
Département GEP	Directeur : M. Hassan HAMMOURI
Département Informatique	Directeur : M. le Professeur S. AKKOUCHE
Département Mathématiques	Directeur : M. le Professeur Georges TOMANOV
Département Mécanique	Directeur : M. le Professeur H. BEN HADID
Département Physique	Directeur : M. Jean-Claude PLENET
UFR Sciences et Techniques des Activités Physiques et Sportives	Directeur : M. Y.VANPOULLE
Observatoire des Sciences de l'Univers de Lyon	Directeur : M. B. GUIDERDONI
Polytech Lyon	Directeur : M. P. FOURNIER
Ecole Supérieure de Chimie Physique Electronique	Directeur : M. G. PIGNAULT
Institut Universitaire de Technologie de Lyon 1	Directeur : M. le Professeur C. VITON
Ecole Supérieure du Professorat et de l'Education	Directeur : M. le Professeur A. MOUGNIOTTE
Institut de Science Financière et d'Assurances	Directeur : M. N. LEBOISNE

Remerciements

Je tiens tout d'abord à remercier les membres du jury Yves Dessaux, Olga Soutourina, Céline Brochier, Guy Condemine et Wafa Achouak pour avoir accepté d'évaluer mon travail de thèse.

Je remercie également Nicole Cotte-Pattat pour m'avoir accueillie au sein du laboratoire Microbiologie Adaptation Pathogénie UMR5240.

Un grand merci à Florence Hommais, pour m'avoir encadrée pendant ces 3 années (et demi) de thèse dans l'équipe ARN Régulateur et Adaptation, et pour la confiance que tu m'as portée. Ca a été un réel plaisir de travailler avec toi, autant au labo, qu'en TP. Merci pour tous les conseils (scientifiques et autres) et pour la formation que tu m'as apportée.

Un grand merci aux membres anciens et présents de l'équipe. Jessica, on n'a pas eu la chance de travailler très longtemps ensemble, mais ça a été un plaisir de te connaître. Tony, c'est en partie grâce à toi que j'ai fait cette thèse, merci pour tous les moments qu'on a partagé, autant au labo qu'en montagne. Benjamin, nouvelle recrue ! Depuis que tu es là, le rez-de-chaussée est toujours plein de vie ! Je te souhaite plein de réussite pour ta thèse. Camille et Julien (tu ne fais pas encore partie de l'équipe, mais c'est tout comme..), après 3 ans passés à vos côtés, vous êtes devenus bien plus que des collègues de travail, vous êtes de véritables amis. Merci pour tous les moments que l'on a pu passer ensemble, pour les franches rigolades pendant les (nombreuses) pauses dans le bureau, pour notre victoire au MAPiades 2014, et pour tout le reste.

Je remercie également les différentes personnes d'Ecologie Microbienne et du MAP avec qui j'ai pu travailler et avec qui j'ai eu la chance d'échanger sur des sujets divers, autant sur le plan scientifique que personnel : Ludovic, Xavier, Zahar, Floriant, William, Sylvie, Anne, Alex, Marc (je te promets que j'essaierai un « vrai art martial » un jour !).

Aux étudiants (et assimilés), il y en a tellement que j'espère ne pas en oublier... un grand merci pour les moments passés ensemble, depuis le master 2 et jusqu'à maintenant (et ça va durer encore quelques mois) : Yohann, Sana, Amélie, Julien C, Kevin, Arnaud, Vincent, Benoît, Juliana, Antoine, Florian, Marie, Yoann (S et BM), Julien G, Quentin, Sébastien, Jordan, Jeanne et David. Un petit plus pour Guillaume, avec qui j'ai partagé de super moments depuis le M2.

Merci également aux différents membres de l'association DocE2M2 avec qui j'ai eu plaisir à organiser plusieurs années consécutives les doctorales et les cafés scientifiques.

Manon bien sûr. On a tout partagé : des répétitions pour l'école doctorale, un travail, une amitié, un appart, des randonnées, du ski, de l'escalade, une rédaction de thèse, des blessures (qui ont un tantinet compliqué notre dernière de thèse), des soirées, des fous rires

et j'en passe. Nous allons bientôt être séparées, mais j'espère sincèrement que notre complicité va durer. Que puis-je te souhaiter ? Et bien, des voyages toujours plus loin, des sommets toujours plus, et un parcours professionnel toujours réussi.

Mes amis de là haut: ma kuik, Emilie, Pierre, Héléna, et Lucile bien évidemment ! Depuis le tout 1er jour de fac, jusqu'à aujourd'hui, la distance n'a rien changé à notre amitié. Je te souhaite le plus grand bonheur pour cet heureux événement qui va bientôt bouleverser ta vie.

Mes camarades de taekwondo qui m'ont permis d'en arriver là et avec qui ça a été un véritable plaisir de s'entrainer (Fred, Estelle, Thierry, Vianney, Caro, et tous les autres). Une pensée particulière pour Aurélie et Pierre, avec qui j'ai passé en plus quelques super mois de colocs. Et puis Pierre, tu sais que je te suis redevable à vie, tu m'as permis de découvrir les joies de Beauduc, du kitesurf, et c'est tout non ?

...Bien sûr que non. Benoît ! Je te suis tellement reconnaissante de partager ma vie depuis plus d'un an, et de tout ce que tu m'apportes. Cette passion à laquelle tu m'as initiée et que l'on partage nous rapproche un peu plus chaque jour. Ça n'a pas toujours été facile sur les derniers mois de thèse, j'ai reporté mon stress et mon manque de sport sur toi! Mais tu as toujours été derrière moi, tu m'as encouragée jusqu'à la derrière ligne droite. Merci pour tout.

Enfin, je pense sincèrement à ma famille. Je dédie tout particulièrement cette thèse à mon frère et à ma mère, qui ont toujours été là pour moi, et à qui je dois tout. Cette thèse m'a permis de me rapprocher de toi Frédéric, et de partager une vraie et belle complicité. Notre voyage et notre arrivée en haut du sommet de l'Afrique sont gravés à jamais dans ma mémoire. Maman, je ne serai jamais parvenue jusqu'ici si tu n'avais pas été là pour nous, si tu n'avais pas tout sacrifié pour nous permettre d'avoir une vie merveilleuse.



La vie ce n'est pas d'attendre que l'orage passe, c'est d'apprendre comment danser sous la pluie.

Sénèque

Table des matières

Liste des abréviations	14
Avant-propos	16
Introduction générale	18
I. Mécanismes de virulence et régulation chez Agrobacterium	20
I.1 Généralités sur <i>Agrobacterium</i>	20
I.1.1 Ecologie	20
I.I.2 Taxonomie	20
I.I.3 Particularités de la souche C58	22
I.2 Mécanismes moléculaires du cycle infectieux	22
I.2.1 Les étapes du cycle infectieux	
I.2.1.1 De la perception de la blessure à l'attachement aux cellules végétales	22
I.2.1.2 Le transfert de l'ADN-T et son intégration dans le génome végétal	24
I.2.1.3 Expression des gènes du brin-T dans la plante et formation de tumeurs	25
I.2.1.4 Création d'une niche écologique particulière pour Agrobacterium	26
I.2.1.5 Transfert conjugatif du plasmide Ti	
I.2.2 Mécanismes de régulation	27
I.2.2.1 Activation de la virulence	27
I.2.2.2 Assimilation des opines	29
I.2.2.3 Conjugaison du plasmide Ti	29
I.2.3 Organisation génétique des gènes de virulence et d'utilisation des opines	
I.2.3.1 Description du plasmide Ti	
I.2.3.2 Rôle accessoire du plasmide At dans la virulence d'Agrobacterium	
I.3 Le plasmide Ti : entre avantages sélectifs et fardeau génétique	33
I.3.1 Avantages sélectifs du plasmide Ti	
I.3.1.1 Utilisation des opines	
I.3.1.2 Utilisation des composés phénoliques	
I.3.2 Fardeau génétique	
I.3.2.1 Le maintien du plasmide Ti peut être coûteux selon l'environnement	
I.3.2.2 et les saisons	
I.3.2.3 Comment expliquer le coût du plasmide Ti ?	
I.3.2.4 Plasmide Ti et « tricherie »	
I.3.3 Mécanismes de maintien	
I.3.3.1 Maitien par la réplication	
I.3.3.2 Maintien par des systèmes d'addiction	
I.3.3.3 Maintien par la conjugaison	

I.3.3.4 Régulation fine et ordonnée de l'expression des gènes	40
II. Régulation par les riborégulateurs et physiologie bactérienne	42
II.1 Généralités	
II.1.1 Structure et localisation	
II.1.1.1 Plusieurs caractéristiques	42
II.1.1.2 Différentes catégories de riborégulateurs	43
II.1.2 Mécanismes d'action	43
II.1.2.1 Action sur protéines	43
II.1.2.2 Cas des riboswitchs et des thermorégulateurs	44
II.1.2.3 Action des riborégulateurs <i>cis</i> et <i>trans</i>	45
II.1.3 La protéine Hfq	
II.1.3.1 Mécanismes d'action	49
II.1.3.2 Hfq est-elle indispensable ?	49
II.1.3.3 Quantité de Hfq dans la cellule	50
II.2 Identification des riborégulateurs et de leur cibles	51
II.2.1 Identification des riborégulateurs	51
II.2.1.1 Identification des riborégulateurs par prédiction bioinformatique	51
II.2.1.2 Identification des riborégulateurs par des méthodes expérimentales	53
II.2.1.3 Combinaisons des deux approches	55
II.2.2 Identification des cibles des riborégulateurs	58
II.2.2.1 Identification des cibles sans <i>a priori</i>	58
II.2.2.2 Identification des cibles <i>a priori</i>	60
II.2.3 Rôle de Hfq dans l'identification des riborégulateurs et de leurs cibles	61
II.3 Fonctions cellulaires et physiologiques des riborégulateurs	61
II.3.I Rôle dans le maintien des plasmides	62
II.3.1.1 Rôle des riborégulateurs cis dans la réplication des plasmides	62
II.3.1.3 Rôle dans les systèmes toxine-antitoxine	63
II.3.1.4 Rôle dans la conjugaison	64
II.3.2 Rôle dans la transition entre deux conditions environnementales	64
II.3.2.1 Changement de température	65
II.3.2.2 Homéostasie du fer	66
II.3.2.3 Milieux carencés en nutriments	67
II.3.2.4 Densité bactérienne	68
II.3.3 Rôle dans l'interaction avec un hôte végétal	69
II.3.3.1 Riborégulateurs et symbiose chez Sinorhizobium meliloti	69
II.3.3.2 Riborégulateurs et virulence chez <i>Erwinia amylovora</i>	70
II.3.3.3 Riborégulateurs et virulence chez <i>Xanthomonas</i>	70

III. Rôle des riborégulateurs chez Agrobacterium	73
III.1 Analyses globales	
III.2 Etudes fonctionnelles	73
III.2.1 Thermorégulateurs	73
III.2.2 Hfq	74
III.2.3 AbcR1	74
III.2.4 RepE	75
Résultats	76
IV. Identification des riborégulateurs et rôle dans le cycle infectieux ch	nez A.
fabrum C58	78
IV.1 Etude globale des riborégulateurs chez <i>A. fabrum</i> C58	
IV.2 Riborégulateurs et rôle dans le cycle infectieux	
IV.2.1 Riborégulateurs chromosomiques	
IV.2.2 Riborégulateurs du plasmide Ti	
IV.2.2.1 Riborégulateurs antisens et UTR	
IV.2.2.2 Caractérisation phénotypique de deux candidats antisens	
IV.2.2.3 Riborégulateurs <i>trans</i>	
IV.2.2.4 Caractérisation de RNA1083 et impact sur la virulence	
V. Rôle de deux riborégulateurs dans le maintien du plasmide Ti	
V.1 RNA1059 et RepE et la réplication du plasmide Ti	
V.2 Rôle potentiel de RNA1051 en tant qu'antitoxine d'un nouveau système	e toxine-
antitoxine du plasmide Ti	139
Discussion et perspectives	158
VI. Discussion générale et perspectives	160
VI.1 Mise au point des approches	160
VI.1.1 Séquençage des transcrits de petite taille	160
VI.1.2 Caractérisation phénotypique des riborégulateurs candidats	161
VI.1.3 Identification des cibles ARNm des riborégulateurs candidats	
VI.2 Rôles physiologiques des riborégulateurs candidats	162
VI.2.1 Régulation de la virulence de la bactérie	
VI.2.2 Rôle dans le maintien du plasmide Ti	165
VI.2.2.1 Contrôle de la réplication	
VI.2.2.2 Rôle potentiel dans un système toxine-antitoxine	
VI.2.2.3 Rôle dans la conjugaison	
VI.2.2.4 Coordination de l'expression des composants du plasmide Ti	

VI.3 Apport de notre étude dans la compréhension des mécanismes de virulence chez
Agrobacterium
Références bibliographiques172
Annexes

Liste des abréviations

°C : degré Celcius σ^{32} : facteur Sigma 32 σ^{70} : facteur Sigma 70 5'UTR : 5' UnTranslated Region A : Adénosine Aa : acide aminé ADN : Acide DésoxyriboNucléique ADNc : ADN complémentaire ADN-T : ADN de transfert AIA : Acide Indole 3-Acétique ARN : Acide RiboNucléique ARNnc : ARN non codant ARNr : ARN ribosomial ARNt : ARN de transfert C: Cytosin **CDS** : CoDing Sequence Cm² : centimètre carré dRNAseq : differential RNA sequencing EDTA : Ethylène Diamine Tétra Acétique G : Guanine GABA : Gamma Aminobutyric Acid GBL : Gamma-ButyroLactone **GFP** : Green Fluorescent Protein Hfq : Host Factor $Q\beta$ HSL : HomoSérine Lactone IGR : InterGenic Sequence ILA : Indole 3-Lactate Kb : Kilobase LB : Luria Bertani LPG : Levure Peptone Glucose (YPG : Yeast Peptone Glucose) Mb : Mégabase

MLSA : MultiLocus Sequence Analysis Mn²⁺ : Manganèse ncRNA : non-coding RNA NLS : Nuclear Localisation Site Nt : nucléotide OC8HSL: Oxo-C8-HomoSérine Lactone PAP : Poly(A)-polymérase Pb : paire de bases PCR : Polymerase Chain Reaction PNP : PolyNucléotide Phosphorylase pTiC58 : plasmide Ti de la souche C58 Pu : Purine Py: Pyrimidine RACE-PCR : Rapid Amplification of cDNA ends PCR **RBS** : Ribosome Binding Site **RFLP** : Restriction Lenght Polymorphysm Rfam : RNA family **RNAP** : **RNA** Polymerase RNAseq : RNA sequencing RPKM : Reads Per Kilobase of transcript per Million reads mapped **ROSE** : Repression Of heat Shock expression RT-qPCR : Reverse Transcription quantitative PCR SD : Shine Dalgarno SELEX : Systematic Evolution of Ligands by EXponential enrichment sRNA : small RNA sRNA-seq : small RNA sequencing T: Thymine TA : Toxine-Antitoxine Ti: Tumor inducing **UFC : Unité Formant Colonies** Xcc : Xanthomonas campestris pathovar campestris Xcv : Xanthomonas campestris pathovar vesicatoria

Avant-propos

L'une des caractéristiques majeures des micro-organismes et donc des bactéries, est qu'ils sont en contact direct avec leur environnement. Or cet environnement est sujet à de constantes variations – physiques, chimiques ou biotiques. Le changement d'un paramètre peut être rapide et pour faire face, la bactérie doit être capable de percevoir les modifications de l'environnement et de mettre en place rapidement une réponse qui permettra à la cellule de survivre et de se multiplier dans le nouvel environnement. Pour cela, elle doit contrôler l'ensemble de ses composants cellulaires et moduler efficacement leurs productions.

Plusieurs niveaux de contrôle existent pour permettre un ajustement constant des composants cellulaires : la régulation transcriptionnelle, la régulation post-transcriptionnelle et les régulations traductionnelle ou post-traductionnelle. La régulation transcriptionnelle a, jusqu'à présent, été la plus largement étudiée. Parmi les différents mécanismes existants, se trouvent les systèmes à deux composants qui permettent la détection des changements environnementaux par des systèmes de perception et la transmission du signal dans la cellule suivie d'une réponse via une régulation ajustée de l'expression des gènes.

De plus en plus, il semble que des mécanismes de régulation post-transcriptionnelle soient aussi particulièrement bien adaptés pour la réponse bactérienne aux changements environnementaux. Ils constitueraient ainsi un degré supplémentaire de contrôle de l'expression des gènes puisqu'ils exercent une action sur les protéines ou sur leur traduction. Parmi les mécanismes existe la régulation par les ARN non codant ou riborégulateurs. Ne nécessitant pas le recrutement de la machinerie de traduction cellulaire et étant de petite taille, leur production par la cellule est économique et rapide, de même que leur dégradation. Ainsi, des modélisations suggèrent une implication majeure de ces riborégulateurs dans la régulation cellulaire face à un changement de l'environnement. Les résultats prédictifs sont soutenus par des exemples de rôles physiologiques de certains riborégulateurs dans l'acclimatation des bactéries à un nouvel environnement

Agrobacterium est une bactérie commune des sols, pouvant interagir avec les plantes au travers d'interactions commensale ou pathogène. Elle est responsable de la maladie dite de la galle du collet qui se traduit par la formation de tumeurs chez les plantes. Tout au long du mécanisme infectieux, depuis la perception de la blessure d'une plante jusqu'à la dissémination du plasmide Ti, cette bactérie est exposée à de nombreux signaux physico-chimiques qui vont diriger la mise en place des différentes étapes de la virulence.

Un système à deux composants VirA/VirG a depuis longtemps été identifié comme le régulateur principal de la virulence et aucun riborégulateur n'avait jusqu'à présent été impliqué dans la mise en

place de la virulence par *Agrobacterium*. Ainsi, nous nous sommes demandés quels sont les riborégulateurs présents chez *Agrobacterium* ? Parmi ceux-ci certains sont-ils impliqués dans la mise en place de la virulence ? Quels sont les mécanismes en jeu de cette potentielle modulation de la virulence?

Mon travail de thèse s'est attaché à répondre à ces questionnements avec l'identification et l'étude des riborégulateurs d'*Agrobacterium fabrum* C58 (anciennement *A. tumefaciens* C58), la recherche de leur implication dans la virulence et l'étude des mécanismes au travers desquels ils modulent la virulence.

Ainsi, la première partie de ce manuscrit correspond à une étude bibliographique détaillant les connaissances actuelles i) sur la virulence d'*Agrobacterium* et son agent majeur le plasmide Ti, sur ii) le fonctionnement et le rôle des riborégulateurs dans la physiologie bactérienne, et iii) et plus précisément dans celle d'*Agrobacterium*.

Dans une seconde partie sont présentés les résultats obtenus au cours de cette thèse, répartis en deux chapitres. Le premier s'intéresse à l'identification globale de l'ensemble des riborégulateurs potentiels exprimés chez *A. fabrum* C58, ceci ayant été réalisé par séquençage des transcrits de petite taille et par prédiction bioinformatique. La caractérisation de candidats modulant potentiellement la virulence est également étudiée. Le second chapitre est consacré à l'étude du rôle de deux riborégulateurs dans le maintien du plasmide Ti dans la cellule, par un mécanisme de contrôle de la réplication, et par un mécanisme impliquant potentiellement un système toxine-antitoxine.

Enfin, dans une dernière partie, nous discutons de l'apport et de la portée de nos travaux.

Introduction générale

I. Mécanismes de virulence et régulation chez Agrobacterium

I.1 Généralités sur Agrobacterium

I.1.1 Ecologie

Agrobacterium est une bactérie commune du sol et de la rhizosphère qui peut également être phytopathogène. Elle est retrouvée quasiment systématiquement dans chaque sol étudié à une densité de 10³ à 10⁷ UFC (Unité Formant Colonie) par gramme de sol (Bouzar et al., 1993; Mougel et al., 2001; New and Kerr, 1972; Puławska and Sobiczewski, 2005; Schroth et al., 1971). Bien qu'*Agrobacterium* soit capable de vivre dans le sol nu, milieu oligotrophe, elle est préférentiellement retrouvée au contact des racines, milieu copiotrophe. Mougel *et al* ont ainsi montré que la densité d'agrobactéries était 100 fois plus importante dans la rhizosphère du maïs et du lotier par rapport à un sol nu (Mougel, 2000). Cette bactérie est cependant plus connue comme agent pathogène des dicotylédones et de quelques monocotylédones, chez qui elle est responsable de la maladie dite de la galle du collet dont les symptômes se traduisent par la formation de tumeurs.

I.I.2 Taxonomie

Chez cette α-protéobactérie de la famille des *Rhizobiaceae*, les premières classifications se sont appuyées sur le caractère de virulence. Trois groupes étaient alors décrits : *Agrobacterium tumefaciens*, responsable de la galle du collet ; *Agrobacterium rhizogenes*, responsable du chevelu racinaire se traduisant par une importante augmentation de la taille des racines secondaires, et enfin *Agrobacterium radiobacter*, avirulent (Conn, 1942; Smith and Townsend, 1907). Cependant, cette classification s'est révélée erronée, puisque nous savons aujourd'hui que le pouvoir pathogène des agrobactéries réside en la présence d'un plasmide de virulence qui est un réplicon accessoire : le plasmide Ti (pour *Tumor inducing*) pour *A. tumefaciens*, et le plasmide Ri (pour *Root inducing*) pour *A. rhizogenes*. La taxonomie s'est alors basée sur des caractères biochimiques dépendant des gènes chromosomiques uniquement (Sneath and Sokal, 1962), et a permis de regrouper les agrobactéries en trois groupes : les biovars 1, 2 et 3 (Keane et al., 1970; Kersters et al., 1973; Panagopoulos and Psallidas, 1973; Popoff et al., 1984). Finalement, la taxonomie moléculaire, fondée sur la similarité des génomes (hybridation ADN/ADN, RFLP, MLSA), a permis de valider et d'affiner cette taxonomie biochimique (Figure 1).



Figure 1: Arbre phylogénétique du genre Agrobacterium. (D'après Costechareyre et al, 2010)

L'arbre phylogénétique a été déterminé par la méthode de *neighbor-joining* et est basé sur la séquence du gène *recA* (969 paires de bases). L'unité de longueur des branches correspond au nombre de substitutions par site nucléotidique. Les nombres G1 à G9 correspondent aux génomovars du complexe d'espèces *Agrobacterium tumefaciens*. La souche de référence C58 est annotée par une étoile rouge.

Le biovar 1 correspond au complexe d'espèces *A. tumefaciens*, et les biovars 2 et 3 correspondent respectivement aux espèces *Rhizobium rhizogenes* et *Agrobacterium vitis* (Costechareyre et al., 2010; Lindström and Young, 2011). Pour le biovar 1, on parle de complexe d'espèces car ce groupe est constitué en réalité de 11 espèces potentielles appelées génomovars. Chaque génomovar (G1-G9, G13 et G14) possède des caractères spécifiques qui font d'eux des espèces à part entière, et certains génomovars ont récemment été élevés au rang d'espèce : la dénomination *A. pusense* a été attribuée au génomovar G2 (Panday et al., 2011), le génomovar G4 a été renommé *A. radiobacter* (Costechareyre et al., 2010), et enfin le génomovar G8 s'appelle maintenant *A. fabrum* (Lassalle et al., 2011). Campillo *et al*, ont ainsi montré que le génome d'*A. fabrum* contenait des gènes permettant la détoxication et l'assimilation de l'acide férulique, un composé phénolique libéré à partir de la paroi végétale suite à une blessure de la plante (Campillo et al., 2014). Ces gènes et cette fonction ne sont retrouvés que chez cette espèce, ce qui pourrait leur conférer un avantage sélectif important par rapport aux autres bactéries présentes avec elles. Des méthodes rapides et efficaces,

basées sur des caractéristiques génétiques (PCR avec des amorces spécifiques) et enzymatiques (coloration du milieu lorsque l'acide caféique est dégradé), permettent la détection et l'identification d'espèces d'*Agrobacterium* à partir de cultures pures ou d'échantillons complexes de l'environnement (Shams et al., 2013).

I.I.3 Particularités de la souche C58

La première séquence génomique d'*Agrobacterium* a été obtenue en 2001 par deux groupes simultanément et correspond à la souche de référence C58 appartenant à l'espèce d'*A. fabrum*, (Goodner et al., 2001; Wood et al., 2001). Son génome est composé de 4 réplicons. Le premier est un chromosome circulaire de 2,84 Mb composé de 2930 gènes, et le second est un chromosome linéaire de 2,08 Mb portant 1954 gènes. Ce second chromosome linéaire est spécifique des agrobactéries (Jumas-Bilak et al., 1998; Slater et al., 2009; Ramírez-Bahena et al., 2014). Les deux autres réplicons sont quant à eux des plasmides accessoires : le plasmide At mesure 543 kb et porte 620 gènes, et le plasmide Ti (pour *Tumor inducing*) mesure 214 kb et porte 234 gènes. Le plasmide Ti est l'agent majeur responsable de la virulence d'*A. tumefaciens*, puisqu'il porte tous les déterminants de la virulence (Figure 5 page 33).

I.2 Mécanismes moléculaires du cycle infectieux

Suite à la blessure d'une plante par un facteur environnemental (vent, animaux) ou une action humaine (bouturage, récolte), *Agrobacterium* met en place des mécanismes conduisant à l'expression du pouvoir pathogène. Celui-ci se caractérise par une capacité unique de transfert de l'information génétique bactérienne directement dans le génome végétal, conduisant à la formation de tumeurs. Cette capacité à former des tumeurs est employée pour déterminer le potentiel infectieux d'une souche. Dans un 1^{er} temps, la capacité de tumorégénicité, c'est-à-dire la capacité de former des tumeurs était utilisée (Vogel and Das, 1992). Aujourd'hui, on s'intéresse à l'aggressivité d'une souche, défini par le nombre moyen de tumeurs observé par plante (Planamente et al., 2012).

I.2.1 Les étapes du cycle infectieux

I.2.1.1 De la perception de la blessure à l'attachement aux cellules végétales

La plante blessée libère des signaux chimiques dans la rhizosphère qui attirent les agrobactéries (Baron and Zambryski, 1995). Une libération de composés issus de la paroi végétale tels que des composés phénoliques comme l'acétosyringone, ou encore des monosaccharides, est



Figure 2 : Etapes de la virulence chez *A. fabrum* C58.

Les composés émis de la blessure de la plante sont perçus par la bactérie (1) qui migre par chimiotactisme vers la blessure (2), puis adhère aux cellules végétales (3). Les gènes de virulence sont ensuite activés et permettent l'excision (4) et le transfert (5) de l'ADN-T directement dans la cellule végétale. L'ADN-T est intégré dans le génome de la plante et ses gènes sont exprimés (6). Ils permettent la production de phytohormones (7) responsables de la formation de tumeurs ; ainsi que la synthèse des opines (8). Ces composés sont assimilés par *Agrobacterium* (9). Les opines de type agrocinopine induisent également l'expression des gènes de conjugaison *tra/trb* et le transfert conjugatif du plasmide Ti (10).

observée (Figure 2 étape 1). De même, des composés pariétaux qui sont pour la plupart des composés acides, vont également entraîner une diminution du pH au niveau de la blessure. Ces différents signaux sont perçus par les agrobactéries qui migrent jusqu'à la zone de blessure par chimiotactisme. Plusieurs gènes sont impliqués : les gènes che (pour Chromosome Encoded) (cheY1, cheY2, cheA, cheB, cheR) et mcp (pour Methyl-accepting Chemotaxis Protein) pour le chimiotactisme ; les gènes fla, fli, flh et flg pour la biosynthèse des flagelles ; et les gènes mot permettant la rotation des flagelles (Chesnokova et al., 1997; Deakin, 1994; Deakin et al., 1999, 1997a, 1997b) (Figure 2 étape 2). Des études ont montré que le chimiotactisme n'était pas nécessaire pour la virulence lorsque les bactéries étaient inoculées directement sur la blessure, mais qu'ils le devenaient lorsque la bactérie était inoculée dans certains sols. Agrobacterium adhère ensuite aux cellules végétales, tout d'abord de manière réversible grâce à la synthèse de polysaccharides acétylés, puis de manière spécifique et durable à l'aide de fibres de cellulose (Matthysse et al, 2014) (Figure 2 étape 3). L'interaction spécifique a lieu grâce à l'expression des gènes chvA, chvB, pscA et exo (Cangelosi et al., 1989; de lannino and Ugalde, 1989; Thomashow et al., 1987; Uttaro et al., 1990) dont les produits sont impliqués dans la biosynthèse et la localisation périplasmique du β 1-2 glucane et dans la biosynthèse d'exopolysaccharides. Par ailleurs, les gènes cel, impliqués dans la biosynthèse de cellulose, semblent importants pour l'attachement des bactéries à la plante (Matthysse, 2014).

I.2.1.2 Le transfert de l'ADN-T et son intégration dans le génome végétal

Une partie du plasmide Ti, appelée ADN-T (pour ADN de transfert) est transféré et intégré dans le génome végétal à l'aide des produits d'un groupe de 25 gènes, les gènes *vir*, dont l'expression est activée suite à la blessure de la plante (figure 2 étape 4). Ces gènes se regroupent en 6 opérons : *virA*, *virB*, *virC*, *virD*, *virE* et *virG*. Les protéines VirC1, VirC2, VirD1 et VirD2 sont responsables de la maturation de l'ADN-T. VirC1 et VirC2 favorisent la formation du relaxosome (Atmakuri et al., 2007) tandis que les protéines VirD1 et VirD2 se fixent au niveau de séquences répétées de 25 nucléotides présentes de part et d'autre de l'ADN-T (appelées *left* et *right border*). Elles clivent et libèrent un ADN simple brin correspondant au brin-T (Padavannil et al., 2014). A l'issue du clivage, VirD2 reste liée de manière covalente au brin-T, formant le complexe-T. Plusieurs unités de VirE2 recouvrent le brin-T, lui assurant une protection (Grange et al., 2008). Ce complexe-T est amené, grâce à la protéine VirC1, vers les pôles de la cellule, puis est transféré dans la cellule végétale. Pour cela, un complexe de transfert, nommé VirB, est mis en place et permettre également le transfert de nombreux effecteurs au travers des membranes et des parois bactériennes et végétales (Figure 2 étape 5). VirB est un système de sécrétion de type IV composé des 12 protéines VirB1-11 et VirD4 (Christie, 2004; Christie et al., 2014). Les protéines VirD4, VirB3, VirB4 et ViB11 sont des ATPases

situées dans la membrane interne de la cellule. Elles fournissent l'énergie nécessaire au transfert du conplexe-T. Les protéines VirB6 à VirB10 s'assemblent dans l'espace périplasmique pour former le cœur du système de sécrétion. Le pilus extracellulaire permettant le passage dans la cellule hôte est constituée des protéines VirB1, VirB2, VirB5 et VirB7 (Subramoni et al., 2014).

De part la présence d'un signal de localisation nucléaire (NLS) dans leurs séquences protéiques, VirE2 et VirD2, associées aux protéines végétales VIP1 et VIP2 (pour VirE2-*Interacting Protein*), interagissent avec des importines α et des cyclophilines qui permettent l'adressage du complexe-T jusqu'au noyau (Figure 2 étape 6). Dans le noyau, le brin-T est relaché du complexe-T. Les protéines VirF et VirD5 vont stimuler la dégradation des protéines le recouvrant par le protéasome (Magori and Citovsky, 2011). Le brin-T est ensuite intégré dans le génome végétal, *a priori* à l'aide de VirE3 (García-Rodríguez et al., 2006) ; de manière aléatoire par recombinaison non homologue. Quelques rares cas d'intégration ciblée par recombinaison homologue ont été observés et deux mécanismes d'intégration ont été proposés. Le premier modèle, basé sur l'intégration du brin-T par microhomologie, suggère que le brin-T serait hybridé au niveau du site de préintégration, l'ADN cible serait clivé, et l'extrémité 5' du brin-T serait ligaturé à l'ADN cible. Le second modèle propose que le brin-T serait converti en une molécule double brin intégrée dans le génome après cassure double brin (Tzfira et al., 2004; Liang and Tzfira, 2013).

I.2.1.3 Expression des gènes du brin-T dans la plante et formation de tumeurs

Dans les cellules végétales transformées, les gènes présents sur le brin-T sont exprimés. Ils possèdent tous les éléments nécessaires à leur expression chez une cellule eucaryote (boîte TATA, boîte CAAT, signal de polyadénylation). L'expression de 5 gènes, appelé oncogènes, perturbe les concentrations d'auxines et de cytokinines au sein des cellules végétales transformées ainsi que la sensibilité de ces cellules à ces hormones. Ceci conduit à une dé-différenciation des cellules végétales, à la prolifération cellulaire et à la formation de tumeurs (Veselov et al., 2003) (Figure 2 étape 7). Le gène *iaaM* (ou *tms1*) code une tryptophane mono-oxygénase qui permet la conversion du tryptophane en indole-3-acétamide. Ce dernier composé est converti en acide indole 3-acétique (AIA) par une indoleacétamide hydrolase, codée par le gène *iaaH* (ou *tms2*). L'AIA est l'une des phytohormones de type auxine la plus répandue chez les plantes. Elle joue un rôle essentiel dans de nombreux processus du développement de la plante, tels que l'organogenèse, la dominance apicale, la croissance des graines ou le développement des fruits et des fleurs (Zhao, 2010). Le gène *ipt* code une isopentényltransférase qui catalyse la condensation d'adénosine-5'-monophosphate et isopenténylpyrophosphate en isopentényle-adésonine-5'-monophosphate (Akiyoshi et al., 1984). Cette étape contrôle la vitesse de production des cytokines, qui participent, comme les auxines, à de

nombreux processus de développement de la plante (Werner and Schmülling, 2009). Le produit du gène *6b* semble avoir un rôle important et complexe dans le développement des tumeurs. Bien qu'il ne soit pas indispensable à la virulence, son expression seule est suffisante pour produire des tumeurs chez un certain nombre de plantes telles que le tabac ou le kalanchoé (Hooykaas et al., 1988). Il possède de multiples fonctions et actions dans la cellule végétale : i) perturbations du transport polaire de l'auxine dans des feuilles de tabac (Kakiuchi et al., 2006), ii) accumulation de sucres et de composés phénoliques (Clément et al., 2007), iii) interaction avec des protéines nucléaires et modification de l'expression de certains gènes (Kitakura et al., 2002; Terakura et al., 2007) et iv) perturbation de la voie des microARN (Wang et al., 2011). Enfin, le gène *5* code une enzyme catalysant la conversion du tryptophane en indole-3-lactate (ILA) qui pourrait entrer en compétition avec l'AIA (Körber et al., 1991). Ceci diminuerait la sensibilité à l'auxine et permettrait d'augmenter le ratio cytokonine/auxine et donc d'augmenter la croissance des tiges infectées (Britton et al., 2008).

I.2.1.4 Création d'une niche écologique particulière pour Agrobacterium

Une deuxième catégorie de gènes portés par le brin-T est impliquée dans la synthèse de molécules spécifiquement produites à partir des tumeurs, appelées opines (Figure 2 étape 8). Il s'agit de molécules de faible poids moléculaire issues le plus souvent de la condensation d'acides aminés et de sucres. Ces composés vont être produits de manière intensive au sein des cellules transformées, et peuvent atteindre 7 % du poids sec des tumeurs (Kim et al., 2001). Plusieurs types d'opines peuvent être produits. Les gènes responsables de la production de ces composés sont : le gène ocs pour la synthèse d'octopine, de lysopine, d'histopine, d'acide octopinique et de sulfopine (Dessaux et al., 1998; Flores-Mireles et al., 2012); les gènes mas pour la synthèse de mannopine et d'acide mannopinique ; le gène ags pour l'agropine ; le gène nos pour la synthèse de nopaline et d'acide nopalinique ; et enfin le gène acs pour la synthèse d'agrocinopines A et B. Les opines ne sont pas toutes produites à partir d'un même plasmide Ti. A partir d'un même plasmide ne vont être produits que quelques types d'opines. Les plasmides Ti ont d'ailleurs été classés en fonction des opines qu'ils produisent, les principaux étant les plasmides à nopaline et les plasmides à octopines. Les plasmides Ti portent également les gènes responsables du transport dans la cellule bactérienne et du catabolisme de leurs opines spécifiques (Figure 2 étape 9). Il s'agit des gènes occ responsables du transport et du catabolisme de l'octopine, de l'acide octopinique et de la lysopine ; des gènes noc permettent d'utiliser la nopaline ; et enfin des gènes acc nécessaires à l'import et à la dégradation des agrocinopines.

I.2.1.5 Transfert conjugatif du plasmide Ti

Les opines stimulent également le transfert conjugatif du pTi d'une cellule bactérienne donneuse à une cellule bactérienne receveuse (Figure 2 étape 10). Les produits des gènes tra et trb sont nécessaires et suffisants pour le transfert (Cook et al., 1997). A partir de l'origine de transfert oriT, le pTi est clivé, dupliqué, puis transféré dans la cellule receveuse grâce à un système de sécrétion de type IV. L'excision au niveau de l'oriT se fait via la protéine TraA qui possède une activité hélicase et clive l'ADN. Son rôle est équivalent à celui de VirD2. Elle est accompagnée des protéines TraC et TraD qui possèdent entre autre une activité de fomation du relaxosome (Cho and Winans, 2007). Les protéines TraG et TraF sont similaires aux protéines du même nom du plasmide RP4. TraG forme un pont entre le relaxosome et le pore dans la membrane (Hamilton et al., 2000). TraF permet la maturation de la sous-unité du pore de conjugaison, la piline, grâce à une activité protéase et protéine cyclase (Eisenbrandt et al., 1999). TraB semble nécessaire pour un transfert conjugatif maximal mais n'est pas indispensable au procédé (Farrand et al., 1996). Les protéines Trb forment le pore de conjugaison du système de sécrétion de type IV (Li et al., 1999). Ce système de sécrétion est similaire à celui utilisé pour le transfert de l'ADN-T. La question s'est longtemps posée de savoir si les transferts du complexe-T et du plasmide Ti utilisaient un système commun. Il a cependant été montré que des souches déficientes dans le système tra étaient capables de transférer l'ADN-T, et à l'inverse des souches déficientes dans le système vir pouvaient disséminer le plasmide Ti par transfert conjugatif (Alt-Mörbe et al., 1996; Cook et al., 1997; von Bodman et al., 1989).

I.2.2 Mécanismes de régulation

L'ensemble des mécanismes décrits précédemment est sujet à une régulation complexe. De nombreux systèmes de régulation sont mis en place et permettent une expression optimale des gènes de virulence, d'assimilation des opines et de la conjugaison.

I.2.2.1 Activation de la virulence

Les signaux issus de blessures des plantes sont perçus via la protéine fixatrice de sucre ChvE (Hu et al., 2013), le régulateur ExoR et les systèmes à 2 composants VirA/VirG et ChvG/ChvI. Certains de ces systèmes régulateurs sont exprimés à un niveau basal et leurs expressions peuvent être stimulées par ces signaux.

Le système à deux composants VirA/VirG est composé du senseur VirA (S. Jin et al., 1990; Lin et al., 2014) et du régulateur de réponse VirG (S. G. Jin et al., 1990) (Figure 3). La perception des composés phénoliques comme l'acétorisyngone et le pH acide conduit à l'autophosphorylation de VirA puis au transfert de son phosphate à VirG. Celui-ci peut alors se fixer en amont de promoteurs au niveau de

séquences particulières appelées « boîtes *vir* ». Ces séquences correspondent aux bases consensus suivantes : 5'dPu(T/A)TDCAATTGHAAPy (H = A,C ou T, D = A,G, ou T, Pu = purine ; Py = pyrimidine) (Pazour and Das, 1990). Ces « boîtes *vir* » sont présentes notamment en amont des gènes de virulence *vir*. Ainsi, la fixation de VirG active l'expression des gènes de virulence. De plus, la transcription de *virG* peut être auto-induite par des composés phénoliques via son promoteur P1, et par un pH acide via son promoteur P2 indépendamment de l'activation par VirA/VirG (Mantis and Winans, 1992).

La protéine ChvE perçoit les sucres libérés de la blessure et collabore avec la sous-unité périplasmique de VirA pour transmettre l'information (He et al., 2009; McCullen and Binns 2006; Shimoda et al., 1993). Le régulateur ExoR, en collaboration avec le système à deux composantes ChvG/ChvI régule l'expression de gènes impliqués dans la mobilité, le chimiotactisme, la synthèse d'exopolysaccharide, la mise en place d'un système de sécrétion de type VI ainsi que dans la virulence (Heckel et al., 2014; Wu et al., 2012). En condition neutre, ExoR interagit avec ChvG et le sequestre, empêchant son action sur ChvI. En condition acide, ExoR est hydrolysé et libère ChvG qui peut alors agir sur ChvI et ainsi permettre l'expression de nombreux gènes. Tomlinsson *et al* ont aussi montré que ExoR pourrait inhiber l'expression des gènes indépendamment de ChvG/ChvI (Tomlinson et al., 2010).



Figure 3 : Régulation des gènes de virulence. (D'après McCullen et Binns, 2006 ; Winans, 1992).

La perception des composés phénoliques par VirA entraîne son autophosphorylation. La perception des sucres et du pH acide par ChvE permet également la phosphorylation de VirA. VirA phosphoryle alors VirG qui induit l'expression des gènes vir nécessaires à la virulence. L'étoile orange en amont de chaque opéron correspond au site de fixation de VirG, appelé « boîte vir » (5'dPu(T/A)TDCAATTGHAAPy (H = A,C ou T, D = A,G, ur T;Pu = purine ; Py = pyrimidine)(Pazour and Das 1990).

I.2.2.2 Assimilation des opines

Les gènes d'assimilation des opines *noc* (pour la nopaline) et *occ* (pour l'octopine) sont régulés respectivement par les régulateurs positifs NocR et OccR qui appartiennent la famille des régulateurs de type LysR (Kreusch et al., 1995; von Lintig et al., 1991). Ils se fixent sur les promoteurs de leurs gènes cibles. En présence de nopaline ou d'octopine, un changement dans la conformation du complexe NocR-ADN, ou OccR-ADN, se produit et permet l'expression des gènes (Marincs and White, 1993). L'assimilation des agrocinopines A et B est contrôlée négativement par le régulateur AccR de type LacR. En absence de ces opines, ce répresseur empêche l'expression des gènes de l'opéron *acc* en se fixant au niveau de l'opérateur (séquence répétée inversée CGCTCATAGTATGAGCG). La fixation des agrocinopines à AccR empêche son appariement avec les promoteurs, et permet la levée de la répression et l'assimilation des agrocinopines (Kim et al., 2008).

I.2.2.3 Conjugaison du plasmide Ti

Enfin, la régulation de la conjugaison du plasmide Ti est contrôlée par la présence de certaines opines et par le *quorum sensing* (Figure 4).

Régulation par le quorum sensing

Le quorum sensing permet un dialogue entre les bactéries et est dépendant de la densité cellulaire (Fuqua et al., 1994). Chez Agrobacterium, les protéines Tral et TraR sont impliquées dans ce mécanisme et sont similaires aux protéines de référence LuxI et LuxR. Tral est responsable de la synthèse de la molécule signal auto-inductrice : la 3-oxo-C8-homosérine lactone (OC8HSL). A forte densité bactérienne, l'OC8HSL est perçue par la protéine régulatrice TraR et interagit avec celle-ci. La présence d'OC8HSL permet notamment la stabilisation de la protéine TraR et diminue sa sensibilité aux protéases (Zhu and Winans, 2001, 1999). Le complexe TraR-OC8HSL se fixe spécifiquement sur l'ADN au niveau de « boîtes tra » (ATGTGCAGATCTGCACGT) et induit l'expression de nombreux gènes : les gènes tra et trb impliqués dans la conjugaison du plasmide Ti et dans sa régulation, et les gènes rep responsables de la réplication et de la répartition du plasmide Ti. Ce système est lui-même régulé par de nombreux facteurs dont la lactonase AiiB (Haudecoeur et al., 2009) et les produits de gènes blcABC qui seraient impliqués dans la dégradation des auto-inducteurs OC8HSL. BlcC est en effet une lactonase capable de cliver le noyau gamma-butyrolactones (GBL) des homosérines lactones (Carlier et al., 2004; Chevrot et al., 2006a). Lorsque l'expression de l'opéron blcABC est induite par des composés végétaux, tels que la molécule de défense végétale GABA, Agrobacterium n'accumule pas les OC8HSL, suggérant ainsi une régulation négative de la conjugaison du plasmide Ti. Cependant, Farrand et al ont montré que la production de BlcC in situ et in planta diminuait la concentration extracellulaire des homosérines lactones mais pas leur concentration intracellulaire ni le transfert conjugatif du plasmide Ti (Khan and Farrand, 2009). Le rôle de l'opéron *blcABC* dans l'inhibition du *quorum sensing* et du transfert conjugatif du plasmide Ti reste donc encore mal connu.



Figure 4 : Régulation du transfert conjugatif du plasmide Ti par le quorum sensing chez la souche A.

fabrum C58. (D'après Lang et al, 2014).

L'expression des gènes du brin-T au sein des cellules végétales permet la production des agrocinopines A et B (1) qui bloquent alors l'action du répresseur AccR et induisent alors la production de TraR (2). TraR accroît la production de Tral, permettant un rétro-contrôle positif de la production des molécules signal OC8HSL (\bigcirc) (3), qui sont relâchées dans le milieu extérieur (4). Plusieurs mécanismes contrecarrent le *quorum sensing* : tout d'abord la répression de TraR par TraM (5), puis la dégradation des OC8HSL (\bigcirc) par les lactonases BlcC et AiiB (6), dont la production est modulée respectivement par la production de GABA (\bigcirc) et d'agrocinopines. Les étoiles jaunes en amont de chaque opéron correspondent au site de fixation de TraR, appelé « boîte *tra* » (ATGTGCAGATCTGCACGT). GABA : γ -amino butyric acid; OC8HSL : 3-oxo-octanoylhomosérine lactone.

Régulation par les opines

L'expression du régulateur transcriptionnel TraR est également contrôlée par la présence des opines dites de conjugaison. En absence d'agrocinopine (ou d'octopine), le régulateur AccR (et OccR) inhibe l'expression de l'opéron au sein duquel *traR* se situe. En présence d'agrocinopine (ou d'octopine), l'inhibition par AccR (et OccR) est levée, permettant l'expression de l'opéron et la production de TraR. La présence d'opines permet aussi l'induction de l'expression du gène *traM* qui code un

inhibiteur du *quorum sensing*. TraM interagit avec le complexe TraR-OC8HSL, empêchant l'activation de l'expression des gènes sous sa dépendance. Il a été montré que des mutations délétaires de *traM* conduisaient à une réponse à de plus faibles concentrations de HSL et à un phénotype hyperconjugatif des souches, tandis que la surexpression de *traM* provoquait un phénotype opposé (Fuqua *et al*, 1995 ; Hwang *et al*, 1995). Chez les plasmides à octopine, la protéine TrIR permet également une inhibition de l'activité de TraR lorsque la mannopine est présente dans le milieu. Il a été montré que la conjugaison du pTi était diminuée en présence de cette opine, mais que des souches inactivées pour *trIR* n'en étaient pas affectées (Oger *et al*, 1998 ; Winans *et al*, 1998). A l'inverse, l'expression constitutive de ce gène provoque une diminution du transfert conjugatif du plasmide Ti (Chai *et al*, 2001).

Enfin, la protéine chaperon GroESL, exprimée à partir du chromosome circulaire, stabilise TraR en augmentant sa solubilité et l'activité transcriptionnelle des promoteurs sous sa dépendance (Chai and Winans, 2009).

I.2.3 Organisation génétique des gènes de virulence et d'utilisation des opines

La totalité des gènes nécessaires à la virulence d'*Agrobacterium* est portée par le plasmide Ti, ce qui fait de lui l'acteur principal de la virulence. Cependant, les protéines impliquées dans les étapes précoces de la virulence permettant la migration et l'attachement des bactéries aux cellules végétales sont codées par le chromosome circulaire, de même que les régulateurs ChvE et ChvG/ChvI.

I.2.3.1 Description du plasmide Ti

En fonction des plasmides, différentes opines sont produites et assimilées. Les plasmides à octopine possèdent ainsi des gènes *occ* pour le catabolisme des octopines, à la place des gènes *noc*. De plus chez les plasmides à octopine, les gènes *acc* ne sont pas présents. Une même cellule ne peut porter les deux plasmides. Les plasmides Ti appartiennent au même groupe d'incompatibilité, qui correspond à un ensemble de plasmides possédant le même système de réplication, et donc incapable de coexister dans une même cellule. Hooykas *et al*, (1980) ont en effet montré que les plasmides à octopine et à nopaline ne pouvaient pas persister dans la même cellule, ou bien que leur cohabitation nécessitait une cointegration.

Sur les plasmides à nopaline, dont le pTiC58, ces gènes de virulence représentent près de 54 % de la séquence et sont regroupés en régions en fonction de leur rôle (figure 5). On distingue sept régions : l'ADN-T de 2,2240 kb ; les gènes *noc* du catabolisme de la nopaline répartis sur 17,979 kb ;

une première région impliquée dans le transfert conjugatif du plasmide Ti de 10,299 kb codant les protéines Trb et le régulateur Tral ; la région de réplication contenant l'origine de réplication et l'opéron *repABC* mesurant 3,991 kb; la seconde région de conjugaison de 12,294 kb, exprimant les gènes *tra* et contenant l'origine de transfert *oriT*; les gènes *acc* pour le catabolisme des agrocinopines de 7,817 kb ; et enfin la région *vir* de 42,164 kb comprenant les gènes de virulence. Ces régions sont séparées les unes des autres par des régions plus ou moins grandes décrites comme non impliquées dans la virulence. Les gènes présents dans ces régions codent des ABC transporteurs, des monooxygénases, et majoritairement des protéines hypothétiques de fonction inconnue.



Figure 5 : Représentation schématique à l'échelle du plasmide TiC58.

Les régions colorées correspondent aux gènes impliqués dans la virulence ou l'interaction avec la plante. Les régions en gris n'ont pas été démontrées comme impliquées dans la virulence. Elles portent des gènes codant des protéines hypothétiques, des protéines de fonctions diverses, ou ne sont pas codantes. *oriT* : origine de transfert ; *oriV* : origine de réplication. Le cercle intérieur (en bleu clair) correspond à la position des séquences codantes tout au long de plasmide Ti.

I.2.3.2 Rôle accessoire du plasmide At dans la virulence d'Agrobacterium

Il semble que le plasmide At accentuerait la virulence d'*Agrobacterium*. Des souches d'*A. fabrum* C58 dépourvues de plasmide At sont capables de former des tumeurs sur plantes mais les tumeurs induites par ces souches sont plus petites que celles obtenues suite à l'inoculation de souches portant le plasmide At (Figure 6) (Hooykaas et al., 1977; Hynes et al., 1985; Nair et al., 2003; Ogawa et al., 2000; Ogawa and Mii, 2001; Rosenberg and Huguet, 1984).



Figure 6 : Impact du plasmide At dans la formation de tumeurs chez d'*A. tumefaciens.* (D'après Nair *et al*, 2003)

Des souches avec ou sans plasmides (At et Ti) ont été testées pour leur capacité à induire la formation de tumeurs. Des inocula à différentes concentrations ont été déposés sur des feuilles de *Kalanchoe diagremontiana* blessées. Les tumeurs ont été dénombrées 3 semaines après infection.

Ce plasmide ne porte pas de gènes essentiels à la virulence. Cependant les gènes attA2ABCDEFGH pourraient conférer à la bactérie une meilleure adhérence sur la plante, facilitant l'infection (Matthysse et al., 2000; Nair et al., 2003). Nair et al ont aussi montré qu'en présence du plasmide At et en conditions d'induction, les gènes vir étaient exprimés de manière plus importante qu'en absence de plasmide At chez les souches AB150-AB153. De plus, les gènes blcABC, potentiellement impliqués dans la régulation du quorum sensing, sont portés pas le plasmide At. Outre la régulation de la conjugaison du plasmide Ti, les HSL pourraient être perçues par la plante et stimuler ses défenses (Chevrot et al., 2006b). De même, l'activation de BlcC par le GABA, suivie de la diminution de la concentration extracellulaire des HSL suggère que cette lactonase empêche la stimulation des réactions de défense de la plante, notamment dans les étapes précoces et délicates de l'infection telles que le transfert de l'ADN-T. Par ailleurs, BlcC est capable de dégrader une large gamme de HSL, et pas uniquement celles produites par Agrobacterium. Elle pourrait donc dégrader les molécules signal d'autres bactéries, perturbant ainsi leur quorum sensing (on parle alors de quorum quenching), et limiter la compétition au niveau de la zone de blessure. Enfin, les gènes blcA, blcB et blcC codent des enzymes de dégradation du GBL produit par la plante. Ce composé peut alors être assimilé par Agrobacterium, lui fournissant une source de carbone supplémentaire (Carlier et Faure, 2004).

I.3 Le plasmide Ti : entre avantages sélectifs et fardeau génétique

Comme la majorité des plasmides, le plasmide Ti confère un avantage à la bactérie dans certaines conditions. Cependant, sa grande taille et les fonctions qu'ils codent peuvent représenter un fardeau pour la bactérie. Plusieurs mécanismes lui permettent alors de diminuer son coût.

I.3.1 Avantages sélectifs du plasmide Ti

I.3.1.1 Utilisation des opines

L'utilisation des opines par les agrobactéries portant un plasmide Ti particulier définit la tumeur comme une véritable niche écologique pour ces souches. Les opines constituent un élément nutritif presque exclusivement pour les bactéries hébergeant le plasmide Ti correspondant. De plus, la stimulation du transfert conjugatif du plasmide Ti par les opines permet la dissémination de ce plasmide, et donc de la virulence au sein des populations d'agrobactéries présentes dans les tumeurs. C'est ce qu'on appelle le concept d'opine, qui a défini les opines comme « des molécules de petite taille dont la présence dans les tumeurs de la galle du collet est déclenchée par A. tumefaciens et qui facilitent sa multiplication et la dissémination des déterminants de sa virulence » (Guyon et al., 1980; Petit et al., 1983). Plusieurs groupes ont montré que la présence d'opines apportait un avantage trophique aux bactéries portant un plasmide Ti. Lors d'essai in planta, Guyon et al ont observé que des plantes produisant des opines stimulaient la croissance d'agrobacteries capables d'assimiler ces molécules (Guyon et al., 1993). Savka et Farrand ont réalisé des tests de compétitions entre des bactéries utilisatrices ou non utilisatrices des opines sur des plants produisant ou non des opines (Savka and Farrand, 1997). Sur les plants ne produisant par d'opine, la distribution des deux types de souches restait constante. Sur les plants produisant les opines, les bactéries utilisatrices de ces composés devenaient vite majoritaires dans la population inoculée. Ces expériences montrent un avantage compétitif des souches capables d'assimiler les opines en leur présence. Cependant, l'avantage trophique conféré par les opines n'a été observé que dans des expériences in planta et le concept d'opine devrait être valider directement au sein des tumeurs.

I.3.1.2 Utilisation des composés phénoliques

Il semble qu'en plus de détecter la présence des composés phénoliques suite à la blessure de la plante, le plasmide Ti soit capable de les dégrader. Il a depuis longtemps été montré que les composés phénoliques sont toxiques pour une large gamme de bactéries, y compris les agrobactéries, et que ces molécules sont produites par les plantes pour lutter contre les pathogènes (Braun, 1962; Rohringer and Samborski, 1967; Sequeira, 1963). Le gène *virH2* code une O-déméthylase capable de déméthyler et de détoxiquer de nombreux composés phénoliques tels que l'acide férulique (Kalogeraki et al., 1999), l'acide sinapinique, l'acide vanillique, l'alcool conyféryl. (Brencic et al., 2004). Le plasmide Ti permet donc de résister à ce mécanisme de défense de la plante.

I.3.2 Fardeau génétique

I.3.2.1 Le maintien du plasmide Ti peut être coûteux selon l'environnement...

Cependant, le plasmide Ti ne paraît pas toujours être bénéfique pour *Agrobacterium*, et semble même représenter un coût important. Guyon *et al* ont montré que des souches portant le plasmide Ti (pTi⁺) étaient défavorisées lorsque les plantes ne produisaient pas d'opines, en comparaison avec des souches ne portant pas le plasmide Ti (pTi⁻) (Guyon et al., 1993). Des souches pTi⁺ ont été co-inoculées avec des souches pTi⁻ sur des plants de lotier sauvages ou transgéniques produisant des opines. Quel que soit le ratio initial d'inoculation pTi⁺:pTi⁻ (50:50 ou 10:90), les souches pTi⁺ représentaient à peine 10% de la population 4 semaines après inoculation sur plants de lotier sauvages, tandis qu'elles représentaient plus de 90% de la population après inoculation sur des plants de lotier transgéniques produisant des opines. D'autres résultats ont mis en évidence qu'en milieu riche, la présence du plasmide Ti ne perturbait pas la croissance des bactéries, mais qu'en milieu carencé, la *fitness* des bactéries pTi⁺ était plus faible, et ce après 35 générations (Platt et al., 2012) (Figure 7A).



Figure 7 : Le plasmide Ti comme fardeau génétique. (D'après Platt et al, 2012)

A/ Impact du plasmide Ti sur la *fitness* d'A. *tumefaciens* 15955 cultivée en milieu riche et en milieu carencé en carbone ou en azote. B/ Impact de l'expression des genes vir sur le temps de doublement d'A. *tumefaciens* 15955, cultivée en milieu supplémenté (gris) ou non (blanc) en acétosyringone connue pour induire fortement l'expression des gènes vir. Les traitements marqués par des lettres différentes sont significativement différents (*p-value* < 0,05)

I.3.2.2 ... et les saisons

Des expériences ont montré que la présence du plasmide Ti représentait une charge importante en hiver et en automne. Après dénombrement sur milieu sélectif de bactéries isolées de sols issus de plusieurs sites en Algérie, ou par évaluation directe du plasmide Ti par PCR dans le sol, il est apparu que seules des souches dépourvues du plasmide Ti avaient été retrouvées au cours de ces saisons, tandis que 40% des souches portaient ce plasmide au printemps et à l'été suivant (Krimi et al., 2002).
I.3.2.3 Comment expliquer le coût du plasmide Ti?

Le coût représenté par le plasmide Ti peut s'expliquer par l'importance de sa taille (5% du génome d'*Agrobacterium*) (Goodner et al., 2001; Wood et al., 2001) (Dahlberg and Chao, 2003; Modi et al., 1991). Pour perdurer dans les cellules, le plasmide doit utiliser des systèmes de réplication et de répartition indépendants de ceux utilisés par le(s) chromosome(s) principal(aux). De plus, le plasmide Ti porte des gènes codant des protéines impliquées dans des fonctions et des structures lourdes et encombrantes, tels que les systèmes de sécrétion de type IV, le système de sécrétion de type VI, et plusieurs ABC transporteur notamment. La mise en place de tous ces systèmes est coûteuse. Platt *et al* ont montré qu'en condition d'induction de la virulence (en présence d'acétosyringone dans le milieu), les bactéries se développaient deux fois moins vite qu'en condition non inductrice (Platt et al., 2012) (Figure 7B). On peut supposer que la mise en place de la virulence et l'expression des gènes *vir* nécessite une dépense d'énergie importante qui retarde la croissance de la bactérie.

I.3.2.4 Plasmide Ti et « tricherie »

L'apparition de souches « tricheuses » dans les tumeurs suggèrent fortement que le plasmide Ti représente un poids important pour la bactérie. Ces tricheurs correspondent à des agrobactéries portant un plasmide Ti tronqué (Bélanger et al., 1995; Fortin et al., 1993). Malgré la réduction de la taille du plasmide Ti et la perte de nombreuses autres fonctions, elles sont toujours capables de tirer profit de la tumeur, puisqu'elle sont toujours capables d'assimiler les opines (Platt et al., 2012). Au sein de la tumeur, elles semblent donc présenter une meilleure *fitness*. Parmi les tricheurs, on retrouve également des souches avirulentes portant un ou des plasmides plus ou moins similaires au plasmide Ti (Merlo and Nester, 1977).

I.3.3 Mécanismes de maintien

Ainsi, bien qu'ils confèrent un avantage à la bactérie, le maintien du plasmide Ti n'en présente pas moins un coût. Dès lors, plusieurs stratégies sont développées pour assurer le maintien de ce plasmide. La réplication et la répartition des plasmides doivent d'abord être assurées. Des systèmes d'addiction permettent également aux plasmides d'être conservés dans les populations. De plus, certains plasmides peuvent être disséminés dans les populations par transfert conjugatif. Enfin, la régulation fine de l'expression des gènes des plasmides permet une réduction maximale du coût.

I.3.3.1 Maitien par la réplication

Mécanismes

Le plasmide Ti est un plasmide à faible nombre de copies (1 à 2 copies par cellule) de la famille des réplicons repABC. Ce type de réplicon est commun aux Rhizobiaceae et c'est également le système de réplication utilisé par le chromosome linéaire et le plasmide At (Cevallos et al., 2008). Ainsi, plusieurs réplicons repABC peuvent co-habiter dans une même cellule. La probabilité pour que ces plasmides soient distribués systématiquement et aléatoirement dans chaque cellule fille est faible. La réplication et la répartition du plasmide Ti doivent donc être finement régulées pour assurer la transmission à toutes les cellules filles. Ces types de réplicons se caractérisent par le regroupement en un même opéron des gènes impliqués dans la répartition et la réplication. Les gènes repA et repB codent les protéines assurant la répartition ou ségrégation des plasmides dans les cellules filles. RepA est une ATPase qui se polymérise en filament et permet la distribution des plasmides suite à la division cellulaire. RepB intéragit avec RepA et se fixe sur l'ADN au niveau de sites parS (séquences palindromiques de 16 bases) situés entre repA et repB. Le gène repC code la protéine initiatrice de réplication. Cette protéine est essentielle et suffisante à la réplication du plasmide, la séquence codante de son gène semblant porter l'origine de réplication oriV du plasmide Ti (Pinto et al., 2011). Aucun effet de l'activité transcriptionnelle de *repC* sur la réplication n'a cependant été montré. De plus, deux autres éléments sont présents dans la majorité des régulons *repABC*, le peptide RepD et le riborégulateur RepE. Le gène repD se situe entre repA et repB et régule la répartition du pTi (Chai and Winans, 2005a). Le gène repE est situé dans l'espace intergénique repB-repC sur le brin complémentaire, et il a été suggéré qu'il régulait l'expression de repC, au niveau transcriptionnel et traductionnel, pour moduler la réplication du plasmide Ti à octopine chez la souche R10 (Chai and Winans, 2005b). La surexpresion de RepE entraîne une diminution du nombre de copies d'un miniplasmide de type repABC. Chez la souche C58, le gène repE a aussi été retrouvé (voir chapitre 2 des résultats).

Régulation

La réplication du plasmide Ti est sujette à une forte régulation (figure 8). En condition normale, les protéines RepA et RepB exercent un rétro-contrôle négatif de l'expression de l'opéron *repABC*. La fixation de RepA, à l'aide de RepB, au niveau du promoteur de l'opéron *repABC* empêche sa transcription. En revanche, l'expression des gènes *repABC* est stimulée par le système VirA/VirG en réponse à la libération d'acétosyringone lorsque la plante est blessée (Cho and Winans, 2005). Il a été montré chez la souche C58 une augmentation d'un facteur 4 du nombre de copies du plasmide Ti en présence d'acétosyringone par rapport à une condition standard. L'induction se fait à partir du promoteur P4 de l'opéron, en amont duquel se trouve une « boîte *vir* ». L'expression des gènes *rep*

est également induite par TraR suite à la détection des opines au niveau de la tumeur (Li and Farrand, 2000; Pappas and Winans, 2003a, 2003b). Dans ce cas, l'expression se fait à partir des promoteurs P1, P2 et P3, en amont desquels se trouvent deux « boîtes *tra* ».

Ainsi, le plasmide Ti est maintenu en faible nombre de copies, sauf lorsque sa présence confère un avantage à la bactérie. L'augmentation du nombre de copies du plasmide Ti accroît par ailleurs le nombre de copies de gènes essentiels au cycle infectieux, et ainsi la quantité de chaque facteur de virulence, sans nécessairement faire intervenir des mécanismes complexes de régulation.



Figure 8 : Régulation de l'expression de l'opéron repABC. (D'après Cho et Winans, 2005).

L'expression de l'opéron *repABC* est fortement réprimée par les protéines RepA et RepB. Le riborégulateur RepE réprime la synthèse de la réplicase RepC (voir page 76) En présence de composés phénoliques, VirG phosphorylé induit l'expression de l'opéron par fixation au niveau d'une « boîte *vir* » (étoile orange) située en amont du promoteur P4. En présence d'homosérine lactone (héxagones jaunes), TraR stimule également l'expression de l'opéron en se fixant au niveau de deux « boîtes *tra*» (étoiles jaunes) positionnées au niveau des promoteurs P1, P2 et P3.

I.3.3.2 Maintien par des systèmes d'addiction

Sur les plasmides TiC58 et Ti-SAKURA, Yamamoto *et al* ont identifié un mécanisme, codé par les opérons *ietAS* et *tiorf24-tiorf25* respectivement, qui facilite la conservation du plasmide Ti au sein des cellules (Yamamoto et al., 2009). Les auteurs ont montré que l'entrée et le maintien forcé d'un plasmide extérieur du même groupe d'incompatibilité entraînent la perte du plasmide Ti résidant, et conduit à la mort de la cellule.

Ce mécanisme résulte d'un système toxine-antitoxine (TA), également appelé système toxineantidote. Il s'agit d'un système d'addiction des cellules bactériennes aux plasmides, qui permet d'éliminer les cellules ayant perdu le plasmide (DeNap and Hergenrother, 2005; Hayes, 2003; Schuster and Bertram, 2013) (figure 9). A partir d'un plasmide sont exprimés des gènes codant une toxine stable et une antitoxine labile. Tant que le plasmide est présent, les deux composants sont produits, et l'antitoxine séquestre la toxine et l'empêche de jouer son rôle néfaste. Lors de la division cellulaire, si la cellule fille conserve le plasmide, les deux éléments continuent à être produits. En revanche, si la cellule fille ne reçoit pas le plasmide, l'antitoxine va rapidement être dégradée, tandis que la toxine perdurera plus longtemps. Elle sera alors libérée et pourra agir, soit en tuant la cellule, soit en ralentissant sa croissance. Ainsi, dans les populations, seules les bactéries ayant conservé le plasmide survivront, assurant le maintien du plasmide. Les toxines peuvent s'attaquer à des fonctions ou des composants majeurs chez les bactéries, comme l'ADN, la paroi, la transcription ou encore la traduction. Cependant, ces toxines sont encore mal caractérisées.



Figure 9 : Principe des systèmes toxine-antitoxine. (D'après DeNap and Hergenrother, 2005)

La conservation du plasmide lors de la division cellulaire permet le maintien de la production de l'antitoxine et donc la survie de la cellule. La perte du plasmide conduit à une perte de l'antitoxine. La toxine alors libérée agit au sein de la cellule et conduit à sa mort.

I.3.3.3 Maintien par la conjugaison

Le plasmide Ti peut cependant être récupéré par le transfert conjugatif d'une cellule donneuse à une cellule receveuse. Par les mécanismes de conjugaison mis en place (pages 28, 30-32), le plasmide Ti peut être disséminé et maintenu dans les populations. L'acquisition d'un plasmide Ti pourrait permettre à une bactérie de coloniser et d'infecter de nouveaux sites, et à nouveau permettre le transfert et la dissémination du plasmide Ti (Lang and Faure, 2014). De plus, le plasmide Ti n'est pas transférable qu'à *Agrobacterium*, puisque des espèces proches telles que *Rhizobium* ou *Sinorhizobium* sont également capables de porter ce plasmide. Nous avons vu que des bactéries tricheuses étaient présentes dans la tumeur. Ces bactéries tirant profit de la tumeur sans porter le fardeau du plasmide Ti pourraient présenter un avantage compétitif vis-à-vis des bactéries portant le plasmide Ti, ce qui conduirait à sa perte. Cependant, la dissémination du plasmide Ti par transfert horizontal permet d'éviter que les bactéries « tricheuses » ne deviennent majoritaires au sein des populations présentes au niveau des tumeurs.

I.3.3.4 Régulation fine et ordonnée de l'expression des gènes

Afin d'éviter une forte consommation d'énergie, l'expression des gènes du plasmide Ti est finement régulée.

Nous avons vu qu'en condition de non induction de la virulence, le plasmide Ti n'est que faiblement répliqué, mais suffisamment pour assurer sa transmission dans les cellules filles. Au contact de composés issus de la blessure d'une plante ou d'opines, le rôle du plasmide Ti devient primordial. L'augmentation de son nombre de copies lui permet alors d'assurer toutes ses fonctions, telles que le transfert de l'ADN-T, la dégradation des opines et le transfert conjugatif (Cho and Winans, 2005). La coordination entre la conjugaison et la réplication du plasmide Ti n'est d'ailleurs pas un évènement isolé. C'est aussi le cas pour les plasmides IncP des *Pseudomonas* (Bingle and Thomas, 2001; Greated et al., 2000).

En absence de signaux, les gènes *vir* ne sont pas exprimés. En revanche, lorsque la bactérie perçoit des composés issus de blessure de plantes, l'expression de ces gènes va être stimulée. Cette régulation permet à *Agrobacterium* de n'activer et de ne mettre en place les mécanismes de sa virulence que lorsqu'elle se trouve à proximité d'une plante et que celle-ci est en état d'être infectée. Liu et Nester ont montré que l'expression des gènes *vir* était éteinte par la production d'AIA à partir des cellules végétales, ce par compétition entre l'AIA et l'acétosyringone pour VirA (Liu and Nester, 2006). La forte synthèse d'AIA à partir des cellules végétales transformées indiquerait ainsi à la bactérie que le brin-T a déjà été transféré, et ainsi que l'expression des gènes n'est plus nécessaire, permettant alors une économie d'énergie.

Le transfert conjugatif du plasmide Ti nécessite la présence de cellules receveuses. Ainsi, il n'est activé qu'à partir d'une certaine densité cellulaire au travers du *quorum sensing*. D'autre part, Chai *et al* ont montré qu'en milieu carencé, l'expression du gène *trlR*, dont le produit de traduction est un inhibiteur du *quorum sensing*, était induite (Chai et al., 2001). L'inhibition de la conjugaison en milieu pauvre permettrait d'éviter de mettre en place un système de sécrétion coûteux pour la cellule.

Enfin, la conjugaison du plasmide Ti et le transfert de l'ADN-T dans les cellules végétales font intervenir des systèmes de sécrétion. Ces deux systèmes impliquent de nombreuses protéines et occupent une place importante dans la membrane bactérienne. Afin d'éviter un encombrement de cette dernière, les deux systèmes ne devraient pas être mis en place en même temps. Il a été montré qu'en présence de composés phénoliques, et donc suite à la blessure d'une plante, la protéine TraM, inhibitrice du *quorum sensing*, était surproduite (Cho and Winans, 2005). D'autre part, il pourrait être envisagé que les gènes *blcABC* portés par le plasmide At, s'ils interviennent effectivement dans la régulation du *quorum sensing*, pourraient également permettre de retarder le transfert conjugatif du

plasmide Ti. Cette hypothèse est argumentée par le fait que l'expression des gènes *blcABC* n'est pas induite par les HSL, mais plutôt dans les phases précoces de la virulence par la libération de GABA.

Ainsi chez *Agrobacterium*, la mise en place de la virulence, ainsi que le maintien de son déterminant majeur, le plasmide Ti, sont soumis à une importante régulation. Plusieurs mécanismes de contrôle sont mis en jeu et sont entre-croisés, tels que des systèmes à deux composantes (dont le système majeur VirA-VirG), des régulateurs transcriptionnels, ou encore le *quorum sensing*. Si la régulation transcriptionnelle a été bien caractérisée chez cette bactérie, encore peu d'informations sont disponibles sur les mécanismes de régulation post-transcriptionnelle, faisant intervenir notamment les riborégulateurs.

II. Régulation par les riborégulateurs et physiologie bactérienne

Les ARN résultent de la transcription d'un gène. En fonction de leur rôle dans la cellule, plusieurs catégories d'ARN ont été définies. Les ARN messagers, comme leur nom l'indique, transmettent un message. Ils sont traduits en protéines. Les ARN issus des autres catégories ne sont pas traduits en protéines et agissent directement en tant que transcrit. Les ARN ribosomaux (ARNr) et de transfert (ARNt) sont impliqués dans la traduction des ARNm. Une dernière catégorie est actuellement très étudiée, il s'agit des ARN non codant (ARNnc), ou petits ARN (sRNA). Ils sont impliqués dans la régulation post-transcriptionnelle de l'expression des gènes. On les retrouve chez les Eucaryotes, les Eubactéries, et chez les Archées (Babski et al., 2014). Nous nous intéresserons ici aux ARN régulateurs bactériens, que nous appellerons riborégulateurs.

II.1 Généralités

II.1.1 Structure et localisation

II.1.1.1 Plusieurs caractéristiques

Les riborégulateurs bactériens possèdent généralement plusieurs caractéristiques : i) ils sont de petite taille - quarante à cinq cents bases- alors que les ARNm mesurent généralement plus de mille bases ; ii) ils possèdent une structure secondaire composée de plusieurs tiges-boucles impliquées dans leur stabilité ou dans leur fonction ; iii) ils ne présentent pas de séquence codante et ne sont pas traduits en protéines. Cependant certains riborégulateurs ne respectent pas ces caractéristiques. En plus de leur rôle de régulateurs, ils peuvent également coder de petits peptides : RNAIII code une δ -hémolysine de 26 acides aminés chez *Staphylococcus aureus* (Balaban and Novick, 1995), SgrS est traduit en un peptide de 43 acides aminés appelé SgrT qui appuie l'action régulatrice du riborégulateur (Wadler and Vanderpool, 2007). Enfin, certains riborégulateurs mesurent plusieurs kilobases (Sesto et al., 2013; Stazic et al., 2011; Toledo-Arana et al., 2009; Wehner et al., 2014). La classification d'ARN « non codant » ou de « petits » ARN ne correspond donc pas à une règle absolue. Ces cas étant cependant rares et ces termes sont toujours appliqués. Ici, nous préfèrerons les appeler riborégulateurs.

II.1.1.2 Différentes catégories de riborégulateurs

Les riborégulateurs peuvent interagir avec des protéines, avec des ARNm ou avec des métabolites. Dans le cas de l'interaction avec des ARNm, deux classes de riborégulateurs peuvent être définies selon la localisation génomique à partir de laquelle ils sont exprimés (Figure 10). On distingue les riborégulateurs antisens agissant en *cis*, dont les gènes sont localisés sur le brin complémentaire de la séquence codante qu'ils contrôlent ; et les riborégulateurs antisens agissant en *trans*, dont les transcrits sont exprimés à partir de séquences intergéniques. Enfin, les riborégulateurs interagissant avec des métabolites correspondent aux parties 5' non traduites des ARNm (5'UTR pour *UnTranslated Region*). Ces régulateurs sont nommés *riboswitch*.





Trois classes de riborégulateurs sont définies : les riborégulateurs codés à partir du brin complémentaire d'une séquence codante appelés ARNnc *cis* ; les riborégulateurs transcrits à partir de régions intergéniques nommés riborégulateur *trans* ; et les éléments positionnés dans les régions 5'UTR des ARNm, que l'on nomme riboswitch. CDS : CoDing Sequence ; UTR : UnTranslated region (localisé en 5' ou 3' des CDS).

II.1.2 Mécanismes d'action

L'interaction des riborégulateurs avec leurs cibles entraîne une régulation de la production de ces cibles, qui peut se faire au travers de différents mécanismes.

II.1.2.1 Action sur des protéines

Les protéines cibles interagissent avec des molécules d'acides nucléiques. Le riborégulateur va mimer, soit par sa structure soit par sa séquence, les acides nucléiques avec lesquels les protéines agissent habituellement. Ils vont alors séquestrer la protéine cible et bloquer son action dans la cellule. Bien que ce mécanisme ne soit pas le plus décrit, trois principaux cas ont été étudiés.

La protéine CsrA (pour *Carbon Storage Regulator*) perturbe la stabilité et la traduction de nombreux ARNm en interagissant avec des motifs GGA situés dans leur région 5'UTR. Les riborégulateurs CrsB et CrsC contiennent plusieurs de ces motifs au sein de leurs séquences. Lorsque leurs concentrations atteignent un seuil, ils séquestrent la protéine CsrA et l'empêchent d'interagir avec ses ARNm cibles (Liu et al., 1997; Weilbacher et al., 2003). Ce système, ainsi que son homologue Rsm (pour *Repression of Secondary Metabolism*), ont été identifiés chez de nombreuses bactéries, dont de nombreux pathogènes humains (Yersinia pestis, Legionella pneumophila, Pseudomonas aeruginosa...) ou des bactéries interagissant avec les plantes (Pectobacterium carotovorum, Pseudomonas fluorescens...) (Cui et al., 2001; Heroven et al., 2012; Kay et al., 2005).

Le riborégulateur 6S est le premier riborégulateur à avoir été identifié (Brownlee *et al*, 1971), chez *Escherichia coli*. Il est maintenant caractérisé chez de très nombreuses bactéries à Gram négatif et à Gram positif. Sa structure secondaire est constituée de deux tiges double brin entourant une large boucle interne appelée « renflement central » (Steuten et al., 2014; Wassarman, 2007) (Figure 11). Cette structure, caractéristique des promoteurs, lui permet d'interagir directement avec la sousunité σ^{70} de l'ARN polymérase (RNAP). Au cours de la phase exponentielle, l'ARN 6S s'accumule dans la cellule et atteint une concentration maximale en phase stationnaire. Dès lors, il séquestre la sousunité σ^{70} associée à l'ARN polymérase et bloque la transcription des gènes sous sa dépendance.



Figure 11 : Interaction avec des protéines.

En prenant la conformation d'un promoteur ouvert, le riborégulateur 6S séquestre l'ARN polymérase et bloque la transcription de nombreux gènes.

Enfin, chez *E. coli*, le riborégulateur GlmZ régule la traduction de l'ARNm *glmS*. Son action est cependant contrecarrée par sa dégradation par l'endonucléase YhbJ. Pour pallier cette dégradation, le riborégulateur GlmY, dont la séquence est similaire à celle de GlmZ, est produit. De part sa séquence plus courte, GlmY n'est pas capable de jouer le rôle régulateur de GlmZ. A forte concentration, il est en revanche la cible majoritaire de l'endonucléase YhbJ et la séquestre. GlmZ est protégé et peut jouer son rôle de régulateur (Göpel et al., 2014; Reichenbach et al., 2008).

II.1.2.2 Cas des riboswitchs et des thermorégulateurs

Situés dans les régions 5' des ARNm, les *riboswitchs* présentent une structure secondaire forte. Leur conformation est différente en fonction des conditions environnementales. Elle leur permet de moduler la traduction et la stabilité des ARNm en *cis*. Lorsqu'ils interagissent avec des molécules spécifiques (vitamines, ions) ou des signaux physico-chimiques (température, pH), leur structure est modifiée, entraînant la modulation de la traduction des ARNm. L'ARNm *prfA* est un régulateur

impliqué dans la virulence de pathogènes d'animaux (Johansson et al., 2002) (Figure 12). A 30°C ce riboswitch, de type thermorégulateur, forme une structure séquestrant le site de fixation des ribosomes (RBS pour *Ribosome Binding Site*) et bloquant la traduction. À 37°C, cette structure est dénaturée, le RBS est relâché, permettant la traduction de l'ARNm *prfA*. De même, la traduction de l'ARNm *mntH*, qui code un transporteur de Mn²⁺, est régulée par la présence de Mn²⁺ intracellulaire chez *Salmonella typhimurium* (Shi et al., 2014). A forte concentration de Mn²⁺, cet ion se fixe au niveau du riboswitch appelé UTR1 et induit la formation d'une structure de terminaison Rho-indépendante stoppant la traduction.



Figure 12 : Thermorégulation.

A basse température, le RBS de l'ARNm *pfrA* est emprisonné dans une structure secondaire. L'augmentation de la température permet de libérer le site d'initiation de la traduction.

II.1.2.3 Action des riborégulateurs cis et trans

Les riborégulateurs agissant en *cis* possèdent une complémentarité de bases parfaite et sur toute leur longueur avec leur ARNm cible. Pour les riborégulateurs agissant en *trans*, la complémentarité de bases avec leur ARNm cible est imparfaite. Cette propriété leur permet d'interagir avec plusieurs ARNm, et permet à un même ARNm d'être régulé par plusieurs riborégulateurs. Cette interaction nécessite, dans la plupart des cas, l'intervention d'une protéine chaperonne appelée Hfq pour faciliter et maintenir l'appariement entre les deux ARN. Ces riborégulateurs peuvent avoir un impact positif ou négatif sur la traduction d'un ARNm et/ou sur sa stabilité.

Inhibition de la traduction

L'inhibition de la traduction est le mécanisme le plus décrit. Le riborégulateur s'apparie avec l'ARNm au niveau du RBS, empêchant ainsi la fixation des ribosomes et l'initiation de la traduction (Figure 13). D'autres mécanismes ont également été identifiés. Le riborégulateur peut entrer en compétition avec les ribosomes au niveau d'une séquence appelé « *ribosome standby site* ». Cette séquence, située en amont du site RBS, permet au ribosome de se fixer sur l'ARNm en attendant la libération de la séquence Shine Dalgarno (SD) séquestrée dans une structure secondaire. Ceci a été observé chez *E. coli* pour le riborégulateur IstR-1 qui se fixe 100 pb en amont du site SD de l'ARNm *tisB* (Darfeuille et al., 2007; Vogel et al., 2004). La fixation d'un riborégulateur dans la séquence codante peut également entraîner un changement de conformation de la séquence d'initiation de la traduction et ainsi empêcher le recrutement des ribosomes. Ce mécanisme a été observé chez *Bacillus subtilis* pour l'interaction entre le riborégulateur SR1 et l'ARNm *ahrC* (Heidrich et al., 2007). Les riborégulateurs peuvent enfin bloquer la traduction des ARNm en interagissant avec des sites facilitateurs de la traduction présents dans leur séquence 5'UTR. Chez *Salmonella enterica*, le riborégulateur GcvB se fixe au niveau de sites riches en bases A/U de nombreux ARNm (Sharma et al., 2011, 2007). Ainsi, les riborégulateurs bloquent généralement la traduction des ARNm tandis que ceux-ci sont en cours de synthèse, empêchant ainsi immédiatement leur traduction. Cependant, l'action du riborégulateur SgrS nécessite que l'ARNm *ptsG*, codant un transporteur majeur du glucose, soit déjà en cours de traduction et que la protéine naissante soit intégrée dans la membrane interne (Kawamoto et al., 2005).





Activation de la traduction

Les riborégulateurs peuvent également activer la traduction des ARNm. Le plus souvent, le riborégulateur interagit avec un ARNm dont le RBS est séquestré dans une structure secondaire et n'est pas accessible aux ribosomes. La fixation du riborégulateur dans la région 5'UTR défait cette structure secondaire, libère le RBS et permet la traduction de l'ARNm. Ce mécanisme a notamment été observé pour les couples riborégulateur/ARNm DsrA/*rpoS*, RprA/*rpoS*, Qrr1-4/*vca0939*, RNAIII/ α -*hla*, RyhB/*shiA* (Repoila and Darfeuille, 2009). Le produit de traduction de *shiA* est impliqué dans l'import de shikimate. En absence du riborégulateur RyhB, les 51 premiers nucléotides de la partie

5'UTR de l'ARNm *shiA* s'apparient avec la séquence contenant le RBS et le codon initiateur. Ceci empêche l'initation de la traduction de *shiA* et entraîne la dégradation rapide de l'ARNm (Prévost et al., 2007). Les auteurs ont montré que RyhB se fixait au niveau de ces 51 bases et empêchait la formation de la structure secondaire, permettant ainsi la traduction de l'ARNm et sa stabilisation chez *E. coli* (Figure 14).



Figure 14 : Activation de la traduction

L'interaction de RyhB avec l'ARNm *shiA* libère le RBS d'une structure secondaire et permet sa traduction.

Dégradation des ARNm cibles

Les riborégulateurs peuvent également faciliter la dégradation de leurs ARNm cibles. Cette dégradation peut être directe, l'interaction entre ARNm et riborégulateur recrutant des RNases (Bandyra et al., 2012), comme chez S. typhimurium, où le riborégulateur MicC recrute directement la RNase E lors de son interaction avec l'ARNm *ompD* (Pfeiffer et al., 2009) (Figure 15). La formation d'une molécule double brin d'ARN issue de l'interaction entre un riborégulateur et un ARNm peut également recruter la RNAse III, comme c'est le cas avec le riborégulateur RNA III chez Staphylococcus aureus (Huntzinger et al., 2005). La dégradation peut également dériver de l'interaction entre le riborégulateur et son ARNm cible suite à l'inhibition de la traduction. En effet, lors de la traduction d'un ARNm, les ribosomes y sont fixés et le recouvrent, ce qui le protège des RNases. En revanche, si la traduction est bloquée par un riborégulateur, l'ARNm est nu et susceptible aux RNases. Ce mécanisme permet de rendre irréversible l'inhibition de la traduction (Morita et al., 2006). La capacité à induire la dégradation des ARNm n'est pas intrinsèque aux riborégulateurs puisque certains riborégulateurs entraînent la dégradation de certains de leurs ARNm cibles mais pas d'autres. C'est le cas du riborégulateur RyhB qui stimule la dégradation de l'ARNm sdhCDAB mais pas celle de l'ARNm cysE, ou encore du riborégulateur Spot42 qui facilite la dégradation des ARNm gltA, maeA et sthA mais pas de l'opéron sdCDAB (Desnoyers et al., 2013).



Figure 15 : Dégradation des ARNm

La fixation de MicC sur l'ARNm *ompD* recrute la RNAse E et conduit à sa dégradation.

Modulation des opérons

Un riborégulateur peut entraîner une expression différentielle des gènes présents au sein d'un même opéron. Le riborégulateur Spot42 régule l'opéron *galETKM*, dont les produits de traduction sont impliqués dans le catabolisme du galactose (Møller et al., 2002). Spot42 se fixe au niveau du RBS de *galK* et bloque sa traduction. Cependant, puisqu'il ne modifie pas la stabilité de l'ARNm *galETKM*, il ne perturbe pas la traduction des autres cistrons (Figure 16). L'activité de GalE, GalT et GalM est nécessaire en permanence dans la cellule, tandis que celle de GalK ne l'est qu'en présence de galactose. Ce mécanisme permettrait d'éviter la production inutile d'une protéine et d'économiser de l'énergie. De même, le riborégulateur GImZ favorise la traduction de l'ARNm *gImS*, qui est issu du clivage par la RNase E de l'ARNm bicistronique *gImUS* impliqué dans le métabolisme des acides aminés. En se fixant entre *gImU* et *gImS*, GImZ va faciliter le clivage, la stabilisation et la traduction de *gImS* (Kalamorz et al., 2007).



Figure 16 : Modulation des opérons

L'interaction de Spot42 au niveau du RBS de *galK* uniquement empêche sa traduction sans perturber la traduction de *galE*, *galT* et *galM* présents sur le même ARNm polycistronique.

II.1.3 La protéine Hfq

La protéine Hfq (pour Host factor Q β) a été découverte en 1968 chez *E. coli*, et était considérée comme un facteur essentiel pour la réplication du bactériophage Q β (Franze de Fernandez et al., 1968). Elle est aujourd'hui connue pour être une protéine chaperon importante dans la régulation par les riborégulateurs. Elle appartient à la famille des « LSm protein » (Like Sm proteins), qui sont aussi présentes chez les Eucaryotes et les Archées et qui sont impliquées dans l'épissage et la dégradation des ARNm. Elle est conservée et répandue chez les bactéries (Wassarman, 2002).

II.1.3.1 Mécanismes d'action

Plusieurs rôles ont été identifiés pour cette protéine. En facilitant l'appariement entre un riborégulateur et son ARNm cible, elle participe à la régulation de la traduction de ce dernier (Figure 17A). Chez les riborégulateurs trans, la complémentarité de bases étant imparfaite et incomplète, l'intervention de Hfq aide à initier et maintenir l'interaction. Elle augmente leur probabilité d'interaction en rapprochant les deux molécules et aide à remodeler les riborégulateurs, très structurés, pour permettre leur interaction avec leurs cibles. Elle peut aussi perturber directement la traduction des ARNm en provoquant un encombrement stérique au niveau des RBS et des « ribosomes standby sites ». Elle est aussi capable de protéger les riborégulateurs, qui possèdent des structures double brin, de l'attaque par des RNases (Moll et al., 2003) (Figure 17B). A l'inverse, elle peut faciliter la dégradation de certains ARNm en rendant accessible leur extrémité 3' (Figure 17C). Celle-ci va alors être rapidement polyadénylée par une poly(A)-polymérase (PAP) puis dégradée par des polynucléotides phosphorylases (PNP) telles que RNase E et RNase II (Mohanty et al., 2004). Des expériences de co-purification ont identifié une interaction directe entre Hfq et les PAP et PNP. Morita et al ont montré que Hfq était capable de recruter la RNase E spécifiquement, sans aucun autre composant du dégradosome, pour faciliter la dégradation de SgrS et RyhB chez E. coli (Morita et al., 2005). Enfin, des études ont mis en évidence qu'elle pouvait réguler la traduction et la stabilité d'ARNm directement et en absence de riborégulateurs associés (Salvail et al., 2013; Vytvytska et al., 2000).

II.1.3.2 Hfq est-elle indispensable ?

Cette protéine Hfq semble indispensable à la riborégulation. Cependant, dans plusieurs cas, la régulation par les riborégulateurs semble se faire indépendamment de cette protéine chaperon. C'est le cas pour l'interaction entre un riborégulateur agissant en *cis* et son ARNm cible. Par ailleurs, il est apparu que l'importance de Hfq était variable en fonction des souches d'une même espèce. En effet Bohn *et al* ont montré que Hfq ne jouait pas de rôle majeur chez les souches RN6390, COL et Newman de *Staphylococcus aureus* (Bohn et al., 2007), tandis que Liu *et al* ont montré qu'elle impactait fortement la virulence de la souche 8325-4 (Liu et al., 2010). Bien qu'elle possède de nombreux homologues, cette protéine n'a pas été identifiée dans toutes les espèces bactériennes étudiées, comme *Helicobacter pylori* ou *Enterococus faecalis*. D'autres protéines pourraient jouer un rôle similaire, comme cela a été montré pour *Sinorhizobium meliloti*, chez qui ont été identifiées des protéines affectant la voie de régulation des riborégulateurs de façon similaire à Hfq (Pandey et al., 2011). Une hypothèse alternative serait que chez les souches dépourvues de Hfq, l'association entre les riborégulateurs et leurs cibles soit plus rapide et plus stable, comme c'est le cas pour les riborégulateurs *cis* et leur cible.



Figure 17 : Modes d'action de la protéine Hfq. (D'après Vogel et al, 2011)

A/ Hfq peut faciliter l'interaction entre un ARNnc et son ARNm cible, et engendre donc l'inhibition ou l'activation de la traduction, ou encore la dégradation de l'ARNm selon la localisation de la zone d'interaction. B/ La présence de la protéine Hfq protège les ARNnc contre la dégradation par la RNAse E en s'appariant au niveau des sites de clivage. C/ Hfq peut enfin interagir directement, en absence d'ARNnc, et induire la dégradation des ARNm par le dégradosome. PAP : Poly-A-Polymérase ; Exo : Exonucléases.

II.1.3.3 Quantité de Hfq dans la cellule

Lorsque l'on additionne l'ensemble des riborégulateurs requérant le chaperon, ce nombre est supérieur au nombre de Hfq par cellule et il semble qu'elle soit limitante pour l'activité des ARNnc (Hussein and Lim, 2011). Ainsi, il a été mis en évidence une compétition entre les ARNnc vis-à-vis de Hfq. *In vitro*, la demi-vie du complexe ARN/Hfq est souvent supérieure à 150 min, alors que lors d'un stress, la réponse par les riborégulateurs est beaucoup plus rapide (1 à 2 minutes). Ceci suggère un échange rapide des riborégulateurs liés à Hfq, qui pourrait s'expliquer par un mécanisme actif de circulation des riborégulateurs au niveau de Hfq. La production d'un nouveau riborégulateur, suite à un stress, provoquerait le départ d'un riborégulateur résidant (Fender et al., 2010). Wagner a ainsi proposé que l'augmentation de la concentration globale en ARN augmenterait la fréquence de dissociation du complexe Hfq-ARN (Wagner, 2013). Ce mécanisme favoriserait la régulation par les riborégulateurs nouvellement produits, et bloquerait la régulation par les riborégulateurs résidents dont la nécessité n'est plus requise.

II.2 Identification des riborégulateurs et de leur cibles

Depuis leur découverte par hasard en 1981 (Stougaard et al., 1981; Tomizawa et al., 1981), de nouveaux riborégulateurs ne cessent d'être identifiés et caractérisés. L'ensemble des riborégulateurs identifiés jusqu'à présent est regroupé dans la base de données Rfam (http://rfam.xfam.org/) (Burge et al., 2013). Cependant, leur analyse est toujours difficile au vu de leurs caractéristiques. En effet, ils ne sont pas pris en compte lors des annotations des génomes puisqu'ils ne possèdent pas les caractéristiques d'un gène classique. Sauf quelques exceptions, ils ne possèdent pas de séquence codante et ne sont donc pas sélectionnés par les logiciels d'annotation. De plus, ils sont de taille en moyenne dix fois plus petite que celle d'un gène classique. A part leur structure secondaire et leur petite taille, rien ne permet de les différencier du reste du génome. Il est donc plus difficile de prédire leur présence et leur fonction. Ainsi, afin mesurer l'importance et le rôle des riborégulateurs d'identification ont été employées et ont du être mises au point.

II.2.1 Identification des riborégulateurs

II.2.1.1 Identification des riborégulateurs par prédiction bioinformatique

Des méthodes d'identification des riborégulateurs reposent sur des prédictions bioinformatiques. Plusieurs modèles de prédiction ont été développés, et se basent sur : i) la prédiction des structures secondaires des riborégulateurs, ii) la similarité de séquences avec d'autres riborégulateurs déjà identifiés, ou iii) la présence de motifs de transcription dans les séquences (site d'initation et de terminaison de la transcription). Cependant, le taux de recouvrement entre les logiciels est extrêmement bas (moins de 10%) (Hershberg et al., 2003). Il a donc été entrepris dès le début des années 2000 de combiner plusieurs méthodes afin d'améliorer les prédictions (Carter et al., 2001). Deux méthodes majeures ont alors été définies : des modèles de prédiction basés sur de la génomique comparative, et des modèles basés sur des logiciels d'apprentissage (Backofen et al., 2014; Li et al., 2012) (Tableau 1).

Type de logiciel	Logiciels	Types de données analysées	Références
Basé sur de la génomique comparative	QRNA	Séquence et structure secondaire Comparaison de deux séquences	Rivas <i>et al</i> , 2001
	RNAz	Séquence et structure secondaire Comparaison de plusieurs séquences	Washietl <i>et al,</i> 2001
	EvoFold	Séquence, structure secondaire et évolution Comparaison de plusieurs séquences	Pedersen <i>et al,</i> 2006
	sRNAPredict	Séquence et terminateur Rho-indépendant	Livny <i>et al,</i> 2005
	SIPHT	Séquence et terminateur Rho-indépendant	Livny <i>et al,</i> 2008
	NAPP	Conservation de séquences; Ne prend pas en compte la structure secondaire	Marchais et al, 2009
Basé sur des méthodes d'apprentissage	Carter <i>el al</i>	Composition en bases et structure secondaire Apprentissage par réseau de neurones.	Carter et al, 2001
	Saestom <i>et al</i>	Découverte automatique de profil de séquence pour prédire des gènes d'ARNnc	Saestrom et al, 2005
	PSoL	Séquence et structure secondaire Apprentissage automatique avec des bases de données	Wang <i>et al,</i> 2006
	Tran <i>et al</i>	Séquence et structure secondaire Apprentissage par réseau de neurones	Tran <i>et al,</i> 2009

Tableau 1 : Exemples de logiciels de prédiction des ARNnc (D'après Li et al, 2012)

Le premier mode de prédiction est actuellement le plus utilisé. Ces logiciels se basent sur la présomption qu'un riborégulateur doit présenter une certaine conservation de séquence et de structure secondaire au sein d'espèces phylogénétiquement proches. Dans un premier temps, la séquence primaire est étudiée. Des motifs caractéristiques de la transcription, tels que des promoteurs ou des sites de terminaison de la transcription indépendants du facteur Rho (ou terminateur Rho-indépendant), sont recherchés. Les terminateurs Rho-indépendants se caractérisent par une structure tige-boucle suivie d'une séquence poly-T en 3' de l'ARN. Cette structure est suffisante pour déstabiliser l'ARN polymérase lors de la transcription et ne nécessite pas l'intervention de la protéine Rho. Cette structure est souvent retrouvée chez les riborégulateurs. La fréquence de mono ou de bi-nucléotides, traduisant la formation de structures double-brin caractéristiques de la structure secondaire des riborégulateurs, est aussi analysée. Le pourcentage GC de la séquence d'intérêt est comparé au reste du génome, et son environnement génomique à celui de la séquence similaire dans d'autres génomes. La structure secondaire du transcrit potentiel

est également prise en compte, et doit être stable et de faible énergie. Enfin, les logiciels recherchent une homologie de structure secondaire et de séquence entre les transcrits potentiels et les transcrits déjà identifiés. Le fait est que les séquences peuvent être différentes mais conduire à une même structure si des mutations compensatrices permettent de contre-balancer une première mutation. Les logiciels QRNA, RNAz, sRNApredict et NAPP (Gruber et al., 2007; Livny et al., 2005; Ott et al., 2012; Rivas and Eddy, 2001), basés sur ces critères, sont fréquemment utilisés. Une étude a comparé ces 4 logiciels et a montré que sRNApredict3 était le plus performant : il engendre le moins de faux positifs, il est le plus rapide et est capable d'identifier le brin d'ADN transcrivant le riborégulateur (Lu et al., 2011). Cependant, ces logiciels présentent plusieurs limites. Ils sont en effet basés sur des comparaisons de séquences de génomes et le problème se pose donc de trouver des génomes suffisamment proches. Chez la souche 1021 de *Sinorhizobium meliloti*, les auteurs n'ont pu détecter que deux riborégulateurs en comparant avec *E.coli* à partir des données de Rfam, et aucun en commun avec *Bacillus* ou *Pseudomonas* (Ulvé et al., 2007). Enfin, puisque ces logiciels ne prennent en compte que les régions intergéniques, seuls les riborégulateurs *trans* peuvent être identifiés, ignorant donc les riboswitchs et les riborégulateurs *cis*.

Pour pallier ces problèmes, une seconde approche peut être utilisée : la technique d'apprentissage automatique (*Machine learning method*). Ces modèles nécessitent le développement, l'analyse et l'implémentation de méthodes automatisables permettant à un logiciel d'évoluer grâce à un processus d'apprentissage, en se fondant sur l'analyse de données empiriques issues de nos connaissances. Les logiciels développés prennent en compte toutes les séquences génomiques (codantes et intergéniques) et permettent d'identifier des riborégulateurs *cis,* des riboswitch, ainsi que des riborégulateurs orphelins présents chez une espèce particulière et absent ches les autres (Li et al., 2012).

Ces logiciels de prédiction sont cependant limités, puisqu'ils sont basés sur nos connaissances des propriétés des riborégulateurs, qui sont elles-mêmes restreintes. Il est donc nécessaire d'analyser plus de données. Pour découvrir de nouvelles caractéristiques, l'utilisation des approches expérimentales d'identification des riborégulateurs est indispensable.

II.2.1.2 Identification des riborégulateurs par des méthodes expérimentales

Plusieurs techniques permettent l'identification expérimentale des riborégulateurs. La méthode de puces à ADN consiste à hybrider sur des lames de verre ou de silice des ADN complémentaires (ADNc) synthétisés à partir d'ARN extraits. Sur ces lames de verre ou de silice ont été fixées au préalable des collections de séquences d'ADN correpondant aux gènes connus ou prédits dans le cas des puces à ADN ou à l'ensemble du génome pour le *tilling array*. Les puces à ADN permettent de

valider la présence de riborégulateurs déjà connus et de définir leurs conditions d'expression, par exemple pour l'étude des riborégulateurs impliqués dans la sporulation de *Bacillus subtilis* (Silvaggi et al., 2006). Le *tilling array* permet une analyse globale et l'identification de nouveaux candidats, comme chez *Listeria monocytogenes* (Toledo-Arana et al., 2009) ou chez *Mycobacterium leprae* (Akama et al., 2009). Cette méthode est également à l'origine de l'identification des longs riborégulateurs (Toledo-Arana et al., 2009). Cette méthode présente cependant plusieurs inconvénients, dont un très grand bruit de fond et des hybridations croisées possibles. De plus, les caractéristiques des riborégulateurs (petite taille et forte structure secondaire) limitent leur marquage et leur hybridation avec les sondes. Des améliorations sont constamment apportées pour augmenter la sensibilité de cette technique.

Les nouvelles techniques de séquençage d'ARN - appelées RNAseq - ont permis d'approfondir nos connaissances sur les riborégulateurs. Après rétrotranscription des ARN en ADN complémentaires (ADNc), ceux-ci peuvent être directement séquencés sans étape de clonage longue et fastidieuse, par les méthodes de pyroséquençage 454, SOLID ou Illumina. Pour cette dernière, les ADNc sont amplifiés en groupe après fixation aléatoire sur une lame de verre. Les groupes de séquence sont ensuite séquencés en utilisant une polymérase et l'ajout de nucléotides marqués à terminaisons réversibles. Ces méthodes sont beaucoup plus rapides, moins chères, et permettent d'étudier des millions de séquences d'ADNc en même temps. De plus, les techniques de séquençage récentes sont brin-spécifiques. Elles permettent de définir le brin d'ADN à partir duquel l'ARN est transcrit, et donc d'identifier les riborégulateurs cis. Des étapes précédant la rétrotranscription sont tout de même nécessaires. En effet, après extraction des ARN, plus de 90% des transcrits correspondent aux ARN ribosomaux (ARNr) et aux ARN de transfert (ARNt). De manière générale, il est donc fortement recommandé d'éliminer ces ARN abondants pour augmenter la profondeur de lecture (Sorek and Cossart, 2010). Pour cela, les ARN abondants peuvent être i) capturés à l'aide de billes magnétiques sur lesquelles sont fixées des amorces complémentaires aux ARN à éliminer, ou ii) dégradés par la RNase H après synthèse d'un ADN complémentaire à partir d'amorces complémentaires aux ARN à dégrader (figure 18A.1). Une autre technique, appelée séquençage différentiel (dRNA-seq), permet l'élimination enzymatique des ARN dégradés, c'est-à-dire possèdant une extrémité 5' monophosphate (Figure 18A.2). De plus, il peut être intéressant d'éliminer les ARNm afin d'enrichir l'échantillon en riborégulateurs. Une sélection par la taille peut être effectuée, les ARNm faisant généralement plus de 1 kb (Figure 18B.1). L'inconvénient de cette technique est que les riborégulateurs de longue taille, récemment mis en évidence, ne sont pas retenus pour le séquençage. Récemment, Lybecker et al ont développé une stratégie pour identifier les riborégulateurs cis et trans chez E. coli (Lybecker et al., 2014). Les riborégulateurs interagissant avec les ARNm, les auteurs ont sélectionné les ARN double brin par immunoprécipitation, à partir d'une souche dépourvue de l'endonucléase RNase III (Figure 18B.2). Trois cents seize riborégulateurs potentiels ont été prédits par cette méthode, dont 21 ont été validés par Northern blot. En revanche la plupart de ces candidats n'ont pas été détectés ou très faiblement chez une souche sauvage exprimant la RNase III, suggérant que ces riborégulateurs sont rapidement dégradés lorsqu'ils interagissent avec un ARNm. Ces résultats montrent que les techniques d'identification peuvent être améliorées et que nos connaissances sur les riborégulateurs ne sont pas exhaustives. Néamoins, l'utilisation de ces méthodes a permis l'identification de plusieurs riborégulateurs chez différentes bactéries, dont voici quelques exemples.



B/ Méthodes d'enrichissement des riborégulateurs



Figure 18 : Enrichissement en ARNm et en riborégulateurs. (D'après Sorek et Cossart, 2009)

A/ Elimination des ARN abondants (en noir): 1. Les ARNr et les ARNt sont retenus sur des billes magnétiques auxquelles sont greffées des amorces spécifiques, ou sont dégradés après synthèse d'un ADNc avec des amorces spécifiques. 2. Les ARNr, maturés en 5', sont dégradés par des exonucléases spécifiques d'extrémités 5' monophosphates. B/ Enrichissement en riborégulateurs : 3. L'extrait est enrichi en riborégulateurs en sélectionnant les transcrits de petites tailles par extraction après migration sur gel. 4. Les riborégulateurs s'appariant avec des ARNm dans les cellules, les molécules d'ARN double brin sont isolées par immunoprécipitation à l'aide d'anticorps spécifiques.

II.2.1.3 Combinaisons des deux approches

Riborégulateurs chez Pseudomonas aeruginosa

Chez *Pseudomonas aeruginosa*, des analyses antérieures avaient permis d'identifier 44 riborégulateurs, par prédiction bioinformatique, par hybridation sur puces à ADN et co-

immuniprécipitation avec Hfq. (Sonnleitner and Haas, 2011). Gomez-lazano *et al* ont identifié de nouveaux riborégulateurs par analyses de dRNAseq réalisées à partir de plusieurs conditions de croissance (Gómez-Lozano et al., 2012). Trois sélections d'ARN ont été effectuées : la première sélection (collection LIB>100), comprend tous les transcrits supérieurs à 100 bases, dont les ARNm. Les deux autres correspondent à des transcrits de moins de 500 bases (collection LIB<500) ou moins de 200 bases (collection LIB<200). Ils ont ainsi découvert plus de 500 riborégulateurs, dont les 44 déjà identifiés, et dont certains étaient présents dans les 3 sélections. 36% ne sont retrouvés que dans une seule collection, ce qui montre la nécessité de diversifier les étapes antérieures au séquençage pour augmenter le nombre de riborégulateurs identifiés. La première collection a permis d'observer environ 8 fois moins de nouveaux candidats que les deux autres. Cette faible proportion de nouveaux candidats est probablement due à l'abondance des ARNm qui masquent les riborégulateurs faiblement produits. Par ailleurs, les collections LIB<500 et LIB<200 permettent de détecter les riborégulateurs *cis* car un séquençage brin spécifique a été entrepris.

Riborégulateurs chez Clostrium difficile

Chez *Clostridium difficile*, les auteurs ont cherché à identifier les riborégulateurs en mettant au point une analyse *in silico* de recherche des terminateurs Rho indépendants (Soutourina et al., 2013). Ils ont prédits 511 candidats, dont 95 ont été identifiés en commun avec une précédente analyse (Chen et al., 2011). La différence du nombre de riborégulateurs potentiels entre les deux analyses s'explique notamment par l'absence d'identification des riborégulateurs *cis* dans l'analyse de Chen *et al.* En parallèle, les auteurs ont déterminé expérimentallement les riborégulateurs par dRNAseq à partir de 3 conditions de croissance différentes. 251 riborégulateurs (*riboswitchs* compris) ont été identifiés, parmi lesquels 126 avaient été prédits, et 71 candidats ont été retrouvés dans la base de données Rfam. 28 riborégulateurs prédits sur les 30 testés ont été validés par qRT-PCR et parmi les 40 candidats identifiés par RNAseq, 35 ont été validés par *Northern blot*.

Riborégulateurs chez Erwinia amylovora

Zeng et Sundin se sont intéressés aux riborégulateurs dépendants de Hfq chez *Erwinia amylovora* (Zeng and Sundin, 2014a). Ils ont combiné une analyse de RNAseq à une prédiction bioinformatique basée sur la recherche de terminateurs Rho-indépendants. Les ARN ont été extraits en condition d'induction de la virulence à partir d'une souche sauvage et d'une souche inactivée dans la production de Hfq. Les transcrits de petite taille et dont l'expression était réduite chez la souche mutante ont été sélectionnés. Trente-huit candidats Hfq-dépendants ont été identifiés, parmi lesquels : i) 8 avaient déjà été prédits précédemment par analyse bioinformatique et par comparaison avec *E. coli*, ii) 4 donnent une correspondance avec Rfam, iii) et 26 candidats n'avaient

jamais été observés. La comparaison des prédictions bioinformatiques et des données des RNA-seq des transcrits de petite taille a mis en évidence que sur 23 candidats prédits, 21 ont également été identifiés par séquençage. De plus, sur 13 candidats, l'expression de 12 a été validée par Northern blot, et le site d'initiation de la transcription de 7 candidats sur 8 a été validé par 5' RACE-PCR.

Ainsi, il semble important de combiner des approches bioinformatiques et de séquencer des transcrits issus de différentes conditions de culture afin d'optimiser l'identification de nouveaux riborégulateurs. En effet, il existe un faible taux de recouvrement entre les approches. Les prédictions bioinformatiques engendrent un très grand nombre de faux positifs, et ne permettent pas d'identifier des riborégulateurs exotiques ne possédant pas les caractéristiques recherchées par les logiciels. D'un autre côté, l'identification par les approches expérimentales comme le RNAseq peut être biaisée en fonction des conditions environnementales testées. Un riborégulateur ne pourra pas être identifié s'il n'est pas exprimé. Les différentes méthodes ainsi que leurs caractéristiques sont résumées dans le Tableau 2.

Tableau 2 : Caractéristiques	des différentes	méthodes d	l'identification	des ARNnc	(D'après	Altuvia,	2007;	Vogel	et al,
2007)									

Méthode d'analyse/Paramètres pris en compte	Bioinformatiques	Puces à ADN	Séquençage haut débit	Co-immuno Précipitation avec des protéines
Conservation de	Oui	Non	Non	Non
séquence				
Localisation du gàng	ICP	Solon los condos	Non bioicó	Non biaisá
Localisation du gene	IGK	Selon les solides	Non blaise	Non blaise
Longueur	50-400 bases	> 50 bases	Selon le protocole	Toutes les longueurs
0				0
Conditions	Non biaisé	Selon le protocole	Selon le protocole	In vitro
d'expression				
Ne prend pas en	ARNnc orphelin	ARNnc fortement	ARNnc fortement	ARNnc
compte		structurés ou	structurés ou modifiés	n'interagissant nas
compte		modifiés	structures ou mountes	
		mournes		avec des proteines
	ARNnc <i>cis</i>	< 50 bases		
	< 50 ; > 400 bases			
Remarques	Méthode globale	Méthode globale	Méthode globale	Nombre limité de candidats
Fxemples	F coli	S meliloti	Pseudomonas	Vihrio cholerae
Exemples	(Arraman at al. 2001)			
	(Argaman et al, 2001)	(Uive et al, 2008)	aeruginosa	(Liu <i>et al</i> , 2009)
	S. meliloti	Listeria	(Gomez-lazano <i>et al ,</i>	CsrB chez <i>E. coli</i>
	(Valverde et al 2008)	monocytogenes	2012)	(Liu et al, 1997)
		(Toledo-Arana et al	Clostridium difficile	Rsm7 chez
		2000 : Camaia at	(Soutouring at al. 2012)	Decudomonas
		2009, currejo el		rseuuomonds
		ai,2009)	E. coli (Lybecker	fluorescens
			et al, 2014)	(Heeb et al, 2002)

II.2.2 Identification des cibles des riborégulateurs

Lorsque le riborégulateur agit en *cis*, l'identification de la cible est rapide, puisqu'elle est codée par le gène situé sur le brin opposé de l'ADN. En revanche, pour les autres catégories de riborégulateurs, d'autres méthodes, *a priori* ou sans *a priori*, doivent être employées.

II.2.2.1 Identification des cibles sans a priori

Par prédiction bioinformatique

Lorsque la cible du riborégulateur est un ARNm, une approche bioinformatique, basée sur la recherche de complémentarité imparfaite de bases, peut être utilisée. Plusieurs algorithmes ont été créés et sont constamment améliorés pour rechercher les ARNm cibles au sein des séguences génomiques totales. Cinq logiciels principaux ont été développés. TargetRNA se base sur le potentiel d'appariement entre un riborégulateur et sa (ses) cible(s), mais ne prend pas en compte les structures secondaires (Tjaden, 2008). sRNAtarget est basé sur des méthodes bayésiennes naïves et considère le profil de structure secondaire du complexe riborégulateur/ARNm comme une caractéristique (Cao et al., 2009). De plus, il ne prédit que les cibles qui sont négativement régulées par le riborégulateur. IntaRNA recherche l'interaction optimale avec une énergie d'hybridation étendue minimale (Busch et al., 2008). Trois facteurs sont considérés : la structure optimale avec l'énergie libre minimale, la somme des énergies nécessaires à l'ouverture des structures initiales des sites de liaison, et enfin l'énergie libre minimale des régions d'initiation de l'interaction. StarPicker recherche tous les duplex pouvant se former entre un riborégulateur et ses cibles potentielles, et sélectionne les duplex les plus stables (Ying et al., 2011). L'énergie d'hybridation est ensuite calculée et étendue à tout le site d'interaction. Il a été montré que StarPicker était plus performant dans la prédiction des cibles et dans la précision des sites d'interaction par rapport à IntaRNA. RNApredator est basé sur l'algorithme de RNAplex, qui mime la compétition entre les interactions intra et intermoléculaires (Eggenhofer et al., 2011). Il considère en plus l'accessibilité de la cible. Cette méthode lui permet d'être plus rapide que des logiciels tels que IntaRNA basés sur la même approche. De plus, il donne des informations sur le rôle physiologique de la cible, grâce à l'annotation GO (gene Ontology), et renseigne sur l'accessibilité du site de fixation des ribosomes suite à l'interaction, permettant ainsi de prédire si le riborégulateur stimule ou prévient la traduction de l'ARNm (Tableau 3).

Par validation expérimentale

Des approches de génétique classique utilisant des souches modifiées dans la production d'un riborégulateur permettent d'évaluer les modifications phénotypiques potentielles – comme la

Type de logiciel	Logiciel	Propriétés analysées	Références
Analyse sans <i>a priori</i>	IntaRNA	Accessibilité des sites de liaison, utilisation de régions d'initiation (<i>seed region</i>)	(Busch et al., 2008)
	StarPicker	Stabilité thermodynamique, accessibilité des sites de l'ARNnc et de la cible ; méthodes Bayesiennes	(Ying et al., 2011)
	RNApredator	Accessibilité de la cible ; basé sur RNAplex	(Eggenhofer et al., 2011)
	Target RNA	Hybridation, ne prend pas en compte la structure secondaire initiale de l'ARNnc et de la cible	(Tjaden, 2008)
	sRNAtarget	Séquence et structure secondaire ; méthodes Bayesiennes	(Cao et al., 2009)
Analyse a priori	RNAcofold	Énergie de conformation minimale étendue de deux séquences	
	RNAhybrid	Énergie de conformation minimale étendue de deux séquences, néglige les appariements intra-moléculaires et les boucles multiples	(Krüger and Rehmsmeier, 2006)
	RNAplex	Énergie de conformation minimale étendue de deux séquences, rapidité de calcul	(Tafer and Hofacker, 2008)
	RNAup inteRNA	Prise en compte de l'accessibilité des sites de liaison Recherche de la structure du complexe ARN/ARN avec	(Mückstein et al., 2006) (Alkan et al., 2006)
		l'énergie libre totale minimale	
	Ripalign	Analyse de la structure du complexe ARN/ARN en considérant la stabilité thermodynamique et la covariation entre la séquence et la structure	(Li et al., 2011)
	PETcofold	Prédiction des interactions et des structures secondaires de deux alignements multiples de séquences d'ARN	(Seemann et al., 2011)

Tableau 3 : Exemples de logiciels de prédiction des cibles des ARNnc (D'après Li et al, 2012, Backofen 2014)

virulence, la mobilité ou encore la croissance de la bactérie. De plus, par insertion aléatoire du gène rapporteur lacZ dépourvu de promoteur, les cibles peuvent être identifiées. Ainsi, chez des souches mutantes, une modification dans la production de β -galactosidase laissera supposer que le gène *lacZ* s'est intégré dans un gène dont le transcrit est régulé par le riborégulateur étudié. Ce gène peut ensuite facilement être identifié par clonage et séquençage. Cette méthode a permis de déterminer les cibles du riborégulateur OxyS chez E. coli (Altuvia et al., 1997). Des analyses de puces à ADN, de RNAseq ou de protéomique peuvent également être entreprises. C'est ainsi que Kim et Kwon ont montré que RyhB-2 régulait l'expression de gènes impliqués dans la mobilité, et que les cibles du riborégulateur FnrS ont été identifiées chez E. coli (Boysen et al., 2010; Kim and Kwon, 2013). La protéomique permet d'observer l'impact de la modification des quantités de riborégulateurs sur le taux d'accumulation des cibles, sans tenir compte de la dégradation de l'ARNm qui n'est pas systématique. Par ces méthodes, un très grand nombre de cibles peut être identifié. Cependant, ces trois méthodes d'analyse ne permettent pas de différencier les cibles directes, qui interagissent physiquement avec le riborégulateur, des cibles indirectes dont l'expression est perturbée suite à l'interaction du riborégulateur avec les cibles directes. Ainsi, l'interaction du riborégulateur 6S avec la sous-unité σ^{70} de l'ARN polymérase, conduit à un effet pléiotropique.

Une méthode de piégeage des cibles peut être mis en place. Le riborégulateur d'intérêt est synthétisé *in vitro*, fixé sur des billes magnétiques et est mis en contact avec des extraits d'ARNm. Les ARNm retenus sont récupérés, rétrotranscrits puis identifiés par hybridation sur puce à ADN ou par RNAseq. Deux cibles du riborégulateur RseX ont été ainsi déterminées (Douchin et al., 2006). Cette méthode permet également d'identifier les cibles protéiques des riborégulateurs. Un extrait protéique est alors incubé avec le riborégulateur synthétisé et fixé, puis les protéines retenues sont séquencées par spectrométrie de masse. Des cibles du riborégulateur 6S ont ainsi été identifiées (Wassarman and Storz, 2000).

II.2.2.2 Identification des cibles a priori

Lorsqu'une cible est pressentie, sa validation fait appel à des approches bioinformatiques et expérimentales *in vivo* et *in vitro*.

Par prédictions bioinformatiques

Plusieurs logiciels de prédiction permettent de définir la zone d'interaction entre un riborégulateur et sa cible prédite. Les logiciels RNAcofold, RNAhybrid, RNAplex prennent seulement en compte la structure des ARNm et riborégulateur au sein de l'interaction. D'autres logiciels considèrent en plus la structure de chacune des deux molécules indépendamment, en dehors de leur interaction. Ces derniers permettent la détection d'interaction boucle-boucle (ou *kissing complex*) entre les deux molécules, et qui correspond à l'interaction entre les nucléotides situés dans les boucles des structures en épingles à cheveux du riborégulateur et de l'ARNm. C'est le cas des logiciels InteRNA, piRNA, RactIP ou encore PETcofold. L'identification précise des zones d'interaction permet de cibler les validations expérimentales.

Par validation expérimentale

Pour valider les régions d'interaction, des approches *in vivo* et *in vitro* sont utilisées. Le riborégulateur est cloné sous contrôle d'un promoteur inductible, et la région d'interaction supposée de l'ARNm cible est fusionnée à un gène rapporteur. L'expression du gène rapporteur peut être suivie selon l'expression du riborégulateur. La modification de l'expression du gène rapporteur, suite à des mutations introduites dans les régions putatives d'appariement, suggère que la zone d'interaction considérée est exacte. Des mutations compensatrices dans la région complémentaire devraient rétablir un phénotype sauvage. Cette approche a permis la validation chez *E. coli* des cibles de nombreux riborégulateurs tels que DsrA, RyhB, Spot42 ou encore SgrS (Urban and Vogel, 2007). *In*

vitro, les zones d'interaction peuvent être validées par retard sur gel. La fixation du riborégulateur sur l'ARNm cible retarde sa migration sur gel d'agarose ou d'acrylamide.

II.2.3 Rôle de Hfq dans l'identification des riborégulateurs et de leurs cibles

Hfq se fixant aux riborégulateurs et aux ARNm cibles, il est possible de les isoler par des expériences de co-immunoprécipiation à l'aide d'anticorps dirigés contre la protéine chaperon. Les ARN associés sont ensuite extraits et identifiés comme précédemment (page 61). La méthode SELEX (pour *Systematic Evolution of Ligands by EXponential enrichment*) peut tout d'abord être utilisée (Lorenz et al., 2006). Des ARN sont synthétisés *in vitro* par la T7 RNA polymérase à partir d'une collection de séquences de 50 à 500 bases correspondant au génome étudié, puis sont incubés avec Hfq, co-immunoprécipités, et identifiés comme précedemment. Cette méthode présente l'avantage de ne pas tenir compte des conditions dans lesquelles les riborégulateurs sont produits, et donc d'identifier un maximum de candidats. Il est cependant nécessaire de vérifier que les riborégulateurs prédits par cette méthode sont réellement produits dans la cellule. Il est également possible d'incuber Hfq directement avec des extraits d'ARN, ce qui permet d'identifier à la fois les riborégulateurs et leurs cibles. Cette méthode a permis d'identifier de nouveaux riborégulateurs et leurs cibles chez plusieures bactéries dont *S. meliloti* (Christiansen et al., 2006; Sittka et al., 2008; Sonnleitner et al., 2008; Torres-Quesada et al., 2014).

D'autre part, des séquences consensus (AAYAAYAA) de sites de liaison de Hfq aux ARNm ont été établies (Link et al., 2009; Lorenz et al., 2010). La recherche de ces séquences sur les ARNm semble également être un moyen d'identifier les cibles potentielles.

La caractérisation des riborégulateurs, ainsi que l'identification de leurs conditions d'expression et de leurs cibles, renseignent sur les fonctions cellulaires et physiologiques qu'ils pourraient moduler.

II.3 Fonctions cellulaires et physiologiques des riborégulateurs

Les riborégulateurs ont été impliqués dans de nombreuses fonctions physiologiques, telles que le maintien des plasmides, l'acclimatation à un facteur de stress environnemental, ou encore l'interaction avec un hôte qui peut être végétal.

II.3.1 Rôle dans le maintien des plasmides

Le 1^{er} rôle attribué aux riborégulateurs concerne le maintien des plasmides (Tomizawa et al., 1981). Depuis, ils ont été impliqués dans de nombreux mécanismes assurant la conservation de ces réplicons.

II.3.1.1 Rôle des riborégulateurs cis dans la réplication des plasmides

Chez le plasmide CoIE1, l'initiation de la réplication ne dépend pas d'une protéine de réplication, mais d'une amorce ARN, RNAII. Celui-ci s'apparie au niveau de l'origine de réplication, et permet l'ancrage de l'ADN polymérase I et la réplication. La synthèse du riborégulateur RNAI prévient la formation du complexe RNAII-ADN en intéragissant directement avec RNAII et en empêchant sa maturation nécessaire à l'initiation de la réplication (Tomizawa et al., 1981) (Figure 19A). La réplication des plasmides Inclα dépend de la production d'une réplicase, RepZ, dont la production nécessite la formation d'une pseudo-boucle en amont du RBS permettant la fixation des ribosomes (Figure 19B). La production d'un riborégulateur nommé RNAI va empêcher la formation de cette boucle en s'appariant au niveau de la séquence la constituant, bloquant ainsi la synthèse de la réplicase et donc la réplication. Enfin, la réplication des plasmides de type Inc18 est contrôlée par un mécanisme d'atténuation transcriptionnelle dépendant du riborégulateur RNAIII (Brantl et al., 1993) (Figure 19C). Celui-ci s'apparie avec l'ARNm de la réplicase RepR, conduisant à la formation d'un site de terminaison de la transcription Rho-indépendant, à l'arrêt de la transcription de *repR*, et au blocage de la réplication.

II.3.1.2 Rôle dans la résolution des multimères

Chez les plasmides à haut nombre de copies tels que les plasmides ColE1, la résolution des multimères en monomères est indispensable pour augmenter le nombre de plasmides dans les cellules, et ainsi assurer leur maintien lors des divisions cellulaires. Patient et Summers ont montré que la synthèse du riborégulateur Rcd conduisait à un ralentissement de croissance chez les bactéries, laissant ainsi du temps pour la résolution des multimères, et facilitant la transmission des plasmides aux cellules filles (Patient and Summers, 1993).



Figure 19 : Exemples de mécanismes de la régulation de la réplication des plasmides par des riborégulateurs

A/ Chez les plasmide CoIE1, le riborégulateur RNAIII s'hybride avec l'amorce RNAI et empêche sa maturation, nécessaire à l'initiation de la réplication. B/ Chez le plasmide Incα, le riborégulateur RNAI s'apparie avec l'ARNm de la réplicase RepZ et bloque sa traduction, empêchant la réplication. C/ Chez le plasmide Inc18, l'appariement du riborégulateur RNAIII sur l'ARNm naissant de la réplicase RepR induit la formation d'un site TIR qui arrête la transcription de *repR*, et empêche ainsi la réplication.

II.3.1.3 Rôle dans les systèmes toxine-antitoxine

Nous avons vu précédement que les systèmes Toxine-Antitoxine assuraient la conservation des plasmides dans les cellules. Dans le système décrit pour le plasmide Ti d'*A. fabrum*, l'antitoxine est une AAA-ATPase et agit donc sous la forme d'une protéine (pages 39-40). Ce système appartient au système TA de type II (figure 20B). D'autres types existent et font intervenir des antitoxines sous forme de riborégulateurs. Dans le système de type I, le riborégulateur antitoxine est transcrit à partir du brin complémentaire du gène codant la toxine. Le riborégulateur s'apparie avec l'ARNm de la toxine, empêche sa traduction et/ou provoque sa dégradation (figure 20A). C'est par exemple le cas du système Hok/Sok (Suppressor Of Killing) décrit pour le plasmide R1 (Gerdes et al., 1986). Dans le système de type III, l'antitoxine correspond à plusieurs sous unités de riborégulateurs qui séquestrent la toxine en prenant la conformation des séquences d'acides nucléiques avec lesquelles s'apparie la toxine (figure 20C). Jusqu'à présent, les systèmes TA de type III ont été montrés comme

impliqués dans la résistance aux phages, mais pas dans le maintien des plasmides (Fineran et al., 2009).



Figure 20 : Mécanismes du système toxine-antitoxine.

Trois mécanismes ont été recensés jusqu'à présent : a) l'antitoxine agit sous forme d'un riborégulateur antisens dont l'interaction avec l'ARNm de la toxine bloque sa traduction et induit sa dégradation ; b) l'antitoxine est sous forme protéique et bloque l'action de la toxine en bloquant son site actif ; c) l'antitoxine correspond à plusieurs unités de riborégulateurs qui séquestrent la toxine.

II.3.1.4 Rôle dans la conjugaison

Enfin, les riborégulateurs interviennent dans la conjugaison des plasmides, mécanisme permettant le transfert horizontal d'un plasmide entre cellules. Chez le plasmide R1, le riborégulateur FinP s'apparie avec l'ARNm *traJ*, qui code l'activateur de la conjugaison, inhibant ainsi sa traduction et le transfert conjugatif du plasmide (Koraimann et al., 1991).

II.3.2 Rôle dans la transition entre deux conditions environnementales

Plusieurs études ont mis en évidence le rôle des riborégulateurs dans la réponse bactérienne aux changements environnementaux, tels que la température, la concentration en fer, la carence en nutriments. Ces modifications peuvent correspondre à l'interaction de la bactérie avec un hôte eucaryote, suggérant le rôle des riborégulateurs dans ces interactions, et principalement lors d'interactions pathogènes. Ainsi, certains riborégulateurs sont identifiés chez des souches pathogènes, comme les riborégulateurs RliA, RliC, RliF, RliG et RliB qui sont retrouvés chez l'espèce pathogène *Listeria monocytogenes* et non chez des souches non pathogènes *Listeria innocua* ni *Listeria ivanovii* (Mandin et al., 2007). Ceci suggère fortement leur rôle dans la virulence des bactéries. Quelques exemples sont présentés ci-dessous.

II.3.2.1 Changement de température

Les riborégulateurs peuvent être impliqués dans l'acclimatation des bactéries à des changements de température. Les riborégulateurs de type thermorégulateur sont particulièrement impliqués. Chez *E. coli,* la régulation des gènes en condition de haute température dépend de la sous-unité o³² (ou RpoH) de l'ARN polymérase. Morita *et al* ont montré que la région 5' UTR de l'ARNm *rpoH* formait une structure secondaire masquant le site de fixation des ribosomes et empêchant sa traduction à basse température (M. Morita et al., 1999; M. T. Morita et al., 1999). Une augmentation de la température permet la dénaturation de la structure secondaire, la fixation des ribosomes et la traduction de l'ARNm. L'augmentation de la quantité de protéines RpoH va alors favoriser l'expression des gènes de réponse à un choc thermique de haute témpérature. Le riborégulateur *trans* DsrA, lui, favorise la croissance de la bactérie à basse température. Sa synthèse, qui se fait uniquement à basse température (<30°C), permet l'expression des gènes *otsA* et *otsB*, qui permettent la production de tréhalose. Cette molécule osmoprotectrice protège la cellule à basse température (Kandror et al., 2002).

Chez les pathogènes de mammifères, l'augmentation de la température (37°C) traduit généralement la présence de l'hôte. Chez *L. monocytogenes*, les gènes de virulence sont sous la dépendance du régulateur PrfA (Johansson et al., 2002). Sa traduction est dépendante de la température (page 46). A basse température, le régulateur n'est pas produit et les gènes de virulence ne sont pas transcrits. En revanche, l'augmentation de la température va permettre l'expression des gènes de virulence tels que le gène codant la Listériolysine O indispensable à l'échappement des phagosomes (Figure 21).



Figure 21: Régulation par PfrA

La présence d'un élément thermorégulateur dans la région 5'UTR de l'ARNm *pfrA* empêche sa traduction à faible température et permet l'activation de la virulence de *L. monocytogenes* seulement à 37°C.

II.3.2.2 Homéostasie du fer

Le fer est un élément indispensable à la croissance bactérienne puisqu'il est un cofacteur de nombreuses protéines impliquées dans la respiration, la photosynthèse ou la synthèse d'ADN. Cependant, à forte quantité, il devient toxique en présence d'oxygène car il engendre la production de radicaux superoxyde. La régulation de la concentration intracellulaire du fer est sous la dépendance du régulateur Fur (pour *Ferric Uptake Regulator*). A faible concentration de fer, le régulateur Fur n'est pas actif et le riborégulateur RyhB empêche la traduction de certains ARNm et favorise leur dégradation (Jacques et al., 2006). Ces ARNm codent des protéines non essentielles de stockage du fer ou des protéines fixant le fer. RyhB stimule également l'expression du gène *shiA* dont le produit de traduction permet l'import en grande quantité de shikimate, composé indispensable pour la biosynthèse de sidérophores (Prévost et al., 2007). RyhB augmente ainsi la concentration de fer libre dans les cellules, d'une part en le rendant plus disponible, et d'autre part en favorisant sa capture et son import. A forte concentration, la production de RyhB est inhibée par Fur. Les protéines capables de stocker le fer intracellulaire sont alors fortement produites, ce qui rend le fer moins accessible et donc moins toxique. De plus, RyhB exerce un contrôle négatif sur la synthèse de Fur, permettant ainsi de l'inactiver quand elle n'est pas nécessaire (Vecerek et al., 2007).

Il a été montré que RyhB était capable de moduler l'expression de gènes de virulence. Lors de l'internalisation dans une cellule eucaryote, la bactérie est en carence de fer. L'activation de l'expression de *shiA* par RyhB et la synthèse de sidérophores vont alors permettre à la bactérie de résister à cette carence. Chez *Shigella dysenteriae*, RyhB réprime VirB qui est un activateur transcriptionnel de nombreux gènes de virulence (Murphy et Payne, 2007). Cette répression de la virulence à l'intérieur des cellules eucaryotes empêche la lyse prématurée des cellules hôtes permettant ainsi la survie de la bactérie. Dans l'environnement extracellulaire riche en fer, comme dans l'intestin, les gènes de virulence sont activés, permettant ainsi à *S. dysenteriae* d'infecter son hôte (Figure 22).



Figure 22: Régulation par RyhB

Chez *S. dysenteriae*, la répression de VirB par RyhB est levée en présence de fer, permettant l'activation de la virulence. En présence de fer, RyhB est produit et réprime la production de VirB et la virulence.

II.3.2.3 Milieux carencés en nutriments

Les régulateurs CsrB et CrsC séquestrent la protéine régulatrice CsrA et empêche son action (pages 44-45). Cette protéine active l'expression des gènes de la phase exponentielle de croissance (comme les gènes de mobilité, de chimiotactisme ou de glycolyse), et inhibe l'expression des gènes relatifs à la phase stationnaire (les gènes impliqués dans la néoglucogenèse ou dans la synthèse et le catabolisme du glycogène) (Romeo, 1996; Timmermans and Van Melderen, 2010). Elle est donc nécessaire lorsque les nutriments sont abondants, et lorsqu'ils deviennent rares, elle est inactivée par CsrB et CsrC, ce qui permet à la bactérie d'utiliser ses réserves (Jonas and Melefors, 2009). De même, le riborégulateur 6S favorise la transition entre la phase exponentielle et la phase stationnaire, qui se caractérise par une carence en nutriments. Il empêche la transcription de la plupart des gènes sous la dépendance du facteur σ^{70} , qui est le régulateur majeur de la phase exponentielle (page 45). 6S permet donc d'inhiber les fonctions majeures de la phase exponentielle et donc d'économiser de l'énergie lors d'une carence en nutriments.

Chez Legionella pneumophila, la virulence se décompose en une phase réplicative non virulente à l'intérieur de macrophages ou d'amibes, et une phase transmissive virulente permettant l'évasion de la bactérie et l'infection de nouvelles cellules. Sahr *et al* ont montré que deux riborégulateurs, RsmY et RsmZ étaient impliqués dans la transition entre ces deux phases (Sahr et al., 2009). Brièvement, suite à une carence en acides aminés et en acides gras à l'intérieur de la cellule eucaryote, ils sont activés par le système à deux composantes LetS/LetA. Ils vont alors séquestrer la protéine CsrA (comme les riborégulateurs CsrB et CsrC chez *E. coli*), ce qui permet i) la répression des gènes correspondant à la phase réplicative et ii) la stimulation de l'expression des gènes de virulence (Figure 23).



Figure 23 : Régulation par RsmY et RsmZ

Lors d'une carence en acides aminés et en acides gras dans la cellule eucaryote, RsmY et RsmZ sont produits chez *L. pneumophila*. Ils bloquent l'action de la protéine CsrA, menant à l'arrêt de la synthèse des protéines de réplication, et à l'augmentation de la production des facteurs de virulence.

II.3.2.4 Densité bactérienne

Chez *S. aureus*, RNAIII active la traduction de l'ARNm *hla* qui code une α -hémolysine, et inhibe la traduction de l'ARNm du facteur de transcription Rot et de l'ARNm de la protéine A (Geisinger et al., 2006; Novick et al., 1993). Dans les étapes précoces de l'infection, ces bactéries adhèrent aux cellules hôtes, via des adhésines telles que la protéine A. Lorsque la densité bactérienne atteint un certain seuil, les bactéries sont à même de pouvoir attaquer leur hôte, ce qui constitue la phase tardive de l'infection. Le riborégulateur RNAIII, accumulé tout au long de la croissance bactérienne, va alors inhiber la synthèse des adhésines, et à l'inverse stimuler la production d' α -hémolysine et du régulateur Rot, permettant l'activation d'autres facteurs de virulence. Une nouvelle cible de RNAIII, Sbi, est indispensable à l'évasion immunitaire. L'expression de son gène est régulée tout au long de la phase exponentielle par le riborégulateur SprD, ce qui permet un niveau d'expression basal du gène et l'évasion immunitaire n'est plus requise (Chabelskaya et al., 2014).

Chez Vibrio cholerae les riborégulateurs Qrr1-4 (pour *Quorum Regulating RNA*) sont impliqués dans la régulation par le *quorum sensing*. A faible densité cellulaire, leur activation permet la traduction du régulateur AphA et inhibe celle du régulateur HapR (Lenz et al., 2005, 2004). AphA permet l'expression de gènes codant des facteurs de virulence tels que la toxine cholérique, ou des gènes impliqués dans la synthèse de biofilms qui vont permettre à la bactérie d'adhérer à l'épithélium intestinal. A forte densité cellulaire, HapR inhibe l'expression de ces gènes et active au contraire la synthèse de protéases et du système de sécretion de type VI. L'arrêt de la production de biofilms et la synthèse des protéases et du système de sécrétion favorisent le détachement de la bactérie de l'épithélium intestinal et son relâchement dans l'environnement (Shao and Bassler, 2014) (Figure 24).



Figure 24 : Régulation par les Qrr1-4

Chez V. cholerae, les Qrr1-4 sont produits à faible densité cellulaire et active AphA et réprime HapR, permettant l'adhésion aux cellules de l'hôte et la virulence. Ils sont inhibés à forte densité cellulaire, permettant l'arrêt de la virulence et le relâchement dans le milieu extérieur.

II.3.3 Rôle dans l'interaction avec un hôte végétal

Le rôle des riborégulateurs a aussi été démontré dans l'interaction de bactéries avec les plantes, qu'il s'agisse d'interactions bénéfiques ou délétères. Cependant, peu de caractérisations fonctionnelles permettant de décrire la mécanistique moléculaire sont disponibles. Quelques exemples chez *Sinorhizobium, Erwinia* et *Xanthomonas* sont présentés ici et sont résumés Tableau 4.

II.3.3.1 Riborégulateurs et symbiose chez Sinorhizobium meliloti

Chez *S. meliloti*, plusieurs études ont montré que l'expression de riborégulateurs était induite dans des conditions suggérant l'interaction symbiotique avec la plante, comme en présence de lutéoline, connue pour activer les gènes *nod* indispensables à la nodulation, ou lors de carence en phosphate et ou en stress acide (del Val et al., 2007; Valverde et al., 2008). L'impact de Hfq sur la symbiose avec la luzerne a aussi été étudié et vérifié *in planta* (Barra-Bily et al., 2010a, 2010b). Des mutants *hfq* se sont révélés incapables de fixer l'azote chez les plantes, et les plantes inoculées sont apparues quatre fois plus petites que celles inoculées avec la souche sauvage. Ceci peut être du à la dérégulation de la production de protéines importantes pour la résistance aux stress et pour la symbiose, puisque lors de la colonisation des plantes et de la mise en place de la nodulation, *S. meliloti* doit faire face à de

nombreux stress, oxydatifs principalement. Si la bactérie n'est pas capable de résister à ces stress, elle ne sera pas capable de développer la nodulation avec son hôte. De plus, Hfq semble nécessaire à la colonisation et à l'invasion des cellules végétales, la plupart des mutants n'étant pas capables de pénétrer les cellules hôtes. Chez la souche mutante, le gène *nifA*, dont le produit permet la fixation de l'azote, et le gène *nfeB* (pour Nodule Formation Efficiency), dont l'expression n'est normalement activée que dans la zone de nodulation, ne sont pas exprimés ou bien de façon abérrante. La majorité des nodules obtenus ne présentaient pas non plus la couleur rose caractéristique de la synthèse de leghémoglobine dans des nodules fonctionnels.

II.3.3.2 Riborégulateurs et virulence chez Erwinia amylovora

La recherche globale de riborégulateurs chez Erwinia amylovora a permis d'identifier quatre riborégulateurs impliqués dans la virulence de cette bactérie : les riborégulateurs ArcZ, Hr6, OmrAB, Hsr21 (Zeng and Sundin, 2014a). En milieu Hrp, qui mime l'interaction avec la plante et stimule la production du système de sécrétion de type III, les quatre gènes présentent un profil d'expression différent par rapport à une condition contrôle. Les gènes arcZ et hsr21 sont surexprimés, tandis que les gènes omrAB et hsr6 sont sousexprimés. Le mutant hsr21 est sensiblement moins virulent que la souche sauvage, et le mutant arcZ présente un fort défaut de virulence, similaire à un mutant hfq. Les auteurs ont par ailleurs montré que ArcZ stimule le système de sécrétion de type III, la mobilité et la formation de biofilm via la synthèse d'amylovorane, mais inhibe l'attachement des bactéries. Tous ces procédés sont des déterminants importants de la virulence. Les riborégulateurs OmrAB et Hsr6 quant à eux régulent positivement la mobilité et négativement la synthèse d'amylovoranes et donc la formation de biofilm. Combinés avec leurs profils d'expression, ces résultats suggèrent qu'ils permettent une régulation temporelle de la virulence. Dans les phases précoces de l'interaction, leur production est activée et permet la mobilité de la bactérie, nécessaire pour la migration de la zone de blessure de la feuille ou de la fleur jusqu'à l'intérieur de la plante. Dans les phases plus tardives de l'infection, ils sont inhibés, ce qui stimule la formation de biofilm, nécessaires pour que la bactérie puisse coloniser toute la plante au travers du xylème et pour causer une infection systémique.

II.3.3.3 Riborégulateurs et virulence chez Xanthomonas

Plusieurs études ont suggéré l'implication des riborégulateurs dans la virulence de *Xanthomonas*. Certains riborégulateurs sont en effet exprimés à partir du brin complémentaire de gènes codant des composants ou des substrats du système de sécrétion de type III ; ou encore voient leur production modulée par les régulateurs majeurs de la virulence. L'expression de huit riborégulateurs est modulée par les régulateurs clé du système de sécrétion de type III, HrpG et HrpX chez *Xanthomonas campestris* pv *vesicatoria* (*Xcv*) (*Schmidtke et al., 2012*) ; et celle de trois candidats est modulée par Rpf (pour *Regulation of Pathogenicity Factors*) chez *Xanthomonas campestris* pv *campestris* (*Xcc*) (An et al., 2013). Une souche inactivée dans la production de ces trois candidats présente un défaut de virulence, tandis que la virulence d'un simple mutant n'est pas affectée. Un riborégulateur particulier, sX13, a été identifié chez *Xcc* et *Xcv* et son rôle dans la virulence a été démontré chez *Xcv* (Schmidtke et al., 2013). L'absence de production de sX13 induit un fort retard dans le développement des symptômes de la maladie. Il a par ailleurs été montré que sX13 facilitait l'expression des gènes du régulon *hrp*, en modulant positivement la production du régulateur HrpX.

Tableau 4 : Exemples de riborégulateurs impliqués dans l'interaction avec un hôte végétale

Espèce bactérienne	Type d'étude	Conclusions	Références
Erwinia amylovora	Séquençage global, approche de génétique classique, analyse phénotypique, condition d'expression	Impact de ArcZ, Hr6, OmrAB et Hsr21 sur la virulence : mobilité, formation de biofilm, synthèse SST3	Zeng et Sudin, 2014
Pectobacterium carotovorum	Conditions d'expression, génétique classique	Impact de RsmB sur la virulence : synthèse d'enzymes extracellulaires	Liu <i>et al,</i> 1998
			Chatterjee <i>et</i> al , 2009
Xanthomonas campestris pv vesicatoria et pv camprestris	Séquençage global, génétique classique, conditions d'expression	Rôle de sX13 dans la virulence	Abendroth <i>et</i> <i>al,</i> 2014
campicotio			Schmidtke <i>et</i> <i>al,</i> 2012, 2013
Rhizobium etli	Séquençage global, recherche in silico, conditions d'expression	Rôle dans la symbiose suggéré ^a	Vercruysse <i>et</i> al, 2010
Sinorhizobium meliloti	Séquençage global, génétique classique, analyse phénotypique, conditions	Impact de Hfq sur la nodulation	Valverde <i>et al,</i> 2007, 2008 ;
	d'expression	Rôle de plusieurs candidats dans la symbiose suggéré ^{a, b}	Ulvé <i>et al,</i> 2007 ; Barra-Billy <i>et</i> <i>al,</i> 2010 ;
Bradyrhizobium japonicum	Comparaison de génomes, séquençage, conditions d'expression	Rôle de plusieurs candidats dans la symbiose suggéré ^a	Madhurigi <i>et</i> al, 2012
Pseudomonas brassicacearum Pseudomonas fluorescens CHAO	Génétique classique, analyse phénotypique, conditions d'expression	Rôle RsmX, Y, Z dans la symbiose : synthèse d'antifongique, d'auxine, formation de biofilm	Lalaouna <i>et al,</i> 2012 Kay <i>et al,</i> 2005

^a : Candidats présentant des profils d'expression différents dans des conditions d'interaction avec la plante

^b: Expression des candidats à partir d'îlots génomiques, souvent impliqué dans l'interaction avec la plante
Ainsi, les riborégulateurs modulent le changement d'état physiologique de la bactérie (transitions entre phase exponentielle/stationnaire, entre état avirulent/virulent...). Leur action est complémentaire d'autres régulateurs protéiques (notamment des sytèmes à deux composants).

La régulation par les riborégulateurs présente plusieurs avantages. De petites tailles et non traduits en protéines, leur production est rapide et peu coûteuse en comparaison avec la synthèse d'un régulateur protéique dont les transcrits sont généralement plus grands et requièrent une étape de traduction. L'action des riborégulateurs est rapide puisqu'ils agissent au niveau post-transcriptionnel, voir directement sur la protéine. Ainsi, si l'action d'une protéine doit être stoppée, elle le sera beaucoup plus rapidement que s'il fallait faire intervenir un facteur de transcription. De même, lorsqu'il s'agit d'un riborégulateur de type riboswitch, le contrôle est immédiat, le régulateur étant situé directement sur l'ARNm. Par ailleurs, puisqu'ils n'agissent pas forcément au niveau de la transcription d'un gène, et que parfois ils ne font que séquestrer une protéine, l'action des riborégulateurs peut rapidement être inversée dès que le signal environnemental disparaît. De plus, lorsqu'ils entraînent la dégradation de leurs ARNm cibles, ils sont souvent eux-mêmes clivés. Ainsi, des modélisations bioinformatiques prédisent une régulation plus rapide par les riborégulateurs (Shimoni et al., 2007). Leur action semble ainsi particulièrement adaptée dans l'acclimation des bactéries à un nouvel environnement.

III. Rôle des riborégulateurs chez Agrobacterium.

III.1 Analyses globales

Deux analyses ont permis l'identification globale des riborégulateurs chez A. fabrum C58 (Lee et al., 2013a; Wilms et al., 2012b). Les ARN ont été extraits à partir de quatre conditions différentes, dont une correspond à une condition d'induction de la virulence (avec acétosyringone dans le milieu) (Lee et al., 2013a). Ils ont ensuite réalisé un séquençage différentiel par la technique Illumina à partir des ADNc produits. Cette analyse a permis d'identifier 475 riborégulateurs répartis sur les 4 réplicons. Une comparaison avec la base de données Rfam a montré que 91 des transcrits identifiés étaient déjà répertoriés. La stratégie employée lors de cette analyse a permis d'identifier beaucoup plus de riborégulateurs que l'analyse précédente, puisque que Wilms et al n'avaient identifié que 228 candidats, dont 81 transcrits en commun. Les différences observées peuvent s'expliquer par plusieurs facteurs. Tout d'abord, deux conditions de culture supplémentaires ont été utilisées dans l'analyse de Lee et al. De plus, chez Lee et al, les ARNr abondants ont été éliminés et un traitement par l'exonucléase TEX a été utilisé pour éliminer les ARN maturés ou dégradés. A l'inverse, Wilms et al n'ont pas éliminé les ARNr. La technique de séquençage diffère également entre les deux études, Lee et al ayant employé un séquençage Illumina, tandis que Wilms et al ont utilisé la technique de pyroséquençage 454. Enfin, les algorithmes de recherche des riborégulateurs par prédiction bioinformatique étaient également différents. Au total, ces deux analyses ont permis d'identifier 621 riborégulateurs. Par ailleurs, plusieurs résultats ont suggéré le rôle des riborégulateurs dans la virulence d'A. fabrum. Il est apparu que 15 riborégulateurs sont induits en conditions d'induction de la virulence (en présence d'acétosyringone) et 7 sont réprimés. De plus, sur les 15 riborégulateurs induits, 14 possèdent une « boîte vir » putative, ce qui suppose qu'ils pourraient être régulés par le système à deux composants VirA/VirG. Les auteurs ont également identifié plusieurs riborégulateurs localisés sur le brin complémentaire de gènes ou d'opérons de virulence tels que les gènes vir. Cette découverte suggère fortement que ces antisens régulent les ARNm vir.

III.2 Etudes fonctionnelles

III.2.1 Thermorégulateurs

Balsiger *et al* ont mis en évidence la présence de thermorégulateurs sur deux gènes exprimés à partir du plasmide At et codant des protéines de choc thermique Hsp (pour *Heat Shock Protein*) (Balsiger et al., 2004). En effet, les séquences et les structures de l'extrémité 5' des ARNm *hspAt1* et *hspAt2* correspondent aux éléments ROSE identifiés chez les *Rhizobiaceae* (Nocker et al., 2001). A basse température, le site Shine Dalgarno (SD) et le codon initiateur sont masqués dans la structure secondaire, empêchant la traduction des ARNm. L'absence de détection de ces transcrits à 30°C suggère qu'ils sont rapidement dégradés. L'augmentation de la température dénature la structure secondaire, ce qui permet la traduction et la stabilisation des ARNm.

III.2.2 Hfq

L'analyse d'un mutant *hfq* a montré qu'une absence de production de cette protéine induisait des défauts de croissance, de division cellulaire (pour 10% des cellules), de mobilité et de virulence (Wilms et al., 2012a). On peut imaginer que le défaut de virulence soit en partie dû à un défaut de mobilité. Cependant, les bactéries ayant été ici directement inoculées sur la feuille, la mobilité n'est pas indispensable à la virulence. D'autre part, l'inoculation d'une quantité de bactéries 100 fois supérieure à celle de la souche sauvage n'a pas permis la restauration de la virulence. Ainsi les auteurs proposent que le défaut de virulence ne résulte pas du défaut de croissance. Une réduction d'un facteur 2 à 3 du transfert du brin-T a par ailleurs été observée. Ce défaut ne semble pas du à une déficience dans la mise en place du système de sécrétion de type IV, celui-ci étant intact chez la souche mutante. Les auteurs ont donc suggéré que l'action de Hfq pourrait se faire dans des étapes plus précoces de la virulence.

Récemment, Möller *et al* ont entrepris d'identifier les ARNm et les riborégulateurs interagissant avec Hfq par co-immunoprécipitation (Möller et al., 2014a). Ils ont montré que de nombreux ARNm codant des protéines de virulence interagissaient avec Hfq, tels que les protéines VirB, C, D et VirE. Les ARNm *flaF, flgL, mcpA* et *cheD* interagissent également avec Hfq, suggérant que la mobilité et le chimitactisme d'*Agrobacterium* sont régulés par les riborégulateurs, comme laissaient supposer les résultats de Wilms *et al* (Wilms et al., 2012a).

III.2.3 AbcR1

Par ailleurs, deux riborégulateurs, AbcR1 et AbcR2 ont été identifiés chez *A. fabrum* C58 (Overlöper et al., 2014; Wilms et al., 2011). AbcR1 bloque la traduction des trois ARNm *atu2422, atu1879* et *frcC*, codant tous les trois des ABC transporteurs. Plus particulièrement, *atu2422* code un transporteur de la proline et du GABA, qui est un composé de défense produit par la plante suite à la perception d'un pathogène. La synthèse de AbcR1 empêche ainsi l'import de GABA, et donc son effet toxique. Chez *Brucella abortus*, AbcR1 et AbcR2 ont également été identifiés, et il a été montré que la virulence d'un double mutant était atténuée lors de l'infection de macrophages ou de souris (Caswell et al., 2012). Cependant, aucun rôle direct dans la virulence n'a été observé pour ces deux

riborégulateurs chez *A. fabrum*, mais il est suggéré qu'ils permettent à la bactérie de résister aux défenses de la plante.

III.2.4 RepE

Enfin, chez la souche R10 d'*A. tumefaciens*, le riborégulateur RepE a été identifié et son rôle dans la régulation de la réplication du plasmide Ti a été étudié (Chai and Winans, 2005c). Les auteurs ont montré qu'une faible production de RepE conduisait à une augmentation du nombre de copies du plasmide Ti, et inversement. Des prédictions bioinformatiques, validées par des approches expérimentales, ont proposé que la fixation de RepE sur l'ARNm *repC* induisait un changement de conformation au niveau du site d'initiation de la traduction de *repC*. Le codon initiateur et le RBS seraient alors masqués, et un site putatif de terminaison de la transcription serait formé, bloquant ainsi la traduction de *repC*. La diminution de la quantité de protéines initiatrices RepC suite à la surproduction de RepE limiterait ainsi la réplication du plasmide Ti.

Etude	Méthodes	Souche utilisée	Conclusions	Références
HspAt1 et HspAt2	Génétique classique	C58	Présence de 2 thermorégulateurs modulant des protéines de choc thermiques	Balsiger <i>et al</i> , 2004
RepE	Génétique classique	R10	Régule la réplication du plasmide Ti	Chai et Winans, 2005
AbcR1	Génétique classique	C58	Régule des ABC transporteurs, notamment transporteur du GABA	Wilms <i>et al,</i> 2011 Overlöper et al, 2014
Hfq	Génétique classique + RIP- RNAseq	C58	Rôle dans la division cellulaire, la croissance, la mobilité, la virulence	Wilms et al,2012
Etude globale	RNA-seq + bioinformatique	C58	Identification de 228 ARNnc potentiels	Wilms <i>et al,</i> 2012
Etude globale	RNA-seq + bioinformatique	C58	Identification de 475 ARNnc potentiels	Lee <i>et al,</i> 2013

Tableau 5 : Etude des ARNnc chez A. tumefaciens

Ainsi chez *Agrobacterium*, plusieurs études ont permis d'identifier des riborégulateurs, et le rôle de certains candidats et de la protéine chaperon Hfq dans l'interaction avec la plante ou dans la virulence a été suggéré (Tableau 5). Cependant, aucun riborégulateur n'a été directement impliqué dans la mise en place de la virulence de la bactérie. Notre étude s'est alors attachée à découvrir si des riborégulateurs pouvaient constituer de nouveaux régulateurs ou acteurs de la virulence chez *A. tumefaciens*.

Résultats

IV. Identification des riborégulateurs et rôle dans le cycle infectieux chez A. fabrum C58

IV.1 Etude globale des riborégulateurs chez A. fabrum C58

La combinaison d'analyses bioinformatiques et expérimentales est de plus en plus utilisée afin d'optimiser l'identification des riborégulateurs. Une telle approche a été utilisée dans notre étude. A l'aide du logiciel sRNApredict, qui se base sur la recherche de structures secondaires et d'éléments de transcription ainsi que sur des comparaisons de séquences avec des génomes proches, 805 candidats potentiels d'A. fabrum C58 ont été prédits. Mille cent huit petits transcrits ont été identifiés expérimentalement par séquençage à haut débit des ARN de petite taille par la technique Illumina. Plusieurs conditions ont été analysées : milieu riche/milieu pauvre, phase exponentielle/phase stationnaire. La répartition des transcrits identifiés s'est montrée homogène tout au long des 4 réplicons. Nous les avons également classés en 4 catégories selon leur localisation vis-à-vis des séquences codantes annotées du génome : en antisens de séquences codantes (AS), dans des régions intergéniques (IGR) ou dans les régions 5' et 3' non traduites (UTR pour UnTranslated Region). Une sur-représentation des riborégulateurs AS sur le pTi a été observée. Ces résultats expérimentaux ont été comparés i) à la base de données Rfam regroupant l'ensemble des riborégulateurs identifiés à ce jour (Sarah W. Burge et al., 2013) ii) aux 805 riborégulateurs prédits et iii) aux données obtenues précédemment lors de deux analyses d'identification des riborégulateurs chez A. fabrum C58 par RNAseq (Lee et al., 2013a; Wilms et al., 2012b). Ainsi, parmi les 1108 identifiés par notre séquençage, 31 candidats ont trouvé une correspondance avec les riborégulateurs de Rfam ; 163 ont aussi été prédits par l'analyse bioinformatique ; et 196 avaient déjà été identifiés par les séquençages précédents. De plus, nous avons validé une vingtaine de nouveaux candidats parmi les 830 identifiés uniquement par notre analyse de transcriptomique. Pour cela, nous avons identifié leurs extrémités 5' et 3' et mesurer leur longueur par RACE-PCR, puis nous avons prédit leur structure avec le logiciel RNAfold, et ainsi nous avons montré que nos candidats étaient fortement structurés.

Pour caractériser le rôle physiologique d'un des nouveaux candidats, RNA1111, des souches modifiées dans sa production (ne le produisant pas ou le surproduisant) ont été construites. L'agressivité sur plantes de ces souches a été déterminée au travers de la mesure du nombre de tumeurs produites suite à l'interaction des souches bactériennes avec une blessure de plants de tomate, de kalanchoé et de tabac. Les cibles potentielles de RNA1111 ont aussi été prédites à l'aide de 3 logiciels et la quantification de leurs transcrits dans les souches mutantes a été entreprise par RT-qPCR. Nous avons montré que l'absence de RNA1111 réduisait la virulence d'*A. fabrum* C58. Six

ARNm cibles ont été prédits en commun par les trois logiciels. Leurs produits de traduction sont impliqués dans la virulence (la protéine 6b, VirC2, VirD3) ou le transfert conjugatif du plasmide Ti (TraA et TrbD), et *atu6072* code une protéine de fonction inconnue. L'absence de RNA1111 induit une diminution de la quantité des ARNm *6b, traA* et *atu6072*, confirmant un impact direct ou indirect de RNA1111 sur ces trois cibles.

Les résultats de cette analyse sont présentés dans une publication sous presse dans le journal MPMI.

Small RNA deep-sequencing analyses reveal a new regulator of virulence in *Agrobacterium* fabrum C58

Dequivre, M.^{1,2,3}, Diel B.^{1,2,3,4,5}, Villard C.^{1,2,3}, Sismeiro, O.⁶, Durot, M.^{7,8}, Coppée, J.Y.⁶, Nesme, X.^{1,2,4,5}, Vial, L.^{1,2,4,5}, and Hommais F.^{1,2,3}

¹Université de Lyon, F-69622, Lyon, France

²Université Lyon 1, F-69622 Villeurbanne, France

³CNRS, UMR 5240 Microbiologie Adaptation et Pathogénie, F-69622 Villeurbanne, France

⁴CNRS, UMR 5557 Ecologie Microbienne, F-69622 Villeurbanne, France

⁵INRA, USC 1364 Ecologie Microbienne, F-69622 Villeurbanne, France

⁶Plate-forme Transcriptome et Epigénome, Département Génomes et Génétique, Institut Pasteur, 25 rue du Dr. Roux, F75015 Paris, France

⁷CEA/DSV/FAR/IG/Genoscope and CNRS UMR8030 Laboratoire d'Analyses Bioinformatiques en Métabolisme et Génomique, 2 rue Gaston Crémieux 91057 Evry Cedex, France

⁸TOTAL New Energies USA, 5858 Horton Street, Emeryville, CA 94608, USA

Corresponding author: Florence Hommais florence.hommais@univ-lyon1.fr

Abstract

Novel ways of regulating Ti plasmid functions were investigated by the study of small RNAs (sRNAs) that are known to act as post-transcriptional regulators in plant pathogenic bacteria. sRNA-seq analyses of *Agrobacterium fabrum* strain C58 allowed us to identify 1108 small transcripts expressed in several growth conditions, which could be sRNAs. A quarter of them were confirmed by bioinformatics or by biological experiments. Antisense RNAs represent 24 % of the candidates and they are over-represented on the pTi (with 62 % of pTi sRNAs), suggesting differences in the regulatory mechanisms between the essential and accessory replicons. Moreover, a large number of these pTi antisense RNAs are transcribed opposite to those genes involved in virulence. Others are 5'- and 3'-UTR RNAs and *trans*-encoded RNAs. We have validated, by RACE-PCR, the transcription of fourteen *trans*-encoded RNAs, among which, RNA1111 is expressed from the pTiC58. Its deletion decreased the aggressiveness of *A. fabrum* C58 on tomatoes, tobaccos and kalanchoe suggesting that this sRNA activates the virulence. The identification of its putative target mRNAs (6b gene, *virC2, virD3* and *traA*) suggests that this sRNA may coordinate two of the major pTi functions: the infection of plants and its dissemination among bacteria.

Introductory statements

Whilst regulation by transcription factors is widespread in plant-bacteria interactions, there is growing evidence that riboregulators (sRNAs) could act as activators or repressors at a post-transcriptional level. The untranslated 5'-ends of genes (5'UTR RNA), acting as riboswitches or RNA thermometers, are known to be involved in post-transcriptional regulation (Bastet et al., 2011). sRNAs usually range from 50 to 500 nucleotides (nt) in length and they mainly act by regulating mRNA translation or stability through base-pairing (Storz et al., 2011). Antisense RNAs (asRNAs) are encoded in the opposite strand of their target and *trans*-encoded RNAs are present in intergenic regions, showing limited complementarity with their multiple targets localized elsewhere in the genome. Biocomputational predictions and experimental approaches, including high throughput sequencing of cDNA (RNA-seq), have led to the discovery of many sRNAs, and hundreds have been shown to be transcribed in different bacteria (Storz et al., 2011). However, overlaps between data are restricted, probably because of differences between the methods or the growth conditions used, and work is needed to improve our knowledge of the sRNA content in bacterial species. Furthermore, attempts to functionally characterize putative sRNAs have been restricted to a small number. The characterization of *trans*-encoded RNA targets is intricate since the post-transcriptional control by sRNA needs only a limited complementary sequence (around only 10 bases), which is difficult to predict (Li et al., 2012; Backofen et al., 2014).

Using a combination of sRNA prediction and deep-sequencing analyses, the sRNA content of several plant-interacting bacteria has been studied (del Val et al., 2007; Schmidtke et al., 2012; Wilms et al., 2012a; Lee et al., 2013; Schlüter et al., 2013; Zeng and Sundin, 2014) and several of these sRNAs are thought to be involved in plant-bacteria interactions. Indeed, various RNA-seq studies of *Sinorhizobium meliloti* revealed a plethora of small untranslated transcripts, suggesting major roles for riboregulators in the adaptive flexibility required for the complex lifestyle of *Rhizobiaceae* (del Val et al., 2007; Schlüter et al., 2010; Schlüter et al., 2013; Becker et al., 2014).

Agrobacterium is a bacterial genus commonly found in rhizospheres, where its members exercise generally benign commensalism with plants. However, when harboring a virulence plasmid called Ti plasmid (pTi), agrobacteria behave as plant pathogens able to induce the formation of tumors (crown galls) in almost all wounded dicots and gymnosperms, and in some monocots. The strain *A. fabrum* C58 is widely used to study the pathogenicity of agrobacteria. C58 was named in former papers as *A. tumefaciens* but this terminology is considered irrelevant because it has been given to several species. It was recently renamed *A. fabrum* (Lassalle et al., 2011; Mousavi et al., 2014). The *A. fabrum* C58 genome is composed of four replicons: two chromosomes and two accessory plasmids, called pAtC58 and pTiC58, the latter encoding major virulence factors (Pitzschke and Hirt, 2010). Pathogenic agrobacteria affect the host plant by exporting T-DNA, a DNA segment of the pTi, through a typical type IV secretion system (encoded by the *vir* region). The T-DNA is integrated into the nuclear plant

genome and expression of its genes, by transformed cells, results in (i) the over-production of phytohormones responsible for the tumor formation and (ii) the synthesis of opines. These compounds are amino-sugars used as nutrient sources by pTi-harboring agrobacteria. Opines also induce the conjugative transfer and dissemination of pTis. To induce virulence, agrobacteria must (i) perceive wound-released chemical signals from plants, (ii) move to the wound site in the plant, (iii) attach themselves to the plant cells and, (iv) transform them through a cascade of events.

Recently, two deep-sequencing studies revealed hundreds of sRNAs in *A. fabrum* C58 (Wilms et al., 2012a; Lee et al., 2013). Some of them were differentially expressed in the presence of acetosyringone and some others were characterized, namely the asRNA RepE, involved in the control of pTiR10 replication (Chai and Winans, 2005); and the sRNA AbcR1, involved in the regulation of the GABA transporter in *A. fabrum* C58 (Wilms et al., 2011; Overlöper et al., 2014). Furthermore, the absence of the RNA chaperone Hfq reduced *A. fabrum* C58 virulence (Wilms et al., 2012b), suggesting an important role for sRNAs during the interaction of agrobacteria with plants. Here, using the *A. fabrum* strain C58 sRNA-seq analyses that specifically target small transcripts, we detected a plethora of putative small RNAs and confirmed more than a quarter of them by bioinformatics predictions or with transcriptomic assays. Several sRNAs were identified and one of them, expressed from pTiC58, was functionally characterized, showing its involvement in the control of virulence.

Results

Identification of small transcripts

Size-selected sRNAs were isolated from A. fabrum C58 cells cultivated in various growth conditions, using different media (rich and minimal) and growth phases (exponential and stationary). They were strand-specific sequenced on an Illumina GAII (Perkins et al., 2009) and coverage depths of the unrepeated genome sequences were calculated for each nucleotide position in the forward and the reverse strands. The resulting coverage plots were visualized on the Artemis genome browser (Rutherford et al., 2000) and genome portions with contiguous sequencing coverage were analyzed manually. Resulting regions, with an average depth at least ten times higher than the background, (estimated by the average depth of the adjacent regions) were considered as transcribed. Finally, signals detected from each library were compared and those transcripts found in all the libraries were considered as putative sRNA candidates. These results were compared to novel small RNAs identified with the Rockhopper software system. 1108 putative sRNAs were selected and were named according to their position on the genome. Small transcripts ranged from 18 nt to 576 nt in length (median=117 nt), in accordance with the size-selection performed. Among the 1108 small transcripts, 602 were found on the circular chromosome, 291 on the linear one, 140 on pAtC58, and 75 on pTiC58, with respect to the size of replicons (Figure 1). With the exception of the repeated regions, small transcripts were homogeneously distributed along genome sequences at a density of about one sRNA candidate

per 5 kb (one sRNA for every 5 proteins) (Figure S1).

Putative sRNAs were also classified according to four categories: (i) 407 putative sRNAs were localized in intergenic regions (IGR) and correspond to putative *trans*-encoded RNAs, (ii) 402 were localized near, or overlapped the translation start of the coding sequence (5'UTR), (iii) 37 were localized near, or overlapped the termination of the coding sequence (3'UTR) and, (iv) 262 were localized in the opposite strand of the coding sequence, corresponding to the antisense sRNA candidates (asRNAs) (Figure 1).

In silico prediction of small RNA candidates in A. fabrum strain C58

Sequences of sRNA candidates were compared to the Rfam database and significant similarities were found for 31 candidates (Table S1). As only a few sRNAs are in the Rfam database, we performed an *in silico* prediction of sRNAs present in the *A. fabrum* C58 genome using the sRNApredict algorithm. Comparative genomics were performed between (i) the *A. fabrum* C58 genome, (ii) the genomes of seventeen *Agrobacterium* strains belonging to different genomospecies (Agrobacterscope database) and (iii) the genomes of *A. vitis* S4, *A. radiobacter* K84 and *S. meliloti* 1021 strains. Using such an approach, 805 regions (from 60 to 550 nt) were predicted to encode putative small RNAs. 469 were localized on the circular chromosome, 214 on the linear one, 91 on pAtC58, and 31 on pTiC58. A comparison with the expressed small transcripts identified above showed that there were 129 features in common (Figure 2). Thus, about 15 % of the transcribed sRNAs were correctly predicted.

Validation of small RNA candidates using comparative transcriptomics

Putative sRNAs were also compared to those identified by previous RNA-seq analyses performed in *A. fabrum* C58 (Wilms et al., 2012a; Lee et al., 2013). 102 had been also identified by Wilms *et al.* and 148 by Lee *et al.* Of these, 45 % of the sRNAs identified by Wilms *et al.* and 31 % of the sRNAs identified by Lee *et al.* were recovered in our analysis (Figure 2). Combining these three RNA-seq analyses, 1560 small transcripts could be identified in *A. fabrum* C58 and, amongst these, 54 were detected in all three of the RNA-seq analyses. They correspond to (i) the sRNAs largely conserved among bacteria phyla and already referred to the Rfam database, such as RNaseP, tmRNA, ssrS, (RNA454, RNA441 and RNA734 respectively); (ii) sRNAs already characterized in various agrobacteria such as AbcR1 (RNA479) (Chai and Winans, 2005; Wilms et al., 2011) or (iii) unknown sRNAs whose presence had been previously validated by northern blot analyses or by 5' and 3'-ends RACE-PCR. Moreover, twelve candidates were similar to the sRNAs previously identified in *S. meliloti* or *Rhizobium* (Ulvé et al., 2007; Valverde et al., 2008; Schlüter et al., 2010) (Table S1). Finally, combining all these RNA-seq analyses, 196 of our 1108 candidates (17.7 %) had been found previously and 828 were identified here for the first time (Figure 2), among those are the 21 shortest transcripts (length \leq 30 nt) (Liu et al., 2009; Behrens et al., 2014).

Characterization of new putative sRNAs

Among the novel small transcripts, 17 were arbitrarily chosen for validation. RACE-PCRs were performed on total RNAs. For every candidate, the 5'- and 3'-ends could be easily identified suggesting that the regions of interest were actually expressed. Evaluation of the 5'- and 3'-ends, and of the length of each transcript, highlighted (i) differences in 3'-end determination (median of difference = 36 bases), likely due to the instability of RNA 3'-end because of the 3'-ribonuclease degradation (Behrens et al., 2014; Hui et al., 2014), (ii) a very good correlation between 5'-end identification (differences \leq 3 bases) and (iii) the small lengths of the resulting transcripts (\leq 270 bases) (Table 1). Thus, the 17 candidates seem to be small transcripts and could be sRNAs.

To gain further insights into the characterization, a search for the presence of coding sequences (CDS) and putative RBSs (Ribosome Binding Site) was performed using ORFfinder. For eleven transcripts, no coding sequence could be identified, indicating that they are untranslated small RNAs. Partial open reading frames (without initiation or stop codons) could be identified for RNA535, RNA834, RNA846, RNA869 and a putative RBS could be predicted upstream RNA846 and RNA869. Nevertheless, no significant similarities could be identified, except for long hypothetical proteins that showed significant similarities with both RNA869 (5'UTR of *atu4669*) and *atu4669*, suggesting a mistake in the annotation of translation initiation. For the last two transcripts, RNA868 and RNA960, a putative RBS and complete CDSs were identified, suggesting that these two genes could encode peptides specific to *A. fabrum* C58 as no significant similarities were found in other bacterial sequences.

Secondary structures of these small transcripts were predicted using the RNAfold algorithm with the calculation of Minimal Free Energy (MFE). Base pair probabilities were analyzed as sRNAs are generally described as strongly structured. For at least 10 structures, base pair probabilities were close to one, suggesting the presence of strong hairpins in most of the transcripts analyzed (Table 1). For the seven other structures, the average of the base pair probabilities was below 0.5; four of them had been previously classified as 5'UTR RNAs and the three others were localized in the intergenic region (IGR sRNA). Sequence conservations of these putative sRNAs were then evaluated and the results showed that their presence is restricted to *Rhizobiaceae*: two of them are largely conserved (RNA834 and RNA846), six seem to be restricted to *Agrobacteria* (RNA115, RNA141, RNA154, RNA275, RNA387 and RNA869), and the others are occasionally present in some *Rhizobiaceae* and some *Agrobacteria*.

Differential expression of sRNAs

We identified the differentially expressed sRNAs using the Rockhopper software system (q-value < 0.01) (McClure et al., 2013). We centered our analysis on the 297 sRNAs also predicted by bioinformatics, detected in other transcriptomic experiments or validated here by RACE-PCR. We compared their expression levels in the stationary phase versus exponential phase and their growth in

rich medium versus minimal medium. Of the total, 58 sRNAs were shown to be differentially expressed between growth-phases: 6 were down-regulated in the exponential phase and 52 were up-regulated. In addition, 56 sRNAs were differentially regulated between growth-media: one was down-regulated in minimal medium and 55 were up-regulated (Table S2). Finally, about 20 % of the analyzed sRNAs were differentially expressed, either according to the growth-phase or the growth-media, and only 5 sRNAs (RNA1066, RNA1077, RNA1094, RNA1096 and RNA1113) were transcribed from pTiC58.

RNA1111, a small transcript expressed from the Ti plasmid, involved in virulence

The RNA1111 is transcribed from the minus strand of the pTiC58, between nucleotide 205,448 and 205,620 (RACE-PCR data) and, consequently, its genomic localization is between gene atu6186 (virE3) and atu6188 (virE0) (Figure S2). A conservation-search showed significant conservation of this sRNA. Nine out of the 15 Ti plasmids presented rna1111 homologous genes completely identical to the C58 rna1111 gene, the synteny was also conserved. All of the nine Ti plasmids are nopalinetype. An additional Ti plasmid presented a *rna1111* homolog. It displays 71% identity and was found in the CFBP5473 strain, which harbors a chrysopine-type Ti plasmid (Figure S3). Since this gene is localized inside the vir region, its expression level was measured in virulence mimicking conditions but no significant changes were observed (data not shown). Its impact on bacterial virulence was also tested. Bacterial aggressiveness was evaluated on tomato plants for mutants deleted for, or overexpressing the genomic region expressing RNA1111. qRT-PCR were performed to measured *rna1111* expression level in the different mutant strains, RNA1111 could not be detected in this deleted strain and when compared to the wild-type strain harboring the empty vector, an increase of 6.5 fold was evaluated in the strain harboring *rna1111* cloned into the pBBR1MCS-5. A similar level of aggressiveness was observed both for the mutant over-expressing rna1111 (AT199) and the wild-type strains (C58 and AT200 the C58 with the empty pBBR1MCS-5). Remarkably, the strain deleted for rnal111 (AT192) showed a drastic decrease in aggressiveness on plants, with an average of two tumors per plant whereas the wild-type presented more than 20 tumors per plant (Figure 3). A significant increase in virulence symptoms was observed when the *rnallll* deleted strain was compared with the deleted strain complemented with *rna1111* carried alone in the plasmid pBBR1MCS-5 (9.5 tumors vs 2 tumors on average, for complemented and deleted strains, respectively; p-value < 0,003). Thus, RNA1111 alone is able to restore at least partially virulence. Similar results were also obtained on tobacco and *Kalanchoe daigremontiana* plants (data not shown). Taken together, these findings indicate the involvement of RNA1111 in A. fabrum C58 virulence on plants.

mRNA targets of RNA1111

RNA1111 belongs to the trans-encoded RNA class. To further investigate its function, its putative

mRNA targets were predicted using a combination of small RNA target prediction algorithms (RNApredator, sTarPicker and IntaRNA). Annotated CDSs and regions upstream of the translation start of CDSs were analyzed. To reduce the number of false positives, we selected only the target genes found by every algorithm. This procedure led to the selection of eight target candidates, only the six localized on pTiC58 were selected for further investigations (Table S3). Putative regions of interaction are located: 91 bases upstream the start codon for *atu6014* (the *6b* gene), 32 bases upstream the start codon for *atu6179* (virC2), 31 bases upstream of the start codon for *atu6039* (trbD), 88 bases downstream of the start codon for the gene atu6072 (encoding an hypothetical protein), 64 bases downstream of the start codon for atu6127 (traA) and, 40 bases downstream of the start codon for atu6183 (virD3) (Figure 4). Hence, three candidates (the 6b gene, virC2 and virD3) are related to virulence functions and two other ones are related to the bacterial conjugative transfer of pTiC58 (traA and trbD). To validate these putative interactions, conservation searches were performed in strains harboring *rna1111* homolog genes (Table S4). In addition, we performed qRT-PCR from bacteria cultivated in rich medium and in medium mimicking virulence using primers localized downstream of predicted region of interaction. In the tested conditions, trbD was not significantly expressed. No significant variation in expression was observed for the atu6183 (virD3) and atu6179 (virC2) genes whatever the strain analyzed, whereas a decrease in expression of 12, 6 and 3.5 fold were observed for atu6127 (traA), atu6072 (hypothetical protein) and atu6014 (the 6b gene), respectively, when we compared the strain deleted for *rna1111* with the wild-type strain (Figure 4). Moreover, their expression levels were completely restored in the *rna1111* deleted strain complemented with *rna1111* carried alone in a plasmid (Figure 4). Conversely, strain over-expressing rna1111 (AT199) showed similar expression levels of target genes when compared to expression levels obtained with the wildtype strain harboring the empty vector (AT200) (data not shown). Taken together, these results suggest that the absence of RNA1111 destabilized, directly or indirectly, the mRNAs of these three genes.

Discussion

Regulation by sRNAs is now widely recognized in bacteria but only a few examples concern pathogenic bacteria interacting with plants. Hence, attempts had been made to increase our knowledge on the *Agrobacterium* RNAomes (Wilms et al., 2011; Lee et al., 2013; Möller et al., 2014). Here, we undertook an exhaustive analysis of small transcripts in the model strain *A. fabrum* C58. Our approach led us to discover numerous small transcripts, including those expressed at low levels, among those are Ti plasmids sRNA candidates. In agreement with previous data (Wilms et al., 2012a), the latter are shown here to be 6 times lower than those of the chromosomes in *A. fabrum* C58. These pTi characteristics could significantly impede an automatic detection of transcripts and consequently necessitate manual detection and validation of the 5'- and 3'-ends of new transcripts. In addition, our data were found to be quantitative as differential growth phase expression analyses were in accordance

with previous quantifications (Wilms et al., 2012a; Lee et al., 2013) and relative quantifications agreed with those performed by northern blot analyses on total RNA (Figures S4 and S5).

The 1108 small transcripts identified might be sRNAs but they could also be processed or degraded RNAs (even if small transcripts from internal mRNA were not taken into account by our procedure). Hence, their validation by other approaches was essential. We compared our sRNA-seq data with sRNApredict data and with previous RNA-seq data. These comparisons allowed the confirmation of a quarter of sRNA candidates. However, small overlaps between results were observed (Figure 2). The small overlap with predicted data likely reflects limitations of RNA prediction algorithms which generally fail to predict a large portion of sRNAs and can also present false positives (Soutourina et al., 2013). Hence, novel small transcripts identified could, therefore, be useful for improving prediction algorithms. Furthermore, the different RNAseq analyses were performed from agrobacteria cultivated in different growth conditions and with different sequencing approaches (Wilms et al., 2011; Lee et al., 2013) and unrecovered sRNAs could be due to their absence of expression in the some of growth conditions tested. Despite these features, a large proportion of previous identified sRNAs (>31 %) could be validated by our analysis, illustrating the quality of our approach.

One third of putative sRNAs are localized upstream of a CDS (5'UTR) and most of them correspond to 5'UTR longer than 100 nt (Table S1). The presence of these small transcripts has been previously described in other bacteria (Kawano et al., 2005; Gómez-Lozano et al., 2012). Long 5'UTR are not unusual in *Rhizobiaceae* (Schlüter et al., 2013). These long mRNA leader sequences may correspond to (i) processed-*cis*-regulatory elements of longer mRNAs, such as metabolite-controlled riboswitches and RNA thermometers (Bastet et al., 2011) or high-affinity protein binding sites that protect the transcripts from exoribonucleolytic degradation (for example Hfq), (ii) *trans*-encoded sRNAs resulting from parallel transcriptional output (Loh et al., 2009), or (iii) artifacts in the sequencing procedures. RACE-PCR experiments on total RNAs showed that at least RNA275, RNA387 and RNA868 are not small transcript artifacts (Table 1) and could encode short peptides (Table S1), known to be a characteristic feature in a mechanism called transcriptional attenuation (Mentz et al., 2013). Generally, small ORFs are found upstream of the CDS of metabolic proteins and 20 % of CDS with a long 5'UTR are actually involved in metabolism, five harbored a putative ORF and three present an ORF on the opposite strand, suggesting that putative sRNAs could also be the asRNAs of those peptide-encoding transcripts (RNA111, RNA573 and RNA575) (Table S1).

The *cis*-encoded RNAs, asRNAs, represented a quarter of the small transcripts (Figure 1). They were found on the opposite strand of ca. 5 % of the CDSs. Surprisingly, the expression of *repE* antisense RNA (Chai and Winans, 2005) was not detected because it was below the threshold in the growth conditions tested, suggesting *cis*-acting RNAs are underestimated. As already observed (Wilms et al., 2011; Schlüter et al., 2013), asRNAs are over-represented in pTiC58 (Figure 1) and 44 % were found to be opposite to genes involved in virulence (Table S1). Hence, the asRNA regulatory mechanisms

might be preferred for these functional categories.

A ratio of 13.8 CDSs per *trans*-encoded RNA candidate was obtained in *A. fabrum* C58 in agreement with the previous data (Mitschke et al., 2011; Schlüter et al., 2013). Sixteen could encoded peptides, since they presented a putative ORF with significant similarities to the already predicted proteins in *Rhizobiaceae* (Table S1). In addition, the conservation of sRNAs seems to be restricted to related alpha-proteobacterial species, as previously observed in *S. meliloti*. However, the primary nucleotide sequence analyses are probably not sufficient for performing an accurate conservation-search and the use of sRNA classification, according to the secondary structure domains, might be more relevant. Among the novel and unpredicted *trans*-encoded sRNA candidates, the transcription, as *trans*-encoded RNAs of fourteen of them, was nevertheless validated by RACE-PCR (Table 1) and, for at least nine of them (> 64 %), the absence of CDSs and their very probable strong secondary structures suggest that they are true *trans*-encoded RNAs.

Interestingly, one trans-encoded RNA, RNA1111, was validated by RACE-PCR and further characterized for its involvement in virulence. A longer transcript named pTi 205576R and corresponding to *rna1111*, has been previously identified, and, in agreement with our data, its expression has previously been shown not modulated in condition mimicking virulence (Lee et al., 2013). Nevertheless, we showed a drastic decrease in virulence symptoms on several plants using the strain deleted for *rna1111*, and the complemented strain restored the virulence. This suggested the involvement of RNA1111 in the virulence. However, as the vir box of virE0 is inside rna1111 gene (Steck et al., 1988), the deletion also removed this vir box, and the virulence decrease could have been due to a polar effect on *virE0*. Nonetheless, the in *trans* complementation with the sole *rna1111* gene is enough to restore -at least partially- the virulence (Figure 3). This establishes the actual involvement of RNA1111 in the virulence phenotype. Moreover, the in trans complementation with the sole rnall11 gene led to recover an expression level of the putative target genes similar to those of the wild-type (Figure 4C). We could also noticed that in addition to the absence of modification in virulence phenotype for the strain over-expressing RNA1111, any significant increase of mRNA target expression levels could be observed in this strain when compared to the wild-type strain. This suggests that in the wild-type strain the RNA1111 copy number could be already sufficient to regulate mRNA targets. Recently, Möller P. et al. showed an enrichment of several mRNAs by Hfq3xFlag, including virC2, virD3 and traA mRNAs (Möller et al., 2014). Interestingly, we presently suspect these mRNAs to be putative RNA1111 targets (Figure 4), strongly suggesting their interactions with the Hfq protein and their putative regulations by sRNA, and potentially by RNA1111. Among the predicted RNA1111 mRNA targets, three candidates are connected with virulence but are not essential for tumorigenicity: two vir genes and the 6b gene (atu6014). VirD3 (Atu6183) was not tested for the modulation of aggressiveness (Vogel and Das, 1992). VirC2 stimulates the translocation of T-DNA into plant cells, and its DNA binding function has been shown to be important for bacterial aggressiveness in different

plants (Lu et al., 2009). The 6b protein is involved in tumor development, with the accumulation of sugars and phenolic compounds (Clément et al., 2007), and the interaction with nuclear proteins (Kitakura et al., 2002; Terakura et al., 2007). It also modulates auxin transport (Kakiuchi et al., 2006) and the microRNA pathway in plants (Wang et al., 2011). The 6b protein is produced from transformed plant cells and its regulation, in the bacterial background, by RNA1111 is surprising. However, the translocation of sRNAs into infected plant cells has been suggested previously and we might speculate that this is the case for RNA1111. Although we cannot rule out that RNA1111 controls other virulence factors, we can assume that the regulation of all these targets together by RNA1111 could have a synergic impact on virulence. The regions of interaction with RNA1111 are localized upstream or downstream of their putative RBSs and could facilitate the binding of ribosomes by modifying mRNA secondary structures and thus, facilitating the translation. Modifying the interaction between one sRNA and its mRNA target could lead to the degradation, or the absence of degradation, of mRNA (Desnoyers et al., 2013). Based on qRT-PCR data, we suspect that RNA1111 could bind to *virD3* and *virC2* mRNA and thus could stimulate their translation without affecting their stability, whereas the putative binding of RNA1111 on atu6014 mRNA stabilizes this transcript. However, to definitively validate a direct regulation of RNA1111 on mRNA targets, RNA-RNA interaction experiments are still required. traA mRNA has also been shown to be down-regulated in the absence of RNA1111. This gene is part of the *tra* operon encoding the type IV secretion system for the conjugative transfer of pTiC58, suggesting that RNA1111 also regulates the conjugation. Noticeably, the virulence and the conjugation of Ti plasmid were actually shown to be regulated, in a coordinated way, by TraM (Cho and Winans, 2005). Our work highlights a potential new level of coordinated regulation of the essential pTi functions by a small trans-encoded RNA that is, furthermore, largely conserved among Ti plasmids. Thus, this suggests that RNA1111, and possibly small RNAs in general, play a major role in the infection of plants and in the dissemination of Ti plasmids.

Materials and methods

Bacterial strains and growth conditions

Bacterial strains used in this study are the *A. fabrum* C58 wild type strain and its derivatives listed in Table S5. Bacteria were grown with shaking (150 rpm) at 28°C in YPG rich media (5 g/L yeast extract; 5 g/L peptone; 10 g/L glucose; pH 7.2) or in AT minimal media (80 mM KH₂PO₄; 0.65 mM MgSO₄, 7H₂O; 18 μ M FeSO₄, 7H₂O; 70 μ M CaCl₂, 2H₂O; MnCl₂, 10 μ M 4H₂O; pH 7.2) supplemented with 10 mM succinate as the carbon source and 10 mM ammonium sulfate as the nitrogen source (Petit et al., 1978). Conditions mimicking virulence are those previously described by (Lee et al., 2013). Media were supplemented, when needed, with appropriate antibiotics (25 μ g/mL gentamicin, 25 μ g/mL neomycin, and 25 μ g/mL kanamycin). *Escherichia coli* DH5 α strain was used for cloning experiments in classical growth conditions.

Vector constructions

The Ti plasmid DNA fragment, from 205,367 bp to 205,626 bp (corresponding to *rna1111*), was amplified by PCR using primers 186 and 187 (Table S6). Fragments were purified and digested using *Sal*I and *BamH*I enzymes, and then cloned into a pBBR1MCS-5 plasmid giving rise to pARA005 (Table S5).

Construction of mutant strain deleted for rna1111

The *A. fabrum* C58 strain was used to construct a mutant, as previously described (Lassalle et al., 2011). The genomic region corresponding to transcript RNA1111 was deleted and replaced by the Neo-Kan resistance gene *nptII* (Table S6), resulting in the strain AT192 (Table S5).

Small RNA selection and abundant sRNA depletions

Total RNAs were extracted as previously described (Hommais et al., 2008), and DNA contamination was eliminated with a DNase I treatment (Roche). 100 µg of total RNAs were separated on a 15 % denaturing polyacrylamide gel (7M urea, 1X TBE) and RNAs, with a size ranging from 25 to 500 nt, were extracted (representing 1.4 % of total RNAs). Briefly, a piece of gel containing RNAs was broken mechanically and incubated, for 4 h, in 500 µL of 10 mM Tris-HCl, 1 mM EDTA, 300 mM NaCl buffer with soft shaking at 15°C. Size-selected RNAs were washed twice and ethanol precipitated. 5S RNA and tRNAs depletions were performed from 2.5 µg of size-selected RNAs by reverse transcription treatment using the First strand cDNA synthesis kit (Life biosciences), with 100 ng of primer 5S and 50 ng of each of the other primers representing the tRNAs (Table S6). An RNAse H treatment was then performed and newly synthesized DNAs were eliminated with a DNAse I treatment (DNA*-free* kit, Ambion). Between each step, RNA quantities and qualities were checked using a Nanodrop 2000 spectrophotometer (Thermo Scientific) and a 2100 bioanalyzer with a, RNA 6000 Nano kit (Agilent). RNAs were stored at – 80°C until further uses.

Library preparation and sequencing

The library preparation was an adaptation of the Illumina directional mRNA-Seq protocol, using the DGE-Small RNA sample preparation kit. 200 ng of selected small RNAs were treated with Antarctic Phosphatase (NEB, cat.#M0289S) and T4 PolyNucleotide Kinase (NEB, cat.#M0201S) and purified on an RNeasy MinElute spin column (Qiagen, cat.#74204). Ligations with Illumina v1.5 sRNA 3' adapter and sRNA 5' adapter, reverse transcription, PCR amplification and purification with Agencourt AMPure beads (Beckman, A63880) were carried out as described in the Illumina directional mRNA-Seq protocol (Illumina). One sequence run per library was performed using the Illumina Genome Analyzer II (Perkins et al., 2009). An average of 24,137,549 reads of 36 bases per run were obtained, corresponding to a total of 193.100 million reads (Table S7).

Data analyses

Raw sequence files were first converted to standard fastq format with Sanger quality encoding and then summary statistics of read quality were generated for each sequencing lane using random samples of 500,000 reads. Read quality of all experiments were found to be sufficiently high for the sequence reads to be directly processed by the mapping software. Reads were mapped on the A. fabrum C58 genome (RefSeq Acc: NC 003062, NC 003063, NC 003064 and NC 003065) using SSAHA2 software, with the '-diff' parameter set to 0 to ensure that only reads mapping to a single location in the genome were retained (Uniquely Mapped Reads, UMR) (Ning et al., 2001). Depending on the sequencing run, 95.24 % to 99.2 % of reads were mapped at least once on the genome. The coverage of UMR (14.71 %) along the genome and the number of UMR located in each genomic object (encompassing CDS, rRNA, tRNA, and predicted sRNA) were then computed using R/Bioconductor packages Rsamtools and GenomicRanges (Gentleman et al., 2004) and visualized in the Agrobacterscope database (Vallenet et al, 2013). Replicon repartition showed that 72.5 % of UMRs mapped on the circular chromosome, 25 % on the linear one, 1.6 % on the pAt and 0.9 % on the pTi. Rockhopper software was used for the detection of novel small transcripts and the differential expression analyses (McClure et al., 2013). The data followed MIAME guidelines and were registered at ArrayExpress under the accession number: E-MTAB-2947.

Determination of 5'- and 3'-ends by RACE-PCR

A phenol/chloroform extraction followed by an ethanol precipitation was performed between each step of the following protocol and enzymatic reactions were performed according to the manufacturer's recommendations. Cleaved mono-phosphate RNAs – isolated from bacteria grown in rich medium - were eliminated from 12 µg of total DNA-free RNAs using the XRN-I exoribonuclease (Biolabs). The resulting tri-phosphate RNAs were treated with Tobacco Acid Pyrophosphatase (TAP) (Epicentre) and ligated using the T4 RNA ligase (Biolabs). Circular RNAs were reverse transcribed using the SuperScript III reverse transcriptase (First strand cDNA synthesis, Life Technologies) and primers specific to the gene of interest. Then, the 5'- and the 3'-end junction was specifically amplified by PCR (Table S6). The resulting RACE-PCR products were cloned into the pGEMT-Easy (Promega) and sequenced using M13fwd primer (Beckman Coulter Genomics). An average of five independent clones per RNA was analyzed.

qRT-PCR analysis

SuperScript III (Life Technologies) was used with random hexamer primers to reverse transcribed 2.5 μ g of the total RNAs, according to the manufacturer's protocol (First strand cDNA synthesis). The DyNAmo Flash SYBR Green qPCR kit (Thermo scientific) was used with 0.1 μ L of the cDNA reaction mixture and gene-specific primers (Table S6). The thermal cycling reactions were performed using the LC480 Lightcycler from Roche according to the following conditions: an initial step at 95°C

for 7 min, followed by 45 cycles at 95°C for 10 s, 58°C for 15 s and 72°C for 20 s. Normalization of data was performed using pAW and genes atu0231 and atu1924 as their expressions are constant in the growth conditions tested. The specificity of the PCR primers was verified with a melting curve analysis.

Virulence assays on plants

The plants inoculated with *A. fabrum* C58 were three week old tomatoes (Marmande) grown in controlled environment chambers with 70 % hygrometry and a 16 h photoperiod, at 25°C, followed by 8 h at 22°C. Virulence assays were performed as previously described (Haudecoeur et al., 2009). Briefly, plants were wounded at the top of the stem just under the second leaves, and infected with around 10⁶ bacterial cells. Tumors were observed and counted 3 weeks post-infection. Ten plants were used for each strain and each plant was classified according to severity symptom categories. Each virulence assay was performed in three independent experiments. Kruskal-Wallis statistic tests were applied to determine significant differences between the virulence assays. These assays were also performed on 6-8 week old *Nicotiana tabacum* burley and 2 month old *Kalanchoe daigemontiana* and symptoms were evaluated 5 weeks and 3-4 weeks post-infection, respectively.

Prediction of bacterial small RNA targets

mRNA target genes of sRNAs were predicted from the *A. fabrum* C58 genome (accession number NC_003062, NC_003063, NC_003064 and NC_003065) using IntaRNA (Busch et al., 2008), sTarPicker (Ying et al., 2011) and RNApredator (Eggenhofer et al., 2011) algorithms. We arbitrarily selected the 350 most significant mRNA targets provided by each algorithm. The mRNA targets, found by the three algorithms, were chosen for further analyses.

Secondary structure predictions and similarity searches

RNA secondary structures were predicted using the RNAfold algorithm. Basic options were chosen for folding. Sequence similarity searches were performed with the BLAST algorithm on the MAGE database (Vallenet et al., 2013).

Acknowledgments

This study was supported by the EcoGenome project of the French Agence Nationale de la Recherche [grant number ANR-BLAN-08-0090], PEPS INSB and PEPII INSB-INEE-INSMI-INS2I " Bio-Math-Info" funding from the CNRS. M. Dequivre received a doctoral grant from the French *Ministère de l'Education nationale de l'Enseignement Supérieur et de la Recherche*. The authors would like to thank J. Baude, T. Campillo and N. Poussereau from UMR5240, C. Oger from PRABI and the LABGeM / MicroScope for skillful technical assistance. We are grateful for the stimulating discussions and comments with W. Nasser, S. Reverchon and C. Lavire.

Literature Cited

- Albrecht, M., Sharma, C.M., Reinhardt, R., Vogel, J., and Rudel, T. 2010. Deep sequencing-based discovery of the Chlamydia trachomatis transcriptome. Nucleic Acids Res 38:868-877.
- Backofen, R., Amman, F., Costa, F., Findeiß, S., Richter, A.S., and Stadler, P.F. 2014. Bioinformatics of prokaryotic RNAs. RNA Biol 11:470-483.
- Bastet, L., Dubé, A., Massé, E., and Lafontaine, D.A. 2011. New insights into riboswitch regulation mechanisms. Mol Microbiol 80:1148-1154.
- Becker, A., Overlöper, A., Schlüter, J.P., Reinkensmeier, J., Robledo, M., Giegerich, R., Narberhaus, F., and Evguenieva-Hackenberg, E. 2014. Riboregulation in plant-associated α-proteobacteria. RNA Biol 11.
- Behrens, S., Widder, S., Mannala, G.K., Qing, X., Madhugiri, R., Kefer, N., Abu Mraheil, M., Rattei, T., and Hain, T. 2014. Ultra deep sequencing of Listeria monocytogenes sRNA transcriptome revealed new antisense RNAs. PLoS One 9:e83979.
- Busch, A., Richter, A.S., and Backofen, R. 2008. IntaRNA: efficient prediction of bacterial sRNA targets incorporating target site accessibility and seed regions. Bioinformatics 24:2849-2856.
- Chai, Y., and Winans, S. 2005. A small antisense RNA downregulates expression of an essential replicase protein of an Agrobacterium tumefaciens Ti plasmid. Mol Microbiol 56:1574-1585.
- Cho, H., and Winans, S.C. 2005. VirA and VirG activate the Ti plasmid repABC operon, elevating plasmid copy number in response to wound-released chemical signals. Proc Natl Acad Sci U S A 102:14843-14848.
- Clément, B., Perot, J., Geoffroy, P., Legrand, M., Zon, J., and Otten, L. 2007. Abnormal accumulation of sugars and phenolics in tobacco roots expressing the Agrobacterium T-6b oncogene and the role of these compounds in 6b-induced growth. Mol Plant Microbe Interact 20:53-62.
- del Val, C., Rivas, E., Torres-Quesada, O., Toro, N., and Jiménez-Zurdo, J.I. 2007. Identification of differentially expressed small non-coding RNAs in the legume endosymbiont Sinorhizobium meliloti by comparative genomics. Mol Microbiol 66:1080-1091.
- Desnoyers, G., Bouchard, M.P., and Massé, E. 2013. New insights into small RNA-dependent translational regulation in prokaryotes. Trends Genet 29:92-98.
- Eggenhofer, F., Tafer, H., Stadler, P.F., and Hofacker, I.L. 2011. RNApredator: fast accessibility-based prediction of sRNA targets. Nucleic Acids Res 39:W149-154.
- Gentleman, R.C., Carey, V.J., Bates, D.M., Bolstad, B., Dettling, M., Dudoit, S., Ellis, B., Gautier, L., Ge, Y., Gentry, J., Hornik, K., Hothorn, T., Huber, W., Iacus, S., Irizarry, R., Leisch, F., Li, C., Maechler, M., Rossini, A.J., Sawitzki, G., Smith, C., Smyth, G., Tierney, L., Yang, J.Y., and Zhang, J. 2004. Bioconductor: open software development for computational biology and bioinformatics. Genome Biol 5:R80.
- Goodner, B., Hinkle, G., Gattung, S., Miller, N., Blanchard, M., Qurollo, B., Goldman, B., Cao, Y., Askenazi, M., Halling, C., Mullin, L., Houmiel, K., Gordon, J., Vaudin, M., Iartchouk, O., Epp, A., Liu, F., Wollam, C., Allinger, M., Doughty, D., Scott, C., Lappas, C., Markelz, B., Flanagan, C., Crowell, C., Gurson, J., Lomo, C., Sear, C., Strub, G., Cielo, C., and Slater, S. 2001. Genome sequence of the plant pathogen and biotechnology agent Agrobacterium tumefaciens C58. Science 294:2323-2328.
- Gómez-Lozano, M., Marvig, R.L., Molin, S., and Long, K.S. 2012. Genome-wide identification of novel small RNAs in Pseudomonas aeruginosa. Environ Microbiol 14:2006-2016.
- Haudecoeur, E., Tannières, M., Cirou, A., Raffoux, A., Dessaux, Y., and Faure, D. 2009. Different regulation and roles of lactonases AiiB and AttM in Agrobacterium tumefaciens C58. Mol Plant Microbe Interact 22:529-537.
- Hommais, F., Oger-Desfeux, C., Van Gijsegem, F., Castang, S., Ligori, S., Expert, D., Nasser, W., and Reverchon, S. 2008. PecS is a global regulator of the symptomatic phase in the phytopathogenic bacterium Erwinia chrysanthemi 3937. J Bacteriol 190:7508-7522.
- Hui, M.P., Foley, P.L., and Belasco, J.G. 2014. Messenger RNA Degradation in Bacterial Cells. Annu Rev Genet.
- Kakiuchi, Y., Gàlis, I., Tamogami, S., and Wabiko, H. 2006. Reduction of polar auxin transport in tobacco by the tumorigenic Agrobacterium tumefaciens AK-6b gene. Planta 223:237-247.
- Kawano, M., Storz, G., Rao, B.S., Rosner, J.L., and Martin, R.G. 2005. Detection of low-level promoter activity within open reading frame sequences of Escherichia coli. Nucleic Acids Res 33:6268-6276.
- Kitakura, S., Fujita, T., Ueno, Y., Terakura, S., Wabiko, H., and Machida, Y. 2002. The protein encoded by oncogene 6b from Agrobacterium tumefaciens interacts with a nuclear protein of tobacco. Plant Cell 14:451-463.
- Kovach, M.E., Elzer, P.H., Hill, D.S., Robertson, G.T., Farris, M.A., Roop, R.M., and Peterson, K.M. 1995. Four new derivatives of the broad-host-range cloning vector pBBR1MCS, carrying different antibioticresistance cassettes. Gene 166:175-176.

- Kröger, C., Dillon, S.C., Cameron, A.D., Papenfort, K., Sivasankaran, S.K., Hokamp, K., Chao, Y., Sittka, A., Hébrard, M., Händler, K., Colgan, A., Leekitcharoenphon, P., Langridge, G.C., Lohan, A.J., Loftus, B., Lucchini, S., Ussery, D.W., Dorman, C.J., Thomson, N.R., Vogel, J., and Hinton, J.C. 2012. The transcriptional landscape and small RNAs of Salmonella enterica serovar Typhimurium. Proc Natl Acad Sci U S A 109:E1277-1286.
- Lassalle, F., Campillo, T., Vial, L., Baude, J., Costechareyre, D., Chapulliot, D., Shams, M., Abrouk, D., Lavire, C., Oger-Desfeux, C., Hommais, F., Guéguen, L., Daubin, V., Muller, D., and Nesme, X. 2011. Genomic species are ecological species as revealed by comparative genomics in Agrobacterium tumefaciens. Genome Biol Evol.
- Lee, K., Huang, X., Yang, C., Lee, D., Ho, V., Nobuta, K., Fan, J.B., and Wang, K. 2013. A genome-wide survey of highly expressed non-coding RNAs and biological validation of selected candidates in Agrobacterium tumefaciens. PLoS One 8:e70720.
- Li, W., Ying, X., Lu, Q., and Chen, L. 2012. Predicting sRNAs and their targets in bacteria. Genomics Proteomics Bioinformatics 10:276-284.
- Liu, J., Livny, J., Lawrence, M., Kimball, M., Waldor, M., and Camilli, A. 2009. Experimental discovery of sRNAs in Vibrio cholerae by direct cloning, 5S/tRNA depletion and parallel sequencing. Nucleic Acids Res 37:e46.
- Loh, E., Dussurget, O., Gripenland, J., Vaitkevicius, K., Tiensuu, T., Mandin, P., Repoila, F., Buchrieser, C., Cossart, P., and Johansson, J. 2009. A trans-acting riboswitch controls expression of the virulence regulator PrfA in Listeria monocytogenes. Cell 139:770-779.
- Lu, J., den Dulk-Ras, A., Hooykaas, P.J., and Glover, J.N. 2009. Agrobacterium tumefaciens VirC2 enhances T-DNA transfer and virulence through its C-terminal ribbon-helix-helix DNA-binding fold. Proc Natl Acad Sci U S A 106:9643-9648.
- McClure, R., Balasubramanian, D., Sun, Y., Bobrovskyy, M., Sumby, P., Genco, C.A., Vanderpool, C.K., and Tjaden, B. 2013. Computational analysis of bacterial RNA-Seq data. Nucleic Acids Res 41:e140.
- Mentz, A., Neshat, A., Pfeifer-Sancar, K., Pühler, A., Rückert, C., and Kalinowski, J. 2013. Comprehensive discovery and characterization of small RNAs in Corynebacterium glutamicum ATCC 13032. BMC Genomics 14:714.
- Mitschke, J., Georg, J., Scholz, I., Sharma, C.M., Dienst, D., Bantscheff, J., Voss, B., Steglich, C., Wilde, A., Vogel, J., and Hess, W.R. 2011. An experimentally anchored map of transcriptional start sites in the model cyanobacterium Synechocystis sp. PCC6803. Proc Natl Acad Sci U S A 108:2124-2129.
- Mousavi, S.A., Österman, J., Wahlberg, N., Nesme, X., Lavire, C., Vial, L., Paulin, L., de Lajudie, P., and Lindström, K. 2014. Phylogeny of the *Rhizobium-Allorhizobium-Agrobacterium* clade supports the delineation of *Neorhizobium* gen. nov. Syst Appl Microbiol.
- Möller, P., Overlöper, A., Förstner, K.U., Wen, T.N., Sharma, C.M., Lai, E.M., and Narberhaus, F. 2014. Profound Impact of Hfq on Nutrient Acquisition, Metabolism and Motility in the Plant Pathogen Agrobacterium tumefaciens. PLoS One 9:e110427.
- Ning, Z., Cox, A.J., and Mullikin, J.C. 2001. SSAHA: a fast search method for large DNA databases. Genome Res 11:1725-1729.
- Overlöper, A., Kraus, A., Gurski, R., Wright, P.R., Georg, J., Hess, W.R., and Narberhaus, F. 2014. Two separate modules of the conserved regulatory RNA AbcR1 address multiple target mRNAs in and outside of the translation initiation region. RNA Biol 11.
- Perkins, T., Kingsley, R., Fookes, M., Gardner, P., James, K., Yu, L., Assefa, S., He, M., Croucher, N., Pickard, D., Maskell, D., Parkhill, J., Choudhary, J., Thomson, N., and Dougan, G. 2009. A strand-specific RNA-Seq analysis of the transcriptome of the typhoid bacillus Salmonella typhi. PLoS Genet 5:e1000569.
- Petit, A., Tempe, J., Kerr, A., Holsters, M., Vanmontagu, M., and Schell, J. 1978. SUBSTRATE INDUCTION OF CONJUGATIVE ACTIVITY OF AGROBACTERIUM-TUMEFACIENS TI PLASMIDS. Nature 271:570-572.
- Pitzschke, A., and Hirt, H. 2010. New insights into an old story: Agrobacterium-induced tumour formation in plants by plant transformation. EMBO J 29:1021-1032.
- Rutherford, K., Parkhill, J., Crook, J., Horsnell, T., Rice, P., Rajandream, M.A., and Barrell, B. 2000. Artemis: sequence visualization and annotation. Bioinformatics 16:944-945.
- Schlüter, J.P., Reinkensmeier, J., Barnett, M.J., Lang, C., Krol, E., Giegerich, R., Long, S.R., and Becker, A. 2013. Global mapping of transcription start sites and promoter motifs in the symbiotic αproteobacterium Sinorhizobium meliloti 1021. BMC Genomics 14:156.
- Schlüter, J.P., Reinkensmeier, J., Daschkey, S., Evguenieva-Hackenberg, E., Janssen, S., Jänicke, S., Becker, J.D., Giegerich, R., and Becker, A. 2010. A genome-wide survey of sRNAs in the symbiotic nitrogenfixing alpha-proteobacterium Sinorhizobium meliloti. BMC Genomics 11:245.

- Schmidtke, C., Findeiss, S., Sharma, C.M., Kuhfuss, J., Hoffmann, S., Vogel, J., Stadler, P.F., and Bonas, U. 2012. Genome-wide transcriptome analysis of the plant pathogen Xanthomonas identifies sRNAs with putative virulence functions. Nucleic Acids Res 40:2020-2031.
- Soutourina, O.A., Monot, M., Boudry, P., Saujet, L., Pichon, C., Sismeiro, O., Semenova, E., Severinov, K., Le Bouguenec, C., Coppée, J.Y., Dupuy, B., and Martin-Verstraete, I. 2013. Genome-wide identification of regulatory RNAs in the human pathogen Clostridium difficile. PLoS Genet 9:e1003493.
- Steck, T.R., Morel, P., and Kado, C.I. 1988. Vir box sequences in Agrobacterium tumefaciens pTiC58 and A6. Nucleic Acids Res 16:8736.
- Storz, G., Vogel, J., and Wassarman, K.M. 2011. Regulation by small RNAs in bacteria: expanding frontiers. Mol Cell 43:880-891.
- Terakura, S., Ueno, Y., Tagami, H., Kitakura, S., Machida, C., Wabiko, H., Aiba, H., Otten, L., Tsukagoshi, H., Nakamura, K., and Machida, Y. 2007. An oncoprotein from the plant pathogen agrobacterium has histone chaperone-like activity. Plant Cell 19:2855-2865.
- Ulvé, V., Sevin, E., Chéron, A., and Barloy-Hubler, F. 2007. Identification of chromosomal alphaproteobacterial small RNAs by comparative genome analysis and detection in Sinorhizobium meliloti strain 1021. BMC Genomics 8:467.
- Uplekar, S., Rougemont, J., Cole, S.T., and Sala, C. 2013. High-resolution transcriptome and genome-wide dynamics of RNA polymerase and NusA in Mycobacterium tuberculosis. Nucleic Acids Res 41:961-977.
- Vallenet, D., Belda, E., Calteau, A., Cruveiller, S., Engelen, S., Lajus, A., Le Fèvre, F., Longin, C., Mornico, D., Roche, D., Rouy, Z., Salvignol, G., Scarpelli, C., Thil Smith, A.A., Weiman, M., and Médigue, C. 2013. MicroScope--an integrated microbial resource for the curation and comparative analysis of genomic and metabolic data. Nucleic Acids Res 41:D636-647.
- Valverde, C., Livny, J., Schlüter, J., Reinkensmeier, J., Becker, A., and Parisi, G. 2008. Prediction of Sinorhizobium meliloti sRNA genes and experimental detection in strain 2011. BMC Genomics 9:416.
- Vogel, A.M., and Das, A. 1992. The Agrobacterium tumefaciens virD3 gene is not essential for tumorigenicity on plants. J Bacteriol 174:5161-5164.
- Wang, M., Soyano, T., Machida, S., Yang, J.Y., Jung, C., Chua, N.H., and Yuan, Y.A. 2011. Molecular insights into plant cell proliferation disturbance by Agrobacterium protein 6b. Genes Dev 25:64-76.
- Wilms, I., Voss, B., Hess, W.R., Leichert, L.I., and Narberhaus, F. 2011. Small RNA-mediated control of the Agrobacterium tumefaciens GABA binding protein. Mol Microbiol 80:492-506.
- Wilms, I., Overlöper, A., Nowrousian, M., Sharma, C.M., and Narberhaus, F. 2012a. Deep sequencing uncovers numerous small RNAs on all four replicons of the plant pathogen Agrobacterium tumefaciens. RNA Biol 9:446-457.
- Wilms, I., Möller, P., Stock, A.M., Gurski, R., Lai, E.M., and Narberhaus, F. 2012b. Hfq influences multiple transport systems and virulence in the plant pathogen Agrobacterium tumefaciens. J Bacteriol 194:5209-5217.
- Wood, D., Setubal, J., Kaul, R., Monks, D., Kitajima, J., Okura, V., Zhou, Y., Chen, L., Wood, G., Almeida, N.J., Woo, L., Chen, Y., Paulsen, I., Eisen, J., Karp, P., Bovee, D.S., Chapman, P., Clendenning, J., Deatherage, G., Gillet, W., Grant, C., Kutyavin, T., Levy, R., Li, M., McClelland, E., Palmieri, A., Raymond, C., Rouse, G., Saenphimmachak, C., Wu, Z., Romero, P., Gordon, D., Zhang, S., Yoo, H., Tao, Y., Biddle, P., Jung, M., Krespan, W., Perry, M., Gordon-Kamm, B., Liao, L., Kim, S., Hendrick, C., Zhao, Z., Dolan, M., Chumley, F., Tingey, S., Tomb, J., Gordon, M., Olson, M., and Nester, E. 2001. The genome of the natural genetic engineer Agrobacterium tumefaciens C58. Science 294:2317-2323.
- Ying, X., Cao, Y., Wu, J., Liu, Q., Cha, L., and Li, W. 2011. sTarPicker: a method for efficient prediction of bacterial sRNA targets based on a two-step model for hybridization. PLoS One 6:e22705.
- Zeng, Q., and Sundin, G.W. 2014. Genome-wide identification of Hfq-regulated small RNAs in the fire blight pathogen Erwinia amylovora discovered small RNAs with virulence regulatory function. BMC Genomics 15:414.

Author-Recommended Internet Resources

Agrobacterscope: https://www.genoscope.cns.fr/agc/microscope/home/index.php RNAfold: http://rna.tbi.univie.ac.at/cgi-bin/RNAfold.cgi ORFfinder: http://www.ncbi.nlm.nih.gov/gorf/gorf.html Rfam database: http://rfam.xfam.org/ RNApredator: http://rna.tbi.univie.ac.at/RNApredator2/target_search.cgi sTarPicker: http://ccb.bmi.ac.cn/starpicker/prediction.php IntaRNA: http://rna.informatik.uni-freiburg.de/IntaRNA/Input.jsp

Tables:

sRNA	Strand	Position sRNA- seq	Position RACE- PCR	Size sRNA- seq	Size RACE- PCR	Secondary structure prediction (RNAfold)
			Circular chromosor	ne		
RNA115	-	534338-534448	534375-534449	110	74	on Contraction a
RNA141	+	693246-693312	693248-693381	66	133	de contrato
RNA154	+	762537-762638	762538-762623	101	85	
RNA275	+	1279489-1279713	1279489-1279716	224	227	Contra Co
RNA387	-	1757820-1757974	1757829-1757973	154	144	or Come Province
RNA535	+	2409060-2409165	2409062-2409193	105	131	Common Common a
RNA576	-	2659267-2659331	2659260-2659323	64	63	General a
			Linear chromosom	e		
RNA779	-	1168075-1168226	1168156-1168227	141	71	

Table 1: Validation of novel sRNAs by RACE-PCR

RNA807	+	1310255-1310367	1310256-1310349	112	93	
RNA834	+	1583219-1583390	1583222-1583346	171	124	a
RNA846	+	1647094-1647233	1647102-1647289	139	187	a
RNA847	-	1648470-1648623	1648521-1648622	153	101	a
RNA868	+	1824963-1825233	1824964-1825182	270	218	a a
RNA869	+	1830152-1830293	1830152-1830354	141	202	soon of the second seco
			At plasmid			
RNA960	+	214071-214308	214074-214288	237	214	A a
RNA1028	-	492222-492349	492186-492346	127	160	a a
			Ti plasmid			
RNA1111	-	205455-205545	205448-205620	90	173	Contrological Control

^a secondary structure with base pair probabilities close to one

Figures:



Figure 1: Distribution of the 1108 small transcripts among the *A fabrum* C58 genome according to sRNA categories. Histograms represent the percentage of each sRNA category per replicon (circular and linear chromosomes, and pAtC58 and pTiC58). Numbers correspond to the number of sRNAs found in each category: 5'UTR-RNA, UTR3'-RNA (UnTranslated Regions), asRNA (antisense RNA) and *trans*-encoded RNA.



Figure 2: Comparison between the different RNA-seq identifications of sRNAs and sRNAs prediction in *A. fabrum* C58. The Venn diagram represents sRNAs identified in common between the prediction data and the sRNA-seq data performed in this study and the RNA-seq data obtained previously (Wilms et al., 2012a; Lee et al., 2013).



Figure 3: Tests of virulence on tomatoes for *rna1111* mutants. Wild-type strain (WT), wild type strain containing empty cloning vector pBBR1MCS-5 (AT200), strains deleted for *rna1111* (AT192), strain overexpressing *rna1111* (AT199) and deleted strain complemented with the pBBR1MCS-5 carrying *rna1111* alone (AT203) were all tested for their ability to induce tumors on 3 week old tomato stems. Bacterial genotypes showing statistical differences (Kruskal-Wallis test) in their aggressiveness are noted with three stars (comparison to wild-type strain) or three hashtags (comparison to deleted strain), corresponding to a p-value ≤ 0.003 .



Figure 4: Regions of interaction between predicted target mRNAs and RNA1111. A. Sequences around the start codon of predicted target genes are presented with the localization of the region of interaction with RNA1111. Fold changes, measured by RT-qPCR in the strain deleted for *rna1111* versus the wild-type strain, are indicated in brackets (culture in rich medium and in virulence induction medium). B. The predicted secondary of RNA1111 is presented with the localization of the predicted regions of interaction for each putative target mRNA. ND Not determined. C. Measurement of differential expression level by quantitative real-time polymerase chain reaction (qRT-PCR) for *atu6014, atu6072, atu6127, atu6179* and *atu6184* compared with the wild-type strain (C58 or AT200, respectively). In grey were fold changes between the deleted strain Δ *rna1111* (AT192) and the wild-type strain C58 and in black were fold changes between the complemented strain Δ *rna1111*/pBBR1MCS5::*rna1111* (AT203) and the wild-type strain (AT200). Fold change was calculated with the 2^{-ΔΔCT} method. Shown here are the mean values and standard error of at least two biological replicates. * p<0.01.

Supplementary data:

Table S1: Lists of sRNA candidates identified from bacteria grown in rich and minimal media in exponential and stationary phases

In a separate file: annexes pages 197-219)

					Stationary vs	s Exponential		
ncRNA ID		RPK	KM		pha	ases	AT vs. YF	G media
					1			
	YPG_	YPG_	AT_	AT_	Fold		Fold	
	Exponential	Stationary	Exponential	Stationary	Change	q_Value	Change	q_Value
Circular								
RNA35	151	21	576	2738	130.38	0	4.75	9.15E-108
RNA68	170	756	200	871	1.15	NS	4.36	1.28E-08
RNA79	504	548	353	1669	3.05	NS	4.73	6.28E-03
RNA84	258	442	173	1567	3.55	7.30E-07	9.06	1.62E-13
RNA86	1489	2858	2074	22432	7.85	3.19E-23	10.82	7.21E-28
RNA98	728	2266	451	2807	1 24	NS	6.22	1 25E-03
RNA118	142	380	452	13	-33 33	9.04E-07	-33 33	3 93E-10
RNA133	305	1034	560	125	-8 33	3 88E-04	-4 54	NS
DNA126	188	004	121	000	1 10	NS	7 56	5 00E 12
DNA160	6187	17608	6600	181634	10.26	1 11E 40	7.50	9.99E-13
$\mathbf{N}\mathbf{N}\mathbf{A}109^{a}$	0107	1/090	0090	101054	10.20	1.11E-40	27.13	0.00E-04
KINA198	83 77	105	443	3432	33.32	0	1.13	IN S
RNA212	//	275	49	2653	9.65	0	54.14	0
RNA216	1401	2645	1902	10/33	4.06	4.09E-07	5.64	3.15E-09
RNA248	762	1079	542	5946	5.51	3.01E-15	10.97	3.30E-23
RNA265	212	317	222	1473	4.65	3.58E-13	6.64	4.25E-17
RNA273 ^a	6755	12560	9348	126341	10.06	1.57E-35	13.54	2.77E-41
RNA275	334	248	387	1878	7.57	2.61E-08	4.85	3.34E-06
RNA277 ^{ab}	2511	5064	3800	32962	6.51	1.80E-21	8.67	1.23E-25
RNA294	515	1213	504	2692	2.22	NS	5.34	1.07E-05
RNA297	64	116	103	1994	17.19	0	19.36	0
RNA313	572	815	444	5731	7.03	6.51E-38	12.91	1.90E-53
RNA324	2203	5594	2162	12566	2.25	NS	5.81	2.07E-05
RNA349	101	446	160	2002	4.49	3.73E-119	12.51	1.06E-231
RNA365 ^a	416	1111	435	104	-11 11	6 64E-03	-4 17	NS
RNA406	300	284	169	3987	14 04	2.94E-104	23 59	2 26E-131
RNA415	123	351	139	573	1 63	NS	4 12	5.69E-07
RNA415	108	332	180	1072	3 23	1 67E 07	5.06	1 12E 12
RNA450	1232	1905	1861	60461	31.74	0	32.40	0
DNA470	520	504	605	2070	2 19	1 92E 02	2.49	5 56E 02
KINA4/2	126	594 607	093	2070	5.48	1.82E-05	2.98	3.30E-03
RINA323	130	007	433	3923	0.40		9.00	0
KNA541	221	438	691 7(0)	2088	4.//	2.66E-26	3.02	6.49E-17
RNA543	631	1866	760	4928	2.64	1.32E-07	6.48	1.88E-17
RNA552	91	132	115	810	6.14	1.25E-28	7.04	1.79E-31
RNA578	719	1313	835	8138	6.20	1.29E-25	9.75	6.05E-34
RNA587	314	1325	394	405	-3.23	3.50E-03	1.03	NS
RNA599	289	75	563	1660	22.13	6.96E-25	2.95	1.03E-06
Linear								
RNA623	540	1029	357	3176	3.09	5.29E-05	8.90	4.40E-11
RNA626	284	567	266	1915	3.38	5.87E-12	7.20	7.68E-22
RNA639	279	623	692	4814	7.73	3.95E-78	6.96	2.38E-73
RNA645	427	1546	288	348	-4.35	1.20E-03	1.21	NS
RNA653	663	701	745	2645	3.77	8.68E-03	3.55	NS
RNA660	1051	1998	288	44683	22.36	0	155.15	0
RNA670	456	804	434	3116	3.88	1 05E-07	7 18	4 15E-12
RNA675	31	53	17	1188	22 42	0	69.88	0
RNA722	331	170	1525	2228	13 11	/ 10E 22	1 16	NS
DNA 766	124	114	1525	6270	55.08		25.78	NG
$\frac{1}{1} \frac{1}{1} \frac{1}$	134	60	1/0 Q1	1542	25.00	0	10.04	0
$\frac{1}{100}$	44 55	00	01	1342	23.70	U 7 20E 41	19.04	U 1.52E 10
KNA//U	33	21	δ4 401	449	21.38	1.39E-41	5.55	1.55E-19
KNA810	4//	039	401	2430	5.80	1.30E-06	0.06	1.94E-09
KNA842	329	1025	524	2126	2.07	2.95E-05	4.06	3.28E-12
RNA857	320	583	315	1510	2.59	NS	4.79	1.45E-04

Table S2: List of putative sRNAs differentially expressed

RNA864 RNA879	329 141	622 336	492 314	3830 2746	6.16 8.17	1.48E-41 1.07E-153	7.78 8.75	1.95E-48 5.25E-160
RNA884	136	426	403	1888	4.43	5.85E-66	4.68	4.50E-69
At plasmid								
RNA909	108	127	140	475	3.74	3.02E-03	3.39	5.85E-03
RNA912 ^a	358	514	727	9921	19.30	0	13.65	0
RNA917	250	855	189	176	-4.76	9.09E-03	-1.07	NS
RNA932	30	19	18	302	15.89	4.19E-39	16.78	1.91E-39
RNA933	72	134	229	2337	17.44	0	10.21	0
RNA953	215	469	582	1219	2.60	1.62E-03	2.09	NS
RNA955	120	82	98	763	9.30	1.29E-10	7.79	1.51E-09
Ti plasmid								
RNA1066	80	218	136	799	3.67	4.64E-16	5.88	5.64E-24
RNA1077	245	350	322	1698	4.85	9.52E-10	5.27	1.86E-10
RNA1094	122	200	160	633	3.17	2.11E-03	3.96	2.55E-04
RNA1096	82	218	135	699	3.21	3.16E-10	5.18	4.33E-16
RNA1113	271	582	412	1762	3.03	1.61E-05	4.28	6.66E-08

NS non significant

^a also previously shown to be differentially expressed in stationary phase medium

^b also previously shown to be differentially expressed in minimal medium

Table S3: Predictions of RNA1111 targets common in the three algorithms IntaRNA, Starpicker and RNApredator (cutoff: 350 first predicted targets)

	Replicon localization	IntaRNA Start & Stop (+300)*	StarPicker*	RNAPredator*	Sum of order	Region	of interaction with IntaRNA	Region	of interaction with StarPicker	Region of interaction with RNApredator
Atu6188	pTi	1	1	1	3					· ·
Atu6127	pTi	12	17	178	207	SRNA	48 GGACGCCGGA 39 :	SRNA	104 aAGGCCGCAGg 114	sRNA: 114 GGACGCCGGA 105
						mRNA	364 UCUGCGGCCU 373	mRNA	74 aUCCGGCGUCu 64	mRNA: 63 UCUGCGGCCU 72
Atu6179	рТі	5	162	40	207	SRNA	160 GGCUU-UCUUAAUGUAAGUG-UGCCGUGGCCAAUGAGCUAGUGAUAGAGG 113 : : : : : : :	SRNA	112 aGGAGAUAGUGAucgaguaaccggugccgug-ugaa 146	sRNA: 133 CCAAUGAGCUAGUGAUAGA 113
						mRNA	222 UCGARCGGAAUCUCAGAAUCGCCAUGGAGGA-ACUCG-UCACUAUCUCC 268	mRNA	-32 aCCUCUAUCACU-gcucaa-ggagguaccgcua -62	mRNA: -50 GGA-ACUCG-UCACUAUCU -32
Atu6039	pTi	19	23	280	322	SRNA	35 CUGCCUCCGAGC 24	SRNA	23 ccgagccuccguca 36	sRNA: 35 CUGCCUCCGAGC 24
						mRNA	233 GAUGGAGGCUCG 244	mRNA	-56 aGCUCGGAGGUAGa -69	mRNA: -67 GAUGGAGGCUCG -56
Atu6014	pTi	25	157	142	324	SRNA	146 AAGUGUGCCGUGGCCAAUGAGCUAGUG 120	SRNA	119 agugaucgaguaaccGGUGCCGUGUGAAu 147	sRNA: 146 AAGUGUGCCGUGGCCAAUGAGCU 120
						MRNA	187 UUCACACGGCAUCAAACGGUCAC 209	mRNA	-91 ccacuggcaaaCUACGGCACACUUu -115	mRNA:-113 UUCACACGGCAUCAAACGG -91
Atu0666	¢C	219	21	114	354	SRNA	139 CCGUGGCCAAUGAGCUAGUGAUAGAGGACG 9	SRNA	109 cGCAGGAGAUAGUGAu 124	sRNA: 123 AGUGAUAGAGGACG 110
						MRNA	299 GGUAUCGCUUGCGCAAACCUCGCUAUCUCUUGC 331	mRNA	-14 uCGUUCUCUAUCGCUc -29	mRNA: -28 UCGCUAUCUCUUGC -15
Atu6072	pTi	29	77	260	366	SRNA	14 AGUGGCUUCCGAG 2	SRNA	1 -gAGCCUUCGGUGAa 14	sRNA: 13 AGUGGCUUCCGAG 1
						mRNA	389 UCGCCGAAGGCUC 401	mRNA	102 gcUCGGAAGCCGCUa 88	mRNA: 89 UCGCCGAAGGCUC 101
Atu6183	pTi	33	40	321	394	SRNA	116 GAGGACGCCGGAAUAAGCUAUUUC 93	SRNA	87 gcacga-cuuuaucgaauaAGGCCGCAGGAg 116	sRNA: 116 GAGGACGCCGGAAUAAGCUAUUUCA-GCA 88
						mRNA	341 UUCCUGUGGCUUCGAAUGAUGAAG 364	mRNA	71 ggugegggaaguaguaageUUCGGUGUCCUu 41	mRNA: 69 UUCCUGUGGCUUCGAAUGAUGAAGGGCGU 40
Atu4174	cL	315	61	201	577	SRNA	160 GGCUUUCUUAAUGUAAGUGUGCCCUGGCCAAUGAGCUAGUG 118 : : : : :	SRNA	119 agugaucgaguaaCCGGUGCCGUGUgaauguaauucuuucgga 161 	sRNA: 143 UGUGCCGUGGCCAAUGAGCU 120
						mRNA	374 CCGAA-GAGCUGAACGCGGCGCUGGACGAUCAC 405	mRNA	-95 acacuagcaGGUCGCGGCGCRagu-cgagaa-gccg -128	mRNA:-115 ACGCGGCGCUGGACGA -96
*number indicated is the order in the list of putative targets										

Strain	Ti plasmid opine-type	Presence of <i>rna1111</i>	Presence of 6b gene	Presence of <i>atu6072</i>	Presence of <i>traA</i>	Presence of <i>trbD</i>	Presence of <i>virC2</i>	Presence of <i>virD3</i>
C58	Nopaline	100% ID	100% ID	100% ID	100% ID	100% ID	100% ID	100% ID
C58-Chbika	Nopaline	100% ID	100% ID	100% ID	100% ID	100% ID	100% ID	99% ID
MAFF 301001	Nopaline	100% ID	99.5% ID	NF	71.8% ID	100% ID	100% ID	98.4% ID
CFBP5473	Chrysopine	71% ID	63% ID	NF	82% ID	77% ID	75% ID	NF
CFBP5480	Nopaline	100% ID	100% ID	NF	75% ID	100% ID	100% ID	100% ID
CFBP5476	Nopaline	100% ID	100% ID	NF	75% ID	100% ID	100% ID	99% ID
LMG294	Nopaline	100% ID	94% ID	NF	84% ID	99% ID	100% ID	99% ID
LMG17935	Nopaline	100% ID	68% ID	99% ID	81% ID	80% ID	99% ID	99% ID
NCPPB1650	Nopaline	100% ID	94% ID	NF	82% ID	99% ID	100% ID	99% ID
S56	Nopaline	100% ID	68% ID	99% ID	75% ID	86% ID	99% ID	99% ID
Zutra3-1	Nopaline	100% ID	68% ID	99% ID	75% ID	86% ID	99% ID	99% ID
ID identiti	es							

Table S4: Conservation of RNA1111 putative targets among genome sequences:

ND Not determined

NF No found (No significant Blast results)

Table S5:	Strains	and p	olasmids	used	in 1	this	study
							-

Strains and plasmids	Genotypes	References
A. fabrum strains	3	
C58	Wild type strain from the G8 genomic species of A. tumefaciens	(Goodner et al., 2001; Wood et al., 2001)
AT192	<i>A. fabrum</i> C58∆ <i>rna1111</i> NeoKm ^R	This study
AT199	A. fabrum C58 containing pARA005, Gm ^R	This study
AT200	A. fabrum C58 containing pBBR1MCS-5, Gm ^R	This study
AT203	A. fabrum C58∆rna1111 containing pARA005, NeoKm ^R Gm ^R	This study
Plasmids		
pBBR1MCS-5	Host range cloning vector, Gm ^R	(Kovach et al., 1995)
pARA005	pBBR1MCS-5:: <i>rna1111</i> Gm ^R	This study

Table S6: Primers used in this study

Primers for	Forward (5'-3')	Reverse (5'-3')
rRNA and tRNA	depletion	
Atu0059	AGACCTGGCAGCGACCTAC	
Atu2540		
Atu3943		
Atu0054	TGGTGGGCCCGGGTAGAC	
Atu2546		
Atu3938		
Atu4181		
Atu0055	TGGTGGAGCTGAGCGGGA	
Atu2545		
Atu3939		
Atu4182		
Atu4102		
Atu0060	TGGTTGCGGGGGGCAGGAT	
Atu2539		
Atu 3944		
Atu418/		
Atu0125		
Atu1017		
Atu1017		
Atu1502		
Δτυ0227	ТССТСССАСААСАССАС	
Atu0227	IGGIGCCCAGAAGAGGAC	
Atu1809	TGGCGCACCCGAAGAGAT	
Atu0414	ТССТССССТТАССТСАС	
Atu2000		
Atu0/35	TGGIGCCGCTIACGIGAC	
Δtu1001	ТССАССССТСААСССАА	
Atu0505		
Atu0505		
Atu3702	TGGAGCGGGTAGCGGGAA	
Atu0541	TGGTGCTGCTAGAGAGAT	
Atu0616	TGGTACGGTTGAGTGGGG	
Atu0808	TGGCTGGGGCGCCAGGAT	
Atu0814	TGGTCGGAGTGAGAGGAT	
Atu1193	TGGTGCCCGCAGCCGGA	
Atu1224	TGGCTCCCCGGGCCGGA	
Atu1265	TGGTGGGCGATGAGAGAC	
Atu1266		
Atu1267	TGGCGCGAGTGACGGGGC	
Atu1555	TGGTGCGGTCGAGAAGAC	
Atu1583-		
Atu1585	TGGCACGCCCTAGGGGAG	
qPCR :		
A121-A122	TCTTGCCAAGTCACGGAGT	GCATTCCGAAAGAATTACATTCAC
(rna1111)		
A167-A168	GCGAACAGAATGCGGTAGAT	AAGATCGGGAAGTTCTGCTG
(<i>atu0231</i>)		
A191-A192	CGACCTTCAACAACACGATG	CGTGTTCCTGAGCCTTCTTC
(atu1924)		
A239-A240	TCCATCCTTCTTTGCCTGTT	CTGTCGTCGAACTCAAAGCA
(atuo1/9)		
A241 - A242	GAAGUUGUTTAUGTUGATAG	GUUTUUTGUAUAAUTUTUTT
(<i>uu</i> 0100) A243-A244	СТСАССССТТСТССААСААС	ሮሮሮͲͲሮሮሮልናናልሞልሞናምሞ
· · · · · · · · · · · · · · · · · · ·		000110000000000011110111

(atu 6127)		
(<i>a</i> 146127) A245-A246	ͲͲͲϷႺϹႺϷͲϹϷͲႺϹϹϷϷႺϹ	ͲሮሮႺሮሮልሬሞሮሞሮሬሞሮሞሞሮ
(atu6014)		10000000000000000
A286-A287	AGTCTCTGTCCGAAACCGTT	GGCTTGAACAATGCTGACGA
(a <i>tu6072</i>)		
A288-A289	AACAAGACAGGGAAGCCAGT	CTTGCATTGCCGGGATACTC
(atu6183)		
A290-A291	CAAGCCCGGTGAGCAATAC	TTCCAACCTTGCGCGTTC
(atu6039)		
Cloning:		
186-187	ACGCGTCGACGCAGTCACTCTTCGGACACA	CGCGGATCCTCGACAAGCGAAGAGTCTGA
(rna1111)		
Inactivation:		
100-1/5 (unstream	CACACTAGICITGAGCGCIGCIAGAAACA	
(upsueani rna1111)		CARITCI
176-177	TAAGGAGGATATTCATATGGCAGCAAATCCC	CACCTCGAGCAACTTGGTGTACGGCAGAA
(downstream	TCATCATCC	
rnal111)		
095-096 (nptII)	TGTGTAGGCTGGAGCTGCTTC	CATATGAATATCCTCCTTAG
Reverse transcript	tion for RACE-PCR	
A251 (rna1111)	GTTACTCGATCACTATCTCC	
A301 (rna275)	TCAATGTCCACCAGTTTGCG	
A302 (rna535)	ACAGATACACGGGGCTCCA	
A303 (rna115)	AATCTGCGTTCGGTCAGCC	
A304 (<i>rplL</i>)	GGTCAGGGTGGAGAGGTCTT	
A305 (rna387)	TCCGTCTCCTGAATGTCTGC	
A306 (rna834)	AGAGAACCCTGTCAGCCCAA	
A307 (rna847)	CGGTCGAGAATGAGGTGACA	
A308 (rna779)	AATCTGCCTTCTCACCCCTG	
A309 (rna960)	CCAAGTTGTTCCCCGATCGTT	
A310 (<i>rna1028</i>)	CGCAGTCCAGTCGGTCGA	
A312 (rna576)	CACGTTCAGGTGCGGATC	
A313 (rna141)	ТТСТСССТСССТТАТТССС	
$\Delta 314 (rnal 54)$		
A316 (rma846)		
A310 (ma807)		
A317 (7nu007) A219 (ma 969)		
A318 (rna808)		
A319 (rna809)		
Amplification of j		
A252-A253	CATAAGCTTCACTATCTCCTGCGGCCTTA	CGCGGATCCCGGTGCCGTGTGAATGTA
(rma1111) 275 276	СССССТСАТАСАСТСТТСТ	CGAAATCCCAAAATCCCCCCC
(rna275)	COCOTCENTRENDICITEI	GAAATCGGAAAATCGCCCC
277-278	CAGATACACGGGGCTCCAG	CACGAAACATGCTGTAGAACTT
(<i>rna535</i>)		
279-280	AATCTGCGTTCGGTCAGCC	GGGATAGAAAGCATAAAGGTCCG
(<i>rna115</i>)		
281-282 (rplL)	GTCAGGGTGGAGAGGTCTTC	CTTGAAGAAAAGTGGGGCGT
283-284	TGCCAGAACAGAACCCTTCA	AAGGCAAACACATTCACCCG
(<i>rna387</i>)		
285 - 286	AGAACCCTGTCAGCCCAAG	CTGGATTCCGGCTCAGGG
(rnuð 54) 287_288	СССТССАСААТСАССТСАСА	ТССАСАСТТССА А АССТС
(rna847)	CCCCCONDUMITION OF LONCA	I CONDITICONARCEIC
289-290	CTGCCTTCTCACCCCTGTAT	TGAAATTTGCGCGCGGAC
(770)		

291-292	CCAAGTTGTTCCCGATCGTT	GCTACCATTGAACGAACGACT
(rna960)		
293-294	CAGTCCAGTCGGTCGAAGG	TAAGGCGGCTGATGTGAGAT
(rna1028)		
297-298	TCAGGTGCGGATCGTCTTTT	CAGAATCTGGGCTCGCC
(rna576)		
299-300	TTCTCGCTCCGTTATTCCCG	GTGATCGCGCTTTCATTTGC
(<i>rna141</i>)		
301-302	CTAGATCAATCACAGACGCAGA	GCGTGAGGGAGAGGGGTAC
(rna154)		
305-306	GTGTCCGCGATGCCAC	CTGCGCGACGGGAAAAG
(rna846)		
307-308	AGCTCTGCCTTAAACGACGA	TGAAATGGCGTACAAAATCATG
(rna807)		
309-310	CAGACCAATGAGCAAGCCTT	ATTGCTGCCTATTTCGCCTG
(rna868)		
311-312	ACTCCTGAACAAAGCCCTCT	CGGCCATGATCGCACTG
(rna869)		

Table S7: Summary of RNA sequencing and alignment onto A. fabrum C58 reference genome

	Rich media				Minimal media				
	Exponential		Stationary		Exponential		Stationary		
Replicates	А	В	А	В	А	В	А	В	Total
Number of reads (x1000)	24,728	23,215	23,241	24,814	23,799	24,925	24,636	23,740	193,100
Total read aligned (x1000)	24,728	22,889	23,055	24,607	22,668	24,253	24,051	23,313	189,117
UMR reads (x1000)	3,960	4,301	3,693	2,663	1,918	3,170	5,116	3,563	28,386
% UMR reads	16.0%	18.52%	15.89%	10.73%	8.06%	12.72%	20.77%	15.0%	14.71 %
% UMR per replicor	1								
Circular	71.7%	74.6%	71.3%	67.6%	69.8%	72.7%	75.3%	72.6%	72.5 %
Linear	25.5%	23.2%	26.3%	29.8%	27.7%	24.5%	22.6%	24.5%	25.0 %
pAt	1.6%	1.3%	1.5%	1.6%	1.5%	1.7%	1.5%	2.15%	1.6 %
pTi	1.2%	0.9%	0.9%	1.0%	1.0%	1.1%	0.6%	0,75%	0.9 %


Repartition of small transcripts along chromosomes and plasmids

Figure S1: Localization of putative small RNAs along chromosomes and plasmids of A fabrum C58.



Figure S2: Genetic organization of *rna1111* locus from the pTi and display of sRNA-seq data. Data from *A. fabrum* sRNA-seq displayed using Artemis. Strand-specific coverages are shown displayed as two plots, as raw reads aligned against the reference strain: in red, coverage displayed from plus strand and in green coverage from minus strand. The genome annotation is displayed underneath. The genomic environment of *rna1111* locus is presented. Grey arrows correspond to genes. 5'-end and 3'-end of *rna1111* gene were determined from RACE-PCR results. Sequences in grey boxes correspond to putative promoter region of *rna1111* the -10 box (TAATAG, with consensus TATNNT), and -35 box (CTTCGG with consensus CTTGN). +1 represents +1 transcription site of *rna1111*. (RPKM: Reads Per Kilobase of transcript per Million reads mapped). Red framed box corresponds to the region deleted in the strain AT194 (inactivation for *rna1111*).

Alignment of *rna1111* genes

RNA1111_C58 RNA1111_CFBP5473	GAGCCTTCGGTGAATCTACCATCCGAGCCTCCGTCATTAAACGAATATCGCGTTTCATAT GAGCCTGTGTTGAACTTATCATCTGGGCCCCCGTCACTAAACGAATATCGCGTTTCATCT ****** * **** ** *** * *** **** ****
RNA1111_C58 RNA1111_CFBP5473	GTAATTCTTGCCAAGTCACGGAGTTAGC <mark>ACGACTTTATCGAAT<u>AAGGCCGCA</u>GG</mark> AGATAG GTAACTCCTGCTAAGTCGCAGAGATAGC <mark>GTGACTTTGTCAAAG<u>AAGGTCGCAGA</u>AAACAG</mark> **** ** *** **** * *** **** **** ****
RNA1111_C58 RNA1111_CFBP5473	TGATCCAGTAACCGGTGCCGTGTGA+TGTAATTCTTTCGGAATGCCACAATGA GCATCAAGTAA-CGGCGCCATCTGCATGCAATT *** ***** *** *** * ** ***

CLUSTAL multiple sequence alignment by MUSCLE (3.8)

In green, the region of interaction with atu6072

In blue, the region of interaction with *virC2*

In red, the region of interaction with virD3

In yellow, the region of interaction with the 6b gene

In Underlined sequence, the region of interaction with *traA*

RNA1111_C58, sequence of the *rna1111* in the strain C58. Sequence is 100% identical to the ones of C58-Chbika, MAFF 301001, CFBP5480, CFBP5476, LMG294, LMG17935, NCPPB1650, S56 and Zutra3-1 that all harbor a nopaline-type Ti plasmid.

RNA1111_CFBP5473, sequence of the *rna1111* in the strain CFBP5473 that harbors a chrysopine-type Ti.

Figure S3: Conservation of *rna1111* and its putative targets among Ti plasmids.



Figure S4: Signal quantifications in sRNA deep-sequencing experiments and Northern blot experiments. sRNA deep-sequencing were performed on size-selected and 5S rRNA/tRNA depleted RNA extracts (after the small RNA separation process). Northern blots were performed on total RNA extracts. Signals were quantified for six different small RNAs: AbcR1, SsrA, RnpB, Atumisc_RNA_13, SsrS and ctRNA_p42d. As illustrated, the quantities of all small RNAs are conserved as compared to data from total RNA extracts (Northern blot analyses) and small RNA separations (RNA deep-sequencing). Data represent the means of four experiments in different growth conditions (rich and minimum media, in exponential and stationary phases).



Figure S5 : Northern blot experiments performed on total RNA extracts for AbcR1, SsrA, RnpB, Atumisc_RNA_13, SsrS and ctRNA_p42d in two different growth conditions: rich and minimum media (A and B).

Figure S5: sRNAs validation by northern blot experiments.

Authors contributions

Magali Dequivre was involved in the analysis and the interpretation of the data

Benjamin Diel was involved in the analysis and the interpretation of the data concerning the RNA1111 analysis

Camille Villard was involved in the conception of the experimental data

Odile Sismeiro was involved in the sequencing of small RNAs

Jean-Yves Coppée was involved in the sequencing of small RNAs

Maxime Durot was involved in the sequencing data analysis

Xavier Nesme was involved in the important critical révision of the manuscript

Ludovic Vial was involved in the conception and the writing of the article

Florence Hommais was involved in the conception, design, analysis and interpretation of the data and the writing of the manuscript

Conclusion : notre approche a permis, à partir d'une analyse globale, de caractériser un nouveau riborégulateur sur le plasmide TiC58, et d'identifier son rôle dans la modulation du cycle infectieux. RNA1111 est un activateur de la virulence, et pourrait contrôler le transfert de l'ADN-T via VirC2 et VirD3, la production de phytohormones via le produit du gène 6b et le transfert conjugatif du plasmide Ti via TraA et TrbD.

IV.2 Riborégulateurs et rôle dans le cycle infectieux

L'analyse fonctionnelle des riborégulateurs nécessite l'identification de leur(s) cible(s). Dans le cas des ARN antisens, la cible attendue est la séquence codante codée sur le brin opposé. Pour les ARN 5'UTR ou 3'UTR, il est probable qu'il s'agisse de la séquence codante située directement en aval ou en amont respectivement. On parle alors de régulation en *cis*. Nous avons recherché la présence de petits transcrits localisés en antisens ou dans les régions UTR de gènes codants des protéines impliquées dans le cycle infectieux d'*A. fabrum* C58, allant de la perception de la blessure jusqu'au transfert conjugatif du plasmide Ti. Parmi les 477 riborégulateurs UTR et les 263 riborégulateurs antisens identifié chez *A. fabrum* C58, 45 pourraient, de par leur localisation, être impliqués dans la régulation des différentes étapes du cycle infectieux. En revanche pour les riborégulateurs agissant en *trans*, l'identification des cibles est plus complexe du fait que l'on ne connaît pas les propriétés exactes de l'interaction.

IV.2.1 Riborégulateurs chromosomiques

Les étapes précoces de l'interaction, correspondant à la migration jusqu'à la blessure et à l'adhésion des bactéries sur les cellules végétales, dépendent des gènes portés par les chromosomes circulaire et linéaire. Bien que non indispensables, ces étapes sont importantes pour la colonisation de la plante et pour un pouvoir pathogène optimal. Nous nous sommes donc intéressés aux riborégulateurs produits à partir de ces régions. Trois riborégulateurs potentiels ont été identifiés au niveau de gènes impliqués dans le chimiotactisme (RNA106 en 5'UTR de *cheA*, RNA569 et RNA570 respectivement en 5'UTR et 3'UTR de *cheW*), 6 au niveau de la synthèse des flagelles (RNA107 en 5'UTR de *fliF*, RNA117 et RNA122 en 5'UTR de *flgB* et *flgE* respectivement, RNA119 en 5'UTR de *flhB*, RNA113 et RNA114 en 5'UTR et en antisens de *fla*), et 3 dans l'adhésion aux cellules végétales (RNA588 en antisens de *chvB*, RNA650 en 5'UTR de *celA*, RNA655 5'UTR de *exoY*). On peut supposer que ces 12 riborégulateurs pourraient moduler l'expression des gènes impliqués dans les étapes précoces nécessaires à la mise en place de la virulence. Pour tester cette hypothèse, il faudrait dans un 1^{er} temps valider la présence de ces petits transcrits par Northern Blot et par RACE-PCR. L'expression des cibles pourrait aussi être suivie par RT-qPCR dans des souches modifiées dans la production des candidats.

IV.2.2 Riborégulateurs du plasmide Ti

Le plasmide Ti étant le déterminant majeur de la virulence, nous nous sommes focalisés sur les riborégulateurs produits à partir de ce réplicon (Figure 25).



Figure 25 : Localisation des gènes codant les riborégulateurs identifiés par notre analyse sur le plasmide TiC58

De l'extérieur vers l'intérieur: le cercle gris représente la position sur le plasmide Ti ; les boîtes bleues claires représentent les séquences codantes ; les barres verticales correspondent aux gènes codant les riborégulateurs : dans des IGR pour les barres bleues foncés, en antisens de séquences codantes pour les barres vertes, en ou 5'UTR ou 3'UTR des séquences codantes pour les barres roses (les étoiles correspondent aux riborégulateurs qui ont été sélectionnés pour une caractérisation fonctionnelle dans l'ordre, depuis la position 0 : RNA1044, RNA1046, RNA1051, RepE, RNA1059, RNA1083, RNA1111) ; l'anneau central représente le plasmide Ti avec la localisation des différentes régions impliquées dans la virulence en couleur.

IV.2.2.1 Riborégulateurs antisens et UTR

Nous avons observé une surreprésentation des riborégulateurs antisens sur le plasmide Ti, puisqu'ils correspondent à 63% des petits transcrits identifiés sur ce réplicon, contre 20% environ sur les chromosomes. Parmi les 57 riborégulateurs du plasmide Ti localisés en UTR ou en AS, 32 sont produits à partir de régions impliquées dans le cycle infectieux (Table S1 pages 197-219). Ces candidats sont transcrits à proximité des gènes : des régulateurs du système à deux composantes VirA/VirG (RNA1099-RNA1100, RNA1104), des protéines Vir (RNA1101-RNA1103, RNA1105-RNA1110, RNA1112- RNA1113), de l'ADN-T (RNA1039, RNA1041-RNA1042, RNA1044-RNA1046), des protéines du catabolisme des opines (RNA1048, RNA1050, RNA1053-RNA1054, RNA1091-RNA1093), des protéines Tra du transfert conjugatif du plasmide Ti (RNA1084, RNA1086, RNA1089) et des protéines Rep de la réplication du plasmide Ti(RNA1056-RNA1057). De plus, certains facteurs en lien

avec le cycle infectieux sont codés en dehors de ces îlots de gènes. C'est le cas du gène *aiiB*, codant une homosérine lactonase impliquée dans la perturbation du *quorum sensing*, et en antisens duquel a été identifié le gène codant le candidat RNA1068.

IV.2.2.2 Caractérisation phénotypique de deux candidats antisens

Nous avons entrepris l'étude de deux candidats antisens exprimés à partir de l'ADN-T : RNA1044 localisé en antisens du gène *tms2* permettant la synthèse de l'AIA; et RNA1046, localisé en antisens du gène *atu6013* codant la protéine 6a qui est une cytokinine glycosidase.

L'impact de ces candidats sur la virulence d'*A. fabrum* C58 a été étudié à l'aide de souches modifiées dans leurs productions. La surexpression des gènes a été obtenue en les clonant dans le vecteur pBBR1MCS-5 et en transformant les plasmides résultants (pARA007 et pARA008 respectivement) dans la souche sauvage C58 (souche AT209 et souche AT201 respectivement). En revanche, pour la construction de la souche déficiente dans la production de ce candidat, nous n'avons pas remplacé les gènes d'intérêt par un gène de résistance aux antibiotiques, comme nous l'avons fait pour l'étude de RNA1111. En effet, les gènes *rna1044* et *rna1046* étant en antisens d'autres gènes, leur délétion induirait une inactivation des gènes situés sur le brin complémentaire. Nous avons développé une méthode similaire à l'interférence à ARN. Cette méthode consiste à séquestrer et à inactiver l'ARN d'intérêt à l'aide d'un transcrit parfaitement complémentaire exprimé à partir d'un plasmide introduit dans la cellule (figure 26).



Figure 26 : Principe de l'ARN à interférence

A/ En absence d'ARN interférent, le riborégulateur (rose) interagit avec l'ARNm cible (vert foncé) et module sa traduction et/ou sa stabilité. B/ En présence d'ARN interférent RNAi (vert clair), le riborégulateur est séquestré et ne peut plus s'apparier avec son ARNm cible.

Nous avons ici amplifié et cloné dans le pBBR1MCS-5 les fragments complémentaires aux gènes *rna1044* et *rna1046* et avons introduit les plasmides résultants pARA009 et pARA010 chez la souche sauvage (souches AT210 et AT202 respectivement).

Après avoir vérifié que les souches mutantes présentaient une croissance identique à la souche contrôle AT200, des tests de virulence ont été effectués en serre par infection au niveau d'incisions réalisées sur la tige des plants de tomate. L'inoculation de la souche contrôle AT200 induit la formation d'une moyenne de 44 tumeurs (figure 27). La modification de la production de *rna1046* ne modifie pas de façon significative l'agressivité d'*A. fabrum*, puisque les souches AT201 et AT202 provoquent la formation de 37 et 41 tumeurs respectivement. De même, les plants infectés par la souche AT210 inactivée pour *rna1044* ne montrent pas de différence significative (48 tumeurs en moyenne). Cependant, la souche AT209 surexprimant *rna1044* semble être plus agressive, puisqu'elle induit la formation en moyenne de 57 tumeurs par plant (p-value $\leq 0,04$).





Les souches surproduisant RNA1044 et RNA1046 (AT209 et AT201) ainsi que les souches inactivées dans leur production (AT210 et AT202) ont été testées pour leur capacité à induire la maladie de la galle du colle. Pour cela, les différentes souches ont été inoculées sur des plants de tomate âgés de 3 semaines et les tumeurs ont été dénombrées 3 semaines après infection. La virulence des différentes souches a été comparée à celle de la souche contrôle (AT200). Les souches présentant un phénotype significativement différent sont annotées par une étoile (p-value > 0,05) (test de Kruskal-Wallis)

IV.2.2.3 Riborégulateurs trans

De nombreux riborégulateurs *trans* identifiés par notre étude sont transcrits à partir des régions relatives au cycle infectieux du plasmide Ti (Figure 25 page 115). Deux sont localisés dans l'ADN-T (RNA1040, RNA1043), 3 dans la région de catabolisme des opines (RNA1049, RNA1051, RNA1052), 2 dans la région de réplication du plasmide Ti, (RNA1058 et RNA1059), 1 dans la région du transfert conjugatif (RNA1085), et 4 dans la région des gènes *vir* (RNA1096-RNA1098, RNA1111). L'analyse

d'un 1^{er} candidat, RNA1111 produit à partir de la région *vir*, a montré son rôle dans la modulation de la virulence (Article 1 pages 81-113). Afin de découvrir si d'autres riborégulateurs *trans* participaient à la régulation de la virulence d'*A. fabrum* C58, 4 autres candidats ont été sélectionnés : RNA1051 produit à partir de la région codant le catabolisme des opines ; RNA1059 et RepE dont les gènes sont localisés dans la zone de réplication du plasmide Ti ; et enfin RNA1083 transcrit en amont de la région codant le transfert conjugatif (Figure 25 page 113).

IV.2.2.4 Caractérisation de RNA1083 et impact sur la virulence

Validation de la présence de RNA1083 chez A. fabrum C58

Le gène du petit transcrit RNA1083 a été identifié en amont de la région de conjugaison du plasmide Ti entre les bases 139 073 et 139 263 sur le brin négatif, entre les gènes *atu6119* et *atu6120* (Table S1 pages 197-219 ; Figure 25 page 115). Nous avons validé la transcription, la taille et les extrémités de ce candidat par la technique de RACE-PCR, à l'aide des amorces A311, 295 et 296 (tableau 6).

Ainsi, l'ARN RNA1083 est long de 187 bases et est produit entre les nucléotides 139075 et 139262 à partir du brin négatif du pTi. Ces résultats sont en accord avec les données obtenues par RNA-seq.

Un promoteur potentiel a été identifié en amont du site d'initiation de la transcription. Le site -35 est défini par les bases CTTCGC (consensus : CTTGNN), et le site -10 par les bases AGAAAT (consensus : TATNNT) (Wilms et al., 2012b) (Figure 28A). Aucune ORF n'a pu être identifiée au sein de la séquence de *rna1053*.

La structure secondaire basée sur le calcul de l'énergie libre minimale de RNA1083 a été prédite à l'aide du logiciel RNAfold (http://rna/tbi.univie.ac.at/cgi-bin/RNAfold.cgi) en utilisant les paramètres par défaut (Figure 28.B1 et B2). Celle-ci est composée de 2 longues structures en épingles à cheveux, l'une de 33 paires de bases entrecoupée de deux « régions simple brin » de 15 et 6 bases, et l'autre de 17 paires de bases entrecoupée de 2 «régions simple brin» de 14 bases ; ainsi que de deux épingles à cheveux centrales de 5 paires de bases chacune. Les bases présentes au sein de ces 4 épingles à cheveux s'apparient totalement avec leurs bases complémentaires respectives (probabilité d'appariement de 1). Par ailleurs, l'entropie ponctuelle/positionnelle est faible (proche de 0), ce qui signifie que l'énergie nécessaire au maintien de cette structure est faible. Ces deux résultats suggèrent donc une grande stabilité de la structure secondaire de RNA1083.

RNA1083 semble donc être un ARN codé en *trans*, fortement structuré, et qui pourrait avoir un rôle régulateur. Sa caractérisation fonctionnelle a été entreprise en utilisant une stratégie similaire à l'étude de RNA1111.



Figure 28 : Caractérisation de RNA1083

A/ Visualisation des données de RNAseq d'A. *fabrum* C58 à partir du logiciel Artemis. La couverture brin spécifique est monté sur deux graphiques séparés, représentants les lectures brutes alignées sur le génome de la souche de référence : en rouge, la couverture sur le brin positif, en vert, la couverture sur le brin négatif. L'environnement génomique de *rna1083* est représenté en dessous. Les flèches grises correspondent aux séquences codantes des gènes. Les extrémités 5' et 3' de *rna1083* ont été déterminées par RACE-PCR. Les séquences comprises dans les boîtes grises correspondent aux séquences promotrices putatives -10 (AGAAAT avec le consensus TATNNT) et -35 (CTTCGC avec le consensus CTTGNN). +1 représente le site d'initiation de la transcription de *rna1083*. RPKM = Reads Per Kilobase of transcript per Million reads mapped. B/ Structure secondaire de RNA1083. La structure secondaire la plus probable de RNA1083 a été prédite par le logiciel RNAfold. 1. Appariement de bases. La couleur des bases correspond à la probabilité d'appariement (bleu : probabilité = 0 ; rouge : probabilité = 1). Pour les régions non appariées, les couleurs représentent la probabilité d'être non appariée. 2. Entropie de la structure secondaire. La couleur des bases correspond à l'entropie positionnelle de la structure secondaire (rouge : entropie = 0, violet : entropie = 1).

Prédiction des ARNm cibles de RNA1083

Les cibles de RNA1083 ont été prédites à l'aide des 3 logiciels IntaRNA (Busch, Richter, Backofen 2008), sTarPicker (Ying *et al.*, 2011) et RNApredator (Eggenhofer *et al.*, 2011). Les régions de 300 paires de bases en amont et en aval du site d'initiation de la traduction de l'ensemble des séquences codantes du génome d'*A. fabrum* C58 ont été analysées. RNA1083 étant situé sur le plasmide, nous nous sommes focalisés sur les cibles préférentiellement localisées sur le plasmide Ti ou annotés comme reliés à la virulence. Les cibles identifiées par la majorité des algorithmes ont été considérées

Souches/plasmides/ amorces	Génotypes	Références
Souches		
DH5a	Souche d'Escherichia coli (utilisée pour les clonages d'ADN)	Stratagene
C58	Souche sauvage de référence d'Agrobacterium fabrum	Wood <i>et al</i> , 2000 ; Goodner <i>et al,</i> 2000
AT200	A. fabrum C58 contenant le vecteur pBBR1MCS-5	Cette étude
AT196	A. fabrum C58 contenant le vecteur pARA006	Cette étude
AT193	A. fabrum C58 inactivée par délétion du gène rna1083 et remplacement par le gène nptll	Cette étude
Plasmides		
pBBR1MCS-5	Vecteur de clonage à large gamme d'hôte, Gm ^R	Kovach <i>et al,</i> 1994
pWSK29	Vecteur de clonageà faible nombre de copies, Amp ^R	Wang <i>et al</i> , 1991
pARA006	Gène <i>rna1083</i> cloné dans le vecteur pBBR1MCS-5entre les sites Sall et BamHI, Gm ^R	Cette étude
Amorces		
182-183 (rna1083)	ACGCGTCGACAAATCGTCATCGTCGTCTCC <u>CGCGGATCC</u> GGTGCCATACGAGGAGTTGT	Cette étude
167-168 (inactivation de <i>rna1083</i>)	CACACTAGTATAGAGCCATGCGTTGAAGC GAAGCAGCTCCAGCCTACACAGGAGCTGGAGATGGTACAGG	Cette étude
169-170 (inactivation de <i>rna1083</i>)	TAAGGAGGATATTCATATGGCCTCGCGACACGAAAAGAG <u>CACCTCGAG</u> CCGAGTTCGTACAGCGACAG	Cette étude
95-96 (gène de résistance nptII)	TGTGTAGGCTGGAGCTGCTTC CATATGAATATCCTCCTTAG	Cette étude
A311	GCAAATCGGTTCCTCCCCT	Cette étude
295-296	<u>CATAAGCTT</u> AAATCGGTTCCTCCCCTCG <u>CGCGGATCC</u> CGGAAGGAGCAAATTGGAGG	Cette étude
M13fwd-M13rev	GTAAAACGACGGCCAG TTCAACGATCAATTCGTCCA	Cette étude

Tableau 6 : Souches,	plasmides et ar	norces utilisés	pour l'étude	de RNA1083
----------------------	-----------------	-----------------	--------------	------------

Les séquences soulignées correspondent aux sites de restriction utilisés pour le clonage. Gm^R : résistant à a gentamycine ; Amp^R : résistant à l'ampicilline.

(figure 29). De manière intéressante, 3 cibles du plasmide Ti ont été prédites en commun par les 3 logiciels : *trbG* (*atu6033*), *trbL* (*atu6035*) et *trbK* (*atu6036*). Ces gènes font partie d'un opéron de 11 gènes codant des protéines impliquées dans le transfert conjugatif du plasmide Ti. Les zones d'interaction se situent : 162 et 83 bases en aval du site d'initiation de la traduction pour *trbK* et *trbG* (interaction de 42 et de 16 bases environ) ; 182 bases en amont du codon initiateur pour *trbL* (interaction de 14 bases environ). Nos analyses ont également prédit plusieurs cibles transcrites à partir du chromosome circulaire et impliquées dans la mobilité bactérienne ou dans le contrôle de la virulence. Ainsi, quatre cibles correspondent à des gènes de structure des flagelles et les régions d'interaction sont localisées : 74 bases en amont du site initiateur de *flgF* (*atu0558*, interaction de 13 bases) ; ou

dans la séquence codante de *flgI* (*atu0550*, interaction de 17 bases) et de *fliL* (*atu0547*, interaction sur 10 bases). De même, l'ARNm *chvI* (*atu0034*) qui code le régulateur du système à deux composants ChvI/ChvG impliqué dans la régulation de *chvE*, interagit potentiellement avec RNA1083, environ 136 bases en amont du site d'initiation de sa traduction. Ainsi, les cibles prédites suggèrent que RNA1083 participerait à la modulation de plusieurs fonctions cellulaires, telles que la virulence (*chvI*), la mobilité (*flgF*, *flgI*, *fliE*, *fliL*), et le transfert conjugatif du plasmide Ti (*trbG*, *trbK et trbL*).



Figure 29 : région d'interaction de RNA1083 sur les ARNm prédits comme ses cibles

Les séquences autour des sites d'initiation de la traduction des cibles qui ont été prédites pour RNA1083 sont présentées. Les régions d'interaction avec RNA1083 sont surlignées. Les gènes *trbK, trbL, trbG* et *fliE* ne présentent pas de RBS car ils ne sont pas les 1^{ers} gènes de l'opéron.

Rôles physiologiques de RNA1083

Des analyses ont été entreprises dans le but de confirmer le rôle de RNA1083 dans ces différentes fonctions cellulaires, et ce à l'aide de souches modifiées dans sa production (Tableau 6). Pour toutes les analyses, la souche AT193 inactivée pour *rna1083* a été comparée à la souche sauvage C58, et la souche AT196 surexprimant *rna1083* à la souche contrôle AT200 portant un vecteur vide.

Rôle de RNA1083 dans la virulence d'A. fabrum C58

L'impact de RNA1083 sur la virulence de la bactérie a tout d'abord été analysé selon un protocole similaire à celui de l'article 1 (Figure 30). La souche AT193 inactivée dans la production de RNA1083 présente une agressivité similaire à celle de la souche sauvage (le pourcentage de plantes développant plus de 15 tumeurs est de 50% pour la souche AT193, et de 65 % pour la souche sauvage). En revanche, il apparaît que la souche AT196 surproduisant RNA1083 est significativement moins virulente que la souche contrôle AT200, puisqu'elle induit le développement de plus de 15 tumeurs chez 22% de plantes, tandis que la souche AT200 produit ce même nombre de tumeurs sur plus de 58% des plants inoculés. Il semble ainsi que l'augmentation de la production de RNA1083 diminue la virulence d'*A. fabrum* C58. L'interaction prédite entre RNA1083 et *chvl* pourrait expliquer l'impact sur la virulence.



Figure 30 : Impact de RNA1083 sur la virulence d'A. fabrum C58

Les souches AT193 (inactivée dans la production de RNA1083) et AT196 (surproduisant RNA1083) ont été testées pour leur agressivité sur plants de tomate. Les résultats ont été comparés à ceux de la souche sauvage (WT) et de la souche sauvage portant un vecteur pBBR1MCS-5 vide (AT200). Les bactéries ont été inoculées sur des tiges blessées de tomate âgées de 3 semaines. Les tumeurs ont été dénombrées 3 semaines post-infection. Les souches présentant des différences significatives (test de Kruskal-Wallis) dans leur agressivité sont annotées avec deux étoiles (p-value \leq 0,01).

Rôle de RNA1083 dans la mobilité d'A. fabrum C58

Nous nous sommes également intéressés au rôle potentiel de RNA1083 dans la mobilité de la bactérie (Figure 31). Pour cela, 5 μ L de cultures liquides (DO_{600nm} = 1) de chacune des différentes souches ont été déposés sur des géloses (0,7% d'agar) composées d'un milieu AB supplémenté de 2 % de mannitol. Les souches ont été incubées à 28°C pendant 48 heures. Les résultats, obtenus à partir de deux réplicas indépendants, sont présentés dans la figure 30 ci-dessous. La mobilité de la souche AT193 n'est pas significativement différente de celle de la souche sauvage (surface des colonies de 48,7 cm² et 45,3 cm² respectivement). En revanche, la souche AT196 est significativement moins mobile que la souche AT200 (p-value \leq 0,04), puisque les colonies issues de la surexpression présentent une surface de 43,5 cm² tandis que les colonies correspondant à la souche contrôle mesurent 50,1 cm² en moyenne.



Figure 31 : Impact de RNA1083 sur la mobilité d'A. fabrum C58

Les souches AT193 (inactivée dans la production de RNA1083) et AT196 (surproduisant RNA1083) ont été testées pour leur mobilité. Les résultats ont été comparés à ceux de la souche sauvage (WT) de la souche sauvage portant un vecteur pBBR1MCS-5 vide (AT200). La mobilité a été déterminée en mesurant la surface de dépôts de suspensions bactériennes sur milieu AB 0,7% agar à 48 h. les souches présentant des différences significatives (test de Student) ont été annotées par une étoile (p-value < 0,05).

Ainsi, l'interaction prédite entre RNA1083 et les gènes de flagelles pourrait avoir un impact posttranscriptionnel négatif sur la production des flagelles, ce qui pourrait expliquer le défaut de mobilité.

Afin de déterminer le rôle potentiel de RNA1083 dans le transfert conjugatif du plasmide Ti, des tests de conjugaison *in planta* ont été initiés. Aucun résultat probant n'a pour le moment été obtenu.

En conclusion, 57 riborégulateurs potentiels sont transcrits à partir de régions génomiques codant des gènes impliqués dans le cycle infectieux. RNA1044, produit à partir de l'ADN-T, semble favoriser l'agressivité de la souche C58. A l'inverse, RNA1083, produit en amont du groupe de gènes impliqués dans lze transfert conjugatif, paraît être un répresseur de la virulence, sa surexepression dans la cellule induisant une diminution de l'agressivité de C58. Il régulerait la mobilité de la bactérie via *flgF, flgI, fliE* et *fliL*, l'induction de la virulence via *chvI* et la conjugaison du plasmide Ti via *trbG, trbK* et *trbL*.

V. Rôle de deux riborégulateurs dans le maintien du plasmide Ti.

Les plasmides sont des réplicons accessoires portés par les bactéries. Bien que non indispensables, ils permettent aux bactéries qui les portent des adaptations particulières à un environnement donné. Ils sont connus pour porter, entre autres, des gènes de résistance aux antibiotiques, ou des gènes permettant l'interaction (bénéfique ou pathogène) avec un hôte. Lorsqu'ils n'apportent pas un avantage sélectif à la bactérie, ils sont considérés comme un fardeau génétique. Afin de perdurer dans les cellules bactériennes, ils doivent assurer leur maintien. Pour cela, des mécanismes prévenant la perte des plasmides et des mécanismes limitant la propagation des bactéries ayant perdu les plasmides peuvent être mis en place. Les mécanismes « préventifs » assurent efficacement la réplication et la répartition des plasmides dans les cellules filles suite à la division cellulaire. Malgré un contrôle drastique de ces systèmes, le plasmide peut être perdu. Deux mécanismes « curatifs » peuvent alors intervenir pour limiter la propagation dans les populations de cellules dépourvues de plasmide. Certains plasmides peuvent d'une part être transmis horizontalement par conjugaison, ce qui permet à une bactérie l'ayant perdu de le récupérer. D'autre part, des systèmes toxine-antitoxine (TA), également appelés système d'addiction, peuvent provoquer la mort (ou empêcher la division cellulaire) des cellules ayant perdu le plasmide, et ainsi favoriser les bactéries porteuses du plasmide. Nous avons vu dans le chapitre précédent que RNA1111 et RNA1083 pourraient jouer un rôle dans la mise en place des mécanismes de conjugaison.

Au cours de notre étude, nous avons aussi identifié deux candidats qui pourraient jouer un rôle dans les mécanismes de maintien des plasmides via la réplication et un système toxine-antitoxine. A partir de nos analyses de RNA-seq, nous nous sommes intéressés aux candidats exprimés à partir de la région *repABC* de réplication du plasmide Ti. De façon surprenante, nous n'avons pas détecté de production du riborégulateur RepE connu pour contrôler la réplication du plasmide Ti de la souche R10 (Chai and Winans, 2005c) (pages 38, 76). En revanche, nous avons identifié et caractérisé un nouveau riborégulateur, nommé RNA1059, dont le gène est situé sur le brin complémentaire au gène *repE* et qui chevauche la région promotrice de *repE* et les 12 premières bases. Nous avons montré que le clonage en *trans* des deux gènes (*repE* et *rna1059*) entrainaît une perte totale de la virulence pour les souches C58 et R10, et que ceci était du à une perte du plasmide Ti. Ainsi, la sur-représentation des deux gènes *rna1059* et *repE* sont nécessaires et suffisants pour induire un phénomène d'incompatibilité avec le plasmide Ti entier.

Nous avons également identifié et caractérisé un 5^{ème} candidat, RNA1051, produit à partir de la région du catabolisme de la nopaline. Nous avons montré que la production de ce candidat était nécessaire à la croissance, des souches C58 portant un plasmide Ti ceci en milieu liquide. En effet, la

121

délétion du gène *rna1051* entraîne un important retard de croissance de la souche, tandis qu'une souche dépourvue de plasmide Ti présente une croissance similaire à celle de la souche sauvage portant le plasmide Ti. Nous proposons que RNA1051 soit l'antitoxine d'un nouveau système TA du plasmide Ti, assurant la persistance de ce plasmide chez la souche C58 d'*A. fabrum*. Nous avons ensuite tenté d'identifier la toxine de ce nouveau système.

Les résultats de ces travaux sont présentés dans deux versions provisoires de publications.

V.1 RNA1059 et RepE et la réplication du plasmide Ti.

RepE requires the presence of a new non coding RNA gene to control the Ti plasmid replication in *Agrobacterium fabrum* C58 and *A. tumefaciens* R10.

Dequivre, Magali^{1,2,3}, Villard Camille^{1,2,3} and Hommais, Florence^{1,2,3}

¹Université de Lyon, F-69622, Lyon, France

²Université Lyon 1, F-69622 Villeurbanne, France

³CNRS, UMR 5240 Microbiologie Adaptation et Pathogénie, F-69622 Villeurbanne, France

Abstract

repABC replicons are widespread among alpha-proteobacteria and their replication are tighly controlled. In *Agrobacterium*, the non-coding RNA RepE had been implicated in the control of Ti plasmid replication by regulating both transcription and translation of *repC* encoding the replication protein. Through a deep sequencing analysis of small transcripts, we identified another non coding RNA, named RNA1059, produced from the intergenic region *repB-repC*. The gene of this transcript is expressed in the opposite strand of repE, and overlaps with it over 12 bases. We showed that those two genes were highly co-conserved among Ti/Ri plasmids. Moreover, we showed that overexpression of *repE* alone or of *rna1059* alone was not sufficient to induce the incompatibility phenotype on the complete Ti plasmid but the overexpression of both *rna1059* and *repE* is necessary. We demonstrated that the plasmid incompatibility is really effective on the complete Ti plasmid and we showed it drastic impacts the pathogenic ability of both the R10 and C58 strains.

Introduction

To sustain in bacteria, plasmids required to be replicated and segregated strictly into each daughter cell. This is particularly critical for plasmids present at low copy number in cells. Among those are replicons from the *repABC* system family (Cevallos, Cervantes-Rivera, and Gutiérrez-Ríos 2008; Pinto, Pappas, and Winans 2012). They are extremely widespread among alpha-proteobacteria, where several of them could even cohabitate in the same cell and thus can belong to different incompatibility group (Inc). In *Agrobacterium tumefaciens*, the linear chromosome and both Ti and At

plasmids indeed belong to different incompatibility group of the *repABC* system (Goodner et al. 2001; Wood et al. 2001).

In *Agrobacterium*, genes involved in plasmid replication and partitioning are transcribed all together in the same operon *repABC* and this organization is unusual (Cevallos, Cervantes-Rivera, and Gutiérrez-Ríos 2008; Pinto, Pappas, and Winans 2012). The two proteins RepA and RepB are involved in the replicon partitioning. RepA, an ATPase, is able to polymerase, and RepB binds to *parS* sites localized between *repA* and *repB* genes. These two proteins work together for the correct partitioning of plasmids they belong to. However, they are not essential for replication whereas *repC* is. This gene codes for the replication protein. Furthermore, the origin of replication (*oriV*) is present in its coding sequence between nucleotides 733 and 754 for the TiR10 plasmid (Pinto et al. 2011).

Transmission of these replicons is strictly controlled. Both proteins RepA and RepB have been involved in a negative feedback regulation with the formation of a complex with the parS sites and the repABC promoter (Pérez-Oseguera and Cevallos 2013). The TraR/OOHL quorum-sensing system and the master regulator of virulence VirA-VirG can activate the replication (Cevallos et al., 2008). Furthermore, an antisense RNA (around 55 to 59 nt in length) is transcribed from the opposite strand of the repB-repC intergenic region. It has been named RepE for the pTiR10 of A. tumefaciens, ctRNA for the pSym of Sinorhizobium meliloti and for the p42d of Rhizobium etli (Venkova-Canova et al. 2004; Chai and Winans 2005; Izquierdo et al. 2005). This ncRNA has been shown to be involved in the regulation of the replication since an increased expression of the region encoding repE blocked the plasmid replication of a repABC-dependent mini-plasmid, and mutations abolishing repE promoter activity cancelled this incompatibility. Furthermore, RepE has been implicated in the control of a Ti mini-plasmid replication by regulating both transcription and translation of repC (Chai and Winans 2005). It has been proposed that, in the absence of RepE, RNA from the intergenic region repB-repC folded into a large stem-loop, releasing the RBS and the start codon of repC whereas, in its presence, RepE bound to this intergenic region preventing the RepC translation. Surprisingly, mini-plasmids harboring repABC operon without repE were not able to replicate and the overexpression of RepC has been shown not lethal, suggesting that a runaway phenotype could not explain this phenotype. It has been thus proposed that RepE could also play a role in the replication initiation complex by an unknown mechanism (Pinto et al. 2011).

Using strand specific small RNA-deep-sequencing analyses, we have previously identified small transcripts present in *A. fabrum* C58 strain (Dequivre et al. 2015). Surprisingly, the RepE antisense was expressed at very low level (below the threshold) whereas a short transcript was expressed at high level from the opposite strand of the *repE* genomic environment. It has been named RNA1059. The involvement of this transcript on the *repABC* regulation in C58 is unknown. Here, we

124

characterized this small transcript, measured its role in the control of the Ti plasmid replication and its impact on the ability of the strain to produce tumors.

Results

Two small RNAs are present in the Ti plasmid replication region of the A. fabrum C58 strain

Our previous sRNA-seq data have identified a very low level of expression of repE in the growth conditions tested (rich and minimum media) whereas a small transcript, named RNA1059, has been detected highly produced (Dequivre et al. 2015). This RNA1059 is transcribed from the opposite strand of the putative promoter of repE between base 54,790 and 54,967 (corresponding to the intergenic region of repB-repC) (Figure 1A). An overlap of 12 bases between sequences of repE and rna1059 also exists (Figure 1A). To confirm the presence of both small transcripts and to determine their 5'- and 3'-ends in the A. fabrum C58 strain, we performed RACE-PCR experiments (Rapid Amplification of cDNA End-PCR). For RepE, the 5'-end, corresponding to the transcription start, was identified at the base 54,801 on the negative strand of the Ti plasmid. Putative repE promoter was then localized with a -10 sequence (GATTCT) and a -35 sequence (CTTGAC) (with consensus TATNNT and CTTGNN respectively) (Wilms et al. 2012). The 5'-end of RNA1059 was localized at the position 54,790 from the positive strand of the Ti plasmid and its 3'-end was localised between bases 54,914 and 54,953 allowing a short transcript of around 160 bases. Search for open reading frame failed to identify a coding sequence and a RBS site, suggesting that this transcript is not translated. Secondary structure predictions using the RNAfold web server suggest the presence of 4 hairpins structures with a high probability of presence (Figure 1B). The RNA1059 transcript is small, probably strongly structured and not translated, suggesting that RNA1059 could be a new small ncRNA produced from the replication region of the Ti plasmid in A. fabrum C58.

mRNA targets amounts are not altered by RNA1059 or RepE alone

RepE has been previously shown to modulate the expression level of *repC* and mRNA target prediction identified the gene *ace* (*atu6143*), localized on the Ti plasmid, as a putative mRNA target of RNA1059. We thus evaluated the involvement of RepE and RNA1059 amounts on the expression level of these Ti plasmid genes by qRT-PCR. Strains overexpressing (i) *repE* alone (strain AT234), (ii) *rna1059* alone (strain AT208), or (iii) both *repE* and *rna1059* (strain AT195) were constructed. Attempts to obtain strain deleted for this region were unsuccessful. When cloned alone into the pBBR1MCS-5, the expression level of *repE* or *rna1059* respectively was increased by four-fold. No significant differences could be highlighted when compared the expression level of *repC* in the wild-type strain and in strains overexpressing *repE* or *rna1059* alone. Similar expression stability was also observed for the accE gene. However, when compared to the control strain (AT200), decreases of

around 10 Ct were evaluated, for bothgene (*accE* and *repC*), in the strain harboring both *repE* and *rna1059* together (Table 3).

Both RepE and RNA1059 are involved in the replication of Ti plasmid

The involvement of RNA1059 in the replication efficiency of the Ti plasmid was then evaluated in the C58 strain and copy number of each replicon were quantified by qPCR using the amplification of sequences specific to each replicon and unique in the genome. An average of one copy per cell was measured for the circular and the linear chromosomes as well as for the At plasmid whatever the overexpression of repE or rna1059 (Table 4). Surprisingly, strain overexpressing repE alone showed a Ti plasmid copy number similar to those of the respective control strains. The same observation was made for strain overexpressing rna1059 alone. On the contrary, when both sRNAs were cloned together in the plasmid pBBR1MCS-5, only one cell over 10,000 still harbored a Ti plasmid, showing a drastic decrease of Ti plasmid amount (Table 4). Similar results were also obtained with the measurement of copy numbers of the circular chromosome and of the pTiR10 in the R10 strains overexpressing repE or rna1059 or overexpressing both genes. Thus, overexpression of both repE and rna1059 seems to be necessary to prevent the Ti plasmid replication. This suggests that the decrease in expression level observed for *repC* and *accE* genes could be due to the loss of the Ti plasmid since both genes were localized in the Ti plasmid. To test this hypothesis, three genes localized in the Ti plasmid were chosen arbitrary and their expression levels were measured in the different genetic contexts. As expected, decreases of 10 Ct were evaluated for virEO (atu6188), virC2 (atu6179) and traA (atu6127) when their mRNA amounts were measured in the strain overexpressing both repE and *rna0159*, supporting the hypothesis (Table 3).

RepE and RNA1059 are co-conserved among Rhizobiaceae.

Conservation of repE and rna1059 genes was evaluated among bacteria. Similarity searches showed a lack of conservation for both genes in genomic sequences outside the Rhizobiceae family whereas twenty-six replicons belonging to the *Rhizobiaceae* family harbored significant similarities with both repE and rna1059 sequence (A. tumefaciens species complex and other strains from the Rhizobiaceae family available the Agrobacterscope database on https://www.genoscope.cns.fr/agc/microscope/about/collabprojects.php?- were analyzed) (Figure 2). Sequence environment analyses showed that every repE and rna1059 homolog genes are localized between the coding sequences of repB and repC. In particular, the poly-A sequence characteristic to repE is highly conserved and a region recovering promoters of repE-like genes is completely conserved (31 bp between repE and repC, from base 54,808 to base 54,838). This region corresponds to the overlap of 12 bases between repE and rna1059. The rna1059 gene is also conserved except the short region between nucleotide 54,845 and 54,852 that are only present in Tilike plasmid and in *Sinorhizobium meliloti* M270 and M22 (genome project in progress).

To enlarge this conservation analysis, the presence of both *repE* and *rna1059* sequences was verified in a collection of 33 *Agrobacterium* strains isolated from different pathogenic environments and previously tested for their aggressiveness on tomato plants and for the presence of Ti-like plasmids (Ludovic Vial personal communication). Amplification analyses were performed by PCR using primers A269 and A270 for *repE* and A115 and A116 for *rna1059*. Among the 33 strains tested, 81.8 % (27 strains) presented an amplification for both *repE* and *rna1059* sequences. For four strains, the sequence of *rna1059* gene was present but no amplification of *repE* could be detected, and for the two other strains, *repE* was amplified but no amplification was observed for *rna1059*. Finally, these results showed a high co-conservation of both *repE* and *rna1059* genes.

Overexpressions of repE and rna1059 in trans drastically decrease the virulence of Agrobacterium

We then measured the involvement of both RepE and RNA1059 on the Agrobacterium tumefaciens virulence on tomato plants (figure 3A and 3B). We used the C58 and the R10 strains and their derivates. Infection assays were performed by the measurement of the aggressiveness degree of the tested strains (number of tumors per plant). We compared the control strains owning an empty vector (AT287 for R10 and AT200 for C58) to strains overexpressing repE alone, rna1059 alone or both repE and rna1059 genes. The wild type R10 strain led to the formation of an average of 10.4 tumors whereas the C58 strain led to an average of 28.7 tumors. Moreover, the presence of the empty vector had no significant impact on the tumor formation (data not shown). Compared to the control strains, the overexpression of repE or rna1059 alone had no significant impact on the pattern of aggressiveness degree whatever the strain analyzed. Averages of 10.9 and 13 tumors per plants were measured for the overexpression of repE in R10 and C58 strains, respectively, and averages of 7.2 and 15 tumors per plants were measured for the overexpression of rna1059 in R10 and C58 strains, respectively. On the contrary, when both *repE* and *rna1059* genes are cloned together in the pBBR1MSC-5, the number of tumors drastically decreased for both R10 and C58 strains. Indeed, no plant developed more than 5 tumors when infected by strain AT289 or by AT195. Thus, the overexpression of both repE and rna1059 acted in synergy to induce a drastic decrease of the aggressiveness of Agrobacterium on plants.

Discussion

In this study, we identified a new small non-coding RNA, named RNA1059, approximately 160 nt in length. When it was cloned together with the *repE* gene in a plasmid and transformed into an *A*. *fabrum* strain harboring a Ti plasmid, it allowed an incompatibility phenotype leading to the loss of

the Ti plasmid and thus the loss of the aggressiveness on plants, giving rise to an avirulent phenotype. Surprisingly, the over-expression of *repE* alone - or *rna1059* alone - was not sufficient to inhibit the Ti plasmid replication. These results were observed either in the C58 and the R10 strains harboring different types of Ti plasmid. Indeed, the C58 strain harbors a nopaline-type Ti plasmid and the R10 strain an octopine-type Ti plasmid. This reinforces the functional importance of the co-conservation of *rna1059* and *repE* observed in the collection of Ti/Ri plasmids.

Taken together, these results suggest that the sole over-expression of *repE* is not able to negatively regulate replication but needs also the presence of the DNA fragment containing *rna1059*. Thus, the increase of both *repE* and *rna1059* levels could allow a runaway phenotype that leads to the incompatibility phenotype between Ti plasmid and plasmid harboring *repE* and *rna1059* genes. If the expression of the *repB-repC* intergenic, including *repE* and *rna1059* genes, or the sole presence of this DNA sequence is responsive for the incompatibility phenotype is still unknown. It has been previously noticed that transcription of RepE is essential for its function in the plasmid replication (Chai and Winans 2005), thus it remains to demonstrate if *rna1059* function on replication depends on its DNA sequence or via the expression of this DNA sequence. Whatever the molecular mechanism involved, functions of both elements seem to be essential. Thereby, it would be interesting to determine the minimal sequence required for the incompatibility phenotype remains also to be determined.

Our results seem unexpected since previous data have demonstrated that increasing the expression of *repE* blocks plasmid replication of a *repABC*-dependent miniplasmid (Chai and Winans 2005), and RepE has been described to promote termination of the *repABC* transcript by its binding to mRNA. However, our results demonstrate that RepE alone has no involvement. Furthermore, that incompatibility phenotype has been observed in the presence of plasmid harboring both *repE* and *rna1059* genes. Indeed, previous studies were performed with a fragment containing the *repE* gene but also the previously unknown gene, *rna1059* as *rna1059* gene is transcribed upstream the *repE* gene from the opposite strand upstream the *repE* gene and it also overlaps the 5'-end and the introduction of mutations in its promoter sequence increased plasmid copy number. In the light of the identification of *rna1059* and we cannot rule out that these mutations also modified the 5'-end of *rna1059* and we cannot rule out that these mutations also modified the function of RNA1059. Thereby, the apparently different results obtained here and previously were finally not antagonist.

The *repABC* plasmids are generally present at very low copy number in cell and their replication systems should be thus reliable to allow the appropriate plasmid repartition. In *R. etli*, the S-element

- proposed to be a transcriptional terminator site - worked together with the ctRNA (RepE homolog), for the replication by a *cis*-regulatory mechanism (Venkova-Canova et al. 2004). Comparison of *rna1059* and the S-element demonstrates their differences since the impact of the S-element and the ctRNA have been shown not additional whereas both *rna1059* and *repE* were shown here necessary for the replication control. Moreover, the over-expression of the S-element alone or ctRNA alone decreased the *repC* expression level whereas neither the over-expression of *repE* alone nor of *rna1059* alone modified the expression level of *repC*. The drastic decrease in *repC* expression level observed in the strain over-expressing both *repE* and *rna1059* seems to be due to the loss of the Ti plasmid. Furthermore, and in accordance with these latter data, it has been previously proposed that, in *A. tumefaciens*, the *repE* promoter has a negative impact on the *repC* promoter via a *cis*-regulatory impact because of a transcription convergence (Chai and Winans 2005).

The smallest region allowing the replication is the repC gene with a fragment containing repE (and also rna1059). repABC-mini-replicons without repE but with repC was indeed unable to replicate. It was proposed that repE is also involved in the replication initiation complex, because (i) the increase in the expression level of RepC does not lead to a lethal runaway phenotype and (ii) the oriV is not localized in repE but in the coding sequence of repC (Pinto et al. 2011). For the oriC of the R6K plasmid in *Escherichia coli*, DNA sequences acting in *cis* could optimize the *oriC* by the sequestration and the translocation of replication initiation protein (Rakowski and Filutowicz 2013). We thus proposed that the incompatibility phenotype could be explained by a titration of replication initiation protein in the sequences of *repE* and *rna1059*. According to this hypothesis, the replication induction and repression phenotypes observed according to the amount of repE could be explained. Moreover, transcriptional activities were also described to induce the ori activity potentially by a modification in the plasmid DNA supercoiling that facilitates the replication (Pinto, Pappas, and Winans 2012). A high transcriptional activity was described for RNA1059 and this could have a positive impact on the RepC replication activities. However, the simple transcriptional activities could not explain the incompatibility phenotype. Such as ColE1, RepE might thus be involved in the maturation of a primer leader that initiated the replication (Tomizawa et al. 1981).

Finally, one of the interesting results obtained here is the impact of the increase of both *repE* and *rna1059* on the complete Ti plasmid. Indeed, all previous analyses were performed on a *repABC*-dependent mini-plasmid. Here, we demonstrated that the plasmid incompatibility is really effective on the complete Ti plasmid and we demonstrated its impact on the pathogenic ability of strains since the loss of the Ti plasmid correlated with the loss of virulence. Thus, in complement to temperature (Rogler 1980), this incompatibility-plasmid could easily be useful to efficiently cure Ti plasmid from every *Agrobacterium* strain. Indeed, similar plasmid-cure was obtained with an octopine (pTiR10)

and a nopaline plasmid (pTiC58). In soil and rhizophere, many *A. tumefaciens* cells were Ti-plasmid less. Thus one could speculate a function of both *repE* and *rna1059* for the cure of Ti plasmid in the ecological environment, when the presence of Ti plasmid in cells is deleterious for the bacterial fitness.

Experimental procedures

Bacterial strains and growth conditions

Bacterial strains used are listed in Table 1. *Escherichia coli* strains were grown with shaking (160 rpm) at 37°C in Luria-Bertani medium, and *Agrobacterium* at 28°C in YPG-rich medium (5 g.L⁻¹ yeast exctact; 5 g.L⁻¹ peptone; 10 g.L⁻¹ glucose; pH 7.2). Antobiotics were supplemented when necessary at the following concentratations: for *E.coli*: 100 μ g.mL⁻¹ ampicillin and 5 μ g.mL⁻¹ gentamicin, and for *A. fabrum* 25 μ g.mL⁻¹ gentamicin.

Vector constructions

PCR amplified fragments of the intergenic region between *repB* and *repC* and of the *repE* gene were cloned into the pBBR1MCS-5 *Sal*I and *BamH*I sites, and the *rna1059* gene into the *Sal*I and *Cla*I sites of the plasmid pBBR1MCS-5 (Kovach et al. 1994), giving rise to plasmids pARA001, pARA002 and pARA003, respectively (Table 1). PCR amplifications were performed using primers 178 and 179 for the *repB* and *repC* intergenic region, primers 197 and 198 for the *repE* gene and 207 and 208 for the *rna1059* gene (Table 2).

RNA extractions

Bacteria were grown in YPG medium until exponential or stationnary phases and RNA were extracted as previously described (Dequivre et al. 2015). Absence of DNA was checked by PCR with primers A58 and A59 (Table 2). RNA purities and quantities were controlled by separating a sample on an agarose gel and by measuring A₂₆₀, A₂₈₀ and A₂₃₀ with a nanodrop spectrometer. RNAs were stored at -80°C.

Determination of RNA 5'- and 3'-ends by RACE-PCR

RNA1059 and RepE 5'- and 3'- ends were determined by RACE-PCR as previously described (Dequivre et al. 2015). Gene specific primers A254 for RepE and A257 for RNA1059 were used for reverse transcription. Junctions between the 5' and the 3'-ends of the RNA were amplified by PCR using primers A255 and A256 for *repE* and A258 and A259 for *rna1059* (Table 2). The resulting RACE-PCR products were cloned into the *Hind*III and *BamH*I sites of the pWSK29 (Wang *et al*, 1991). DNA fragments from five independent clones were then sequenced using M13fwd and M13rev primers (Invitrogen).

Genomic DNA extraction

4.10⁹ cells in stationnary phase were centrifuged at 7000 rpm for 10 min at 4°C, resuspended in 500 μ L of 50 mM Tris-HCl, 50 mM EDTA pH 8 and then incubated for 15 min at room temperature with 2 mg of lysozyme. A gently shaken incubation was performed for 1 h 30 at 37°C with 1 mg of proteinase K and 1 mL of 10 % SDS. Genomic DNA was phenol/chloroform extracted and was then propanol-2 precipitated and resuspended in 500 μ L of 10 mM Tris-HCl, 1 mM EDTA, pH 7.6 with 1 mg of RNase. DNA purities and quantities were controlled by separating a sample on an agarose gel and by measuring A₂₆₀, A₂₈₀ and A₂₃₀ with a nanodrop spectrometer.

Quantitative DNA amplification by PCR (qPCR)

The DyNAmo Flash SYBR Green qPCR kit (Thermo scientific) was used with gene-specific primers (Table 2) to quantify the corresponding gene from 1 ng of genomic DNA. Only circular chromosome was quantified for R10 strain, as no primer was available for the two other replicons. Single copy genes were selected to represent each replicon. For R10, genes *atu1Cv2-l10001 (recQ)* represented the circular chromosome, and *atu1Cv2-pll0012 (ophA)* the Ti plasmid. For C58, genes *atu0043* and *atu1823, atu3327* and *atu3916, atu5311* and *atu5449, atu6143* and *atu6151* were selected to represent respectively circular chromosome, linear chromosome, At plasmid and Ti plasmid. The thermal cycling reactions were performed using the LC480 Lightcycler from Roche according to the following conditions: an initial step at 95°C for 7 min, followed by 45 cycles at 95°C for 10 s, 58°C for 15 s and 72°C for 20 s. Specificities of the PCR primers were verified with melting curve analyses.

Reverse transcription and quantitative PCR (RT-qPCR)

Reverse transcription and quantitative PCR were performed as previously described (Dequivre et al. 2015) using gene-specific primers (Table 2). Genes *atu0231* and *atu1924* were chosen for normalization as their expressions are constant in the growth conditions tested. Specificities of the PCR primers were verified with melting curve analyses.

Plants assays

Virulence assays were performed as previously described (Dequivre et al. 2015). Tumors were observed and counted 3 weeks post infection for C58 strains, and 4 weeks post infection for R10 strains. Ten plants were used for each strain and each plant was classified according to following severity symptom categories: 0-4; 5-9; 10-14; 15-19; \geq 20 tumors per plant for C58, and 0-1; 2-3; 4-5; 6-7; \geq 8. Each virulence assay was performed in three independent experiments. Kruskal-Wallis statistic tests were applied to determine significant differences between virulence assays.

Target predictions

Target genes of small RNAs were predicted from *A. fabrum* genome as previously described (Dequivre et al., 2015), (accession number NC_003062, NC_003063, NC_003064 and NC_003065) using IntaRNA (Busch, Richter, and Backofen 2008), sTarPicker (Ying et al. 2011) and RNApredator (Eggenhofer et al. 2011) algorithms. For RNApredator the first 300 targets were selected for further analysis.

Sequence aligment and secondary structure prediction

Sequence similarity searches were performed with the BLAST program on the MAGE database (http://www.genoscope.cns.fr/agc/microscope/about/collabprojects.ph) and multiple aligments for nucleotide sequences were performed using the Mafft program version 7 (http://mafft.cbrc.jp/alignment/server/). RNA secondary structures were predicted using the RNAfold algorithm (http://rna.tbi.univie.ac.at/cgi-bin/RNAfold.cgi).

Acknowledgements

M. Dequivre received a doctoral grant from the French *Ministère de l'Education nationale de l'Enseignement Supérieure et de la Recherche.* The authors would like to thank Benjamin Diel and Céline deVos for the technical assistance. We also thank the DTAMB and the greenhouse platform of the *federation de recherche* FR41.

Literature Cited

Busch, Anke, Andreas S. Richter, and Rolf Backofen. 2008. "IntaRNA: Efficient Prediction of Bacterial sRNA Targets Incorporating Target Site Accessibility and Seed Regions." *Bioinformatics (Oxford, England)* 24 (24): 2849–56. doi:10.1093/bioinformatics/btn544.

Cevallos, Miguel A., Ramón Cervantes-Rivera, and Rosa María Gutiérrez-Ríos. 2008. "The repABC Plasmid Family." *Plasmid* 60 (1): 19–37. doi:10.1016/j.plasmid.2008.03.001.

Chai, Yunrong, and Stephen C Winans. 2005. "A Small Antisense RNA Downregulates Expression of an Essential Replicase Protein of an *Agrobacterium Tumefaciens* Ti Plasmid." *Molecular Microbiology* 56 (6): 1574–85. doi:10.1111/j.1365-2958.2005.04636.x.

Dequivre, Magali, Benjamin Diel, Camille Villard, Odile Sismeiro, Maxime Durot, Jean-Yves Coppée, Xavier Nesme, Ludovic Vial, and Florence Hommais. 2015. "Small RNA Deep-Sequencing Analyses Reveal a New Regulator of Virulence in Agrobacterium Fabrum C58." *MPMI*.

Eggenhofer, Florian, Hakim Tafer, Peter F. Stadler, and Ivo L. Hofacker. 2011. "RNApredator: Fast Accessibility-Based Prediction of sRNA Targets." *Nucleic Acids Research* 39 (Web Server issue): W149–54. doi:10.1093/nar/gkr467.

Goodner, B, G Hinkle, S Gattung, N Miller, M Blanchard, B Qurollo, B S Goldman, et al. 2001. "Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium Tumefaciens C58." *Science (New York, N.Y.)* 294 (5550): 2323–28. doi:10.1126/science.1066803. Izquierdo, Javier, Tatiana Venkova-Canova, Miguel A. Ramírez-Romero, Juan Téllez-Sosa, Ismael Hernández-Lucas, Juan Sanjuan, and Miguel A. Cevallos. 2005. "An Antisense RNA Plays a Central Role in the Replication Control of a repC Plasmid." *Plasmid* 54 (3): 259–77. doi:10.1016/j.plasmid.2005.05.003.

Kovach, M. E., R. W. Phillips, P. H. Elzer, R. M. Roop, and K. M. Peterson. 1994. "pBBR1MCS: A Broad-Host-Range Cloning Vector." *BioTechniques* 16 (5): 800–802.

Pérez-Oseguera, Angeles, and Miguel A. Cevallos. 2013. "RepA and RepB Exert Plasmid Incompatibility Repressing the Transcription of the repABC Operon." *Plasmid* 70 (3): 362–76. doi:10.1016/j.plasmid.2013.08.001.

Pinto, Uelinton M., Ana L. Flores-Mireles, Esther D. Costa, and Stephen C. Winans. 2011. "RepC Protein of the Octopine-Type Ti Plasmid Binds to the Probable Origin of Replication within repC and Functions Only in Cis." *Molecular Microbiology* 81 (6): 1593–1606. doi:10.1111/j.1365-2958.2011.07789.x.

Pinto, Uelinton M., Katherine M. Pappas, and Stephen C. Winans. 2012. "The ABCs of Plasmid Replication and Segregation." *Nature Reviews. Microbiology* 10 (11): 755–65. doi:10.1038/nrmicro2882.

Rakowski, Sheryl A., and Marcin Filutowicz. 2013. "Plasmid R6K Replication Control." *Plasmid* 69 (3): 231–42. doi:10.1016/j.plasmid.2013.02.003.

Rogler, C. E. 1980. "Plasmid-Dependent Temperature-Sensitive Phase in Crown Gall Tumorigenesis." *Proceedings of the National Academy of Sciences of the United States of America* 77 (5): 2688–92.

Tomizawa, J., T. Itoh, G. Selzer, and T. Som. 1981. "Inhibition of ColE1 RNA Primer Formation by a Plasmid-Specified Small RNA." *Proceedings of the National Academy of Sciences of the United States of America* 78 (3): 1421–25.

Venkova-Canova, Tatiana, Nora E. Soberón, Miguel A. Ramírez-Romero, and Miguel A. Cevallos. 2004. "Two Discrete Elements Are Required for the Replication of a repABC Plasmid: An Antisense RNA and a Stem-Loop Structure." *Molecular Microbiology* 54 (5): 1431–44. doi:10.1111/j.1365-2958.2004.04366.x.

Wilms, Ina, Aaron Overlöper, Minou Nowrousian, Cynthia M. Sharma, and Franz Narberhaus. 2012. "Deep Sequencing Uncovers Numerous Small RNAs on All Four Replicons of the Plant Pathogen Agrobacterium Tumefaciens." *RNA Biology* 9 (4): 446–57. doi:10.4161/rna.17212.

Wood, D W, J C Setubal, R Kaul, D E Monks, J P Kitajima, V K Okura, Y Zhou, et al. 2001. "The Genome of the Natural Genetic Engineer *Agrobacterium Tumefaciens* C58." *Science (New York, N.Y.)* 294 (5550): 2317–23. doi:10.1126/science.1066804.

Ying, Xiaomin, Yuan Cao, Jiayao Wu, Qian Liu, Lei Cha, and Wuju Li. 2011. "sTarPicker: A Method for Efficient Prediction of Bacterial sRNA Targets Based on a Two-Step Model for Hybridization." *PloS One* 6 (7): e22705. doi:10.1371/journal.pone.0022705.

Figures



Figure 1: Characterization of RNA1059 and RepE

A. Data from *A. fabrum* sRNA-seq displayed using Artemis. Strand-specific coverages are shown displayed as two plots, as raw reads aligned against the reference strain: in red, coverage displayed from plus strand and in green coverage from minus strand. The genome annotation is displayed underneath. The genomic environment of *rna1059* and *repE* locus is presented. Grey arrows correspond to genes. 5'-end and 3'-end of *rna1059* and *repE* gene was determined from RACE-PCR results. Narrows represent transcription start of *rna1059* and *repE*. (RPKM: Reads Per Kilobase of transcript per Million reads mapped). B. The most probable secondary structrure of RNA1059 (B.1) and RepE (B.2) were predicted using RNAfold. Colors of bases correspond to the probability of base pairing (violet: probability=0; red: probability=1). For unpaired regions, colors represent the probabolity of being unpaired.



Figure 2: DNA alignment of repB-repC intergenic sequences belonging to members of the repABC plasmid family.

Replicon and strains names are indicated on the left part of the figure. pTiC58 and pAtC58 correspond to Ti and At plasmids of the C58 strain, pAtk84b corresponds to pAt plasmid of *A. radiobacter k84*. pTiZutra3-1, pTiS56, pTiNCPPB925, pTiTT111, pTiB6 correspond to Ti plasmids of the strains Zutra3-1, S56, NCPPB925, TT111, B6 respectively; F2, M270 and S22 to sequences of draft genome of *A. tumefaciens* F2 and *Sinorhizobium meliloti* strains M270 and M22, respectively. Sequence p42a corresponds to the *Rhizobium* plasmid; pAtJ-07, pAtB6, pAtNCPPB925, pAtS56, pAtTT111, pAtCFBP5771, pAtCFBP5494, pAtRV3, pKerr14 and pAtS4c to At plasmid from strains J-07, B6, NCPPB925, S56, TT111, CFBP5771, CFBP5494, RV3, Kerr14 and *A. vitis* S4. Finally, pRL8 corresponds to the eighth plasmid of *R. leguminosarum* bv viciae 3841.

Α.



Β.



Figure 3: impact of repB-repC intergenique region on pathogenicity of C58 strains

The *repE* and/or *rna1059* genes were cloned into pBBR1MCS-5 plasmid. The ability to induce tumoregenicity on 3 weeksold tomato stems is evaluated in the wild type strain C58 (3.A) and R10 (3.B) and strains overexpressing *repE* (AT234/AT288), *rna1059* (AT208/AT290) or both (AT195/AT289). Bacterial genotypes showing statistical differences (Kruskall Wallis test) in their aggresiveness are noted by three stars corresponding to a p-value ≤ 0.001 .

Tables

Strains or plasmids	Genotypes	References
Strains		
DH5 α	E. coli strain (using for DNA cloning)	Stratagene
A. fabrum C58	Wild type strain from the G8 genomic species of A. tumefaciens	Wood <i>et al</i> (2001),
		Goodner <i>et al</i> (2001)
A. tumefaciens R10	Wild type strain from the G1 genomic species of A. tumefaciens	Dessaux <i>et a</i> l (1989)
AT200	A. fabrum C58 containing pBBR1MCS-5	This study
AT195	A. fabrum C58 containing pARA001	This study
AT208	A. fabrum C58 containing pARA003	This study
AT234	A. fabrum C58 containing pARA002	This study
AT287	A. tumefaciens R10 containing pBBR1MCS-5	This study
AT288	A tumefaciens R10 containing pARA002	This study
AT289	A tumefaciens R10 containing pARA001	This study
AT290	A tumefaciens R10 containing pARA003	This study
Plasmids		
pBBR1MCS-5	Host range cloning vector, Gm ^R	Kovach <i>et al</i> (1994)
pWSK29	Host range cloning vector, Amp [®]	Wang <i>et al</i> , 1991
pARA001	<i>repB-repC</i> intergenic region cloned into the pBBR1MCS-5 <i>Sal</i> I and <i>BamH</i> I sites, Gm ^R	This study
pARA002	<i>repE</i> gene cloned into the pBBR1MCS-5 <i>Sal</i> I and <i>BamH</i> I sites, Gm ^R	This study
pARA003	<i>rna1059</i> gene cloned into the pBBR1MCS-5 <i>Sal</i> I and <i>BamH</i> I sites, Gm ^R	This study

Table 1 : Strains and plasmids used in this study

Gm^R represents a resistance of the strains to gentamycin, Amp^R represents a resistance of the strains to ampicillin.

Table 2 : Primers used in this study

Names	Forward (5'-3')	Reverse (5'-3')
Cloning primers		
178-179 (repB-repC)	ACGCGTCGACCTCACCGCAAAAGAAAAAGG	CGCGGATCCCTGGAAGACCTGCCATTTGT
197-198 (<i>repE</i>)	ACGCGTCGACGCGAGACTACAGAGAAGGTC	CGCGGATCCCTGAGATGGACTGGCTCTCA
207-208 (rna1059)	ACGCGTCGACAGCTAAGAGAATCGCATTTCCA	CGCGGATCCGCAATATGGCCGAGTGTCAT
A255-A256 (<i>repE</i> RACE-PCR product)	CATAAGCTTCGTTCCGGAAGACCTTCTCT	CGCGGATCCCGGGGGGCCTTTTTCTTTT
A258-A259 (rna1059 RACE-PCR product)	CATAAGCTTCCTGGAAATGCGATTCTCTTA	CGCGGATCCTCGTAGTCAAGGGTCCCGTA
qPCR primers		
A58-A59 (atu1420)	TTCCAGTCTGACCGAGGTTT	TTCAACGATCAATTCGTCCA
A115-A116 (<i>rna1059</i>)	TCTCGCAGCTAAGAGAATCG	CCTTTTGTTAGGCAATTGAAATC
A127-A128 (atu6177)	GAACGTGTTTCAACGGTTCA	GCTAGCTTGGAAGATCGCAC
A157-A158 (atu0043)	GCATATCGATGACCACCACA	TTGACGTTCCACACAAGGAA
A167-A168 (atu0231)	GCGAACAGAATGCGGTAGAT	AAGATCGGGAAGTTCTGCTG
A191-A192 (atu1924)	CGACCTTCAACAACACGATG	CGTGTTCCTGAGCCTTCTTC
A207-A208 (atu3327)	TCGGCTGAACAGACACTCAG	TCGGAGAGTTTCACCAGGAC
A213-A214 (atu3916)	ACTGGTCGCTTTCGCATACT	CGAGAACGGTGTTACGGATT
A231-A232 (atu6151)	CAGTGATGCGAACGTTTCTG	TTAGCCATGGCCATTCTTTC
A235-A236 (atu1823)	CAAGTCGACGCTGTCCTACA	AACCTTCAGGAACTGCATGG
A247-A248 (atu6143)	TTGTCGTCTTCGTGGACATC	CGACAAGAGCGATCATCAAA
A267-A268 (atu6045)	GAGAGCGTTCCTGGAAACTG	GCTCTTGCCCAATACTCTCG
A269-A270 (repE)	GCGAGACTACAGAGAAGGTCT	CTGAGATGGACTGGCTCTC
A293-A294 (recQ)	CTGGATGCGCTTCAACCTAT	GTGTCTAGCTGGCGGTAAGG
A299-A300 (ophA)	GGCTCGATCCAGCATTCTAC	AGCACGGAGAGTGCCATC
Deverse transcription primers for DACE		
Reverse transcription primers for RACE		
experiments:		
A254 (repE)	AAACGGCGTTCCGGAAGACC	
A257 (rna1059)	CCTGGAAATGCGATTCTCT	

Underlined sequences correspond to enzymatic restriction sites

Table 3: Impact of rna109 and/or repE on the amount of mRNA

Expression factor of genes expression compared to the control strain				Expre	ssion of :	
accE	repC	virE0	virC2	traA	repE	rna1059
0.89	0.94	0.94	0.48	0.65	+	+++
1.041	1.022	0.94	0.95	1.12	+++	+
1 10 ⁻³	9.5 10 ⁻⁴	2.66 10 ⁻⁴	1.9 10 ⁻⁴	4.8 10 ⁻³	+++	+++

Expression fold change in strain overexpressing rna1059 and/or repE vs. wild type strain for 5 genes expressed from the Ti plasmid.

Table 4 : Impact of the expression levels of repE and rna1059 on the copy number of the Ti plasmid measured by qPCR.

Amount per cell of :				Overexp	ression of :
Ti	Cc	Lc	At	repE	rna1059
	C58 s				
0.68	2.48	0.42	0.16	+	+
0.4	1.34	0.26	0.09	+	+++
0.63	1.79	0.29	0.14	+++	+
9.02 10 ⁻⁵	1.49	0.27	0.1	+++	+++

Ti: correspond to the Ti plasmid; Cc: Circular chromosome; Lc: Linear chromosome; At : At plamid. *repE* and *rna1059* (alone or together) were cloned into the pBBR1MCS-5 leading to their overexpression. Similar results were obtained with the R10 strain.

En conclusion, en plus de RepE, un second riborégulateur, RNA1059, a été identifié dans la région portant les gènes impliqués dans la réplication du plasmide Ti, entre les gènes *repB* et *repC*. Ces deux candidats semblent nécessaires à la régulation de la réplication du plasmide Ti, et semblent fonctionner en synergie. Leur co-surexpression en *trans* entraîne la perte du plasmide Ti résidant. Les mécanismes exacts n'ont pas été élucidés.

V.2 Rôle potentiel de RNA1051 en tant qu'antitoxine d'un nouveau système toxine-antitoxine du plasmide Ti

Identification of a new small non-coding RNA involved in a potential toxin-antitoxin system in the Ti plasmid in the C58 *Agrobacterium fabrum* strain.

Dequivre, Magali^{1,2,3}, Villard, Camille^{1,2,3} and Hommais Florence^{1,2,3}. ¹Université de Lyon, F-69622, Lyon, France ²Université Lyon 1, F-69622 Villeurbanne, France ³CNRS, UMR 5240 Microbiologie Adaptation et Pathogénie, F-69622 Villeurbanne, France Running title : New toxin-antitoxin system in *A. fabrum* Keyword : *Agrobacterium tumefaciens,* sRNA, toxin-antitoxin, RNA1051

Abstract

Plasmids are facultative autonomous-replication DNA molecules carried by bacteria of which persistence required several regulatory mechanisms. In the plant pathogen *Agrobacterium fabrum* C58 strain (anciently called *Agrobacterium tumefaciens* C58), the virulence and the opine catabolism only depend on the presence of the Ti plasmid. We characterized here a new small non-coding RNA (ncRNA) encoded on the Ti plasmid from the region encoding genes implicated in the nopaline catabolism. It was named RNA1051. Surprisingly, we demonstrated that RNA1051 is essential to the growth of bacteria harboring Ti plasmids since growth delays were observed with strains deleted for *rna1051* whereas wild type strains deprived for or harboring pTi grew similarly. We proposed RNA1051 as a ncRNA antitoxin of a new Toxin-Antitoxin (TA) system of the Ti plasmid. To identify the toxin associated with RNA1051, we analyzed the *rna1051* genomic environment, predicted its mRNA targets and identified mutations responsible for the suppression of the growth-delayed phenotype. The most convincing toxin candidate was NoxB of which gene was localized close to *rna1051* and of which mRNA could interact with RNA1051. Thus, we identified here a new toxin-antitoxin system encoded in the nopaline catabolism cluster of genes and conserved among nopaline-type Ti plasmids.

Introduction

Non-coding RNAs (ncRNA) are small, highly structured and non-translated transcripts that function as post-transcriptional regulatory molecules. They can act by base-pairing with mRNA, modulating their stability and/or their translation; or by sequestering DNA/RNA-binding proteins (Repoila and Darfeuille, 2009). They have been implicated in the regulation of multiple physiological functions, such as iron homeostasis (Massé et al., 2007), stress response (Hoe et al., 2013) or bacterial virulence (Zhou and Xie, 2011), and the earliest described examples of ncRNAs were antisense RNAs involved in plasmid replication and persistence. The antisens RNAI has been shown to control the replication of the ColE1 plasmid in *Escherichia coli* (Tomizawa et al., 1981).

Plasmids are autonomous-replication DNA molecules carried by bacteria. They generally do not code for essential components in bacteria, but often allow the adaptation and the survival of bacteria to variations of their environment. Some of them carry numerous genes involved in antibiotic resistance (Carattoli, 2013; Saunders, 1984), symbiotic interaction with plant (Goldmann et al., 1991; Holloway et al., 1996) or virulence toward mammal (Cornelis et al., 1998) or plant hosts (Goodner et al., 2001; Wood et al., 2001). Nevertheless, their maintenances are cost-effective for the cell and sthey could be sometimes easily lost. Several mechanisms ensure their persistence in bacteria. These mechanisms are particularly important for low copy number plasmids. Active partitioning system helps the plasmid distribution among each daughter cell after the cellular division (Gordon and Wright, 2000; Nordström and Austin, 1989).⁻ Furthermore, toxin-antitoxin (TA) systems (also called addiction systems) act as post-segregation killing systems, which lead to the death of plasmid-less daughter cells, ensuring the persistence of the sole bacteria harboring plasmids (Hayes, 2003). Several ncRNAs have been demonstrated to regulate such mechanisms (Brantl, 2002). In type I TA system, the antitoxin is a ncRNA preventing the translation of the toxin protein, while in type III, the ncRNA antitoxin directly sequester the toxin protein.

In the plant pathogen *Agrobacterium tumefaciens*, the virulence strictly depends on the presence of a low copy number plasmid named Ti plasmid (for Tumor inducing). This plasmid encodes all the determinants necessary to infect plant hosts. Persistence of Ti plasmid in cells is ensured by the replication and partitioning system *repABC* where RepA and RepB are partitioning proteins and RepC is the replication protein. The letAS Toxin-Antitoxin (TA) system has been identified in the *Agrobacterium fabrum* C58 strain (anciently called *A. tumefaciens* C58) strain. It is encoded by the Ti plasmid and authors have shown that the absence of the antitoxin letA leads to a toxic expression of *ietS* (*atu6083*). The toxin, letS, is a subtilisin-like serine protease and has been involved in the inability of cells to harbor two different Ti plasmids, also named the plasmid incompatibility (Yamamoto et al., 2009).

Recently, small RNAs present in the *A. fabrum* C58 strain have been identified by several studies (Dequivre et al., 2015; Lee et al., 2013a; Möller et al., 2014a; Wilms et al., 2012c), and one antisense RNA, named RepE, has been showed to be involved the Ti plasmid replication (Chai and Winans, 2005c). Here, we attempted to characterize a new small RNA candidate, named RNA1051, and showed that it could be part of a new TA system allowing the persistence of the Ti plasmid in the C58 strain.

<u>Results</u>

Identification of a small ncRNA in the Ti plasmid of the A. fabrum C58 strain.

A small transcript, named RNA1051, has been identified previously (Dequivre et al., 2015). This transcript was of 183 bases in length and its gene was localized on the minus strand of the Ti plasmid between nucleotides 31,156 and 31,339. The *atu6021 (hyuA*) gene was localized 461 bases upstream the 5'-end of *rna0151* on the opposite strand, and the *atu6022 (noxB*) gene was localized 70 nucleotides downstream the 3'-end of *rna1051* on the same strand (Figure 1). A small transcript, named RNA1052, has been also identified between *hyuA* and *rna1051* gene, in the same strand as *rna1051*. RACE-PCR (Rapid Amplification of CDNA End) analysis was performed, from total RNA of the *A. fabrum* C58 strain cultivated in YPG medium, to validate the presence of RNA1051. Four of the six independent sequences of the 5'-3' junctions determined the transcription start site at the nucleotide 31,236 and 31,165 and differences among 3'-ends suggest the instability of this RNA (Behrens et al., 2014). When selecting the position giving the longest transcript (nucleotide 31,165), the small RNA was of 178 bases in length. This confirms the transcription of RNA1051 and results obtained for the 5'- and 3' mappings with RACE-PCR matched to those obtained previously from RNA-seq data (Dequivre et al. 2015).

In accordance with the consensus established previously (Wilms et al., 2012c) (TATNNT and CTTGNN), a putative promoter was identified between the nucleotide 31,349 and 31,356 for the -10 box (TATCCC) and between nucleotide 31,374 and 31,379 for the -35 box (CTTGTA) (Figure 1). A putative "vir box" (TNCAATTGAAPy), which corresponds to the fixation site of the virulence regulator VirG, was predicted 52 nucleotides upstream the transcription start (TAACATTTGAAC) (Figure 1) (Steck *et al* (1988).

Nucleotide sequence conservation of *rna1051* gene was searched among genomic sequences available on the NCBI database and on the Agrobacterscope database gathering several *Rhizobiaceae*

genomes. Genes homologous to *rna1051* were shown restricted to strains carrying Ti plasmid. A specific search was then performed on a collection of 21 Ti/Ri plasmid sequences, and ten harbored *rna1051* homologous genes, showing with *rna1051* C58 strain more than 97 % identities (Figure 2A). Remarkably, all plasmids were nopaline-type. To get further inside into *rna1051* conservation, the presence of this gene was also verified on a collection of 18 nopaline-type plasmids (Ludovic Vial personal communication). For that, PCR amplifications using *rna1051* specific primers were performed, and ten plasmids amplified a fragment corresponding to *rna1051*.

Secondary structure of RNA1051 was predicted using RNAalifold software from the aligned sequences (Figure 2B). It was composed of three highly probable hairpins of 9, 11 and 20 base pairs, suggesting that this small transcript is strongly structured. RNA1051 is an untranslated RNA since the scanning of its nucleotide sequence to identify ORF and RBS sequence remained unsuccessful. Taken together, these results showed RNA1051 as a small non-translated transcript structured, suggesting that RNA1051 is a new small ncRNA produced from the pTi of *A. fabrum* C58.

Inactivation of rna1051 leads to a growth decrease but this phenotype is unstable

To characterize RNA1051 cellular function, we constructed strains modified in its production by overexpressing or deleting its gene (strains AT197 and AT194, respectively). First, kinetics and rates growth were evaluated (Figure 3). Strain harboring the Ti plasmid (C58) or strain overexpressing RNA1051 (AT197) grew similarly in rich medium. Surprisingly, a drastic delay of growth was observed for the strain harboring the Ti plasmid without the *rna1051* gene (AT194). However, the strain deleted for *rna1051* gene recovered a growth rate similar to those of the wild type after at least 10 cellular division cycles in liquid medium, suggesting the instability of the phenotype. These phenomena were observed for the four mutants constructed independently, which gave rise to four reverting strains, named AT194, AT236, AT302 and AT303, respectively. The growth delay observed was unanticipated since Ti plasmid (C58) (figure 3). Specific amplification PCR allowed us to verify the presence of the Ti plasmid in the reverting strains (data not shown).

RNA1051 could interact with several mRNA targets

We predicted putative mRNA targets of RNA1051 using three predictive algorithms as previously described and we selected for further investigation several candidates encoded in the Ti plasmid (Dequivre et al., 2015). Among those were *atu6021* (*noxB*), *atu6129* (*traB*) and *atu6174* (*virB8*). For *atu6129* (*traB*), the predicted interaction region was 32 bases upstream the start codon; for *atu6174* (*virB8*), it overlapped the RBS and the start codon (12 bases upstream and 14 bases downstream the start codon); and for *noxB* the interaction was in the coding sequence, around 84 bases downstream

the start codon (Figure 4). To evaluate the impact of RNA1051 on the expression levels of these genes, RT-qPCR were performed on the four reverting strains (AT194, AT236, AT302 and AT303), on the strain overexpressing *rna1051* (AT196), and on the wild-type strain (C58). Expression levels of *hyuA*, which surrounds *rna1051*, were also analyzed. The mRNA amounts of *virB8, traB* and *hyuA* were similar in the reverting strains, in the overexpressing strain and in the wild type strains (C58 and AT200 respectively), whereas the amount of *noxB* was increased significantly by a factor of 3.82 in reverting strains when compared to the wild type strain.

Reverting strains presented several substitutions in their Ti plasmid and in their chromosomes

We sequenced the genomes of the four reverting strains (AT194, AT236, AT302 and AT303) and compared them to the genomic sequences of the parental strain and of the A. fabrum C58 reference strain available at the NCBI database (NC_003062, NC_003063, NC_003064 and NC_003065) (Goodner et al., 2001; Wood et al., 2001). We observed that the parental strain carried nine mutations compared to the wild type reference strain (Table S1). We also observed mutations only present in the reverting strains (Table S1). Among those, three mutations were localized on the Ti plasmid. Two mutations were localized in the coding sequence of the noxB gene (atu6021). The first mutation was a silent substitution at the nucleotide 30,974 (A->G) and was common to the four reverting strains. The second mutation was a substitution at the nucleotide 30,325 (C->T), leading to the conversion of the glycine 257 into a serine and was present in reverting strains AT194, AT302 and AT303. Finally, the third mutation was a substitution of a thymine into a cytosine at the nucleotide 31,512 (intergenic region in rna1052) and was present in the reverting strains AT194, AT236 and AT302. However, all these three mutations were also present in the fragment we constructed to perform the homologous recombinations and to obtain the *rna1051* deleted strains. They were thus present before the reversion event. However, the second mutation (30,325) was not observed for the reverting strain AT236, and the third mutation (31,512) was not observed for reverting strain AT303. This suggests that supplementary mutations appeared at these positions in those reverting strains (Table 3). No other mutation (deletion, insertion or substitution) was identified on the Ti plasmid (Table S1).

For the reverting strain AT303, one substitution was also identified at nucleotide 2,790,917 of the circular chromosome (G->C). It was localized in the coding sequence of the *citE* gene (*atu2788*), coding for a citrate lyase, and led to the conversion of the arginine 66 into a glycine (Table 3). For the reverting strain AT236, one substitution was identified at the nucleotide 1,954,816 on the linear chromosome (A->C). It was localized in the coding sequence of the *afuA2* gene (*atu4783*) coding for an ABC transporter and led to the conversion of the tyrosine 251 into a cysteine (Table 3).

143
Reverting strains AT194, AT236 and AT302 presented a 30kb duplication of the circular chromosome

For the reverting strains AT194, AT236 and AT302, a two-fold increase in the number of sequence reads was measured for the reads mapped in a region of 30 kb on the circular chromosome, suggesting the duplication of this region. It was localized between nucleotides 430,476 and 460,950 which are surrounded by two inverted repeated sequences of 46 bases. 49 genes were annotated in this region (from *atu0435* to *atu0471*) and most were conserved uncharacterized proteins (corresponding to 24 genes), 15 encoded phage related proteins, and several could be involved in important cellular functions (Table S2). Indeed, Atu0436 harbors the superfamily I DNA and RNA helicase the domain, Atu0446 a N-6 adenine-specific DNA methylase domain, and NusG (Atu0450) encodes a transcription antitermination protein. However, no increase in their expression levels was observed in the reverting strains (AT194, AT236, AT302) when compared to the parental strain or to reverting strain AT303 that doesn't present the duplication (data not shown). Finally, one gene showed similarity to a GTPase (Atu0452) and three other genes showed similarity to genes related to cell wall (AERS4k1_041 similar to the D-alanyl-D-alanine carboxypeptidase DacA, Atu0453 similar to yeast cell wall protein and Atu0462 similar to L-alanyl-D-glutamate peptidase).

Discussion

In this study, we attempted to functionally characterize a small transcript that is produced from the Ti plasmid of the *A. fabrum* C58 strain. We characterized this short transcript as a small non-coding RNA, highly structured, of which the encoded gene is conserved among nopaline-type Ti plasmids. We proposed that this ncRNA, named RNA1051, could play an essential function in the maintenance of the Ti plasmid among cells.

Using strains modified for the production of RNA1051 (strain deleted for or overexpressing *rna1051* gene), we showed that the absence of *rna1051* drastically decreased the growth kinetics and the growth rate of bacteria. This phenotype was shown unstable since strain deleted for *rna1051* recovered growth kinetics and rate similar to those of the wild type after more than 10 cellular divisions in liquid medium. This suggests an important role of RNA1051 for the bacterial survival. The Ti plasmid does not carry any essential gene and is accessory to *A. fabrum* since several *A. fabrum* strains are pTi-less. Furthermore, growth rates and growth kinetics of Ti plasmid-less cells (AB150) were comparable to those of strains harboring the Ti plasmid (C58) in rich media (Platt et al., 2012b). The growth decrease observed in the absence of RNA1051 was unanticipated since the *rna1051* gene is transcribed from the Ti plasmid and thus could not be essential for bacteria. Hence, RNA1051 production seems important for growth of strains carrying the Ti plasmid. This is a typical property of an antitoxin molecule from a Toxin-Antitoxin system (TA system) (Hayes, 2003) and these prompted

us to hypothesize that RNA1051 is the antitoxin of a new TA system. One TA system has already been identified in *A. fabrum* C58 (Yamamoto et al., 2009), and it is not unusual for a big plasmid to carry several TA system (Kopfmann and Hess, 2013; Milunovic et al., 2014). TA systems ensure plasmid persistence in bacterial population. Upon cell division, when a daughter cell did not received the plasmid, the labile antitoxin would be degraded, while the stable toxin would persist and would be able to kill plasmid-less cells. Hence, a stabilisation of the toxin - in the absence of RNA1051- could be responsible for the growth delay observed here with the *rna1051* deleted mutants. Among TA system types, the antitoxin molecules could be RNAs of which interaction with the toxin mRNA leads to the degradation or the translation inhibition of the latter (type I) (Gerdes et al., 1986) or could be RNAs of which interaction with the toxin protein sequesters and inhibits its activity (type III) (Fineran et al., 2009).

We then attempted to identify the associated toxin gene. To achieve this, we selected toxin candidates according to properties defined for TA system using a ncRNA as the antitoxin: (i) genes encoding the antitoxin and the toxin are generally localized close to each other, sometimes they are in the same operon or they overlap each other on different DNA strands (Hayes, 2003), (ii) mRNA toxin and RNA antitoxin interaction occurs by base-pairing, and (iii) toxins are generally small protein rather hydrophobic (Fozo et al., 2010). We predicted and selected sequences that could base-paired with RNA1051 sequence, and genes surrounding *rna1051* as *rna1051* gene is expressed from an intergenic region. Among selected candidates, we also analysed the hydrophobicity of the encoded proteins. Four candidates (*virB8, hyuA, traB* and *noxB*) were highlighted, although none of the four-selected candidates harbors all previously defined properties.

Since the identification of the toxin is not obvious, we took advantages of the instability of the growth delay observed for the deleted strains. The phenotype reversion could be due to the suppression of the harmful effect of the toxin on cells. The Ti plasmid was still present in the reverting strains (data not shown), suggesting the suppression phenotype cannot be due to the loss of the toxin following the loss of the Ti plasmid. Mutations in the toxin gene, in the toxin target gene or in any genes of which products could indirectly counteract the toxic action of the toxin, could be responsible for that phenotype suppression. To identify such mutations, we sequenced the genome of the four reverting strains. Only a few mutations were identified, and three of the four genomes showed a duplication of 30 kb between bases 430,476 and 460,950 on the circular chromosome. For two of the reverting strains, it was the sole genomic modifications identified. Since these mutants were constructed independently, the duplication events were also independent and this could have been facilitated by the presence of the two inverted repeated sequences (of 47 bp) present at the ends of the duplicated fragment. We hypothesized that there is at least one gene encoding the toxin-

target on the duplicated fragment and that the duplication of this gene could increase quantities of their products, disrupting the effect of the toxin. Among the 49 genes present, three were involved in transcription and translation function, which could be disturbed by toxin. However, no modification in their expression level was noticed. Three other genes could be interesting as they were annotated as involved in the cell wall function and could thus be targets of the toxin. The quantification of their expression levels in reverting strains could reinforce our hypothesis. Another possibility is that an anti-addiction module could be encoded in this region. This system has been described in *Dickeya dadantii* (Saavedra De Bast et al., 2008), where an homolog of a *cdd_F* TA system is encoded in the chromosome, and protects the cell from the post-segregation killing mediated by the toxin.

The duplication was not found in the fourth reverting strain. A substitution, giving rise to the modification of the CitE protein (citrate lyase), was nevertheless identified on the circular chromosome. It has been demonstrated in *Cryptococcus neoformans* that inactivation of an ATP-citrate lyase (*acl1*) led to a delay in the bacterial growth (Griffiths et al., 2012). CitE could be targeted and sequestered by the toxin, and the loss of the antitoxin could bring to an inhibition of CitE. The mutation obtained in the reverting strain could prevent the interaction with the toxin, and thus able the reverting strain AT303 to recover the parental growth rate and growth kinetics. One mutation was also localized inside the *rna1052* gene, and we showed that RNA1051 and the mRNA of *rna1052* could base pair over 11 nucleotides from nucleotide 55 to 66 for *rna1052* mRNA, and 44 to 55 for RNA1051 (data not shown). However, how this gene could encode a toxin is unknown and further investigation is needed to decipher the putative mechanism of toxicity.

Among previous putative toxin candidates, no mutation could be observed in *virB8, traB* or *hyuA*. Furthermore, none of these genes was differentially expressed in the wild type strain versus the reverting strains. Of course, we could not rule out that the interaction between the mRNA toxin and RNA1051 allowed the inhibition of translation without affecting the expression level (Desnoyers et al., 2013), but it does not seem that any of these genes could encode a protein with toxin properties. Interestingly, we identified mutations in the *noxB* gene. However, most of these mutations were also present in the pARA011 that allowed the construction of the deleted strain. The reverting strain AT236 harbours a substitution that recovered the wild type sequence at position 30,325. One can imagine that this mutation has a toxic impact on cell, but this hypothesis does not explain the observation for the three others reverting strains that still harbored the mutation at position 30,325. Furthermore, the increase in the expression levels of *noxB* observed in the four reverting strains was probably due to the modification of the promoter because of the deletion of *rna1051*. Nevertheless, NoxB harbors several interesting properties. Its gene is localized exactly downstream the *rna1051*

gene, its mRNA was predicted as a putative target of RNA1051, and *rna1051* was conserved among nopaline-type plasmid (that could be due to genomic constraint or because of functional constraint due to its interaction with noxB). Moreover, a portion of noxB was deleted with the deletion of rna1051 (45 nucleotides of the 5'extremity of noxB). It has been shown in Sinorhizobium meliloti (Milunovic et al., 2014) that the deletion of a complete toxin-antitoxin system led in a delay in the growth of the bacteria. Deletion of the TA system results in the fast loss of the antitoxin, allowing the liberation of the toxin. When the toxin has completely disappeared from the bacterial cell, growth could be restored. Here, we could speculate that the modification of noxB (or rna1052) gene could induce the formation of an altered and ineffective toxin, allowing the recovery of the growth after all the native toxin has been eliminated from the cell, suggesting also that the toxin has a bacteriostatic effect rather than a bactericide effect. NoxB codes for the D-nopaline dehydrogenase implicated in the assimilation of nopaline (Zanker et al., 1994), and is not known to be toxic for the cell. Its role as a toxin is thus surprising. One could speculate that, in the absence of nopaline, its enzymatic activity could produce a compound toxic for the bacteria or that the accumulation of NoxB in the cell allowed its translocation in the membrane leading to its destabilization (a trans- membrane helice was predicted (amino-acid 7 to 24) by TMhmm algorithm (Krogh et al., 2001)). To test the putative toxic effect of the toxin candidate, the impact of its overexpression have to be tested on the bacterial survival.

The Ti plasmid is the main actor of the virulence and its persistence in *Agrobacterium* cells is important when the bacteria is in interaction with a plant it can infect. The antitoxin function of RNA1051 would be essential in the virulence context, as the aggressiveness of reverting strains was shown comparable to those of the wild type (data not shown), suggesting the correct maintenance of the plasmid. Nevertheless, this TA system might not be always functional. In the environment, Agrobacteria are mainly found without the Ti plasmid in the soil, in the rhizosphere and in the tumor environment, where strains harboring truncated Ti plasmids could be also identified. It is thus interesting that the TA system is encoded from the nopaline catabolism cluster of genes essential for the bacterial fitness in tumors, according to the opine concept (Dessaux et al, 1998). Thus this TA system could have an important role in the maintenance of the nopaline catabolism function of the Ti plasmid.

Experimental procedures

Bacterial strains and growth conditions

Bacterial strains used in this study were listed in Table 1. *Escherichia coli* strains were grown with shaking (160 rpm) at 37°C in Luria-Bertani media (Miller, 1992) and *A. fabrum* at 28°C in YPG-rich

media (yeast extract 5 g.L⁻¹; peptone 5 g.L⁻¹; glucose 10 g.L⁻¹; pH=7.2). Media were supplemented when necessary with antibiotics at the following concentrations: 100 μ g.mL⁻¹ ampicillin and 5 μ g.mL⁻¹ gentamicin, for *E. coli* and 25 μ g.mL⁻¹ gentamicin, 25 μ g.mL⁻¹ neomycin and 25 μ g.mL⁻¹ kanamycin for *A. fabrum*.

Growth kinetics

C58 derivative strains were inoculated at OD_{600} 0.05 in YPG rich medium and growth kinetics were followed by the measure of the OD_{600} every two hours post-inoculation.

Vector construction

DNA fragment between nucleotides 31,070 and 31,427 of the C58 Ti plasmid (corresponding to *rna1051* gene) was amplified by PCR using primers 184 and 185 (Table 2). Fragments were purified and digested using *Sal*I and *Bam*HI enzymes, and cloned into pBBR1MCS-5, giving rise to pARA004 (Table 1).

Constructions of mutant strain deleted for rna1051 gene

The *A. fabrum* C58 strain was used to construct a mutants deleted for *rna1051* gene as previously described (Lassalle et al., 2011). Genomic region between nucleotides 31,041 and 31,480 was deleted and replaced by the neomycine-kanamycine resistance gene *nptll*, giving rise to strain AT194 (Table 1). For that, sequences adjacent to *rna1051* gene were amplified and cloned with the *nptll* gene instead of *rna1051* gene in the pJQ200sk, giving rise to pARA011 (Table 1).

RNA extractions

RNA extractions were performed as as previously described (Dequivre et al., 2015) from bacteria grown in YPG medium until exponential or stationary phases. Absence of DNA was checked by PCR with primers A58 and A59 (Table 2). RNA purities and quantities were controlled by separating a sample on an agarose gel and by measuring A260, A280 and A230 with a nanodrop spectrometer.

Determination of RNA 5'- and 3'-ends by RACE-PCR

RACE-PCRs were performed as previously described (Dequivre et al., 2015). Junctions between the 5'-end and the 3'-end of the RNA were then amplified by PCR using primers A261 and A262 (Table 2). The resulting RACE-PCR products were cloned into the *Hind*III and *BamH*I sites of the pWSK29. DNA fragments were then sequenced using the M13fwd and the M13rev primers (Invitrogen).

Genomic DNA extraction

3.9 10^9 cells in stationary phase were centrifuged at 7000 rpm for 10 min at 4°C, resuspended in 500 μ L of 50 mM Tris-HCl, 50 mM EDTA pH 8 and then incubated for 15 min at room temperature with 2

mg of lysozymes. A gently shaken incubation was performed for 1 h 30 at 37°C with 1 mg of proteinase K and 1 mL of 10 % SDS. Phenol/chloroform extracted DNA was then propanol-2 precipitated and resuspended in 500 μ L of 10 mM Tris-HCl, 1 mM EDTA, pH 7.6 with 100 μ g.mL⁻¹ of RNase. DNA purities and quantities were controlled by separating a sample on an agarose gel and by measuring A260, A280 and A230 with a nanodrop spectrometer.

Genome sequencing

Genomic DNA of wild-type C58 strain and its derivates were extracted and purified as described above. The 300 to 400 bp paired-end libraries were constructed using the TruSeq DNA kit (Illumina) from 1 µg of genomic DNA and sequenced on the Illumina MiSeq (ProfilExpert platform). Paired-end sequencing of 151 bp was performed (kit v2) from the five genomic libraries on one run. Analyses were performed using the Illumina pipeline. The resulting sequences were mapped on the published genome of *A. fabrum* C58 (NC_003062, NC_003063, NC_003064 and NC_003065) with the bowtie algorithm (Langmead et al., 2009). SNP and Indel searches were performed using the Integrated Genome browser IGB (Nicol et al., 2009) and the MPpileup analyses (Li et al., 2009) on the galaxy web-based Platform (Blankenberg et al., 2010).

Reverse transcription and quantitative PCR (RT-qPCR)

Reverse transcriptions and quantitative PCR were performed using primers listed in Table 2, as previously described (Dequivre et al., 2015). Genes *atu0231* and *atu1924* were chosen for normalization as their expressions are constant in the growth conditions tested. The specificity of the PCR primers was verified with a melting curve analysis.

Secondary structure predictions and similarity searches

RNA secondary structures were predicted using the RNAalifold algorithm (http://rna.tbi.univie.ac.at/cgi-bin/RNAalifold.cgi). Basic options were chosen for folding. Sequence similarity searches were performed using the BLAST algorithm (Altschul et al., 1990). Sequences similar to the *rna1051* gene were searched among Ti plasmid sequences available on the Genbank database and among a collection of 21 Ti/Ri plasmids those sequences were available privately.

Bacterial small RNA target predictions

mRNA target genes of RNA1051 were selected as previously described (Dequivre et al., 2015). They were predicted from *A. fabrum* C58 genome using IntaRNA (Busch et al., 2008), sTarPicker (Ying et al., 2011) and RNApredator (Eggenhofer et al., 2011) algorithms. The first 300 targets were selected for further analysis and predicted targets found by every algorithm and localized in the Ti plasmid were chosen for further investigation.

Acknowledgements

M. Dequivre received a doctoral grant from the French *Ministère de l'Education nationale de l'Enseignement Supérieure et de la Recherche*. The authors would like to thank Benjamin Diel, Javier Crespo Manzo and Mélodie Bancal for the technical assistance, and ProfilExpert for the assistance with the genomic sequences. We also thank the DTAMB and the greenhouse platform of the *fédération de recherche* FR41.

Literature Cited

Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman. 1990. "Basic Local Alignment Search Tool." *Journal of Molecular Biology* 215 (3): 403–10. doi:10.1016/S0022-2836(05)80360-2.

Behrens, Sebastian, Stefanie Widder, Gopala Krishna Mannala, Xiaoxing Qing, Ramakanth Madhugiri, Nathalie Kefer, Mobarak Abu Mraheil, Thomas Rattei, and Torsten Hain. 2014. "Ultra Deep Sequencing of Listeria Monocytogenes sRNA Transcriptome Revealed New Antisense RNAs." Edited by Stefan Bereswill. *PLoS ONE* 9 (2): e83979. doi:10.1371/journal.pone.0083979.

Blankenberg, Daniel, Gregory Von Kuster, Nathaniel Coraor, Guruprasad Ananda, Ross Lazarus, Mary Mangan, Anton Nekrutenko, and James Taylor. 2010. "Galaxy: A Web-Based Genome Analysis Tool for Experimentalists." *Current Protocols in Molecular Biology / Edited by Frederick M. Ausubel ... [et Al.]* Chapter 19 (January): Unit 19.10.1–21. doi:10.1002/0471142727.mb1910s89.

Brantl, Sabine. 2002. "Antisense RNAs in Plasmids: Control of Replication and Maintenance." *Plasmid* 48 (3): 165–73. doi:10.1016/S0147-619X(02)00108-7.

Carattoli, Alessandra. 2013. "Plasmids and the Spread of Resistance." *International Journal of Medical Microbiology* 303 (6-7): 298–304. doi:10.1016/j.ijmm.2013.02.001.

Chai, Yunrong, and Stephen C Winans. 2005. "A Small Antisense RNA Downregulates Expression of an Essential Replicase Protein of an *Agrobacterium Tumefaciens* Ti Plasmid." *Molecular Microbiology* 56 (6): 1574–85. doi:10.1111/j.1365-2958.2005.04636.x.

Cornelis, G. R., A. Boland, A. P. Boyd, C. Geuijen, M. Iriarte, C. Neyt, M. P. Sory, and I. Stainier. 1998. "The Virulence Plasmid of Yersinia, an Antihost Genome." *Microbiology and Molecular Biology Reviews: MMBR* 62 (4): 1315–52.

Dequivre, Magali, Benjamin Diel, Camille Villard, Odile Sismeiro, Maxime Durot, Jean-Yves Coppée, Xavier Nesme, Ludovic Vial, and Florence Hommais. 2015. "Small RNA Deep-Sequencing Analyses Reveal a New Regulator of Virulence in Agrobacterium Fabrum C58." *MPMI*.

Desnoyers, Guillaume, Marie-Pier Bouchard, and Eric Massé. 2013. "New Insights into Small RNA-Dependent Translational Regulation in Prokaryotes." *Trends in Genetics: TIG* 29 (2): 92–98. doi:10.1016/j.tig.2012.10.004.

Dessaux, Y., Petit, A., Farrand, S.K., Murphy, P.J., 1998. Opines and Opine-Like Molecules Involved in Plant-Rhizobiaceae Interactions, in: Spaink, H.P., Kondorosi, A., Hooykaas, P.J.J. (Eds.), The Rhizobiaceae. Springer Netherlands, Dordrecht, pp. 173–197.

Fineran, Peter C., Tim R. Blower, Ian J. Foulds, David P. Humphreys, Kathryn S. Lilley, and George P. C. Salmond. 2009. "The Phage Abortive Infection System, ToxIN, Functions as a Protein-RNA Toxin-

Antitoxin Pair." *Proceedings of the National Academy of Sciences of the United States of America* 106 (3): 894–99. doi:10.1073/pnas.0808832106.

Fozo, Elizabeth M., Kira S. Makarova, Svetlana A. Shabalina, Natalya Yutin, Eugene V. Koonin, and Gisela Storz. 2010. "Abundance of Type I Toxin-Antitoxin Systems in Bacteria: Searches for New Candidates and Discovery of Novel Families." *Nucleic Acids Research* 38 (11): 3743–59. doi:10.1093/nar/gkq054.

Gerdes, K., P. B. Rasmussen, and S. Molin. 1986. "Unique Type of Plasmid Maintenance Function: Postsegregational Killing of Plasmid-Free Cells." *Proceedings of the National Academy of Sciences of the United States of America* 83 (10): 3116–20.

Goldmann, A., C. Boivin, V. Fleury, B. Message, L. Lecoeur, M. Maille, and D. Tepfer. 1991. "Betaine Use by Rhizosphere Bacteria: Genes Essential for Trigonelline, Stachydrine, and Carnitine Catabolism in Rhizobium Meliloti Are Located on pSym in the Symbiotic Region." *Molecular Plant-Microbe Interactions: MPMI* 4 (6): 571–78.

Goodner, B, G Hinkle, S Gattung, N Miller, M Blanchard, B Qurollo, B S Goldman, et al. 2001. "Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium Tumefaciens C58." *Science (New York, N.Y.)* 294 (5550): 2323–28. doi:10.1126/science.1066803.

Gordon, G. S., and A. Wright. 2000. "DNA Segregation in Bacteria." *Annual Review of Microbiology* 54: 681–708. doi:10.1146/annurev.micro.54.1.681.

Griffiths, Emma J., Guanggan Hu, Bettina Fries, Mélissa Caza, Joyce Wang, Joerg Gsponer, Marcellene A. Gates-Hollingsworth, Thomas R. Kozel, Louis De Repentigny, and James W. Kronstad. 2012. "A Defect in ATP-Citrate Lyase Links Acetyl-CoA Production, Virulence Factor Elaboration and Virulence in Cryptococcus Neoformans." *Molecular Microbiology* 86 (6): 1404–23. doi:10.1111/mmi.12065.

Hayes, Finbarr. 2003. "Toxins-Antitoxins: Plasmid Maintenance, Programmed Cell Death, and Cell Cycle Arrest." *Science (New York, N.Y.)* 301 (5639): 1496–99. doi:10.1126/science.1088157.

Hoe, Chee-Hock, Carsten A Raabe, Timofey S Rozhdestvensky, and Thean-Hock Tang. 2013. "Bacterial sRNAs: Regulation in Stress." *International Journal of Medical Microbiology: IJMM* 303 (5): 217–29. doi:10.1016/j.ijmm.2013.04.002.

Holloway, P., W. McCormick, R. J. Watson, and Y. K. Chan. 1996. "Identification and Analysis of the Dissimilatory Nitrous Oxide Reduction Genes, nosRZDFY, of Rhizobium Meliloti." *Journal of Bacteriology* 178 (6): 1505–14.

Kopfmann, Stefan, and Wolfgang R. Hess. 2013. "Toxin-Antitoxin Systems on the Large Defense Plasmid pSYSA of Synechocystis Sp. PCC 6803." *The Journal of Biological Chemistry* 288 (10): 7399–7409. doi:10.1074/jbc.M112.434100.

Krogh, A., B. Larsson, G. von Heijne, and E. L. Sonnhammer. 2001. "Predicting Transmembrane Protein Topology with a Hidden Markov Model: Application to Complete Genomes." *Journal of Molecular Biology* 305 (3): 567–80. doi:10.1006/jmbi.2000.4315.

Langmead, Ben, Cole Trapnell, Mihai Pop, and Steven L. Salzberg. 2009. "Ultrafast and Memory-Efficient Alignment of Short DNA Sequences to the Human Genome." *Genome Biology* 10 (3): R25. doi:10.1186/gb-2009-10-3-r25.

Lassalle, Florent, Tony Campillo, Ludovic Vial, Jessica Baude, Denis Costechareyre, David Chapulliot, Malek Shams, et al. 2011. "Genomic Species Are Ecological Species as Revealed by Comparative Genomics in Agrobacterium Tumefaciens." *Genome Biology and Evolution* 3: 762–81. doi:10.1093/gbe/evr070.

Lee, Keunsub, Xiaoqiu Huang, Chichun Yang, Danny Lee, Vincent Ho, Kan Nobuta, Jian-Bing Fan, and Kan Wang. 2013. "A Genome-Wide Survey of Highly Expressed Non-Coding RNAs and Biological

Validation of Selected Candidates in Agrobacterium Tumefaciens." *PloS One* 8 (8): e70720. doi:10.1371/journal.pone.0070720.

Li, Heng, Bob Handsaker, Alec Wysoker, Tim Fennell, Jue Ruan, Nils Homer, Gabor Marth, Goncalo Abecasis, Richard Durbin, and 1000 Genome Project Data Processing Subgroup. 2009. "The Sequence Alignment/Map Format and SAMtools." *Bioinformatics (Oxford, England)* 25 (16): 2078–79. doi:10.1093/bioinformatics/btp352.

Massé, Eric, Hubert Salvail, Guillaume Desnoyers, and Mélina Arguin. 2007. "Small RNAs Controlling Iron Metabolism." *Current Opinion in Microbiology* 10 (2): 140–45. doi:10.1016/j.mib.2007.03.013.

Milunovic, Branislava, George C. diCenzo, Richard A. Morton, and Turlough M. Finan. 2014. "Cell Growth Inhibition upon Deletion of Four Toxin-Antitoxin Loci from the Megaplasmids of Sinorhizobium Meliloti." *Journal of Bacteriology* 196 (4): 811–24. doi:10.1128/JB.01104-13.

Möller, Philip, Aaron Overlöper, Konrad U. Förstner, Tuan-Nan Wen, Cynthia M. Sharma, Erh-Min Lai, and Franz Narberhaus. 2014. "Profound Impact of Hfq on Nutrient Acquisition, Metabolism and Motility in the Plant Pathogen Agrobacterium Tumefaciens." *PloS One* 9 (10): e110427. doi:10.1371/journal.pone.0110427.

Nicol, John W., Gregg A. Helt, Steven G. Blanchard, Archana Raja, and Ann E. Loraine. 2009. "The Integrated Genome Browser: Free Software for Distribution and Exploration of Genome-Scale Datasets." *Bioinformatics (Oxford, England)* 25 (20): 2730–31. doi:10.1093/bioinformatics/btp472.

Nordström, K., and S. J. Austin. 1989. "Mechanisms That Contribute to the Stable Segregation of Plasmids." *Annual Review of Genetics* 23: 37–69. doi:10.1146/annurev.ge.23.120189.000345.

Platt, Thomas G., James D. Bever, and Clay Fuqua. 2012. "A Cooperative Virulence Plasmid Imposes a High Fitness Cost under Conditions That Induce Pathogenesis." *Proceedings. Biological Sciences / The Royal Society* 279 (1734): 1691–99. doi:10.1098/rspb.2011.2002.

Repoila, Francis, and Fabien Darfeuille. 2009. "Small Regulatory Non-Coding RNAs in Bacteria: Physiology and Mechanistic Aspects." *Biology of the Cell / Under the Auspices of the European Cell Biology Organization* 101 (2): 117–31. doi:10.1042/BC20070137.

Saavedra De Bast, Manuel, Natacha Mine, and Laurence Van Melderen. 2008. "Chromosomal Toxin-Antitoxin Systems May Act as Antiaddiction Modules." *Journal of Bacteriology* 190 (13): 4603–9. doi:10.1128/JB.00357-08.

Saunders, J. R. 1984. "Genetics and Evolution of Antibiotic Resistance." *British Medical Bulletin* 40 (1): 54–60.

Tomizawa, J., T. Itoh, G. Selzer, and T. Som. 1981. "Inhibition of ColE1 RNA Primer Formation by a Plasmid-Specified Small RNA." *Proceedings of the National Academy of Sciences of the United States of America* 78 (3): 1421–25.

Wilms, Ina, Aaron Overlöper, Minou Nowrousian, Cynthia M. Sharma, and Franz Narberhaus. 2012. "Deep Sequencing Uncovers Numerous Small RNAs on All Four Replicons of the Plant Pathogen Agrobacterium Tumefaciens." *RNA Biology* 9 (4): 446–57. doi:10.4161/rna.17212.

Wood, D W, J C Setubal, R Kaul, D E Monks, J P Kitajima, V K Okura, Y Zhou, et al. 2001. "The Genome of the Natural Genetic Engineer *Agrobacterium Tumefaciens* C58." *Science (New York, N.Y.)* 294 (5550): 2317–23. doi:10.1126/science.1066804.

Yamamoto, S., K. Kiyokawa, K. Tanaka, K. Moriguchi, and K. Suzuki. 2009. "Novel Toxin-Antitoxin System Composed of Serine Protease and AAA-ATPase Homologues Determines the High Level of Stability and Incompatibility of the Tumor-Inducing Plasmid pTiC58." *Journal of Bacteriology* 191 (14): 4656–66. doi:10.1128/JB.00124-09.

Zanker, H., G. Lurz, U. Langridge, P. Langridge, D. Kreusch, and J. Schröder. 1994. "Octopine and

Nopaline Oxidases from Ti Plasmids of Agrobacterium Tumefaciens: Molecular Analysis, Relationship, and Functional Characterization." *Journal of Bacteriology* 176 (15): 4511–17.

Zhou, Yexin, and Jianping Xie. 2011. "The Roles of Pathogen Small RNAs." *Journal of Cellular Physiology* 226 (4): 968–73. doi:10.1002/jcp.22483.



Figures

Data from A. fabrum sRNA-seq displayed using Artemis. Strand-specific coverages are shown displayed as two plots, as raw reads aligned against the reference strain: in red, coverage displayed from plus strand and in green coverage from minus strand. The genome annotation is displayed underneath. The genomic environment of *rna1051* locus is presented. Grey arrows correspond to genes. 5'-end and 3'-end of *rna1051* gene was determined from RACE-PCR results. Sequences in grey boxes correspond to putative the -10 box (TATCCC, with consensus TATNNT), and -35 box (CTTGTA with consensus CTTGN of the promoter region of *rna1051*, and to the putative *vir* box (TAACATTTGAAC with the consensus TNCAATTGAAPy). +1 represents transcription start of *rna1051* gene. (RPKM: Reads Per Kilobase of transcript per Million reads mapped)

CLUSTAL multiple sequence alignment by NUSCLE (3.8)
B1 AMGGMMGATTCAMAGTACCAGACCTGAMTGACCCGCTTCACHTGCGTTCACHTGCGTACCATTCAGACGAGCAAAAAAAGAGTCCCTCAAAACTGCGCCGCCAAAAAAAA
CFBP5473 -AAGGAAGAATTCCAAAGTAGCCAGACCTGGAATTCGACGCGGTTCGACTGGTCGATTTTTTTCAAGAAGAGAATCCGAAGCAAAGAAGTTCTCAAAAACGGCGGGCGCAGAGAAGAATTCCAAGATTAGCCAGACCGATGGCGGTTCGACTGGGTCGACTCAGTGGTGGACATCGGACGGA
CSS GAAGGAAGANTTCCAAGANTAGCCAARTTCCACATRAACCGCGTTCCACATGGTGCTCACTGGTCACGACGAAGAAGGATCCTCAAAAATTGGCCGATAAACGGCG GFB5580 GAAGGAAGANTTCCAAGANTAGCCAARTTGACCGCTTCCAATGGGGTCACTCGGTCCAAGACGGAAGAAGGATCCTCAAAAAGGGCCTCAAAAACGGCG GFB5570 GAAGGAAGANTTCCAAGANGCCAGATCGCAATGAAGCCGGTTCCACATGGGGTCACTCGGTCCAAGACGGAAGAAGGTCCTCAAAAACGGCG GAAGGAAGAAGATTCCAAGANAGCCAGATCGCAATGAACGCGGTCCACATGGGGTCACTCGGTCCAAGACGGAAGAAGGACCAAGAAGGACCTCAAAAACGGCG GAAGGAAGAAGATTCCAAGANAGCCAGATCGCAATGAACGCGCGTCCACATGGGGTCACTCGGTCCAAGACGGAAGAAGGTCCTCCAAAACGGGCG JUE17935 GAAGGAAGAATTCCAAGANAGCCAGATCGCAATGAAGCCGGCGTCCACATGGGGTCACTCGGTCCAAGACGGAAGAAGGTCCTCAAAACGGGCG JUE17937 GAAGGAAGAATTCCAAGANAGCCAGATCGCAATGAAGCCGCGCTTCAACATGGGGTCACATCGGTCGAAGACGGAAGAAGGACCAAGAAGGACCTCCAAAACGGGCG JUE17937 ACAGCTGGAAACACTGTAAGCCAGACCGCGCTCAAATGAGGCGGTCGACATCGGGTCGAACAGGGACGAAGAAGGTCCCCAAAGAAGAGGCCCCCGAAGAAGAGCCAAGAGGATCGAAGAGGACGCAGGAGGAAGAAGGCCACCGCGCTCAAACGGGC AAAGGCAAGAAACACTGGAAGACCCGGCGCGAAGAAGAATGAAGGATCGAGTTGGGCTCAA- LMS294 ACAACTGGACACCCGGCGCAAAGAGGATGAAGGATGAGGTCGGCTCGGCCCACT CSD54 ACAGCTGGGACACCTGTATAGCCGACCGCGCGCAGAGGAATGAAGGATGCGTCGGCCCCACT CSD54 ACAGCTGGGACACCCTGTATAGCCGACCGCGCAGGGAATGAAGGATGCGCTGGGCCCACT CSD54 ACAGCTGGGACACCTGTATAGCCGACCGCGCGCGAGGGAATGAAGGATGCGCTCGGCCCCACT CSD54 ACAGCTGGGACACCCTGTATAGCCGACCGCGCGAGGGAATGAAGGATGCGCTCGGCCCCACT CSD54 ACAGCTGGGACACCCTGTATAGCCGACCCGCACGAGGAATGAAGGATGCGCTCGGCCCCACT CSD54 ACAGCTGGGACACCCTGTATAGCCGACCCGCACGAGGAATGAAGGATGCGCTCGGCCCCACT CSD54 ACAGCTGGGACACCCTGTATAGCCGACCCGCACGAGGAATGAAGGATGCGCTCGGCCCCACT CSD54 ACAGCTGGGACACCCTGTATAGCCGACCGCACGAGGAATGAAGGATGCGCTCGGCCCCACT CSD5540 ACAGCTGGGACACCCTGTATAGCCGACCCGCACGAGGAATGAAGGATGCGCTCGGCCCCACT CSD5540 ACAGCTGGACACCCTGTATAGCCGACCCGCACGAGGAATGAAGGATGCGCTCGGCCCCACT CSD5540 ACAGCTGGACACCCTGTATAGCCGACCGCACGAGGAATGAAGGATGCGCTCGGCCCCACT CSD5540 ACAGCTGGACACCCTGTATAGCCGACCCGCACGAGGAATGAAGGATGCGCTCGGCCCCACT CSD5540 ACAGCTGGACACCCTGTATAGCCGACCCGCGCACGAGGAATGAAGGATCGCCTCGGCCCCACT CCACGTGGACACACCTGTATAGCCGACCCGCACGAGGAATGAAGGATGCGCTCGGCCCCCC CCACGTGGCCACCCCGCGCACGCGCGCGCGAGGAATGAGGCTCGGC
CFBP5488 GAAGGAAGAATTCCAAGATAGCCAGATCTGCAATAGACCGGCTTCACATGGGTGCATCTGTTCAAGACGTGGAAAGAAGGAGCTCCTCAAAACTGGCCGAAAAAGGGCCTCCAAAAACGGCG GAAGGCAAGAATTCCAAGATAGCCAGATCTGCAATGACGCGGTTCAACAGGGGTGGGT
LkG17953 GAAGGAAGAATTCCAAGATAGCAAGTTCCAAGTAGCCAACTCGACCGCCTTCAATGGGGATCATCGTTCTGTCTAAGACTGATGAAACGAAGAAGGTCCTCAAAACTGGCCGTCAATGGGCGATCAATGGCCGGTTCAATGGGCGGTTCAATGGGCGGTTCAATGGGCGGAAGAAGGTCCTCAAAACTGGCCGGCAAGAAGGTCCTCAAAACTGGCCGGCAAGAAGGTCCTCAAAACTGGCCGGCAAGAAGGTCCTCAAAACTGGCCGGCAAGAAGGTCCTCAAAACTGGCCGGCAAGAAGGTCCTCAAAACTGGCCGGCAAGAAGGTCCTCAAAACTGGCCGGCAAGAAGGTCCTCAAAACTGGCCGGCAAGAAGGTCCCTCAAAACTGGCCGGCAAGAAGGTCCCTCAAAGCAGGTCGGCAAGAAGGTCCCTCAAAACTGGCCGTCAACTGGCGCGCAAGAAGGTCCCAAGAGGATCGAAGAAGGTCCCTCAAAGAGGTCCCTCAAAGACGGAAGAAGGTCCCTCAAAGAGGTCCCAAAGAAGGTCCCTCAAAGACGGAAGAAGGTCCCTCAAAGACGGACG
Zutra3-1 GAAGGCAAGAATTCCAAGATTAGCCAGATCTSCAATCGACCGCCTTCACATGGTGATCATCTGTTCAAGACTGATGAAACGGAAGAAGGTCCTCCAAAACTGGCCGTCAAACGGGCG MAFF301001 GAAGGCAAGAATTCCAAGATTAGCCAGATCTGCCAATGGGCGTTCAACTGGGTGATCATCTGTTCAAGACTGATGAAACGGAAGCAAAGAGGTCCTCAAAACTGGCCGTCAAACGGGC CFBP5473 ACAGCTGGAAACACTGTATAGCCGGCCGCCAATAGGGATGAAGATTCAGTTTGGCTCA LMC294 ACAACTGGACACCGCGTATAGCCGACCGCACGAGGGATTGAAGGATCCAGTCTGGGCTCACT CSB ACAGCTGGGACACACTGTATAGCCGACCGCACGGAGGATTGAAGGATGCAGTCGGCTCGCTC
GF8P5473 ACAGCTGGAAACACTGTATAGCCGGCCGCATAAGGGATGAAGATTCAGTTGGGCTCA LMG294 ACAGCTGGGACACCTGTATAGCCGACCGCACAGGGATTGAAGGATGCGTCTGGCTCACT Chbika ACAGCTGGGACACCTGTATAGCCGACCGCACAGGGATTGAAGGATGCGTCTGGCTCACT CF8P5476 ACAGCTGGACACACTGTATAGCCGACCGCACAGGGATTGAAGGATGCGTCTGGCTCACT CF8P5476 ACAGCTGGACACACTGTATAGCCGACCGCACAGGGATTGAAGGATGCGTCTGGCTCACT LMG17935 ACAGCTGGACACACTGTATAGCCGACCGCACAGGGATTGAAGGATGCGTCTGGCTCACT LMG17936 ACAGCTGGACACACTGTATAGCCGACCGCACAGGGATTGAAGGATGCGTCTGGCTCACT S56 ACAGCTGGACACACTGTATAGCCGACCGCACAGGGATTGAAGGATGCGTCTGGCTCACT MAFF301001 ACAGCTGGACACACTGTATAGCCGACCGCACAGGGATTGAAGGATGCGTCTGGCTCACT MAFF301001 ACAGCTGGACACACTGTATAGCCGACCGCACGA-GATTGAAGGATGCGTCTGGCTCACT MAFF301001 ACAGCTGGACACACTGTATAGCCGACCGCACGA-GATTGAAGGATGCGTCTGGCTCACT
B.
Chbika ACAGCTGGACACACTGTATAGCCGACCGACGAGAGATTGAAGGATGCGTCAGCT CFBF5476 ACAGCTGGACACACTGTATAGCCGACCGACGAGATTGAAGGATGCGTCTGGCTCACT CFBF54776 ACAGCTGGACACACTGTATAGCCGACCGACGAGATTGAAGGATGCGTCGGCTCACT S56 ACAGCTGGACACACTGTATAGCCGACCGACGAGGATTGAAGGATGCATCTGGGCTCACT ACAGCTGGACACACTGTATAGCCGACCGACGAGGATTGAAGGATGCGTCGGCTCACT MAFF301001 ACAGCTGGACACACTGTATAGCCGACCGACGAGATTGAAGGATGCGTCTGGCTCACT BAFF301001 ACAGCTGGACACACTGTATAGCCGACCGACGACGAGATTGAAGGATGCGTCTGGCTCACT ACAGCTGGACACACTGTATAGCCGACCGACGACGACGAGGATTGAAGGATGCGTCTGGCTCACT BAFF301001 ACAGCTGGACACACTGTATAGCCGACCGACGACGACGAGGATTGAAGGATGCGTCTGGCTCACT ACAGCTGGACACACTGTATAGCCGACCGACGACGACGACGAGGATTGAAGGATGCGTCTGGCTCACT BAFF301001 ACAGCTGGACACACTGTATAGCCGACCGACGACGACGACGACGACGACTGGTCGGCTCACT ACAGCTGGACACACTGTATAGCCGACCGACGACGACGACGACGACGAGGATGGAT
CPBP5480 ACAGCTGGACACACTGTATAGCCGACCGACGAGGATTGAAGGATGCATGAGGTCGCTCTGGCTCACT CPBP5476 ACAGCTGGACACACTGTATAGCCGACCGACGAGGATTGAAGGATGCGTCTGGCTCACT LMG17935 ACAGCTGGACACACTGTATAGCCGACCGACGAGGATTGAAGGATGCGTCTGGCTCACT Zutra3-1 ACAGCTGGACACACTGTATAGCCGACCGACGAGGATTGAAGGATGCATCTGGCTCACT MAFF301001 ACAGCTGGACACACTGTATAGCCGACCGACGAGGATTGAAGGATGCATCTGGCTCACT *********************************
LMG17935 ACAGCTGGACACACTGTATAGCCGACCGCAAGGAATTGAAGGATGCGTCTGGCTCAGCT S5 ACAGCTGGACACACTGTATAGCCGACCGCACGGAGGAATTGAAGGATGCGTCTGGCTCACT ZULFa3-1 ACAGCTGGACACACTGTATAGCCGACCGCACGAGGATTGAAGGATGCGTCTGGCTCACT NAFF301001 ACAGCTGGACACACTGTATAGCCGACCGCACGAGGATTGAAGGATGCGTCTGGCTCACT B. Image: State of the
355 ACAGCTGGACACACTGTATAGCCGACCGACGGATTGAAGGATGCGTCTGGCTCACT zutra-1 NAGCTGGACACACTGTATAGCCGACCGACGGACTGAGGATTGAAGGATGCGTCTGGCTCACT NAFF301001 ACAGCTGGACACACTGTATAGCCGACCGACGACGACGACGACTGGATGCGTCTGGCTCACT B. Image: State of the
B.
в.
14
L. L
S. Martin Commence
All the second

Figure 2 : Conservation of RNA1051 and secondary structure.

A. Among a collection of 21 Ti/Ri plasmid sequences, the 10 harboring nucleotide sequences similar to *rna1051* gene were aligned using Muscle (http://www.ebi.ac.uk/Tools/msa/muscle). Among those are sequences from strains CFBP5473,

Figure 1: Genetic organization of rna1051 locus on Ti plasmid and display of sRNA-seq data.

LMG294, C58, Chbika, CFPB5480, CFPB5476, LMG17935, S56, Zutra3-1 and MAFF301001. Multiple alignment result has been used to predict the secondary structure of RNA1051 with more accuracy. B. The most probable secondary structure of RNA1051 was predicted using RNAalifold. Colors of bases correspond to the probability of base pairing (violet: probability=0; red: probability=1). For unpaired regions, colors represent the probability of being unpaired.



Figure 3: growth kinetics of strains modified for the production of RNA1051

Bacteria were grown in rich medium and kinetics of growth were followed by the measurement of the OD at 600 nm. Strains analysed were the wild type with or without Ti plasmid (C58 and AT009, respectively), the strain deleted for *rna1051* (C58 Δ *rna1051*), and the reverting strain deleted for *rna1051* (AT194). Results obtained with the reverting strain AT194 were similar to those obtained with the other reverting strains AT236, AT302 and AT303.



Figure 4: Putative targets of RNA1051

A. Predicted regions of interaction between the mRNAs targets and RNA1051. The predictions were performed with three algorithms, RNAPredator, sTarPicker and IntaRNA. Sequences around the start codon (black boxes) of predicted target genes are presented with the localization of the regions of interaction with RNA1051. Putative RBS are framed with dotted

lines. B. Expression levels measurement of predicted targets by qRT-PCR in *A. fabrum* C58 strains. Results presented correspond to 3 independent experiments obtained with the reverting strain AT194. Identical results were found with the other reverting strains.

Tables

Table 1 : Strains and plasmids used in this study

Strains or plasmids	Genotypes	References		
Strains				
A. fabrum C58	Wild type strain from the G8 genomic species of A. tumefaciens	Wood <i>et al,</i> 2001,		
		Goodner <i>et al,</i> 2001		
AT009	A. fabrum C58 cured of Ti plasmid	This study		
AT200	A. fabrum C58 containing pBBR1MCS-5	This study		
AT197	A. fabrum C58 containing pARA004	This study		
AT194	A. fabrum C58 deleted for rna1051 gene with reverting phenotype	This study		
AT236	A. fabrum C58 deleted for rna1051 gene with reverting phenotype	This study		
AT302	A. fabrum C58 deleted for rna1051 gene with reverting phenotype	This study		
AT303	A. fabrum C58 deleted for rna1051 gene with reverting phenotype	This study		
E.coli DH5 $lpha$	E. coli strain (using for DNA cloning)	Stratagene		
Plasmids		Kovach <i>et al,</i> 1994		
pBBR1MCS-5	Host range cloning vector, Gm [®]	Wang et al, 1991		
pWSK29	Host range cloning vector, Amp ^R	This study		
pJQ200SK	Suicide vector used for gene replacement, Gm ^R	ATCC®77483 [™]		
pARA004	<i>rna1051</i> gene cloned into the pBBR1MCS-5 <i>Sal</i> I and <i>BamH</i> I sites, Gm ^R	This study		
pARA011	pJQ200SK::∆rna1051, Neo [®] Kan [®] Gm [®]			

Table 2 : Primers used in this study

Names	Forward (5'-3')	Reverse (5'-3')
Cloning primers		
184-185 (rna1051)	ACGCGTCGACGTGCCGCCTTCATAACATTT	CGCGGATCCAATGGCTGTCTCGAACTCGT
171-171(<i>rna1051</i> inactivation)	CACACTAGTCAACCGTTCAGCCACATAAG	GAAGCAGCTCCAGCCTACACATCGTCGCCCTCAGTTTCTAT
173-174 (rna1051 inactivation)	TAAGGAGGATATTCATATGGTTGGATACGGGTTAGCCAAG	CACCTCGAGCATGCGACGACGTGGATAC
95-96 (nptII)	TGTGTAGGCTGGAGCTGCTTC	CATATGAATATCCTCCTTAG
qPCR primers		
A123-A124 (rna1051)	CTGCAATAGACCCGCTTCA	TAGCCTCCAGTGAGCCAGA
A167-A168 (atu0231)	GCGAACAGAATGCGGTAGAT	AAGATCGGGAAGTTCTGCTG
A191-A192 (atu1924)	CGACCTTCAACAACACGATG	CGTGTTCCTGAGCCTTCTTC
A265-A266 (noxB)	TTGGATACGGGTTAGCCAAG	AGATCAGGCCGAAATTCGTT
A273-A274 (traB)	GACCTCGAACAGGGTCAAAC	CTTTCGGGAAGAACGATGAC
A275-A276 (virB8)	GCGCCAAAATTATCAGCAAT	ACCCGGAGTAACATCATTCG
A279-A280 (hyuA)	AGGAGAGAGACAGGATCGTCA	GCCGTTCCAGTAGTTTTTCG
RACE experiments:		
A260 (rna1051)	GTCTTGAACAGATGATCACC	
A261-A262	<u>CATAAGCTT</u> CCATGTGAAGCGGGTCTATT	CGCGGATCCTCGAAGCAAAGAAGGTCCTC
M13fwd-M13rev	GTAAAACGACGGCCAG	CAGGAAACAGCTATGACC

Control of DNA absence for qPCR

A58-A59

)

TTCAACGATCAATTCGTCCA

Underlined sequences correspond to enzyme restriction sites

Table 3 : Modifications in genomes of the growth+ spontaneous reverting strains

TTCCAGTCTGACCGAGGTTT

	Insertion and SNP	Gene	Position	Replicon
AT194	30 kb duplication	<i>atu0435</i> to <i>atu0471</i>	430,476 - 460,950	Сс
AT236	T -> C (S257G)	noxB	30,325	рТі
	A -> C (Y251C)	afuA2 (atu4783)	1,954,816	Lc
	30 kb duplication	<i>atu0435</i> to <i>atu0471</i>	430,476 - 460,950	Сс
AT302	30 kb duplication	<i>atu0435</i> to <i>atu0471</i>	430,476 - 460,950	Сс
AT303	C -> T	IGR	31,512	рТі
	G -> C (R66G)	citE (atu2788)	2,790,917	Сс

The duplication corresponds to a region of 30 kb of the circular chromosome (nucleotides 430,476 to 460,950). SNP are compared between reverting strains and mutant strains *C58*Δ*rna*1051.

Supplementary tables

Table S1: Comparison of the SNP in the wild type strain and in the reverting strains

Replicon	gene	Position	Reference	WT	AT194	AT236	AT302	AT303	pJQ200SK::∆rna1051
Circular chromosome	atu0653	647236	С	G	G	G	G	G	G
Circular chromosome	IGR (atu8026-atu2200)	2173701	т	С	С	С	С	С	С
Circular chromosome	IGR (atu2518-atu2519)	2488221	G	А	А	А	А	А	A
Circular chromosome	atu2525	2498774	С	т	т	т	т	т	т
Circular chromosome	atu2730 (chvB)	2722952	G	С	С	С	С	С	С
Circular chromosome	atu2788 (citE)	2790917	G	G	G	G	G	С	G
Circular chromosome	atu2824	2828943	G	А	А	А	А	А	A
Linear chromosome	atu4160	1279388	А	С	С	С	С	С	C
Linear chromosome	atu4163	1281476	А	G	G	G	G	G	G
Linear chromosome	atu4783 (afuA2)	1954816	А	А	А	С	А	А	А
Plasmid At	atu5267	266698	Т	G	G	G	G	G	G
Plasmid Ti	atu6021 (noxB)	30325	С	С	т	С	т	т	т
Plasmid Ti	atu6021 (noxB)	30974	А	А	G	G	G	G	G
Plasmid Ti	rna1052	31512	Т	Т	С	С	С	т	С

En conclusion, de part son rôle dans la croissance d'*A. fabrum* C58 en présence du plasmide Ti, RNA0151 pourrait être l'antitoxine d'un nouveau système toxine antitoxine du plasmide Ti. La toxine associée pourrait être la protéine NoxB. La récupération d'une croissance similaire à celle de la souche sauvage après 48 heures pourrait résulter en l'inactivation de la toxine potentielle (NoxB), ou encore de la mutation ou de la duplication de la cible de la toxine, contrecarrant l'effet de cette dernière.

Discussion et perspectives

VI. Discussion générale et perspectives

Chez A. tumefaciens, plusieurs mécanismes de régulation transcriptionnelle impliqués dans la mise en place de la virulence ont été caractérisés, tels que la régulation par le système à deux composants VirA/VirG (Rogowsky et al., 1987; Stachel and Zambryski, 1986; Winans et al., 1994). Notre étude a permis de montrer que des mécanismes de régulation post-transcriptionnelle, impliquant des riborégulateurs, prenaient également part au contrôle direct ou indirect de la virulence de la souche de référence A. fabrum C58. Nous avons pour cela mis au point une approche originale. Dans un premier temps, nous avons identifié l'ensemble des riborégulateurs candidats exprimés chez C58, en utilisant des prédictions bioinformatiques et en les validant expérimentalement par séquençage des transcrits de petite taille. Nous nous sommes ensuite focalisés sur les transcrits exprimés à partir du plasmide Ti. Après avoir caractérisé leurs tailles et leurs extrémités 5' et 3' par RACE-PCR, nous avons entrepris de caractériser leur rôle physiologique. Des souches modifiées dans la production de nos candidats ont été construites et testées vis-à-vis de plusieurs phénotypes tels que la croissance et la virulence. Enfin, les cibles potentielles de ces riborégulateurs ont été prédites, et l'expression de certaines d'entre elles a été quantifiée par RT-PCR quantitative (RT-qPCR) dans les différents contextes génétiques.

Ainsi, nous avons pu étudier 6 candidats exprimés à partir du plasmide Ti. Certains candidats réguleraient des cibles relatives à la virulence ou aux étapes précoces de l'interaction bactérie-hôte (RNA1111, RNA1083, RNA1044), tandis que d'autres assureraient le maintien du plasmide Ti au travers de différents mécanismes : i) le contrôle de sa réplication (RepE et RNA1059), ii) le contrôle de sa dissémination par conjugaison (RNA1083 et RNA1111), iii) ou de sa persistance dans les bactéries à l'aide d'un système toxine-antitoxine (RNA1051).

VI.1 Mise au point des approches

VI.1.1 Séquençage des transcrits de petite taille

Une approche de RNA-seq par la méthode Illumina nous a permis d'identifier les riborégulateurs candidats chez la souche C58. Nous avons entrepris de sélectionner les transcrits dont la taille est comprise entre 18 à 576 nucléotides. Les ARN abondants, tels que les ARNr et les ARNt, ont été éliminés. Ces traitements ont permis d'enrichir nos extraits en transcrits de petite taille, et ainsi

d'augmenter la probabilité d'identification des riborégulateurs. De plus, l'élimination des ARN abondants, selon un protocole original utilisant des oligonucléotides spécialement dessinés complémentaires aux séquences des ARNr et ARNt d'A. fabrum C58, a permis d'obtenir une meilleure élimination des ARN abondants que l'élimination à l'aide de kit (MICROB*Express*[™] kit Ambion, USA). Nous estimons que notre méthode a permis d'éliminer 85% des ARN abondants, contre 55% avec le kit utilisé par Lee et al. Notre analyse a ainsi permis la détection de transcrits très faiblement exprimés, et notamment ceux du plasmide Ti. De plus, la réalisation d'un séquençage brin spécifique nous a permis d'identifier les riborégulateurs antisens. Cependant, contrairement aux précédentes analyses chez Agrobacterium (Lee et al., 2013; Wilms et al., 2012b), nous n'avons pas traité nos échantillons de manière à éliminer les ARN dégradés en 5' (à l'aide d'une exonucléase spécifique des extrémités 5' mono-phosphates) et à ne conserver que les ARN intègres (5' triphosphates). Il est donc possible qu'un certain nombre des candidats identifiés corresponde en réalité à des produits de dégradation. Par ailleurs, le faible taux d'expression des candidats identifiés et localisés sur le plasmide Ti pourrait provenir des conditions de croissance utilisées. Des cultures dans des conditions d'induction de la virulence, en présence d'acétosyringone par exemple, pourraient être envisagées.

Pour compléter l'analyse expérimentale, nous avons utilisé une approche bioinformatique à l'aide du logiciel sRNApredict basé sur la comparaison de séquences et la recherche de terminateur Rhoindépendants. L'utilisation d'une quinzaine de séquences de la famille des *Rhizobiaceae* – la majorité provenant de séquences d'*A. tumefaciens* - a permis une prédiction sensiblement plus fiable des riborégulateurs.

VI.1.2 Caractérisation phénotypique des riborégulateurs candidats

L'identification du rôle des riborégulateurs est facilitée par l'identification de phénotypes associés à la modulation de leur production. Cependant, les modulations générées par les riborégulateurs sont quelques fois modérées. L'utilisation de souches modifiées dans la production des candidats sélectionnés nous a permis d'étudier leur impact sur la physiologie d'*Agrobacterium*, notamment sur la croissance et la virulence, les riborégulateurs étudiés étant exprimés à partir du plasmide Ti.

Nos tests de virulence ont été réalisés sur des plants de tomates. Plusieurs améliorations ont été apportées au protocole original afin d'augmenter la reproductibilité (Planamente et al., 2012). Nous avons effectué nos expériences dans un phytotron, à l'intérieur duquel les conditions sont très contrôlées (contrôle des périodes diurnes, de la température et de l'humidité). Les inoculations étaient au préalable réalisées au niveau de blessures de quelques centimètres le long de la tige de pieds de tomate âgés de trois semaines. Cependant, la taille et la profondeur des incisions peuvent introduire de la variabilité, de même que le stade de développement des pieds de tomate. Nous avons donc réalisé nos tests d'infection sur des plants de tomate au stade d'apparition des deuxièmes feuilles. Les bactéries ont été déposées au sommet de la tige au-dessous des deuxièmes feuilles. Ceci a permis une homogénéité de la taille de la surface d'inoculation. Ces modifications nous ont permis d'obtenir des résultats répétables et reproductibles, et de mettre en évidence des modulations - même modérées - de l'agressivité des souches, comme pour la surexpression de RNA1083. Ainsi, nous envisageons de refaire des tests de virulence dans ces conditions pour mieux caractériser RNA1044 et d'autres candidats.

VI.1.3 Identification des cibles ARNm des riborégulateurs candidats

L'une des difficultés majeures de l'étude des riborégulateurs est l'identification de leurs cibles. Ici, nous avons développé une stratégie originale et ainsi utilisé trois logiciels capables de prédire les zones d'interaction entre un riborégulateur et l'ensemble d'un génome: IntaRNA, sTarPicker et RNApredator. Cependant, les algorithmes ne sont pas encore suffisamment performants. Nos connaissances sur les riborégulateurs et leurs mécanismes d'action étant encore limitées, peu de caractéristiques peuvent être prises en compte. En utilisant ces 3 logiciels basés sur des algorithmes différents et en ne sélectionnant que les cibles prédites en commun, nous supposons avoir limité le nombre de cibles candidates et les faux positifs. Avec l'hypothèse que l'interaction d'un riborégulateur impacte la stabilité de son ARNm cible, nous avons quantifié par RT-qPCR les cibles prédites en comparant les souches modifiées dans la production des riborégulateurs et la souche sauvage. Cependant, cette approche n'est valable que si le riborégulateur module la stabilité de sa cible, ce qui n'est pas nécessairement le cas. Il peut en effet ne moduler que la traduction de l'ARNm. Dans ce cas, des analyses protéomiques pourraient être envisagées afin d'observer l'impact de la modulation des quantités d'un riborégulateur sur la quantité de protéines cibles.

VI.2 Rôles physiologiques des riborégulateurs candidats

L'analyse phénotypique de souches modifiées dans la production de nos candidats, ainsi que la recherche de leurs ARNm cibles, ont permis de mettre en évidence l'impact des riborégulateurs sur la virulence et sur le maintien du plasmide Ti.

VI.2.1 Régulation de la virulence de la bactérie

Plusieurs candidats sembleraient impliqués dans la virulence d'*A. fabrum* C58. L'absence de production de RNA1111 et la surproduction de RNA1083 induisent en effet une diminution de l'agressivité de la bactérie, tandis que la surproduction de RNA1044 stimulerait la virulence de C58. Par ailleurs, la surproduction de RNA1059 et de RepE conduit à la perte de la virulence de la bactérie suite à la perte du plasmide Ti.

La prédiction des ARNm cibles de RNA1083 et RNA1111 semble montrer que ces deux riborégulateurs agissent à différentes étapes du cycle infectieux. RNA1111 cible en effet des facteurs impliqués dans le transfert de l'ADN-T (VirC2 et VirD3) et dans la formation de tumeurs chez la plante (6b). RNA1083 semble agir au niveau de la perception de la blessure de la plante par la bactérie (ChvI) et au niveau de la mobilité permettant la migration jusqu'à la zone de blessure (FlgF, FlgI, FliE et FliF). Pour le moment, seule la validation des cibles de RNA1111 par RT-qPCR a été entreprise. Nos résultats ont permis de montrer un impact de cet ARNnc sur les ARNm 6b et atu6072, l'absence de RNA1111 conduisant à une diminution de la quantité de ces ARNm cibles. Nous envisageons de confirmer les interactions entre RNA1111 et ses cibles par des approches in vivo, en clonant le gène du riborégulateur en aval d'un promoteur inductible, et le promoteur et la région d'interaction du gène cible en amont d'un gène rapporteur. L'expression du gène rapporteur est ensuite suivie selon l'expression du riborégulateur (Urban and Vogel, 2007). L'absence de modification de la quantité des autres ARNm cibles, virC2 et virD3, n'exclut pas la possibilité que RNA1111 interagisse tout de même avec eux. De plus, ces deux cibles de RNA1111 semblent interagir avec la protéine chaperon Hfq (Möller et al., 2014b). Ceci va dans le sens d'une régulation de ces ARNm par des riborégulateurs, et potentiellement par RNA1111. Enfin, les cibles prédites pour RNA1111 ont été montrées comme non indispensables pour la virulence de C58. La diminution de la virulence résultant de la délétion de rna1111 semble ne pas pouvoir s'expliquer simplement par l'action sur ces cibles, hormis par un effet cumulatif. La mesure de l'agressivité de souches déficientes pour rna1111 et complémentées par une ou plusieurs cibles permettrait de vérifier cette hypothèse. D'autre part, RNA1111 semble cibler l'ARNm de virD3, dont le gène est en opéron avec le gène virD4 codant une protéine essentielle à l'export de l'ADN-T. On pourrait imaginer que l'interaction de RNA1111 avec virD3 perturbe également la traduction et/ou la stabilité de virD4. Une autre possibilité serait que RNA1111 interagisse avec d'autres cibles, non identifiées par notre étude, et notamment avec des protéines. Seules des analyses de piégage permettraient d'identifier les protéines cibles (voir page 61). Par ailleurs, la délétion de rna1111 a entrainé un effet polaire sur l'expression du gène virEO, qui n'explique cependant pas à lui seul la perte de la virulence. Pour valider le rôle de RNA1111 dans la virulence, un mutant délété uniquement pour *rna1111* pourra être reconstruit.

Nos résultats suggèrent également un impact potentiel de RNA1044 sur la virulence. RNA1044 étant un ARNnc antisens, on peut facilement envisager qu'il cible le produit du gène issu du brin complémentaire, l'ARNm tms2. Ce simple mécanisme d'action ne concorde cependant pas avec l'impact de la surexpression de RNA1044 sur la virulence. L'expression du gène tms2, codant une indole-3-acétamide hydrolase permettant la formation de phytohormones, n'est nécessaire que dans les cellules végétales transformées. On imagine que l'interaction entre RNA1044 et l'ARNm tms2 dans la bactérie pourrait permettre le blocage de la production de la protéine Tms2 au sein de celleci. Il serait aussi possible que RNA1044 interagisse avec d'autres ARNm ou même directement avec des protéines, l'action en trans de riborégulateur antisens ayant déjà été suggérée. En effet, chez Mycobacterium tuberculosis et chez L. monocytogenes, les riborégulateurs ASdes et ASpks d'une part, et RliE, d'autre part, ont été prédits comme pouvant interagir non seulement avec l'ARNm produit à partir de la séquence codante complémentaire, mais aussi avec des ARNm produits à partir d'autres régions génomiques (Arnvig and Young, 2009; Mandin et al., 2007). L'interaction physique entre RliE et ces ARNm cibles a été montrée. La recherche d'autres cibles par prédiction bioinformatique permettrait de tester cette hypothèse. Il est également envisageable que ce candidat antisens, ainsi que les autres riborégulateurs de l'ADN-T, soient produits dans les cellules végétales transformées. A l'intérieur de celles-ci, ils pourraient faciliter ou empêcher la production de leurs cibles, afin de ne permettre leurs expressions qu'au moment opportun. L'absence d'impact sur la virulence de la surexpression ou de la délétion de *rna1046* dans la bactérie semble aller dans ce sens.

Nous avons également montré que l'ARNm 6b était prédit comme cible potentielle de RNA1111, et que sa quantité était diminuée dans des souches déficientes pour le riborégulateur. Le gène 6b étant localisé sur l'ADN-T, on peut se demander si RNA1111 pourrait également s'apparier avec lui dans les cellules végétales transformées. Pour cela, RNA1111 devrait être transféré dans la plante. Aucun système de sécrétion n'a pour l'instant été montré capable de transférer des ARN. Cependant, une étude récente a montré que le champignon pathogène *Botrytis cinerea* était capable d'injecter des micro-ARN directement dans des cellules végétales, bien que les mécanismes de transfert n'aient pas encore été mis en évidence (Weiberg et al., 2013). Weiberg *et al* ont également montré que les riborégulateurs de *B. cinerea* agissaient comme des effecteurs dans la cellule végétale, en interagissant avec la protéine AGO, qui permet l'interaction entre les riborégulateurs et leurs cibles. L'interaction des riborégulateurs de *B. cinerea* avec cette protéine chaperon redirige son action et

permet l'inhibition de l'expression des gènes de l'immunité de la plante. Bien que les riborégulateurs bactériens et eucaryotes soient différents, on peut imaginer une action similaire des riborégulateurs issus de l'ADN-T et potentiellement de RNA1111. Ceci permettrait de contrecarrer les défenses de la plante et d'accroître le pouvoir pathogène d'*A. fabrum* C58. Cette hypothèse est d'autant plus intéressante que la protéine 6b est connue pour interférer dans la voie des micro-ARN chez la plante (Wang et al., 2011).

L'équipe d'Olivier Voinnet a co-transformé un ADN-T dépourvu de ses gènes de virulence et exprimant à la place le gène codant la GFP et des gènes viraux permettant l'extinction des gènes de défense de la plante. Ils ont ainsi montré une production plus importante de la fluorescence par comparaison avec la transformation par l'ADN-T modifié sans les gènes viraux (Gelvin, 2003). Il semble donc que la présence d'inhibiteurs viraux des défenses de la plante permette une meilleure expression des gènes issus de l'ADN-T. Ainsi, on pourrait supposer que l'absence des gènes des riborégulateurs de l'ADN-T, comme c'est le cas pour les souches utilisées pour les OGM, conduirait à une efficacité de transformation et d'insertion de l'ADN-T plus faible qu'en leur présence. Il serait intéressant de rechercher la présence ou l'expression des riborégulateurs exprimés à partir de l'ADN-T dans les cellules végétales transformées, mais également d'étudier l'impact de leur surexpression ou de l'absence de leur production dans la plante, afin de vérifier ces hypothèses.

VI.2.2 Rôle dans le maintien du plasmide Ti

Le plasmide Ti étant le déterminant majeur de la pathogénie d'*Agrobacterium*, sa présence au sein de la bactérie est primordiale pour assurer l'infection. Les caractérisations phénotypiques de plusieurs de nos candidats exprimés à partir du plasmide Ti et sélectionnés sans *a priori*, a montré de façon surprenante qu'ils étaient impliqués dans le maintien du plasmide Ti dans les cellules.

VI.2.2.1 Contrôle de la réplication

Les riborégulateurs RepE et RNA1059 semblent impliqués dans la réplication du plasmide Ti : leurs expressions en *trans* sur un plasmide d'expression induisent une perte de ce réplicon. Les mécanismes par lesquels ils agissent sont encore mal compris, cependant leurs présences mutuelles semblent essentielles. On peut imaginer qu'ils agiraient conjointement, ou en opposition, pour assurer une réplication optimale du plasmide Ti. Il est également possible que la séquence d'ADN de cette région intergénique soit responsable du phénotype d'incompatibilité. Des mutations dans la séquence, ne perturbant pas l'expression de *rna1059* et de *repE* pourraient permettre de mieux comprendre les mécanismes mis en jeu.

Il serait également intéressant d'étendre nos analyses aux régions intergéniques *repB-repC* du plasmide At et du chromosome linéaire, afin de mieux comprendre comment plusieurs réplicons *repABC* peuvent cohabiter dans une même cellule. Des séquences similaires à *repE* et à *rna1059* sont retrouvées sur ces deux réplicons (RNA770 et RNA771 pour le chromosome linéaire, et RNA990 et RNA991 pour le plasmide At). Cependant, nos analyses de RNA-seq ont montré des profils d'expression de ces gènes différents. Ainsi, pour le chromosome linéaire, les deux riborégulateurs sont identifiés à des taux d'expression similaires, tandis que pour le plasmide At, RNA990 est plus abondant que RNA991. De plus, nous avons montré que les expressions en *trans* de *repE* et *rna1059* n'impactaient pas les quantités du chromosome linéaire ni du plasmide At. Ces résultats préliminaires suggèrent donc des mécanismes de contrôle de la réplication différents entre le chromosome linéaire, le plasmide At et le plasmide Ti, et que d'autres éléments pourraient intervenir.

VI.2.2.2 Rôle potentiel dans un système toxine-antitoxine

Nous avons par ailleurs observé que RNA1051 est indispensable à la croissance de C58 en présence du plasmide Ti, et serait donc l'antitoxine d'un nouveau système toxine-antitoxine de ce réplicon. L'ensemble de nos résultats suggère que la toxine associée serait la protéine NoxB. Afin de vérifier cette hypothèse, nous envisageons de surexprimer *noxB* et d'observer l'impact sur la croissance de la bactérie. Des analyses *in vivo* et *in vitro* seront également entreprises pour valider l'interaction de RNA1051 avec l'ARNm de *noxB*. Le rôle de NoxB comme toxine n'est pas compris. On pourrait supposer que son activité enzymatique, en absence de nopaline, conduirait à la production de composés toxiques, ou que sa surproduction entraînerait sa translocation dans la membrane et la destabilisation de celle-ci. Un autre fait étonnant est que l'effet toxique de la délétion de *rna1051* ne semble être apparu qu'en milieu liquide. Cette observation pourrait nous renseigner quant à la nature de la toxine, et quant au fonctionnement de ce nouveau système TA.

Concernant l'instabilité du phénotype, nous proposons deux hypothèses non exclusives. La première se baserait sur la perte de la toxine fonctionnelle, et son remplacement par une toxine tronquée ou mutée non fonctionnelle. Nous avons en effet observé que la région délétée pour l'inactivation de *rna1051* contenait également l'extrémité 5' de *noxB*. Ainsi, suite à la perte de RNA1051, les protéines natives NoxB seraient produites à partir de l'ARNm de *noxB* libéré. Après une dizaine de division cellulaire, la protéine native serait remplacée par une protéine non produite ou tronquée, incapable de jouer son rôle toxique. La seconde hypothèse réside en la duplication d'une région de 30 kb du chromosome circulaire. On pourrait imaginer que l'un des gènes portés par cette région coderait la cible de la toxine, et que la duplication de ce gène entraînerait une augmentation de la protéine cible

dans la cellule, contrecarrant l'effet de la toxine. Il est également possible que cette région code pour un système d'anti-addiction, réduisant l'impact du système toxine-antitoxine, comme cela a été montré chez *Dickeya dadantii* (Saavedra De Bast et al., 2008). Cette hypothèse n'est cependant pas valable pour le révertant AT303 chez qui nous n'avons pas observé la duplication.

La perte de l'antitoxine dans une cellule, conduisant à la mort ou au ralentissement de croissance de celle-ci, s'effectue généralement lorsque le plasmide portant le système TA est perdu suite à la division cellulaire notamment. Cependant, nous n'avons pas observé de ralentissement de croissance chez les souches ayant perdu le plasmide Ti suite à la surexpression de *rna1059* et de *repE*. Ce résultat est surprenant. L'existence d'un dialogue entre les deux systèmes pourrait être envisagé. Une première possibilité serait que RepE et/ou RNA1059 inhibent l'effet toxique, dans la cellule, de la toxine malgré la perte de l'antitoxine suite à la disparition du plasmide Ti. Une autre hypothèse serait que RNA1059 et/ou RepE stabiliserai(en)t RNA1051 et le rendrai(en)t moins labile. Il pourrait alors perdurer dans la cellule même après l'arrêt de sa production, et contrecarrer l'action de la toxine jusqu'à sa disparition totale suite à la perte du plasmide Ti. Ces deux hypothèses pourraient facilement être vérifiées en détectant *in vitro* et *in vivo* l'interaction de RNA1059 ou de RepE avec RNA1051 et avec la toxine (ou son ARNm). A partir des résultats observés et en se basant sur cette hypothèse, la toxine associée à RNA1051 pourrait par ailleurs être identifiée en approfondissant nos recherches sur les cibles de RNA1059 et RepE.

VI.2.2.3 Rôle dans la conjugaison

Le transfert conjugatif du plasmide Ti permet sa dissémination dans les populations d'*Agrobacterium*. La prédiction des cibles de RNA1111 et RNA1083 suggère que ces candidats sont impliqués dans ce processus. En effet, RNA1111 ciblerait l'ARNm *traA*, dont le produit de traduction permet le clivage du plasmide Ti au niveau de l'origine de transfert, tandis que RNA1083 interagirait avec les ARNm *trbG*, *trbK* et *trbL* codant des protéines du pore de conjugaison. L'impact de RNA1083 sur la stabilité de ses cibles n'a pas encore été testé, mais des analyses de RT-qPCR sont prévues, ainsi que la validation de l'interaction *in vivo* par fusion de gènes. De plus des expériences *in planta* sont en cours pour déterminer si la surexpression ou l'absence de production de *rna1083* a un impact sur le taux de transfert conjugatif du plasmide Ti.

D'autre part, puisque RNA1111 et RNA1083 paraissent tous les deux être impliqués dans la modulation du transfert conjugatif du plasmide Ti, il serait intéressant de tester s'ils exercent une régulation l'un sur l'autre. Des quantifications par RT-qPCR et des analyses par fusion de gènes pourraient être envisagées.

167

VI.2.2.4 Coordination de l'expression des composants du plasmide Ti

Le plasmide Ti est un plasmide de grande taille (214 kb) qui code des fonctions cellulaires coûteuses en énergie (notamment 2 systèmes de sécrétion de type IV). Il est important de contrôler l'expression des gènes qu'il porte afin de réduire sa charge pour la bactérie. Plusieurs observations faites au cours de nos analyses suggèrent que des riborégulateurs pourraient permettre la réduction du coût du plasmide Ti pour *A. fabrum* C58.

RNA1111 et RNA1083 semblent réguler la mise en place de plusieurs étapes du cycle infectieux. La régulation de plusieurs étapes de la virulence par des riborégulateurs a déjà été mise en évidence, notamment, chez *Erwinia amylovora* (Zeng and Sundin, 2014b). Chez cette bactérie, les riborégulateurs OmrAB et Hsr6 stimulent directement la mobilité de la bactérie au cours des premières étapes de la virulence, et indirectement la formation de biofilms dans les phases plus tardives. Cette capacité des riborégulateurs de réguler plusieurs cibles est particulièrement adaptée pour l'interaction des bactéries avec leurs hôtes. Un seul régulateur est en effet capable de moduler de façon positive ou négative plusieurs facteurs. La synthèse ou la dégradation des riborégulateurs est rapide comparée à un régulateur protéique, et l'activation ou l'arrêt de plusieurs fonctions cellulaires peut être coordonnée.

RNA1111 semblerait réguler à la fois le transfert de l'ADN-T (au travers de VirC2 et VirD3) et celui du plasmide Ti (TraA). La mise en place de ces deux systèmes de sécrétion est régulée par TraM. On peut imaginer que RNA1111 représenterait un degré supplémentaire dans la coordination du transfert de l'ADN-T et du plasmide Ti, le contrôle de la mise en place de ces deux systèmes à des étapes distinctes étant primordial pour ne pas induire un important encombrement stérique de la membrane.

Par ailleurs, notre analyse a révélé une prédominance des riborégulateurs antisens sur les plasmides de la souche C58, et principalement sur le plasmide Ti, puisque cette catégorie de riborégulateurs représente 63 % des candidats de ce réplicon contre seulement 15 à 27 % sur les chromosomes circulaire et linéaire. Cette forte représentation des riborégulateurs antisens sur les réplicons accessoires a déjà été observée dans d'autres génomes bactériens, tels que *S. meliloti* (Schlüter et al., 2010), chez qui les riborégulateurs 5'UTR et IGR sont prévalents sur le chromosome tandis que les riborégulateurs antisens dominent sur les mégaplasmides, et principalement sur le mégaplasmide pSymB (34 %). Chez ces bactéries, et comme beaucoup d'autres bactéries en interaction symbiotique ou pathogène avec des plantes, les plasmides portent des gènes codant des fonctions nécessaires à l'interaction, mais qui ne sont pas requises en absence des hôtes. La régulation par les riborégulateurs antisens pourrait permettre d'activer rapidement l'expression des gènes portés par

ces plasmides au contact de l'hôte, ou à l'inverse d'éteindre leur expression lorsque ceux-ci ne sont plus nécessaires. Ce mécanisme permettrait : i) à la bactérie de s'acclimater rapidement à son hôte, et ii) de diminuer le coût que représentent les plasmides pour la cellule bactérienne en réprimant la production de leurs composants sans nécessiter le recrutement de la machinerie de traduction pour la production de facteur de transcription. De plus, les gènes portés par les plasmides sont souvent issus de transferts horizontaux, et de par leur GC% différent du core-génome, ils pourraient être surexprimés par rapport au reste du génome. Des mécanismes d'atténuation de l'expression de ces gènes issus de transferts horizontaux et à faible GC% ont déjà été mis en évidence, notamment via leur répression par la protéine H-NS (Navarre et al., 2006). H-NS est absente chez Agrobacterium. Nous pouvons émettre l'hypothèse que les riborégulateurs antisens permettraient d'éteindre l'expression des gènes issus de transferts horizontaux quand leur expression n'est pas nécessaire. Des riborégulateurs antisens (RNA1041-RNA1042, RNA1044-RNA1046) ont été identifiés sur le brin complémentaire de gènes de l'ADN-T, qui possèdent un GC % de 47 % (contre 58 % en moyenne pour le reste du génome) (Wood et al., 2001). Ces candidats pourraient permettre l'extinction de l'expression des gènes de l'ADN-T au sein de la bactérie, leur action n'étant nécessaire que dans la cellule végétale.

VI.3 Apport de notre étude dans la compréhension des mécanismes de virulence chez Agrobacterium

La démarche que nous avons developpée, depuis l'identification d'un riborégulateur à partir d'une analyse globale, jusqu'à sa caractérisation fonctionnelle, a permis de mieux comprendre le rôle d'au moins un candidat (RNA1111). Cette méthode efficace peut être appliquée à d'autres candidats chez *Agrobacterium*, mais également chez d'autres bactéries.

Nous avons montré que RNA1111 était impliqué dans la virulence d'*A. fabrum* C58. Plusieurs études précédentes avaient suggéré le rôle de certains riborégulateurs dans le contrôle de la virulence : i) une différence d'expression de certains riborégulateurs a été mesurée en présence de composés phénoliques, ii) des gènes de riborégulateurs sont localisés sur le brin complémentaire de gènes *vir*, et iii) la co-imunoprécipitation d'ARNm *vir* avec la protéine Hfq a été observée (Lee et al., 2013; Möller et al., 2014; Wilms et al., 2012b). Cependant, notre étude est la 1^{ère} à avoir démontré l'impact de riborégulateurs directement dans la virulence de cette bactérie.

Il semble ainsi qu'en plus des mécanismes de régulation faisant intervenir principalement des systèmes à deux composants, la régulation par les riborégulateurs permettrait une mise en place et un développement optimal de la virulence. La question reste de savoir à quel niveau se positionne ce

nouveau mécanisme par rapport au réseau de régulation. Les riborégulateurs sont-ils régulés par VirA/VirG ? Par le *quorum sensing* ? Où agissent-ils au contraire en amont de ces systèmes? La présence de « boîtes *vir* » ou de « boîtes *tra* » sur certains riborégulateurs, tels que RNA1051, laisse supposer qu'ils pourraient agir sous la dépendance de ces régulateurs transcriptionnels. Nous envisageons de quantifier l'expression de nos candidats chez des mutants *virA/virG*. Il serait aussi intéressant de découvrir si la présence d'extraits de plantes (comme des composés phénoliques) ou de tumeurs (comme les opines) a un impact sur la production de nos candidats. Des analyses de quantification de l'expression de nos candidats dans ces conditions, par fusion de gènes et par RT-qPCR, sont d'ailleurs en cours. Ces études nous permettraient de mieux comprendre le fonctionnement et le positionnement des riborégulateurs dans le réseau de régulation contrôlant de la virulence.

Références bibliographiques

Akama, T., Suzuki, K., Tanigawa, K., Kawashima, A., Wu, H., Nakata, N., Osana, Y., Sakakibara, Y., Ishii, N., 2009. Whole-genome tiling array analysis of Mycobacterium leprae RNA reveals high expression of pseudogenes and noncoding regions. J. Bacteriol. 191, 3321–3327. doi:10.1128/JB.00120-09

Akiyoshi, D.E., Klee, H., Amasino, R.M., Nester, E.W., Gordon, M.P., 1984. T-DNA of Agrobacterium tumefaciens encodes an enzyme of cytokinin biosynthesis. Proc. Natl. Acad. Sci. U. S. A. 81, 5994–5998.

Alkan, C., Karakoç, E., Nadeau, J.H., Sahinalp, S.C., Zhang, K., 2006. RNA-RNA interaction prediction and antisense RNA target search. J. Comput. Biol. J. Comput. Mol. Cell Biol. 13, 267–282. doi:10.1089/cmb.2006.13.267

Alt-Mörbe, J., Stryker, J.L., Fuqua, C., Li, P.L., Farrand, S.K., Winans, S.C., 1996. The conjugal transfer system of Agrobacterium tumefaciens octopine-type Ti plasmids is closely related to the transfer system of an IncP plasmid and distantly related to Ti plasmid vir genes. J. Bacteriol. 178, 4248–4257.

Altuvia, S., Weinstein-Fischer, D., Zhang, A., Postow, L., Storz, G., 1997. A small, stable RNA induced by oxidative stress: role as a pleiotropic regulator and antimutator. Cell 90, 43–53.

An, S.-Q., Febrer, M., McCarthy, Y., Tang, D.-J., Clissold, L., Kaithakottil, G., Swarbreck, D., Tang, J.-L., Rogers, J., Dow, J.M., Ryan, R.P., 2013. High-resolution transcriptional analysis of the regulatory influence of cell-to-cell signalling reveals novel genes that contribute to Xanthomonas phytopathogenesis. Mol. Microbiol. 88, 1058–1069. doi:10.1111/mmi.12229

Arnvig, K.B., Young, D.B., 2009. Identification of small RNAs in *Mycobacterium tuberculosis*. Mol. Microbiol. 73, 397–408. doi:10.1111/j.1365-2958.2009.06777.x

Atmakuri, K., Cascales, E., Burton, O.T., Banta, L.M., Christie, P.J., 2007. Agrobacterium ParA/MinDlike VirC1 spatially coordinates early conjugative DNA transfer reactions. EMBO J. 26, 2540–2551. doi:10.1038/sj.emboj.7601696

Babski, J., Maier, L.-K., Heyer, R., Jaschinski, K., Prasse, D., Jäger, D., Randau, L., Schmitz, R.A., Marchfelder, A., Soppa, J., 2014. Small regulatory RNAs in Archaea. RNA Biol. 11, 484–493. doi:10.4161/rna.28452

Backofen, R., Amman, F., Costa, F., Findeiß, S., Richter, A.S., Stadler, P.F., 2014. Bioinformatics of prokaryotic RNAs. RNA Biol. 11, 470–483. doi:10.4161/rna.28647

Balaban, N., Novick, R.P., 1995. Translation of RNAIII, the Staphylococcus aureus agr regulatory RNA molecule, can be activated by a 3'-end deletion. FEMS Microbiol. Lett. 133, 155–161.

Balsiger, S., Ragaz, C., Baron, C., Narberhaus, F., 2004. Replicon-specific regulation of small heat shock genes in Agrobacterium tumefaciens. J. Bacteriol. 186, 6824–6829. doi:10.1128/JB.186.20.6824-6829.2004

Bandyra, K.J., Said, N., Pfeiffer, V., Górna, M.W., Vogel, J., Luisi, B.F., 2012. The seed region of a small RNA drives the controlled destruction of the target mRNA by the endoribonuclease RNase E. Mol. Cell 47, 943–953. doi:10.1016/j.molcel.2012.07.015

Baron, C., Zambryski, P.C., 1995. The plant response in pathogenesis, symbiosis, and wounding: variations on a common theme? Annu. Rev. Genet. 29, 107–129. doi:10.1146/annurev.ge.29.120195.000543

Barra-Bily, L., Fontenelle, C., Jan, G., Flechard, M., Trautwetter, A., Pandey, S.P., Walker, G.C., Blanco,

C., 2010a. Proteomic alterations explain phenotypic changes in Sinorhizobium meliloti lacking the RNA chaperone Hfq. J. Bacteriol. 192, 1719–1729. doi:10.1128/JB.01429-09

Barra-Bily, L., Pandey, S.P., Trautwetter, A., Blanco, C., Walker, G.C., 2010b. The Sinorhizobium meliloti RNA chaperone Hfq mediates symbiosis of S. meliloti and alfalfa. J. Bacteriol. 192, 1710–1718. doi:10.1128/JB.01427-09

Bélanger, C., Canfield, M.L., Moore, L.W., Dion, P., 1995. Genetic analysis of nonpathogenic Agrobacterium tumefaciens mutants arising in crown gall tumors. J. Bacteriol. 177, 3752–3757.

Bingle, L.E., Thomas, C.M., 2001. Regulatory circuits for plasmid survival. Curr. Opin. Microbiol. 4, 194–200.

Bohn, C., Rigoulay, C., Bouloc, P., 2007. No detectable effect of RNA-binding protein Hfq absence in Staphylococcus aureus. BMC Microbiol. 7, 10. doi:10.1186/1471-2180-7-10

Bouzar, H., Ouadah, D., Krimi, Z., Jones, J.B., Trovato, M., Petit, A., Dessaux, Y., 1993. Correlative Association between Resident Plasmids and the Host Chromosome in a Diverse Agrobacterium Soil Population. Appl. Environ. Microbiol. 59, 1310–1317.

Boysen, A., Moller-Jensen, J., Kallipolitis, B., Valentin-Hansen, P., Overgaard, M., 2010. Translational Regulation of Gene Expression by an Anaerobically Induced Small Non-coding RNA in Escherichia coli. J. Biol. Chem. 285, 10690–10702. doi:10.1074/jbc.M109.089755

Brantl, S., Birch-Hirschfeld, E., Behnke, D., 1993. RepR protein expression on plasmid pIP501 is controlled by an antisense RNA-mediated transcription attenuation mechanism. J. Bacteriol. 175, 4052–4061.

Braun, A.C., 1962. Tumor Inception and Development in the Crown Gall Disease. Annu. Rev. Plant Physiol. 13, 533–558. doi:10.1146/annurev.pp.13.060162.002533

Brencic, A., Eberhard, A., Winans, S.C., 2004. Signal quenching, detoxification and mineralization of vir gene-inducing phenolics by the VirH2 protein of Agrobacterium tumefaciens. Mol. Microbiol. 51, 1103–1115.

Britton, M.T., Escobar, M.A., Dandekar, A.M., 2008. The oncogenes of Agrobacterium tumefaciens and Agrobacterium rhizogenes, in: From Biology to Biothechnology. New York, pp. 523–563.

Burge, S.W., Daub, J., Eberhardt, R., Tate, J., Barquist, L., Nawrocki, E.P., Eddy, S.R., Gardner, P.P., Bateman, A., 2013. Rfam 11.0: 10 years of RNA families. Nucleic Acids Res. 41, D226–D232. doi:10.1093/nar/gks1005

Busch, A., Richter, A.S., Backofen, R., 2008. IntaRNA: efficient prediction of bacterial sRNA targets incorporating target site accessibility and seed regions. Bioinforma. Oxf. Engl. 24, 2849–2856. doi:10.1093/bioinformatics/btn544

Campillo, T., Renoud, S., Kerzaon, I., Vial, L., Baude, J., Gaillard, V., Bellvert, F., Chamignon, C., Comte, G., Nesme, X., Lavire, C., Hommais, F., 2014. Analysis of hydroxycinnamic acid degradation in Agrobacterium fabrum reveals a coenzyme A-dependent, beta-oxidative deacetylation pathway. Appl. Environ. Microbiol. 80, 3341–3349. doi:10.1128/AEM.00475-14

Cangelosi, G.A., Martinetti, G., Leigh, J.A., Lee, C.C., Thienes, C., Theines, C., Nester, E.W., 1989. Role for [corrected] Agrobacterium tumefaciens ChvA protein in export of beta-1,2-glucan. J. Bacteriol. 171, 1609–1615.

Cao, Y., Zhao, Y., Cha, L., Ying, X., Wang, L., Shao, N., Li, W., 2009. sRNATarget: a web server for prediction of bacterial sRNA targets. Bioinformation 3, 364–366.

Carlier, A., Chevrot, R., Dessaux, Y., Faure, D., 2004. The assimilation of gamma-butyrolactone in Agrobacterium tumefaciens C58 interferes with the accumulation of the N-acyl-homoserine lactone signal. Mol. Plant-Microbe Interact. MPMI 17, 951–957. doi:10.1094/MPMI.2004.17.9.951

Carter, R.J., Dubchak, I., Holbrook, S.R., 2001. A computational approach to identify genes for functional RNAs in genomic sequences. Nucleic Acids Res. 29, 3928–3938.

Caswell, C.C., Gaines, J.M., Ciborowski, P., Smith, D., Borchers, C.H., Roux, C.M., Sayood, K., Dunman, P.M., Roop Ii, R.M., 2012. Identification of two small regulatory RNAs linked to virulence in Brucella abortus 2308. Mol. Microbiol. 85, 345–360. doi:10.1111/j.1365-2958.2012.08117.x

Cevallos, M.A., Cervantes-Rivera, R., Gutiérrez-Ríos, R.M., 2008. The repABC plasmid family. Plasmid 60, 19–37. doi:10.1016/j.plasmid.2008.03.001

Chabelskaya, S., Bordeau, V., Felden, B., 2014. Dual RNA regulatory control of a Staphylococcus aureus virulence factor. Nucleic Acids Res. 42, 4847–4858. doi:10.1093/nar/gku119

Chai, Y., Winans, S.C., 2005a. RepB protein of an Agrobacterium tumefaciens Ti plasmid binds to two adjacent sites between repA and repB for plasmid partitioning and autorepression. Mol. Microbiol. 58, 1114–1129. doi:10.1111/j.1365-2958.2005.04886.x

Chai, Y., Winans, S.C., 2005b. A small antisense RNA downregulates expression of an essential replicase protein of an *Agrobacterium tumefaciens* Ti plasmid. Mol. Microbiol. 56, 1574–1585. doi:10.1111/j.1365-2958.2005.04636.x

Chai, Y., Winans, S.C., 2009. The chaperone GroESL enhances the accumulation of soluble, active TraR protein, a quorum-sensing transcription factor from Agrobacterium tumefaciens. J. Bacteriol. 191, 3706–3711. doi:10.1128/JB.01434-08

Chai, Y., Zhu, J., Winans, S.C., 2001. TrIR, a defective TraR-like protein of Agrobacterium tumefaciens, blocks TraR function in vitro by forming inactive TrIR:TraR dimers. Mol. Microbiol. 40, 414–421.

Chen, Y., Indurthi, D.C., Jones, S.W., Papoutsakis, E.T., 2011. Small RNAs in the genus Clostridium. mBio 2, e00340–00310. doi:10.1128/mBio.00340-10

Chesnokova, O., Coutinho, J.B., Khan, I.H., Mikhail, M.S., Kado, C.I., 1997. Characterization of flagella genes of Agrobacterium tumefaciens, and the effect of a bald strain on virulence. Mol. Microbiol. 23, 579–590.

Chevrot, R., Rosen, R., Haudecoeur, E., Cirou, A., Shelp, B.J., Ron, E., Faure, D., 2006a. GABA controls the level of quorum-sensing signal in *Agrobacterium tumefaciens*. Proc. Natl. Acad. Sci. 103, 7460–7464.

Chevrot, R., Rosen, R., Haudecoeur, E., Cirou, A., Shelp, B.J., Ron, E., Faure, D., 2006b. GABA controls the level of quorum-sensing signal in *Agrobacterium tumefaciens*. Proc. Natl. Acad. Sci. 103, 7460–7464.

Cho, H., Winans, S.C., 2005. VirA and VirG activate the Ti plasmid repABC operon, elevating plasmid copy number in response to wound-released chemical signals. Proc. Natl. Acad. Sci. U. S. A. 102, 14843–14848. doi:10.1073/pnas.0503458102

Cho, H., Winans, S.C., 2007. TraA, TraC and TraD autorepress two divergent quorum-regulated promoters near the transfer origin of the Ti plasmid of Agrobacterium tumefaciens. Mol. Microbiol. 63, 1769–1782. doi:10.1111/j.1365-2958.2007.05624.x

Christiansen, J.K., Nielsen, J.S., Ebersbach, T., Valentin-Hansen, P., Søgaard-Andersen, L., Kallipolitis, B.H., 2006. Identification of small Hfq-binding RNAs in Listeria monocytogenes. RNA N. Y. N 12, 1383–1396. doi:10.1261/rna.49706

Christie, P.J., 2004. Type IV secretion: the Agrobacterium VirB/D4 and related conjugation systems. Biochim. Biophys. Acta 1694, 219–234. doi:10.1016/j.bbamcr.2004.02.013

Christie, P.J., Whitaker, N., González-Rivera, C., 2014. Mechanism and structure of the bacterial type IV secretion systems. Biochim. Biophys. Acta 1843, 1578–1591. doi:10.1016/j.bbamcr.2013.12.019

Clément, B., Perot, J., Geoffroy, P., Legrand, M., Zon, J., Otten, L., 2007. Abnormal accumulation of sugars and phenolics in tobacco roots expressing the Agrobacterium T-6b oncogene and the role of these compounds in 6b-induced growth. Mol. Plant-Microbe Interact. MPMI 20, 53–62. doi:10.1094/MPMI-20-0053

Conn, H.J., 1942. Validity of the Genus Alcaligenes. J. Bacteriol. 44, 353–360.

Cook, D.M., Li, P.L., Ruchaud, F., Padden, S., Farrand, S.K., 1997. Ti plasmid conjugation is independent of vir: reconstitution of the tra functions from pTiC58 as a binary system. J. Bacteriol. 179, 1291–1297.

Costechareyre, D., Rhouma, A., Lavire, C., Portier, P., Chapulliot, D., Bertolla, F., Boubaker, A., Dessaux, Y., Nesme, X., 2010. Rapid and efficient identification of Agrobacterium species by recA allele analysis: Agrobacterium recA diversity. Microb. Ecol. 60, 862–872. doi:10.1007/s00248-010-9685-7

Cui, Y., Chatterjee, A., Chatterjee, A.K., 2001. Effects of the two-component system comprising GacA and GacS of Erwinia carotovora subsp. carotovora on the production of global regulatory rsmB RNA, extracellular enzymes, and harpinEcc. Mol. Plant-Microbe Interact. MPMI 14, 516–526. doi:10.1094/MPMI.2001.14.4.516

Dahlberg, C., Chao, L., 2003. Amelioration of the cost of conjugative plasmid carriage in Eschericha coli K12. Genetics 165, 1641–1649.

Darfeuille, F., Unoson, C., Vogel, J., Wagner, E.G.H., 2007. An antisense RNA inhibits translation by competing with standby ribosomes. Mol. Cell 26, 381–392. doi:10.1016/j.molcel.2007.04.003

Deakin, W.J., 1994. Molecular characterisation of flagellar genes from agrobacterium tumefaciens.

Deakin, W.J., Furniss, C.S., Parker, V.E., Shaw, C.H., 1997a. Isolation and characterisation of a linked cluster of genes from Agrobacterium tumefaciens encoding proteins involved in flagellar basal-body structure. Gene 189, 135–137.

Deakin, W.J., Parker, V.E., Wright, E.L., Ashcroft, K.J., Loake, G.J., Shaw, C.H., 1999. Agrobacterium tumefaciens possesses a fourth flagellin gene located in a large gene cluster concerned with flagellar structure, assembly and motility. Microbiology 145, 1397–1407. doi:10.1099/13500872-145-6-1397

Deakin, W.J., Sanderson, J.L., Goswami, T., Shaw, C.H., 1997b. The Agrobacterium tumefaciens motor gene, motA, is in a linked cluster with the flagellar switch protein genes, fliG, fliM and fliN. Gene 189, 139–141.

De Iannino, N.I., Ugalde, R.A., 1989. Biochemical characterization of avirulent Agrobacterium tumefaciens chvA mutants: synthesis and excretion of beta-(1-2)glucan. J. Bacteriol. 171, 2842–2849.

Del Val, C., Rivas, E., Torres-Quesada, O., Toro, N., Jiménez-Zurdo, J.I., 2007. Identification of differentially expressed small non-coding RNAs in the legume endosymbiont Sinorhizobium meliloti by comparative genomics. Mol. Microbiol. 66, 1080–1091. doi:10.1111/j.1365-2958.2007.05978.x

DeNap, J.C.B., Hergenrother, P.J., 2005. Bacterial death comes full circle: targeting plasmid replication in drug-resistant bacteria. Org. Biomol. Chem. 3, 959–966. doi:10.1039/b500182j

Desnoyers, G., Bouchard, M.-P., Massé, E., 2013. New insights into small RNA-dependent translational regulation in prokaryotes. Trends Genet. TIG 29, 92–98. doi:10.1016/j.tig.2012.10.004

Dessaux, Y., Petit, A., Farrand, S.K., Murphy, P.J., 1998. Opines and Opine-Like Molecules Involved in Plant-Rhizobiaceae Interactions, in: Spaink, H.P., Kondorosi, A., Hooykaas, P.J.J. (Eds.), The Rhizobiaceae. Springer Netherlands, Dordrecht, pp. 173–197.

Douchin, V., Bohn, C., Bouloc, P., 2006. Down-regulation of porins by a small RNA bypasses the essentiality of the regulated intramembrane proteolysis protease RseP in Escherichia coli. J. Biol. Chem. 281, 12253–12259. doi:10.1074/jbc.M600819200

Eggenhofer, F., Tafer, H., Stadler, P.F., Hofacker, I.L., 2011. RNApredator: fast accessibility-based prediction of sRNA targets. Nucleic Acids Res. 39, W149–154. doi:10.1093/nar/gkr467

Eisenbrandt, R., Kalkum, M., Lai, E.M., Lurz, R., Kado, C.I., Lanka, E., 1999. Conjugative pili of IncP plasmids, and the Ti plasmid T pilus are composed of cyclic subunits. J. Biol. Chem. 274, 22548–22555.

Farrand, S.K., Hwang, I., Cook, D.M., 1996. The tra region of the nopaline-type Ti plasmid is a chimera with elements related to the transfer systems of RSF1010, RP4, and F. J. Bacteriol. 178, 4233–4247.

Fender, A., Elf, J., Hampel, K., Zimmermann, B., Wagner, E.G.H., 2010. RNAs actively cycle on the Smlike protein Hfq. Genes Dev. 24, 2621–2626. doi:10.1101/gad.591310

Fineran, P.C., Blower, T.R., Foulds, I.J., Humphreys, D.P., Lilley, K.S., Salmond, G.P.C., 2009. The phage abortive infection system, ToxIN, functions as a protein-RNA toxin-antitoxin pair. Proc. Natl. Acad. Sci. U. S. A. 106, 894–899. doi:10.1073/pnas.0808832106

Flores-Mireles, A.L., Eberhard, A., Winans, S.C., 2012. Agrobacterium tumefaciens can obtain sulphur from an opine that is synthesized by octopine synthase using S-methylmethionine as a substrate. Mol. Microbiol. 84, 845–856. doi:10.1111/j.1365-2958.2012.08061.x

Fortin, C., Marquis, C., Nester, E.W., Dion, P., 1993. Dynamic structure of Agrobacterium tumefaciens Ti plasmids. J. Bacteriol. 175, 4790–4799.

FRANZE de FERNANDEZ, M.T., Eoyang, L., August, J.T., 1968. Factor Fraction required for the Synthesis of Bacteriophage Q β -RNA. Nature 219, 588–590. doi:10.1038/219588a0

Fuqua, W.C., Winans, S.C., Greenberg, E.P., 1994. Quorum sensing in bacteria: the LuxR-LuxI family of cell density-responsive transcriptional regulators. J. Bacteriol. 176, 269–275.

García-Rodríguez, F.M., Schrammeijer, B., Hooykaas, P.J.J., 2006. The Agrobacterium VirE3 effector protein: a potential plant transcriptional activator. Nucleic Acids Res. 34, 6496–6504. doi:10.1093/nar/gkl877

Geisinger, E., Adhikari, R.P., Jin, R., Ross, H.F., Novick, R.P., 2006. Inhibition of rot translation by RNAIII, a key feature of agr function. Mol. Microbiol. 61, 1038–1048. doi:10.1111/j.1365-2958.2006.05292.x

Gelvin, S.B., 2003. Agrobacterium-Mediated Plant Transformation: the Biology behind the "Gene-Jockeying" Tool. Microbiol. Mol. Biol. Rev. 67, 16–37. doi:10.1128/MMBR.67.1.16-37.2003

Gerdes, K., Rasmussen, P.B., Molin, S., 1986. Unique type of plasmid maintenance function: postsegregational killing of plasmid-free cells. Proc. Natl. Acad. Sci. U. S. A. 83, 3116–3120.

Gómez-Lozano, M., Marvig, R.L., Molin, S., Long, K.S., 2012. Genome-wide identification of novel small RNAs in Pseudomonas aeruginosa. Environ. Microbiol. 14, 2006–2016. doi:10.1111/j.1462-2920.2012.02759.x

Goodner, B., Hinkle, G., Gattung, S., Miller, N., Blanchard, M., Qurollo, B., Goldman, B.S., Cao, Y., Askenazi, M., Halling, C., Mullin, L., Houmiel, K., Gordon, J., Vaudin, M., Iartchouk, O., Epp, A., Liu, F., Wollam, C., Allinger, M., Doughty, D., Scott, C., Lappas, C., Markelz, B., Flanagan, C., Crowell, C., Gurson, J., Lomo, C., Sear, C., Strub, G., Cielo, C., Slater, S., 2001. Genome sequence of the plant pathogen and biotechnology agent Agrobacterium tumefaciens C58. Science 294, 2323–2328. doi:10.1126/science.1066803

Göpel, Y., Khan, M.A., Görke, B., 2014. Ménage à trois: post-transcriptional control of the key enzyme for cell envelope synthesis by a base-pairing small RNA, an RNase adaptor protein, and a small RNA mimic. RNA Biol. 11, 433–442. doi:10.4161/rna.28301

Grange, W., Duckely, M., Husale, S., Jacob, S., Engel, A., Hegner, M., 2008. VirE2: a unique ssDNAcompacting molecular machine. PLoS Biol. 6, e44. doi:10.1371/journal.pbio.0060044

Greated, A., Titok, M., Krasowiak, R., Fairclough, R.J., Thomas, C.M., 2000. The replication and stableinheritance functions of IncP-9 plasmid pM3. Microbiol. Read. Engl. 146 (Pt 9), 2249–2258.

Gruber, A.R., Neuböck, R., Hofacker, I.L., Washietl, S., 2007. The RNAz web server: prediction of thermodynamically stable and evolutionarily conserved RNA structures. Nucleic Acids Res. 35, W335–338. doi:10.1093/nar/gkm222

Guyon, P., 1993. Transformed Plants Producing Opines Specifically Promote Growth of Opine-Degrading Agrobacteria. Mol. Plant. Microbe Interact. 6, 92. doi:10.1094/MPMI-6-092

Guyon, P., Chilton, M.D., Petit, A., Tempé, J., 1980. Agropine in "null-type" crown gall tumors: Evidence for generality of the opine concept. Proc. Natl. Acad. Sci. U. S. A. 77, 2693–2697.

Hamilton, C.M., Lee, H., Li, P.L., Cook, D.M., Piper, K.R., von Bodman, S.B., Lanka, E., Ream, W., Farrand, S.K., 2000. TraG from RP4 and TraG and VirD4 from Ti plasmids confer relaxosome specificity to the conjugal transfer system of pTiC58. J. Bacteriol. 182, 1541–1548.

Haudecoeur, E., Tannières, M., Cirou, A., Raffoux, A., Dessaux, Y., Faure, D., 2009. Different regulation and roles of lactonases AiiB and AttM in Agrobacterium tumefaciens C58. Mol. Plant-Microbe Interact. MPMI 22, 529–537. doi:10.1094/MPMI-22-5-0529

Hayes, F., 2003. Toxins-antitoxins: plasmid maintenance, programmed cell death, and cell cycle arrest. Science 301, 1496–1499. doi:10.1126/science.1088157

Heckel, B.C., Tomlinson, A.D., Morton, E.R., Choi, J.-H., Fuqua, C., 2014. Agrobacterium tumefaciens exoR controls acid response genes and impacts exopolysaccharide synthesis, horizontal gene
transfer, and virulence gene expression. J. Bacteriol. 196, 3221–3233. doi:10.1128/JB.01751-14

He, F., Nair, G.R., Soto, C.S., Chang, Y., Hsu, L., Ronzone, E., DeGrado, W.F., Binns, A.N., 2009. Molecular Basis of ChvE Function in Sugar Binding, Sugar Utilization, and Virulence in Agrobacterium tumefaciens. J. Bacteriol. 191, 5802–5813. doi:10.1128/JB.00451-09

Heidrich, N., Moll, I., Brantl, S., 2007. In vitro analysis of the interaction between the small RNA SR1 and its primary target ahrC mRNA. Nucleic Acids Res. 35, 4331–4346. doi:10.1093/nar/gkm439

Heroven, A.K., Böhme, K., Dersch, P., 2012. The Csr/Rsm system of Yersinia and related pathogens: a post-transcriptional strategy for managing virulence. RNA Biol. 9, 379–391. doi:10.4161/rna.19333

Hershberg, R., Altuvia, S., Margalit, H., 2003. A survey of small RNA-encoding genes in Escherichia coli. Nucleic Acids Res. 31, 1813–1820.

Hooykaas, P.J., den Dulk-Ras, H., Schilperoort, R.A., 1988. The Agrobacterium tumefaciens T-DNA gene 6(b) is an onc gene. Plant Mol. Biol. 11, 791–794. doi:10.1007/BF00019519

Hooykaas, P.J.J., Klapwijk, P.M., Nuti, M.P., Schilperoort, R.A., Rorsch, A., 1977. Transfer of the Agrobacterium tumefaciens TI Plasmid to Avirulent Agrobacteria and to Rhizobium ex planta. J. Gen. Microbiol. 98, 477–484. doi:10.1099/00221287-98-2-477

Huntzinger, E., Boisset, S., Saveanu, C., Benito, Y., Geissmann, T., Namane, A., Lina, G., Etienne, J., Ehresmann, B., Ehresmann, C., Jacquier, A., Vandenesch, F., Romby, P., 2005. Staphylococcus aureus RNAIII and the endoribonuclease III coordinately regulate spa gene expression. EMBO J. 24, 824–835. doi:10.1038/sj.emboj.7600572

Hussein, R., Lim, H.N., 2011. Disruption of small RNA signaling caused by competition for Hfq. Proc. Natl. Acad. Sci. U. S. A. 108, 1110–1115. doi:10.1073/pnas.1010082108

Hu, X., Zhao, J., DeGrado, W.F., Binns, A.N., 2013. Agrobacterium tumefaciens recognizes its host environment using ChvE to bind diverse plant sugars as virulence signals. Proc. Natl. Acad. Sci. U. S. A. 110, 678–683. doi:10.1073/pnas.1215033110

Hynes, M.F., Simon, R., Pühler, A., 1985. The development of plasmid-free strains of Agrobacterium tumefaciens by using incompatibility with a Rhizobium meliloti plasmid to eliminate pAtC58. Plasmid 13, 99–105.

Jacques, J.-F., Jang, S., Prévost, K., Desnoyers, G., Desmarais, M., Imlay, J., Massé, E., 2006. RyhB small RNA modulates the free intracellular iron pool and is essential for normal growth during iron limitation in Escherichia coli. Mol. Microbiol. 62, 1181–1190. doi:10.1111/j.1365-2958.2006.05439.x

Jin, S.G., Prusti, R.K., Roitsch, T., Ankenbauer, R.G., Nester, E.W., 1990. Phosphorylation of the VirG protein of Agrobacterium tumefaciens by the autophosphorylated VirA protein: essential role in biological activity of VirG. J. Bacteriol. 172, 4945–4950.

Jin, S., Roitsch, T., Ankenbauer, R.G., Gordon, M.P., Nester, E.W., 1990. The VirA protein of Agrobacterium tumefaciens is autophosphorylated and is essential for vir gene regulation. J. Bacteriol. 172, 525–530.

Johansson, J., Mandin, P., Renzoni, A., Chiaruttini, C., Springer, M., Cossart, P., 2002. An RNA thermosensor controls expression of virulence genes in Listeria monocytogenes. Cell 110, 551–561.

Jonas, K., Melefors, O., 2009. The Escherichia coli CsrB and CsrC small RNAs are strongly induced

during growth in nutrient-poor medium. FEMS Microbiol. Lett. 297, 80–86. doi:10.1111/j.1574-6968.2009.01661.x

Jumas-Bilak, E., Michaux-Charachon, S., Bourg, G., Ramuz, M., Allardet-Servent, A., 1998. Unconventional genomic organization in the alpha subgroup of the Proteobacteria. J. Bacteriol. 180, 2749–2755.

Kakiuchi, Y., Gàlis, I., Tamogami, S., Wabiko, H., 2006. Reduction of polar auxin transport in tobacco by the tumorigenic Agrobacterium tumefaciens AK-6b gene. Planta 223, 237–247. doi:10.1007/s00425-005-0080-4

Kalamorz, F., Reichenbach, B., März, W., Rak, B., Görke, B., 2007. Feedback control of glucosamine-6phosphate synthase GlmS expression depends on the small RNA GlmZ and involves the novel protein YhbJ in Escherichia coli. Mol. Microbiol. 65, 1518–1533. doi:10.1111/j.1365-2958.2007.05888.x

Kalogeraki, V.S., Zhu, J., Eberhard, A., Madsen, E.L., Winans, S.C., 1999. The phenolic vir gene inducer ferulic acid is O-demethylated by the VirH2 protein of an Agrobacterium tumefaciens Ti plasmid. Mol. Microbiol. 34, 512–522.

Kandror, O., DeLeon, A., Goldberg, A.L., 2002. Trehalose synthesis is induced upon exposure of Escherichia coli to cold and is essential for viability at low temperatures. Proc. Natl. Acad. Sci. U. S. A. 99, 9727–9732. doi:10.1073/pnas.142314099

Kawamoto, H., Morita, T., Shimizu, A., Inada, T., Aiba, H., 2005. Implication of membrane localization of target mRNA in the action of a small RNA: mechanism of post-transcriptional regulation of glucose transporter in Escherichia coli. Genes Dev. 19, 328–338. doi:10.1101/gad.1270605

Kay, E., Dubuis, C., Haas, D., 2005. Three small RNAs jointly ensure secondary metabolism and biocontrol in Pseudomonas fluorescens CHA0. Proc. Natl. Acad. Sci. U. S. A. 102, 17136–17141. doi:10.1073/pnas.0505673102

Keane, P.J., Kerr, A., New, P.B., 1970. Crown gall of stone fruits II. Identification and nomenclature of Agrobacterium isolates. Aust. J. Biol. Sci. 23, 585–596.

Kersters, K., De Ley, J., Sneath, P.H.A., Sackin, M., 1973. Numerical Taxonomic Analysis of Agrobacterium. J. Gen. Microbiol. 78, 227–239. doi:10.1099/00221287-78-2-227

Khan, S.R., Farrand, S.K., 2009. The BlcC (AttM) lactonase of Agrobacterium tumefaciens does not quench the quorum-sensing system that regulates Ti plasmid conjugative transfer. J. Bacteriol. 191, 1320–1329. doi:10.1128/JB.01304-08

Kim, H.S., Yi, H., Myung, J., Piper, K.R., Farrand, S.K., 2008. Opine-based Agrobacterium competitiveness: dual expression control of the agrocinopine catabolism (acc) operon by agrocinopines and phosphate levels. J. Bacteriol. 190, 3700–3711. doi:10.1128/JB.00067-08

Kim, J.N., Kwon, Y.M., 2013. Identification of target transcripts regulated by small RNA RyhB homologs in Salmonella: RyhB-2 regulates motility phenotype. Microbiol. Res. 168, 621–629. doi:10.1016/j.micres.2013.06.002

Kim, K.S., Baek, C.H., Lee, J.K., Yang, J.M., Farrand, S.K., 2001. Intracellular accumulation of mannopine, an opine produced by crown gall tumors, transiently inhibits growth of Agrobacterium tumefaciens. Mol. Plant-Microbe Interact. MPMI 14, 793–803. doi:10.1094/MPMI.2001.14.6.793

Kitakura, S., Fujita, T., Ueno, Y., Terakura, S., Wabiko, H., Machida, Y., 2002. The protein encoded by

oncogene 6b from Agrobacterium tumefaciens interacts with a nuclear protein of tobacco. Plant Cell 14, 451–463.

Koraimann, G., Koraimann, C., Koronakis, V., Schlager, S., Högenauer, G., 1991. Repression and derepression of conjugation of plasmid R1 by wild-type and mutated finP antisense RNA. Mol. Microbiol. 5, 77–87.

Körber, H., Strizhov, N., Staiger, D., Feldwisch, J., Olsson, O., Sandberg, G., Palme, K., Schell, J., Koncz, C., 1991. T-DNA gene 5 of Agrobacterium modulates auxin response by autoregulated synthesis of a growth hormone antagonist in plants. EMBO J. 10, 3983–3991.

Kreusch, D., von Lintig, J., Schröder, J., 1995. Ti plasmid-encoded octopine and nopaline catabolism in Agrobacterium: specificities of the LysR-type regulators OccR and NocR, and protein-induced DNA bending. Mol. Gen. Genet. MGG 249, 102–110.

Krimi, Z., Petit, A., Mougel, C., Dessaux, Y., Nesme, X., 2002. Seasonal fluctuations and long-term persistence of pathogenic populations of Agrobacterium spp. in soils. Appl. Environ. Microbiol. 68, 3358–3365.

Krüger, J., Rehmsmeier, M., 2006. RNAhybrid: microRNA target prediction easy, fast and flexible. Nucleic Acids Res. 34, W451–454. doi:10.1093/nar/gkl243

Lang, J., Faure, D., 2014. Functions and regulation of quorum-sensing in Agrobacterium tumefaciens. Front. Plant Sci. 5, 14. doi:10.3389/fpls.2014.00014

Lassalle, F., Campillo, T., Vial, L., Baude, J., Costechareyre, D., Chapulliot, D., Shams, M., Abrouk, D., Lavire, C., Oger-Desfeux, C., Hommais, F., Guéguen, L., Daubin, V., Muller, D., Nesme, X., 2011. Genomic species are ecological species as revealed by comparative genomics in Agrobacterium tumefaciens. Genome Biol. Evol. 3, 762–781. doi:10.1093/gbe/evr070

Lee, K., Huang, X., Yang, C., Lee, D., Ho, V., Nobuta, K., Fan, J.-B., Wang, K., 2013a. A genome-wide survey of highly expressed non-coding RNAs and biological validation of selected candidates in Agrobacterium tumefaciens. PloS One 8, e70720. doi:10.1371/journal.pone.0070720

Lenz, D.H., Miller, M.B., Zhu, J., Kulkarni, R.V., Bassler, B.L., 2005. CsrA and three redundant small RNAs regulate quorum sensing in Vibrio cholerae. Mol. Microbiol. 58, 1186–1202. doi:10.1111/j.1365-2958.2005.04902.x

Lenz, D.H., Mok, K.C., Lilley, B.N., Kulkarni, R.V., Wingreen, N.S., Bassler, B.L., 2004. The small RNA chaperone Hfq and multiple small RNAs control quorum sensing in Vibrio harveyi and Vibrio cholerae. Cell 118, 69–82. doi:10.1016/j.cell.2004.06.009

Liang, Z., Tzfira, T., 2013. In vivo formation of double-stranded T-DNA molecules by T-strand priming. Nat. Commun. 4, 2253. doi:10.1038/ncomms3253

Li, A.X., Marz, M., Qin, J., Reidys, C.M., 2011. RNA-RNA interaction prediction based on multiple sequence alignments. Bioinforma. Oxf. Engl. 27, 456–463. doi:10.1093/bioinformatics/btq659

Lindström, K., Young, J.P.W., 2011. International Committee on Systematics of Prokaryotes Subcommittee on the taxonomy of Agrobacterium and Rhizobium: minutes of the meeting, 7 September 2010, Geneva, Switzerland. Int. J. Syst. Evol. Microbiol. 61, 3089–3093. doi:10.1099/ijs.0.036913-0

Link, T.M., Valentin-Hansen, P., Brennan, R.G., 2009. Structure of Escherichia coli Hfq bound to

polyriboadenylate RNA. Proc. Natl. Acad. Sci. U. S. A. 106, 19292–19297. doi:10.1073/pnas.0908744106

Lin, Y.-H., Pierce, B.D., Fang, F., Wise, A., Binns, A.N., Lynn, D.G., 2014. Role of the VirA histidine autokinase of Agrobacterium tumefaciens in the initial steps of pathogenesis. Front. Plant Sci. 5, 195. doi:10.3389/fpls.2014.00195

Li, P.L., Farrand, S.K., 2000. The replicator of the nopaline-type Ti plasmid pTiC58 is a member of the repABC family and is influenced by the TraR-dependent quorum-sensing regulatory system. J. Bacteriol. 182, 179–188.

Li, P.L., Hwang, I., Miyagi, H., True, H., Farrand, S.K., 1999. Essential components of the Ti plasmid trb system, a type IV macromolecular transporter. J. Bacteriol. 181, 5033–5041.

Liu, M.Y., Gui, G., Wei, B., Preston, J.F., Oakford, L., Yüksel, U., Giedroc, D.P., Romeo, T., 1997. The RNA molecule CsrB binds to the global regulatory protein CsrA and antagonizes its activity in Escherichia coli. J. Biol. Chem. 272, 17502–17510.

Liu, P., Nester, E.W., 2006. Indoleacetic acid, a product of transferred DNA, inhibits vir gene expression and growth of Agrobacterium tumefaciens C58. Proc. Natl. Acad. Sci. U. S. A. 103, 4658–4662. doi:10.1073/pnas.0600366103

Liu, Y., Wu, N., Dong, J., Gao, Y., Zhang, X., Mu, C., Shao, N., Yang, G., 2010. Hfq is a global regulator that controls the pathogenicity of Staphylococcus aureus. PloS One 5. doi:10.1371/journal.pone.0013069

Livny, J., Fogel, M.A., Davis, B.M., Waldor, M.K., 2005. sRNAPredict: an integrative computational approach to identify sRNAs in bacterial genomes. Nucleic Acids Res. 33, 4096–4105. doi:10.1093/nar/gki715

Li, W., Ying, X., Lu, Q., Chen, L., 2012. Predicting sRNAs and their targets in bacteria. Genomics Proteomics Bioinformatics 10, 276–284. doi:10.1016/j.gpb.2012.09.004

Lorenz, C., Gesell, T., Zimmermann, B., Schoeberl, U., Bilusic, I., Rajkowitsch, L., Waldsich, C., von Haeseler, A., Schroeder, R., 2010. Genomic SELEX for Hfq-binding RNAs identifies genomic aptamers predominantly in antisense transcripts. Nucleic Acids Res. 38, 3794–3808. doi:10.1093/nar/gkq032

Lorenz, C., von Pelchrzim, F., Schroeder, R., 2006. Genomic systematic evolution of ligands by exponential enrichment (Genomic SELEX) for the identification of protein-binding RNAs independent of their expression levels. Nat. Protoc. 1, 2204–2212. doi:10.1038/nprot.2006.372

Lu, X., Goodrich-Blair, H., Tjaden, B., 2011. Assessing computational tools for the discovery of small RNA genes in bacteria. RNA N. Y. N 17, 1635–1647. doi:10.1261/rna.2689811

Lybecker, M., Zimmermann, B., Bilusic, I., Tukhtubaeva, N., Schroeder, R., 2014. The double-stranded transcriptome of Escherichia coli. Proc. Natl. Acad. Sci. U. S. A. 111, 3134–3139. doi:10.1073/pnas.1315974111

Magori, S., Citovsky, V., 2011. Agrobacterium counteracts host-induced degradation of its effector Fbox protein. Sci. Signal. 4, ra69. doi:10.1126/scisignal.2002124

Mandin, P., Repoila, F., Vergassola, M., Geissmann, T., Cossart, P., 2007. Identification of new noncoding RNAs in Listeria monocytogenes and prediction of mRNA targets. Nucleic Acids Res. 35, 962–974. doi:10.1093/nar/gkl1096

Mantis, N.J., Winans, S.C., 1992. The Agrobacterium tumefaciens vir gene transcriptional activator virG is transcriptionally induced by acid pH and other stress stimuli. J. Bacteriol. 174, 1189–1196.

Marincs, F., White, D.W., 1993. Nopaline causes a conformational change in the NocR regulatory protein-nocR promoter complex of Agrobacterium tumefaciens Ti plasmid pTiT37. Mol. Gen. Genet. MGG 241, 65–72.

Matthysse, A.G., 2014. Attachment of Agrobacterium to plant surfaces. Front. Plant Sci. 5, 252. doi:10.3389/fpls.2014.00252

Matthysse, A.G., Yarnall, H., Boles, S.B., McMahan, S., 2000. A region of the Agrobacterium tumefaciens chromosome containing genes required for virulence and attachment to host cells. Biochim. Biophys. Acta 1490, 208–212.

Merlo, D.J., Nester, E.W., 1977. Plasmids in avirulent strains of Agrobacterium. J. Bacteriol. 129, 76–80.

Modi, R.I., Wilke, C.M., Rosenzweig, R.F., Adams, J., 1991. Plasmid macro-evolution: selection of deletions during adaptation in a nutrient-limited environment. Genetica 84, 195–202.

Mohanty, B.K., Maples, V.F., Kushner, S.R., 2004. The Sm-like protein Hfq regulates polyadenylation dependent mRNA decay in Escherichia coli. Mol. Microbiol. 54, 905–920. doi:10.1111/j.1365-2958.2004.04337.x

Möller, P., Overlöper, A., Förstner, K.U., Wen, T.-N., Sharma, C.M., Lai, E.-M., Narberhaus, F., 2014a. Profound Impact of Hfq on Nutrient Acquisition, Metabolism and Motility in the Plant Pathogen Agrobacterium tumefaciens. PloS One 9, e110427. doi:10.1371/journal.pone.0110427

Möller, P., Overlöper, A., Förstner, K.U., Wen, T.-N., Sharma, C.M., Lai, E.-M., Narberhaus, F., 2014b. Profound Impact of Hfq on Nutrient Acquisition, Metabolism and Motility in the Plant Pathogen Agrobacterium tumefaciens. PloS One 9, e110427. doi:10.1371/journal.pone.0110427

Møller, T., Franch, T., Udesen, C., Gerdes, K., Valentin-Hansen, P., 2002. Spot 42 RNA mediates discoordinate expression of the E. coli galactose operon. Genes Dev. 16, 1696–1706. doi:10.1101/gad.231702

Moll, I., Afonyushkin, T., Vytvytska, O., Kaberdin, V.R., Bläsi, U., 2003. Coincident Hfq binding and RNase E cleavage sites on mRNA and small regulatory RNAs. RNA N. Y. N 9, 1308–1314.

Morita, M., Kanemori, M., Yanagi, H., Yura, T., 1999. Heat-induced synthesis of sigma32 in Escherichia coli: structural and functional dissection of rpoH mRNA secondary structure. J. Bacteriol. 181, 401–410.

Morita, M.T., Tanaka, Y., Kodama, T.S., Kyogoku, Y., Yanagi, H., Yura, T., 1999. Translational induction of heat shock transcription factor sigma32: evidence for a built-in RNA thermosensor. Genes Dev. 13, 655–665.

Morita, T., Maki, K., Aiba, H., 2005. RNase E-based ribonucleoprotein complexes: mechanical basis of mRNA destabilization mediated by bacterial noncoding RNAs. Genes Dev. 19, 2176–2186. doi:10.1101/gad.1330405

Morita, T., Mochizuki, Y., Aiba, H., 2006. Translational repression is sufficient for gene silencing by bacterial small noncoding RNAs in the absence of mRNA destruction. Proc. Natl. Acad. Sci. U. S. A. 103, 4858–4863. doi:10.1073/pnas.0509638103

Mougel, C., 2000. Structure génétique des populations d'Agrobacterium spp : Effet séléctif de la plante et implication dans la diffusion conjugative du plasmide Ti. Université Lyon 1.

Mougel, C., Cournoyer, B., Nesme, X., 2001. Novel tellurite-amended media and specific chromosomal and Ti plasmid probes for direct analysis of soil populations of Agrobacterium biovars 1 and 2. Appl. Environ. Microbiol. 67, 65–74. doi:10.1128/AEM.67.1.65-74.2001

Mückstein, U., Tafer, H., Hackermüller, J., Bernhart, S.H., Stadler, P.F., Hofacker, I.L., 2006. Thermodynamics of RNA-RNA binding. Bioinforma. Oxf. Engl. 22, 1177–1182. doi:10.1093/bioinformatics/btl024

Nair, G.R., Liu, Z., Binns, A.N., 2003. Reexamining the role of the accessory plasmid pAtC58 in the virulence of Agrobacterium tumefaciens strain C58. Plant Physiol. 133, 989–999. doi:10.1104/pp.103.030262

Navarre, W.W., Porwollik, S., Wang, Y., McClelland, M., Rosen, H., Libby, S.J., Fang, F.C., 2006. Selective silencing of foreign DNA with low GC content by the H-NS protein in Salmonella. Science 313, 236–238. doi:10.1126/science.1128794

New, P.B., Kerr, A., 1972. Biological Control of Crown Gall: Field Measurements and Glasshouse Experiments. J. Appl. Bacteriol. 35, 279–287. doi:10.1111/j.1365-2672.1972.tb03699.x

Nocker, A., Krstulovic, N.P., Perret, X., Narberhaus, F., 2001. ROSE elements occur in disparate rhizobia and are functionally interchangeable between species. Arch. Microbiol. 176, 44–51.

Novick, R.P., Ross, H.F., Projan, S.J., Kornblum, J., Kreiswirth, B., Moghazeh, S., 1993. Synthesis of staphylococcal virulence factors is controlled by a regulatory RNA molecule. EMBO J. 12, 3967–3975.

Ogawa, Y., Ishikawa, K., Mii, M., 2000. Highly tumorigenic Agrobacterium tumefaciens strain from crown gall tumors of chrysanthemum. Arch. Microbiol. 173, 311–315.

Ogawa, Y., Mii, M., 2001. Ti- and cryptic-plasmid-borne virulence of wild-type Agrobacterium tumefaciens strain CNI5 isolated from chrysanthemum (Dendranthema grandiflora Tzvelev). Arch. Microbiol. 176, 315–322. doi:10.1007/s002030100331

Ott, A., Idali, A., Marchais, A., Gautheret, D., 2012. NAPP: the Nucleic Acid Phylogenetic Profile Database. Nucleic Acids Res. 40, D205–209. doi:10.1093/nar/gkr807

Overlöper, A., Kraus, A., Gurski, R., Wright, P.R., Georg, J., Hess, W.R., Narberhaus, F., 2014. Two separate modules of the conserved regulatory RNA AbcR1 address multiple target mRNAs in and outside of the translation initiation region. RNA Biol. 11, 624–640.

Padavannil, A., Jobichen, C., Qinghua, Y., Seetharaman, J., Velazquez-Campoy, A., Yang, L., Pan, S.Q., Sivaraman, J., 2014. Dimerization of VirD2 binding protein is essential for Agrobacterium induced tumor formation in plants. PLoS Pathog. 10, e1003948. doi:10.1371/journal.ppat.1003948

Panagopoulos, C.G., Psallidas, P.G., 1973. Characteristics of Greek Isolates of Agrobacterium tumefaciens (E. F. Smith & amp; Townsend) Conn. J. Appl. Bacteriol. 36, 233–240. doi:10.1111/j.1365-2672.1973.tb04096.x

Panday, D., Schumann, P., Das, S.K., 2011. Rhizobium pusense sp. nov., isolated from the rhizosphere of chickpea (Cicer arietinum L.). Int. J. Syst. Evol. Microbiol. 61, 2632–2639. doi:10.1099/ijs.0.028407-0

Pandey, S.P., Minesinger, B.K., Kumar, J., Walker, G.C., 2011. A highly conserved protein of unknown function in Sinorhizobium meliloti affects sRNA regulation similar to Hfq. Nucleic Acids Res. 39, 4691–4708. doi:10.1093/nar/gkr060

Pappas, K.M., Winans, S.C., 2003a. A LuxR-type regulator from Agrobacterium tumefaciens elevates Ti plasmid copy number by activating transcription of plasmid replication genes. Mol. Microbiol. 48, 1059–1073.

Pappas, K.M., Winans, S.C., 2003b. The RepA and RepB autorepressors and TraR play opposing roles in the regulation of a Ti plasmid repABC operon. Mol. Microbiol. 49, 441–455.

Patient, M.E., Summers, D.K., 1993. ColE1 multimer formation triggers inhibition of Escherichia coli cell division. Mol. Microbiol. 9, 1089–1095.

Pazour, G.J., Das, A., 1990. Characterization of the VirG binding site of Agrobacterium tumefaciens. Nucleic Acids Res. 18, 6909–6913.

Petit, A., David, C., Dahl, G.A., Ellis, J.G., Guyon, P., Casse-Delbart, F., Temp, J., 1983. Further extension of the opine concept: Plasmids in Agrobacterium rhizogenes cooperate for opine degradation. MGG Mol. Gen. Genet. 190, 204–214. doi:10.1007/BF00330641

Pfeiffer, V., Papenfort, K., Lucchini, S., Hinton, J.C.D., Vogel, J., 2009. Coding sequence targeting by MicC RNA reveals bacterial mRNA silencing downstream of translational initiation. Nat. Struct. Mol. Biol. 16, 840–846. doi:10.1038/nsmb.1631

Pinto, U.M., Flores-Mireles, A.L., Costa, E.D., Winans, S.C., 2011. RepC protein of the octopine-type Ti plasmid binds to the probable origin of replication within repC and functions only in cis. Mol. Microbiol. 81, 1593–1606. doi:10.1111/j.1365-2958.2011.07789.x

Planamente, S., Mondy, S., Hommais, F., Vigouroux, A., Moréra, S., Faure, D., 2012. Structural basis for selective GABA binding in bacterial pathogens: Structure and function of a selective GABA sensor. Mol. Microbiol. 86, 1085–1099. doi:10.1111/mmi.12043

Platt, T.G., Bever, J.D., Fuqua, C., 2012. A cooperative virulence plasmid imposes a high fitness cost under conditions that induce pathogenesis. Proc. Biol. Sci. 279, 1691–1699. doi:10.1098/rspb.2011.2002

Popoff, M.Y., Kersters, K., Kiredjian, M., Miras, I., Coynault, C., 1984. [Taxonomic position of Agrobacterium strains of hospital origin]. Ann. Microbiol. (Paris) 135A, 427–442.

Prévost, K., Salvail, H., Desnoyers, G., Jacques, J.-F., Phaneuf, E., Massé, E., 2007. The small RNA RyhB activates the translation of shiA mRNA encoding a permease of shikimate, a compound involved in siderophore synthesis. Mol. Microbiol. 64, 1260–1273. doi:10.1111/j.1365-2958.2007.05733.x

Puławska, J., Sobiczewski, P., 2005. Development of a semi-nested PCR based method for sensitive detection of tumorigenic Agrobacterium in soil. J. Appl. Microbiol. 98, 710–721. doi:10.1111/j.1365-2672.2004.02503.x

Ramírez-Bahena, M.H., Vial, L., Lassalle, F., Diel, B., Chapulliot, D., Daubin, V., Nesme, X., Muller, D., 2014. Single acquisition of protelomerase gave rise to speciation of a large and diverse clade within the Agrobacterium/Rhizobium supercluster characterized by the presence of a linear chromid. Mol. Phylogenet. Evol. 73, 202–207. doi:10.1016/j.ympev.2014.01.005

Reichenbach, B., Maes, A., Kalamorz, F., Hajnsdorf, E., Görke, B., 2008. The small RNA GlmY acts

upstream of the sRNA GImZ in the activation of gImS expression and is subject to regulation by polyadenylation in Escherichia coli. Nucleic Acids Res. 36, 2570–2580. doi:10.1093/nar/gkn091

Repoila, F., Darfeuille, F., 2009. Small regulatory non-coding RNAs in bacteria: physiology and mechanistic aspects. Biol. Cell Auspices Eur. Cell Biol. Organ. 101, 117–131. doi:10.1042/BC20070137

Rivas, E., Eddy, S.R., 2001. Noncoding RNA gene detection using comparative sequence analysis. BMC Bioinformatics 2, 8.

Rogowsky, P.M., Close, T.J., Chimera, J.A., Shaw, J.J., Kado, C.I., 1987. Regulation of the vir genes of Agrobacterium tumefaciens plasmid pTiC58. J. Bacteriol. 169, 5101–5112.

Rohringer, R., Samborski, D.J., 1967. Aromatic Compounds in the Host-Parasite Interaction. Annu. Rev. Phytopathol. 5, 77–86. doi:10.1146/annurev.py.05.090167.000453

Romeo, T., 1996. Post-transcriptional regulation of bacterial carbohydrate metabolism: evidence that the gene product CsrA is a global mRNA decay factor. Res. Microbiol. 147, 505–512.

Rosenberg, C., Huguet, T., 1984. The pAtC58 plasmid of Agrobacterium tumefaciens is not essential for tumour induction. MGG Mol. Gen. Genet. 196, 533–536. doi:10.1007/BF00436205

Saavedra De Bast, M., Mine, N., Van Melderen, L., 2008. Chromosomal toxin-antitoxin systems may act as antiaddiction modules. J. Bacteriol. 190, 4603–4609. doi:10.1128/JB.00357-08

Sahr, T., Brüggemann, H., Jules, M., Lomma, M., Albert-Weissenberger, C., Cazalet, C., Buchrieser, C., 2009. Two small ncRNAs jointly govern virulence and transmission in Legionella pneumophila. Mol. Microbiol. 72, 741–762. doi:10.1111/j.1365-2958.2009.06677.x

Salvail, H., Caron, M.-P., Bélanger, J., Massé, E., 2013. Antagonistic functions between the RNA chaperone Hfq and an sRNA regulate sensitivity to the antibiotic colicin. EMBO J. 32, 2764–2778. doi:10.1038/emboj.2013.205

Savka, M.A., Farrand, S.K., 1997. Modification of rhizobacterial populations by engineering bacterium utilization of a novel plant-produced resource. Nat. Biotechnol. 15, 363–368. doi:10.1038/nbt0497-363

Schlüter, J.-P., Reinkensmeier, J., Daschkey, S., Evguenieva-Hackenberg, E., Janssen, S., Jänicke, S., Becker, J.D., Giegerich, R., Becker, A., 2010. A genome-wide survey of sRNAs in the symbiotic nitrogen-fixing alpha-proteobacterium Sinorhizobium meliloti. BMC Genomics 11, 245. doi:10.1186/1471-2164-11-245

Schmidtke, C., Abendroth, U., Brock, J., Serrania, J., Becker, A., Bonas, U., 2013. Small RNA sX13: a multifaceted regulator of virulence in the plant pathogen Xanthomonas. PLoS Pathog. 9, e1003626. doi:10.1371/journal.ppat.1003626

Schmidtke, C., Findeiss, S., Sharma, C.M., Kuhfuss, J., Hoffmann, S., Vogel, J., Stadler, P.F., Bonas, U., 2012. Genome-wide transcriptome analysis of the plant pathogen Xanthomonas identifies sRNAs with putative virulence functions. Nucleic Acids Res. 40, 2020–2031. doi:10.1093/nar/gkr904

Schroth, M.N., Weinhold, A.R., McCain, A.H., Hildebrand, D.C., Ross, N., 1971. Biology and control of *Agrobacterium tumefaciens*. Hilgardia 40, 537–552. doi:10.3733/hilg.v40n15p537

Schuster, C.F., Bertram, R., 2013. Toxin-antitoxin systems are ubiquitous and versatile modulators of prokaryotic cell fate. FEMS Microbiol. Lett. 340, 73–85. doi:10.1111/1574-6968.12074

Seemann, S.E., Richter, A.S., Gesell, T., Backofen, R., Gorodkin, J., 2011. PETcofold: predicting conserved interactions and structures of two multiple alignments of RNA sequences. Bioinforma. Oxf. Engl. 27, 211–219. doi:10.1093/bioinformatics/btq634

Sequeira, L., 1963. Growth Regulators in Plant Disease. Annu. Rev. Phytopathol. 1, 5–30. doi:10.1146/annurev.py.01.090163.000253

Sesto, N., Wurtzel, O., Archambaud, C., Sorek, R., Cossart, P., 2013. The excludon: a new concept in bacterial antisense RNA-mediated gene regulation. Nat. Rev. Microbiol. 11, 75–82. doi:10.1038/nrmicro2934

Shams, M., Vial, L., Chapulliot, D., Nesme, X., Lavire, C., 2013. Rapid and accurate species and genomic species identification and exhaustive population diversity assessment of Agrobacterium spp. using recA-based PCR. Syst. Appl. Microbiol. 36, 351–358. doi:10.1016/j.syapm.2013.03.002

Shao, Y., Bassler, B.L., 2014. Quorum regulatory small RNAs repress type VI secretion in Vibrio cholerae. Mol. Microbiol. 92, 921–930. doi:10.1111/mmi.12599

Sharma, C.M., Darfeuille, F., Plantinga, T.H., Vogel, J., 2007. A small RNA regulates multiple ABC transporter mRNAs by targeting C/A-rich elements inside and upstream of ribosome-binding sites. Genes Dev. 21, 2804–2817. doi:10.1101/gad.447207

Sharma, C.M., Papenfort, K., Pernitzsch, S.R., Mollenkopf, H.-J., Hinton, J.C.D., Vogel, J., 2011. Pervasive post-transcriptional control of genes involved in amino acid metabolism by the Hfq-dependent GcvB small RNA. Mol. Microbiol. 81, 1144–1165. doi:10.1111/j.1365-2958.2011.07751.x

Shimoda, N., Toyoda-Yamamoto, A., Aoki, S., Machida, Y., 1993. Genetic evidence for an interaction between the VirA sensor protein and the ChvE sugar-binding protein of Agrobacterium. J. Biol. Chem. 268, 26552–26558.

Shimoni, Y., Friedlander, G., Hetzroni, G., Niv, G., Altuvia, S., Biham, O., Margalit, H., 2007. Regulation of gene expression by small non-coding RNAs: a quantitative view. Mol. Syst. Biol. 3, 138. doi:10.1038/msb4100181

Shi, Y., Zhao, G., Kong, W., 2014. Genetic analysis of riboswitch-mediated transcriptional regulation responding to Mn2+ in Salmonella. J. Biol. Chem. 289, 11353–11366. doi:10.1074/jbc.M113.517516

Silvaggi, J.M., Perkins, J.B., Losick, R., 2006. Genes for small, noncoding RNAs under sporulation control in Bacillus subtilis. J. Bacteriol. 188, 532–541. doi:10.1128/JB.188.2.532-541.2006

Sittka, A., Lucchini, S., Papenfort, K., Sharma, C.M., Rolle, K., Binnewies, T.T., Hinton, J.C.D., Vogel, J., 2008. Deep sequencing analysis of small noncoding RNA and mRNA targets of the global post-transcriptional regulator, Hfq. PLoS Genet. 4, e1000163. doi:10.1371/journal.pgen.1000163

Slater, S.C., Goldman, B.S., Goodner, B., Setubal, J.C., Farrand, S.K., Nester, E.W., Burr, T.J., Banta, L., Dickerman, A.W., Paulsen, I., Otten, L., Suen, G., Welch, R., Almeida, N.F., Arnold, F., Burton, O.T., Du, Z., Ewing, A., Godsy, E., Heisel, S., Houmiel, K.L., Jhaveri, J., Lu, J., Miller, N.M., Norton, S., Chen, Q., Phoolcharoen, W., Ohlin, V., Ondrusek, D., Pride, N., Stricklin, S.L., Sun, J., Wheeler, C., Wilson, L., Zhu, H., Wood, D.W., 2009. Genome sequences of three agrobacterium biovars help elucidate the evolution of multichromosome genomes in bacteria. J. Bacteriol. 191, 2501–2511. doi:10.1128/JB.01779-08

Sledjeski, D.D., Gupta, A., Gottesman, S., 1996. The small RNA, DsrA, is essential for the low

temperature expression of RpoS during exponential growth in Escherichia coli. EMBO J. 15, 3993–4000.

Smith, E.F., Townsend, C.O., 1907. A PLANT-TUMOR OF BACTERIAL ORIGIN. Science 25, 671–673. doi:10.1126/science.25.643.671

Sneath, P.H.A., Sokal, R.R., 1962. Numerical Taxonomy. Nature 193, 855–860. doi:10.1038/193855a0

Sonnleitner, E., Haas, D., 2011. Small RNAs as regulators of primary and secondary metabolism in Pseudomonas species. Appl. Microbiol. Biotechnol. 91, 63–79. doi:10.1007/s00253-011-3332-1

Sonnleitner, E., Sorger-Domenigg, T., Madej, M.J., Findeiss, S., Hackermüller, J., Hüttenhofer, A., Stadler, P.F., Bläsi, U., Moll, I., 2008. Detection of small RNAs in Pseudomonas aeruginosa by RNomics and structure-based bioinformatic tools. Microbiol. Read. Engl. 154, 3175–3187. doi:10.1099/mic.0.2008/019703-0

Sorek, R., Cossart, P., 2010. Prokaryotic transcriptomics: a new view on regulation, physiology and pathogenicity. Nat. Rev. Genet. 11, 9–16. doi:10.1038/nrg2695

Soutourina, O.A., Monot, M., Boudry, P., Saujet, L., Pichon, C., Sismeiro, O., Semenova, E., Severinov, K., Le Bouguenec, C., Coppée, J.-Y., Dupuy, B., Martin-Verstraete, I., 2013. Genome-wide identification of regulatory RNAs in the human pathogen Clostridium difficile. PLoS Genet. 9, e1003493. doi:10.1371/journal.pgen.1003493

Stachel, S.E., Zambryski, P.C., 1986. virA and virG control the plant-induced activation of the T-DNA transfer process of A. tumefaciens. Cell 46, 325–333.

Stazic, D., Lindell, D., Steglich, C., 2011. Antisense RNA protects mRNA from RNase E degradation by RNA-RNA duplex formation during phage infection. Nucleic Acids Res. 39, 4890–4899. doi:10.1093/nar/gkr037

Steuten, B., Schneider, S., Wagner, R., 2014. 6S RNA: recent answers--future questions. Mol. Microbiol. 91, 641–648. doi:10.1111/mmi.12484

Stougaard, P., Molin, S., Nordström, K., 1981. RNAs involved in copy-number control and incompatibility of plasmid R1. Proc. Natl. Acad. Sci. U. S. A. 78, 6008–6012.

Subramoni, S., Nathoo, N., Klimov, E., Yuan, Z.-C., 2014. Agrobacterium tumefaciens responses to plant-derived signaling molecules. Front. Plant Sci. 5, 322. doi:10.3389/fpls.2014.00322

Tafer, H., Hofacker, I.L., 2008. RNAplex: a fast tool for RNA-RNA interaction search. Bioinforma. Oxf. Engl. 24, 2657–2663. doi:10.1093/bioinformatics/btn193

Terakura, S., Ueno, Y., Tagami, H., Kitakura, S., Machida, C., Wabiko, H., Aiba, H., Otten, L., Tsukagoshi, H., Nakamura, K., Machida, Y., 2007. An oncoprotein from the plant pathogen agrobacterium has histone chaperone-like activity. Plant Cell 19, 2855–2865. doi:10.1105/tpc.106.049551

Thomashow, M.F., Karlinsey, J.E., Marks, J.R., Hurlbert, R.E., 1987. Identification of a new virulence locus in Agrobacterium tumefaciens that affects polysaccharide composition and plant cell attachment. J. Bacteriol. 169, 3209–3216.

Timmermans, J., Van Melderen, L., 2010. Post-transcriptional global regulation by CsrA in bacteria. Cell. Mol. Life Sci. CMLS 67, 2897–2908. doi:10.1007/s00018-010-0381-z

Tjaden, B., 2008. TargetRNA: a tool for predicting targets of small RNA action in bacteria. Nucleic Acids Res. 36, W109–113. doi:10.1093/nar/gkn264

Toledo-Arana, A., Dussurget, O., Nikitas, G., Sesto, N., Guet-Revillet, H., Balestrino, D., Loh, E., Gripenland, J., Tiensuu, T., Vaitkevicius, K., Barthelemy, M., Vergassola, M., Nahori, M.-A., Soubigou, G., Régnault, B., Coppée, J.-Y., Lecuit, M., Johansson, J., Cossart, P., 2009. The Listeria transcriptional landscape from saprophytism to virulence. Nature 459, 950–956. doi:10.1038/nature08080

Tomizawa, J., Itoh, T., Selzer, G., Som, T., 1981. Inhibition of ColE1 RNA primer formation by a plasmid-specified small RNA. Proc. Natl. Acad. Sci. U. S. A. 78, 1421–1425.

Tomlinson, A.D., Ramey-Hartung, B., Day, T.W., Merritt, P.M., Fuqua, C., 2010. Agrobacterium tumefaciens ExoR represses succinoglycan biosynthesis and is required for biofilm formation and motility. Microbiol. Read. Engl. 156, 2670–2681. doi:10.1099/mic.0.039032-0

Torres-Quesada, O., Reinkensmeier, J., Schlüter, J.-P., Robledo, M., Peregrina, A., Giegerich, R., Toro, N., Becker, A., Jiménez-Zurdo, J.I., 2014. Genome-wide profiling of Hfq-binding RNAs uncovers extensive post-transcriptional rewiring of major stress response and symbiotic regulons in Sinorhizobium meliloti. RNA Biol. 11, 563–579.

Tzfira, T., Li, J., Lacroix, B., Citovsky, V., 2004. Agrobacterium T-DNA integration: molecules and models. Trends Genet. TIG 20, 375–383. doi:10.1016/j.tig.2004.06.004

Ulvé, V.M., Sevin, E.W., Chéron, A., Barloy-Hubler, F., 2007. Identification of chromosomal alphaproteobacterial small RNAs by comparative genome analysis and detection in Sinorhizobium meliloti strain 1021. BMC Genomics 8, 467. doi:10.1186/1471-2164-8-467

Urban, J.H., Vogel, J., 2007. Translational control and target recognition by Escherichia coli small RNAs in vivo. Nucleic Acids Res. 35, 1018–1037. doi:10.1093/nar/gkl1040

Uttaro, A.D., Cangelosi, G.A., Geremia, R.A., Nester, E.W., Ugalde, R.A., 1990. Biochemical characterization of avirulent exoC mutants of Agrobacterium tumefaciens. J. Bacteriol. 172, 1640–1646.

Valverde, C., Livny, J., Schlüter, J.-P., Reinkensmeier, J., Becker, A., Parisi, G., 2008. Prediction of Sinorhizobium meliloti sRNA genes and experimental detection in strain 2011. BMC Genomics 9, 416. doi:10.1186/1471-2164-9-416

Vecerek, B., Moll, I., Bläsi, U., 2007. Control of Fur synthesis by the non-coding RNA RyhB and iron-responsive decoding. EMBO J. 26, 965–975. doi:10.1038/sj.emboj.7601553

Veselov, D., Langhans, M., Hartung, W., Aloni, R., Feussner, I., Götz, C., Veselova, S., Schlomski, S., Dickler, C., Bächmann, K., Ullrich, C.I., 2003. Development of Agrobacterium tumefaciens C58induced plant tumors and impact on host shoots are controlled by a cascade of jasmonic acid, auxin, cytokinin, ethylene and abscisic acid. Planta 216, 512–522. doi:10.1007/s00425-002-0883-5

Vogel, A.M., Das, A., 1992. The Agrobacterium tumefaciens virD3 gene is not essential for tumorigenicity on plants. J. Bacteriol. 174, 5161–5164.

Vogel, J., Argaman, L., Wagner, E.G.H., Altuvia, S., 2004. The small RNA IstR inhibits synthesis of an SOS-induced toxic peptide. Curr. Biol. CB 14, 2271–2276. doi:10.1016/j.cub.2004.12.003

Von Bodman, S.B., McCutchan, J.E., Farrand, S.K., 1989. Characterization of conjugal transfer functions of Agrobacterium tumefaciens Ti plasmid pTiC58. J. Bacteriol. 171, 5281–5289.

Von Lintig, J., Zanker, H., Schröder, J., 1991. Positive regulators of opine-inducible promoters in the nopaline and octopine catabolism regions of Ti plasmids. Mol. Plant-Microbe Interact. MPMI 4, 370–378.

Vytvytska, O., Moll, I., Kaberdin, V.R., von Gabain, A., Bläsi, U., 2000. Hfq (HF1) stimulates ompA mRNA decay by interfering with ribosome binding. Genes Dev. 14, 1109–1118.

Wadler, C.S., Vanderpool, C.K., 2007. A dual function for a bacterial small RNA: SgrS performs base pairing-dependent regulation and encodes a functional polypeptide. Proc. Natl. Acad. Sci. U. S. A. 104, 20454–20459. doi:10.1073/pnas.0708102104

Wagner, E.G.H., 2013. Cycling of RNAs on Hfq. RNA Biol. 10, 619–626. doi:10.4161/rna.24044

Wang, M., Soyano, T., Machida, S., Yang, J.-Y., Jung, C., Chua, N.-H., Yuan, Y.A., 2011. Molecular insights into plant cell proliferation disturbance by Agrobacterium protein 6b. Genes Dev. 25, 64–76. doi:10.1101/gad.1985511

Wassarman, K.M., 2002. Small RNAs in Bacteria. Cell 109, 141–144. doi:10.1016/S0092-8674(02)00717-1

Wassarman, K.M., 2007. 6S RNA: a small RNA regulator of transcription. Curr. Opin. Microbiol. 10, 164–168. doi:10.1016/j.mib.2007.03.008

Wassarman, K.M., Storz, G., 2000. 6S RNA regulates E. coli RNA polymerase activity. Cell 101, 613–623.

Wehner, S., Mannala, G.K., Qing, X., Madhugiri, R., Chakraborty, T., Mraheil, M.A., Hain, T., Marz, M., 2014. Detection of Very Long Antisense Transcripts by Whole Transcriptome RNA-Seq Analysis of Listeria monocytogenes by Semiconductor Sequencing Technology. PloS One 9, e108639. doi:10.1371/journal.pone.0108639

Weiberg, A., Wang, M., Lin, F.-M., Zhao, H., Zhang, Z., Kaloshian, I., Huang, H.-D., Jin, H., 2013. Fungal small RNAs suppress plant immunity by hijacking host RNA interference pathways. Science 342, 118–123. doi:10.1126/science.1239705

Weilbacher, T., Suzuki, K., Dubey, A.K., Wang, X., Gudapaty, S., Morozov, I., Baker, C.S., Georgellis, D., Babitzke, P., Romeo, T., 2003. A novel sRNA component of the carbon storage regulatory system of Escherichia coli. Mol. Microbiol. 48, 657–670.

Werner, T., Schmülling, T., 2009. Cytokinin action in plant development. Curr. Opin. Plant Biol. 12, 527–538. doi:10.1016/j.pbi.2009.07.002

Wilms, I., Möller, P., Stock, A.-M., Gurski, R., Lai, E.-M., Narberhaus, F., 2012a. Hfq influences multiple transport systems and virulence in the plant pathogen Agrobacterium tumefaciens. J. Bacteriol. 194, 5209–5217. doi:10.1128/JB.00510-12

Wilms, I., Overlöper, A., Nowrousian, M., Sharma, C.M., Narberhaus, F., 2012b. Deep sequencing uncovers numerous small RNAs on all four replicons of the plant pathogen Agrobacterium tumefaciens. RNA Biol. 9, 446–457. doi:10.4161/rna.17212

Wilms, I., Voss, B., Hess, W.R., Leichert, L.I., Narberhaus, F., 2011. Small RNA-mediated control of the </i>Agrobacterium tumefaciens<i> GABA binding protein. Mol. Microbiol. 80, 492–506. doi:10.1111/j.1365-2958.2011.07589.x

Winans, S.C., Mantis, N.J., Chen, C.Y., Chang, C.H., Han, D.C., 1994. Host recognition by the VirA, VirG two-component regulatory proteins of agrobacterium tumefaciens. Res. Microbiol. 145, 461–473.

Wood, D.W., Setubal, J.C., Kaul, R., Monks, D.E., Kitajima, J.P., Okura, V.K., Zhou, Y., Chen, L., Wood, G.E., Almeida, N.F., Jr, Woo, L., Chen, Y., Paulsen, I.T., Eisen, J.A., Karp, P.D., Bovee, D., Sr, Chapman, P., Clendenning, J., Deatherage, G., Gillet, W., Grant, C., Kutyavin, T., Levy, R., Li, M.J., McClelland, E., Palmieri, A., Raymond, C., Rouse, G., Saenphimmachak, C., Wu, Z., Romero, P., Gordon, D., Zhang, S., Yoo, H., Tao, Y., Biddle, P., Jung, M., Krespan, W., Perry, M., Gordon-Kamm, B., Liao, L., Kim, S., Hendrick, C., Zhao, Z.Y., Dolan, M., Chumley, F., Tingey, S.V., Tomb, J.F., Gordon, M.P., Olson, M.V., Nester, E.W., 2001. The genome of the natural genetic engineer *Agrobacterium tumefaciens* C58. Science 294, 2317–2323. doi:10.1126/science.1066804

Wu, C.-F., Lin, J.-S., Shaw, G.-C., Lai, E.-M., 2012. Acid-induced type VI secretion system is regulated by ExoR-ChvG/ChvI signaling cascade in Agrobacterium tumefaciens. PLoS Pathog. 8, e1002938. doi:10.1371/journal.ppat.1002938

Ying, X., Cao, Y., Wu, J., Liu, Q., Cha, L., Li, W., 2011. sTarPicker: a method for efficient prediction of bacterial sRNA targets based on a two-step model for hybridization. PloS One 6, e22705. doi:10.1371/journal.pone.0022705

Zeng, Q., Sundin, G.W., 2014a. Genome-wide identification of Hfq-regulated small RNAs in the fire blight pathogen Erwinia amylovora discovered small RNAs with virulence regulatory function. BMC Genomics 15, 414. doi:10.1186/1471-2164-15-414

Zeng, Q., Sundin, G.W., 2014b. Genome-wide identification of Hfq-regulated small RNAs in the fire blight pathogen Erwinia amylovora discovered small RNAs with virulence regulatory function. BMC Genomics 15, 414. doi:10.1186/1471-2164-15-414

Zhao, Y., 2010. Auxin biosynthesis and its role in plant development. Annu. Rev. Plant Biol. 61, 49–64. doi:10.1146/annurev-arplant-042809-112308

Zhu, J., Winans, S.C., 1999. Autoinducer binding by the quorum-sensing regulator TraR increases affinity for target promoters in vitro and decreases TraR turnover rates in whole cells. Proc. Natl. Acad. Sci. U. S. A. 96, 4832–4837.

Zhu, J., Winans, S.C., 2001. The quorum-sensing transcriptional regulator TraR requires its cognate signaling ligand for protein folding, protease resistance, and dimerization. Proc. Natl. Acad. Sci. U. S. A. 98, 1507–1512. doi:10.1073/pnas.98.4.1507

Annexes

Table S1 Article

	Predicted with sRNApredict (x) or Rfam data (RF) or	Identified by	Identified by					
Replicon ^a	homologous in other rhizobia	Wilms et al. 2012	Lee <i>et al</i> 2013	5'-end	3'-end	Length	sRNA type	Strand
Circular						•	**	
RNA1 ^a		х		7924	8131	207	IGR	+
RNA2				12267	12355	88	IGR	+
RNA3				18182	18245	63	IGR	+
RNA4				19824	20005	181	UTR5' <i>trxA</i>	-
RNA5				27913	27964	51	AS atu0026	+
RNA6				33083	33335	252	UTR5' ahcY	-
RNA7				33971	34177	206	atu0031	-
RNA8				36400	36566	166	IGR	-
RNA9				37548	37748	200	UTR5' pckA	+
RNA10				46626	46686	60	IGR	+
RNA11				50547	50736	189	AS atu0048	+
RNA12				66258	66390	132	IGR	-
RNA13				66471	66563	92	IGR	+
RNA14				67574	67818	244	UTR3' frcB	+
RNA15 [°]	RF00174			70431	70659	228	IGR	+
RNA16				70735	70871	136	IGR	+
RNA17				74723	74836	113	AS atu0072	+
RNA18				77467	77486	19	AS moeB	-
RNA19				86866	87235	369	<i>atu0088</i> UTR5'	-
RNA20				91388	91781	393	atu0087	-
RNA21 ^a	х	х		94646	94684	38	IGR	-
RNA22 ^a	RF00169	х	C1_101545R ^b	101444	101546	102	IGR	-
RNA23				103162	103294	132	AS pheA	+
RNA24				104483	104711	228	UTR5' kdsB	-
RNA25				108033	108111	78	AS pbpC	+
RNA26				108717	108834	117	UTR5' pbpC	-
RNA27 ^a	SmrC14	x (C1)	C1_109477F	109471	109577	106	AS atu0105	+
RNA28 ^a			C1 109596F	109692	110050	358	UTR5' cspA	+
RNA29 ^ª	x	x	C1_110258F ^b	110160	110358	198	IGR	+
RNA30				112398	112488	90	atu0109	-
RNA31 ^a	x; SmrC7, sra03	x (C2)	C1 112676R	112538	112676	138	IGR	-
RNA32	, ,		-	117019	117105	86	IGR	+
RNA33				124118	124196	78	UTR5' <i>slyX</i>	+
RNA34				146980	147151	171	atu0143	+
RNA35 ^a		х		166545	166626	81	IGR	+
RNA36 ^ª		х		182638	182737	99	IGR	+
RNA37				184993	185071	78	atu0182	+
RNA38				188143	188299	156	IGR	-
RNA39				189264	189360	96	UTR5' <i>mepA</i>	-
RNA40				191412	191661	249	atu0188	+
RNA41				200097	200220	123	UTR5' gntK	-
RNA42				205593	206094	501	UTR5' rirA	-
RNA43				212538	212592	54	IGR	+
RNA44 ^a	х			213447	213678	231	IGR	+
RNA45				216444	216609	165	UTR5'	+

							atu0212	
RNA46				217031	217119	88	IGR	+
RNA47				229764	229881	117	IGR	-
RNA48				230832	230910	78	AS <i>atu0227</i> UTR3'	+
RNA49 ^a	х			233025	233094	69	<i>atu0232</i> UTR5'	+
RNA50				235274	235347	73	atu0235	+
RNA51				239508	239622	114	AS <i>atu0240</i> UTR5'	-
RNA52				243920	244064	144	atu0245	+
RNA53				248880	249046	166	IGR	-
RNA54 ^a		х		250839	250953	114	IGR	-
RNA55				251939	252429	490	UTR5' infC	+
RNA56				252651	252774	123	IGR	+
RNA57				264900	264941	41	IGR	-
RNA58				270067	270176	109	AS pecM	-
RNA59				277965	278037	72	AS tspO	+
RNA60				284829	285013	184	UTR5' sun	-
RNA61				289750	289904	154	IGR	-
RNA62				289976	290099	123	AS atu0295	-
RNA63 ^ª			C1_293634F	293648	293732	84	IGR UTR5'	+
RNA64				295103	295202	99	atu00302	+
RNA65 ^a			C1_298719R	298588	298789	201	AS ugpA	-
RNA66				315454	315589	135	IGR	+
RNA67				316387	316747	360	UTR5' rpsT	+
RNA68 ^a	х			316873	316948	75	IGR	+
RNA69 ^a	х			317345	317570	225	UTR5' dnaA	+
RNA70				321863	321989	126	UTR5' <i>hrcA</i> UTR3'	+
RNA71				324320	324554	234	<i>atu0331</i> UTR5'	-
RNA72				324843	324969	126	atu0331	-
RNA73 ^a			C1_326637F	326607	326714	107	AS rpoN	+
RNA74				336685	336874	189	IGR	+
RNA75				348187	348295	108	UTR5' fur	+
RNA76				357763	357856	93	UTR5' trmB	+
RNA77				370442	370526	84	IGR	+
RNA78				370912	371020	108	IGR	+
RNA79 ^ª		х		374910	375004	94	IGR	+
RNA80				380954	381062	108	IGR UTR5'	+
RNA81				389313	389490	177	atu0396	+
RNA82				395599	395722	123	UTR5' scrK	+
RNA83				400009	400090	81	UTR5' <i>pgi</i>	-
RNA84 [°]	х			401920	402109	189	IGR	-
RNA85				407956	407977	21	AS atu0409 UTR5'	-
RNA86				410284	410524	240	<i>atu8120</i> UTR5'	-
KNA87				414725	414839	114	atuU419	+
RNA88				414986	415004	18	AS atu0419	-
KNA89				416344	416405	61 	AS atu0420	-
RNA90°		х	C1_423713R	423431	423722	291	IGR	-
KNA91				427682	427925	243	UTR5' apaG	-
RNA92 [°]	x; RF00521			429351	429429	78	IGR	-
RNA93				429489	429591	102	UTR5' dcd	+
RNA94				433680	433734	54	AS atu0437	-
RNA95 ^a	х			434829	434919	90	IGR	-

RNA96 ^a	х		C1_445306F	445302	445482	180	IGR LITR5'	+
RNA97				458715	458787	72	atu0466	-
RNA98 ^a		x	C1_459703R	459606	459711	105	AS atu0468	-
RNA99				460704	460776	72	UTR5' <i>atu0471</i> UTR5'	-
RNA100				477357	477489	132	atu0490	+
RNA101				477934	478150	216	IGR	+
RNA102				489351	489371	20	AS atu0499	+
RNA103 ^a	х			493052	493121	69	IGR	+
RNA104				503405	503693	288	UTR5' <i>atu0513</i> UTR5'	-
RNA105				504941	505181	240	atu0514	+
RNA106				506397	506504	107	UTR5' cheA	+
RNA107				512068	512179	111	UTR5' <i>fliF</i>	+
RNA108 ^a	x	x		513886	514119	233	IGR	+
RNA109	~	~		521482	521650	168	UTR5' mscl	+
RNA110				523579	523717	138	UTR5' exoB	+
RNA111				529074	529245	171	UTR5'	_
				525071	525215	1/1	UTR5'	
RNA112				529628	529850	222	atu0538	-
RNA113 ^a	х			531847	531970	123	UTR3' <i>fla</i>	-
RNA114 ^a			C1_532642F	532637	532745	108	AS fla	+
RNA115 ^a				534338	534448	110	IGR	-
RNA116				535468	535542	74	IGR	-
RNA117				541820	541895	75	UTR5' <i>flgB</i>	-
RNA118 ^a	x			545504	545612	108	IGR	-
RNA119 ^a	х			549330	549437	107	UTR5' <i>flhB</i>	+
RNA120				552584	552686	102	IGR UTR5'	+
RNA121				558123	558295	172	atu0573	+
RNA122				558833	558989	156	UTR5' <i>flgE</i>	+
RNA123				567950	568037	87	IGR	+
RNA124				574184	574295	111	UTR5' aglR	-
RNA125				574439	574565	126	IGR	+
RNA126				575820	576017	197	UTR3' <i>aglE</i> UTR5'	+
RNA127				581144	581240	96	<i>atu8133</i> UTR5'	+
RNA128				590303	590364	61	atu0603	+
RNA129				592856	592928	72	AS atu0605	-
RNA130				601133	601194	61	AS atu0613	+
RNA131				614489	614612	123	UTR5' adhP	+
RNA132				624985	625064	79	IGR UTR5'	+
RNA133 ^a		х		649259	649427	168	<i>atu8135</i> UTR5'	+
RNA134				654152	654269	117	atu0659	-
RNA135				660168	660300	132	UTR5' glcD	+
RNA136 ^a		x		661697	661796	99	IGR	+
RNA137 ^ª	х	х	C1_665234R	665129	665216	87	IGR UTR5'	-
RNA138 ^a	х			670496	670772	276	atu0675	-
RNA139				675765	675852	87	IGR UTR5'	-
RNA140 ^a		x		683819	684038	219	atu0687	+
RNA141 ^a				693246	693510	264	IGR	+
RNA142				695649	695736	87	UTR5' <i>lpxK</i> UTR5'	+
RNA143				696986	697137	151	atu0698	-

						UTR5'	
RNA144			701028	701145	117	<i>atu0701</i> UTR5'	+
RNA145			709209	709293	84	atu0709	+
RNA146			715356	715578	222	UTR5' atpl	+
RNA147			716547	716979	432	UTR5' atpC	+
RNA148			717917	717941	24	AS <i>atpC</i> UTR5'	-
RNA149			718330	718450	120	<i>atu0718</i> UTR5'	+
RNA150			728586	728712	126	<i>atu0728</i> UTR5'	+
RNA151			729189	729648	459	<i>atu0730</i> UTR5'	+
RNA152			756846	756990	144	atu0759	-
RNA153			761541	761640	99	UTR5'tidD	-
RNA154 ^a			762536	762591	55	IGR	+
RNA155			765315	765471	156	IGR	+
RNA156			765525	765675	150	UTR5' ctaB	+
RNA157			777192	777309	117	IGR UTR5'	+
RNA158			777924	778095	171	atu0782	+
RNA159			790722	790878	156	UTR5' ccrM	+
RNA160 ^a	х		791951	792033	82	IGR	+
RNA161 ^ª		C1 792114F	792090	792180	90	AS atu0795	+
RNA162 ^a	x	_	800042	800087	45	IGR	-
RNA163			800081	800120	39	IGR	+
RNA164 ^a	x		801383	801557	174	UTR5' nndK	+
RNA165	~		806978	807059	81	AS atu0805	-
RNA166			813196	813216	20	AS atu0814	_
$RNA167^{a}$	v	C1 8135468 ^b	813/77	813564	87	IGR	_
RNA169	*	C1_813340K	813477	013304	87		-
RNA169 ^a	RE01867: sra25		822234	877888	180	UTR5'	-
RNA170			825736	825973	237	UTR5' atu0826	-
RNA171			826032	826347	315	UTR5'	_
RNA172			826362	826491	129	IGR	_
RNA173			828777	828885	108	IGR	+
RNA174			828927	828987	60	UTR5' betA	+
RNA175			831069	831219	150	atu8138	-
RNA176			835428	835470	42	IGR UTR5'	-
RNA177			836694	836832	138	atu0837	-
RNA178			846114	846213	99	IGR UTR5'	-
RNA179			847350	847431	81	atu0845	+
RNA180 ^a	x	C1_852289R	852135	852291	156	IGR UTR5'	-
RNA181 ^ª	x		860163	860286	123	Atu0862	+
RNA182			872799	872895	96	UTR5' <i>sodBI</i> UTR5'	+
RNA183			876069	876231	162	atu0881	+
RNA184 ^ª	x; RF00174		877638	877773	135	IGR	+
RNA185			891927	892017	90	IGR UTR5'	+
RNA186			895833	896031	198	atu0904	+
RNA187			896865	896904	39	IGR	-
RNA188			905031	905337	306	IGR	+
						UTR5'	

RNA190				907752	907821	69	IGR	+
RNA191				908310	908445	135	atu0920	-
RNA192				012883	012070	96	IGR	_
RNA192				912003	912979	30		
KNA193				913194	913329	165	UTR5'	+
RNA194				922585	922732	147	<i>atu0932</i> UTR5'	+
RNA195				929984	930063	79	atu0942	+
RNA196				937338	937425	87	IGR	-
RNA197				937779	937887	108	AS atu0949	-
RNA198 ^a		х	C1_942251R ^b	942027	942258	231	IGR UTR5'	-
RNA199				958452	958534	82	<i>atu8142</i> UTR5'	+
RNA200				958791	958898	107	atu0970	+
RNA201				960939	961128	189	IGR	+
RNA202				961599	961686	87	IGR	+
RNA203				967722	967878	156	IGR LITR5'	+
RNA204				976791	976937	146	atu0982	-
RNA205 ^a	v			977391	977474	83	IGR	_
	~		C1 0020200 b	001747	002025	200		
RNA206	X	x	C1_982030R	981/4/	982035	288	AS atuo986	-
RNA207				989397	989706	309	UTR5' <i>atu994</i> UTR5'	+
RNA208				991611	991786	1/5	UTR5'	+
RNA209				993627	993699	72	atu0997	+
RNA210°	х			996489	996651	162	IGR	-
RNA211				1005552	1005708	156	IGR	+
RNA212 ^a	х			1005981	1006077	96	IGR	+
RNA213				1008150	1008267	117	IGR	+
RNA214				1012209	1012338	129	AS <i>atu1019</i> UTR5'	+
RNA215				1013867	1013958	91	atu1021	+
RNA216 ^a		х		1021920	1022037	117	IGR UTR5'	+
RNA217				1025004	1025184	180	atu1031	+
RNA218				1026522	1026729	207	IGR	+
RNA219				1030621	1030677	56	UTR5' <i>recO</i>	+
RNA220				1032819	1032900	81	<i>atu1041</i> UTR5'	+
RNA221				1037838	1038009	171	<i>atu1044</i> UTR5'	-
RNA222				1044645	1044702	57	<i>atu1050</i> UTR5'	-
RNA223				1047792	1047945	153	atu1055	-
RNA224 ^ª		x	C1_1052093R	1051824	1051930	106	IGR	-
RNA225 ^a			C1 1052093R	1051965	1052115	150	IGR	-
RNA226				1074288	1074363	75	AS atu1083	-
RNA227				1076154	1076235	81	UTR5' dnaC	_
	X			1070024	1070077	E2		
	X			10/9824	T013811	53		+
KNA229 °	sra29	х		1082478	1082703	225	AS atu1092	-
RNA230 ^a	х			1085961	1086363	402	UTR5' acpP	+
RNA231 ^a	x			1094598	1094799	201	IGR UTR5'	-
RNA232				1101912	1102056	144	atu1112	+
RNA233				1113363	1113552	189	UTR5' ndk	+
RNA234				1113918	1114104	186	AS <i>atu1123</i> UTR5'	+
RNA235				1119300	1119645	345	atu1129	-

RNA236 ^ª	х			1122894	1122942	48	IGR	+
RNA237				1130562	1130727	165	UTR5' <i>purM</i>	-
RNA238				1132896	1133013	117	UTR5' <i>ppk</i> UTR5'	+
RNA239				1137774	1137903	129	atu8089	-
RNA240				1138437	1138539	102	IGR	+
RNA241				1151619	1151697	78	IGR	-
RNA242				1154327	1154390	63	AS ate	+
RNA243				1155570	1155666	96	IGR	-
RNA244				1156647	1156719	72	UTR5' <i>hemB</i> UTR5'	-
RNA245				1169919	1170057	138	atu1175	-
RNA246				1173735	1173846	111	UTR5' <i>ihfA</i>	+
RNA247 ^a	x		C1 1175776F	1175775	1175907	132	IGR	+
RNA248 ^a	x		C1 1182639R	1182576	1182651	75	IGR	_
RNA249	X		C1_11020351	1200333	1200477	144	IGR	+
RNA250				1208601	1208697	96	IGR	_
RNA251				1214852	1214933	81	AS atu1221	_
RNA252 ^a			C1 12155/11	1215/38	1215573	135	AS atu1222	_
		Y	CI_1215541N	1215450	1213075	21		-
		X		1210970	1217057	81	IGK	+
RINA254				121/122	121/10/	05	AS ULUIZZ4	Ŧ
RINA255				1218451	1218532	81	IGR	-
RINA256				1225830	1225974	144		-
RNA257				1231947	1232040	93		+
RNA258				1234848	1234926	/8	UTR5 argc	-
RNA259				1236900	1237095	195	IGR	-
RNA260				1246200	1246317	117	UTR5' <i>CIPP2</i>	+
RNA261 °		х		1248609	1248783	174	AS atu1260	+
RNA262				1248792	1248885	93	IGR	+
RNA263				1251513	1251654	141	UTR5' hupA	+
RNA264 ^a	х			1251933	1252074	141	IGR	+
RNA265 ^a			C1_1254774R	1254601	1254778	177	AS atu1264	-
RNA266				1254831	1254922	91	IGR	+
RNA267				1255533	1255575	42	IGR	+
RNA268 ^a	х			1255886	1256006	120	IGR	-
RNA269 ^a	х			1256139	1256244	105	UTR5' <i>nuoA</i>	+
RNA270				1262048	1262279	231	IGR	+
RNA271				1265906	1266029	123	IGR	+
RNA272				1270397	1270469	72	AS nuoM	-
RNA273 ^a	x; SmrC9, sra32	x (C5)	C1 1275443R	1275350	1275443	93	IGR	-
RNA274		. ,	_	1279433	1279522	89	AS atu1291	+
RNA275 ^a				1279489	1279713	224	UTR5' dnaE	+
RNA276				1286324	1286462	138	UTR5' divK	+
RNA277 ^a	v. cr233	x (C6)	C1 1288586F	1288580	1288601	111		_
	λ, 51α55	X (CO)	C1_1288380F	1206571	1206704	172	LITES' tonA	т
				1200071	1290794	57		-
				1206270	1206556	177		
RNA280		x		1306379	1306556	1//	IGK	+
RNA281				1307588	1307699	111	UTR5'	-
KNA282				1311458	1311538	80	atu1322	+
RNA283				1313024	1313048	24	AS gatB UTR5'	-
RNA284				1315190	1315373	183	atu1327	+
RNA285				1318949	1319177	228	UTR5' aroQ1	-
RNA286				1322936	1323074	138	UTR5' fabG	+
RNA287				1326863	1326983	120	IGR	-
RNA288				1327224	1327419	195	UTR5' amiA	+
RNA289				1328762	1328837	75	IGR	+

RNA290				1331462	1331810	348	UTR5' <i>prfB</i> UTR5'	+
RNA291				1335765	1335933	168	atu1346	+
RNA292				1343857	1343905	48	AS atu1348	+
				10.0007	10.0000	10	UTR5'	
RNA293				1344254	1344509	255	atu1348	-
RNA294 ^a	х	x (C7)	C1 1345805R	1345661	1345814	153	IGR	-
RNA295 ^a			 C11352134F	1352156	1352243	87	AS atu1358	+
1111255			01_13321341	1552150	1332243	07	UTR5'	•
RNA296 [°]	x			1355027	1355174	147	atu1362	+
RNA297 ^a			C1 1357826R	1357763	1357862	99	IGR	-
RNA298				1357892	1358045	153	IGR	+
RNA299				1366703	1366778	75	AS atu1372	+
RNA300				1367882	1368017	135	IGR	+
				1260052	1260222	190		_
RNA301				1309032	1309232	130		т
RNA302				1374440	1374566	126		+
RNA303				1386398	1386539	141	UTR5 gitA	-
RNA304				1398065	1398170	105	AS atu1403	+
RNA305				1400282	1400456	174	atu1405	_
RNA306				1415402	1415645	243	LITR5' vanR	+
				1422610	1/122001	291	UTR5' ndhA	_
KNA307				1423010	1423991	301	UTR5'	т
RNA308				1429727	1430096	369	atu1435	+
RNA309				1431374	1431519	145	UTR3' <i>lipA</i> UTR5'	+
RNA310 ^ª	х			1433525	1433687	162	atu1438	-
RNA311				1440788	1440895	107	IGR	+
RNA312				1445975	1446161	186	IGR	+
RNA313 ^ª			C1 1449058F	1449047	1449104	57	AS atu1453	+
RNA314			_	1453904	1454048	144	UTR5' mvrA	+
RNA315 ^a	RE00504		C1 1462469B	1462286	1462472	186	IGR	_
RNA315 ^a	RF00504		C1_1462469R	1462286	1462472	186	IGR	-
RNA315 ^a RNA316	RF00504		C1_1462469R	1462286 1462681	1462472 1462807	186 126	IGR AS atu1466	-
RNA315 ^a RNA316 RNA317	RF00504		C1_1462469R	1462286 1462681 1463011	1462472 1462807 1463077	186 126 66	IGR AS atu1466 IGR	- - +
RNA315 ^ª RNA316 RNA317 RNA318	RF00504		C1_1462469R	1462286 1462681 1463011 1465615	1462472 1462807 1463077 1465699	186 126 66 84	IGR AS atu1466 IGR UTRS' rluC	- - + +
RNA315 ^a RNA316 RNA317 RNA318 RNA319 ^a	RF00504		C1_1462469R C1_1474436R	1462286 1462681 1463011 1465615 1474297	1462472 1462807 1463077 1465699 1474447	186 126 66 84 150	IGR AS atu1466 IGR UTR5' rluC IGR	- + +
RNA315 [°] RNA316 RNA317 RNA318 RNA319 [°] RNA320	RF00504		C1_1462469R C1_1474436R	1462286 1462681 1463011 1465615 1474297 1490830	1462472 1462807 1463077 1465699 1474447 1491130	186 126 66 84 150 300	IGR AS atu1466 IGR UTRS' rluC IGR UTRS' ripA	- + + -
RNA315 ^a RNA316 RNA317 RNA318 RNA319 ^a RNA320 RNA321	RF00504		C1_1462469R C1_1474436R	1462286 1462681 1463011 1465615 1474297 1490830 1498654	1462472 1462807 1463077 1465699 1474447 1491130 1498783	186 126 66 84 150 300 129	IGR AS atu1466 IGR UTR5' rluC IGR UTR5' ripA UTR5' gyrA UTR5'	- + + -
RNA315 ^a RNA316 RNA317 RNA318 RNA319 ^a RNA320 RNA321 RNA322	RF00504		C1_1462469R C1_1474436R	1462286 1462681 1463011 1465615 1474297 1490830 1498654 1501432	1462472 1462807 1465699 1474447 1491130 1498783	186 126 66 84 150 300 129 120	IGR AS atu1466 IGR UTRS' rluC IGR UTRS' ripA UTRS' gyrA UTRS' atu1511	- + - -
RNA315 ^a RNA316 RNA317 RNA318 RNA319 ^a RNA320 RNA321 RNA322 RNA323	RF00504		C1_1462469R C1_1474436R	1462286 1462681 1463011 1465615 1474297 1490830 1498654 1501432 1502035	1462472 1462807 1465699 1474447 1491130 1498783 1501552 1502137	186 126 66 84 150 300 129 120 102	IGR AS atu1466 IGR UTR5' rluC IGR UTR5' ripA UTR5' gyrA UTR5' atu1511 UTR5' ssb	- + - -
RNA315 ^a RNA316 RNA317 RNA318 RNA319 ^a RNA320 RNA321 RNA322 RNA323 RNA324 ^a	RF00504 X		C1_1462469R C1_1474436R	1462286 1463011 1465615 1474297 1490830 1498654 1501432 1502035 1512856	1462472 1463077 1465699 1474447 1491130 1498783 1501552 1502137 1512928	186 126 66 84 150 300 129 120 102 72	IGR AS atu1466 IGR UTRS' rluC IGR UTRS' ripA UTRS' gyrA UTRS' atu1511 UTRS' ssb IGR	- + + - -
RNA315 ^a RNA316 RNA317 RNA318 RNA319 ^a RNA320 RNA321 RNA322 RNA323 RNA324 ^a RNA325	RF00504 X		C1_1462469R C1_1474436R	1462286 1463011 1465615 1474297 1490830 1498654 1501432 1502035 1512856 1518964	1462472 1463077 1465699 1474447 1491130 1498783 1501552 1502137 1512928 1519069	186 126 66 84 150 300 129 120 102 72 105	IGR AS atu1466 IGR UTRS' rluC IGR UTRS' ripA UTRS' gyrA UTRS' gyrA UTRS' ssb IGR UTRS' gntZ	- + +
RNA315 ^a RNA316 RNA317 RNA318 RNA319 ^a RNA320 RNA321 RNA322 RNA323 RNA324 ^a RNA325 RNA326	RF00504 x		C1_1462469R C1_1474436R	1462286 1463011 1465615 1474297 1490830 1498654 1502035 1512856 1518964 1523665	1462472 1463077 1465699 1474447 1491130 1498783 1501552 1502137 1512928 1519069 1523740	186 126 66 84 150 300 129 120 102 72 105 75	IGR AS atu1466 IGR UTRS' rluC IGR UTRS' ripA UTRS' gyrA UTRS' gyrA UTRS' ssb IGR UTRS' gntZ UTRS' fixG	- + +
RNA315 ^a RNA316 RNA317 RNA318 RNA319 ^a RNA320 RNA321 RNA322 RNA323 RNA324 ^a RNA325 RNA326 RNA327	RF00504 x		C1_1462469R C1_1474436R	1462286 1462681 1463011 1465615 1474297 1490830 1498654 1502035 1512856 1512856 1518964 1523665 1528489	1462472 1463077 1465699 1474447 1491130 1498783 1501552 1502137 1512928 1519069 1523740 1528600	186 126 66 84 150 300 129 120 102 72 105 75 111	IGR AS atu1466 IGR UTR5' rluC IGR UTR5' ripA UTR5' ripA UTR5' gyrA UTR5' ssb IGR UTR5' gntZ UTR5' fixG IGR	- + +
RNA315 ^a RNA316 RNA317 RNA319 ^a RNA320 RNA321 RNA322 RNA323 RNA324 ^a RNA325 RNA326 RNA327 RNA328	RF00504 x		C1_1462469R C1_1474436R	1462286 1462681 1465615 1474297 1490830 1498654 1501432 1502035 1512856 1518964 1523665 1528489 1528822	1462472 1463077 1465699 1474447 1491130 1498783 1501552 1502137 1512928 1519069 1523740 1528600 1528870	186 126 66 84 150 300 129 120 102 72 105 75 111 48	IGR AS atu1466 IGR UTR5' rluC IGR UTR5' ripA UTR5' gyrA UTR5' gyrA UTR5' ssb IGR UTR5' gntZ UTR5' fixG IGR IGR	+ + +
RNA315 ^a RNA316 RNA317 RNA318 RNA320 RNA320 RNA321 RNA322 RNA323 RNA324 ^a RNA325 RNA326 RNA327 RNA328 RNA329 ^a	RF00504 x	x	C1_1462469R C1_1474436R C1_1533961R	1462286 1462681 1463011 1465615 1474297 1490830 1498654 1502035 1512856 1518964 1523665 1528489 1528822 1533772	1462472 1463077 1465699 1474447 1491130 1498783 1501552 1502137 1512928 1519069 1523740 1528600 1528870 1533961	186 126 66 84 150 300 129 120 102 72 105 75 111 48 189	IGR AS atu1466 IGR UTRS' rluC IGR UTRS' ripA UTRS' gyrA UTRS' gyrA UTRS' ssb IGR UTRS' fixG IGR IGR	+ + + -
RNA315 ^a RNA316 RNA317 RNA318 RNA319 ^a RNA320 RNA321 RNA322 RNA323 RNA324 ^a RNA325 RNA325 RNA326 RNA327 RNA328 RNA329 ^a	RF00504 ×	x	C1_1462469R C1_1474436R C1_1533961R	1462286 1462681 1463011 1465615 1474297 1490830 1498654 1502035 1512856 1518964 1523665 1528489 1528822 1533772	1462472 1463077 1465699 1474447 1491130 1498783 1501552 1502137 1512928 1519069 1523740 1528600 1528870 1533961	186 126 66 84 150 300 129 120 102 72 105 75 111 48 189	IGR AS atu1466 IGR UTRS' rluC IGR UTRS' ripA UTRS' gyrA UTRS' gyrA UTRS' ssb IGR UTRS' fixG IGR IGR IGR IGR UTRS' fixG	- + + + -
RNA315 ^a RNA316 RNA317 RNA318 RNA319 ^a RNA320 RNA321 RNA322 RNA323 RNA324 ^a RNA325 RNA326 RNA326 RNA327 RNA328 RNA329 ^a RNA329 ^a	RF00504 x	X	C1_1462469R C1_1474436R C1_1533961R	1462286 1463011 1465615 1474297 1490830 1498654 1502035 1512856 1518964 1523665 1528489 1528822 1533772	1462472 1463077 1465699 1474447 1491130 1498783 1501552 1502137 1512928 1519069 1523740 1528600 1528870 1533961	186 126 66 84 150 300 129 120 102 72 105 75 111 48 189 62	IGR AS atu1466 IGR UTR5' rluC IGR UTR5' ripA UTR5' gyrA UTR5' gyrA UTR5' ssb IGR UTR5' gntZ UTR5' fixG IGR IGR IGR IGR UTR5' atu1546 AS atu156	- + + +
RNA315 ^a RNA316 RNA317 RNA319 ^a RNA320 RNA321 RNA322 RNA323 RNA325 RNA326 RNA325 RNA326 RNA327 RNA328 RNA329 ^a RNA329 ^a	RF00504 x	x	C1_1462469R C1_1474436R C1_1533961R	1462286 1462681 1463011 1465615 1474297 1490830 1498654 1501432 1502035 1512856 1518964 1523665 1528429 1528822 1533772	1462472 1462807 1465699 147447 1491130 1498783 1501552 1502137 1512928 1519069 1523740 1528700 1528870 1533961	186 126 66 84 150 300 129 120 102 72 105 75 111 48 189 62 51	IGR AS atu1466 IGR UTRS' rluC IGR UTRS' ripA UTRS' gyrA UTRS' gyrA UTRS' ssb IGR UTRS' gntZ UTRS' fixG IGR IGR IGR IGR IGR UTRS' atu1546 AS atu1566	- + + + +
RNA315 ^a RNA316 RNA317 RNA318 RNA319 ^a RNA320 RNA321 RNA322 RNA323 RNA325 RNA326 RNA326 RNA327 RNA326 RNA327 RNA328 RNA329 ^a RNA329 ^a	RF00504 x	x	C1_1462469R C1_1474436R C1_1533961R	1462286 1462681 1463011 1465615 1474297 1490830 1498654 1501432 1502035 1512856 1518964 1523665 1528489 1528822 1533772 1536526 1553317	1462472 1462807 1465699 147447 1491130 1498783 1501552 1502137 1512928 1519069 1523740 1528600 1528870 1533961 1536588 1553688	186 126 66 84 150 300 129 120 102 72 105 75 111 48 189 62 51 62 62	IGR AS atu1466 IGR UTR5' rluC IGR UTR5' ripA UTR5' gyrA UTR5' gyrA UTR5' ssb IGR UTR5' gntZ UTR5' fixG IGR IGR IGR IGR IGR UTR5' atu1566 AS atu1566	+ + + + -
RNA315 ^a RNA316 RNA317 RNA318 RNA319 ^a RNA320 RNA321 RNA322 RNA323 RNA324 ^a RNA325 RNA326 RNA326 RNA327 RNA328 RNA329 ^a RNA330 RNA331 RNA331 RNA332 RNA333	RF00504 x	X	C1_1462469R C1_1474436R C1_1533961R	1462286 1463011 1465615 1474297 1490830 1498654 1502035 1512856 1518964 1523665 1528489 1528822 1533772 1536526 1553317 1556576 1558153	1462472 1463077 1465699 1474447 1491130 1498783 1501552 1502137 1512928 1519069 1523740 1528600 1528870 1533961 1533961 1553368 1553368 155638	186 126 66 84 150 300 129 120 102 72 105 75 111 48 189 62 51 62 99	IGR AS atu1466 IGR UTRS' rluC IGR UTRS' ripA UTRS' gyrA UTRS' gyrA UTRS' ssb IGR UTRS' gntZ UTRS' fixG IGR IGR IGR IGR IGR UTRS' fixG AS atu1546 AS atu1566 AS cyc UTRS' cysE UTRS' cysE	+ + + + +
RNA315 ^a RNA316 RNA317 RNA318 RNA319 ^a RNA320 RNA321 RNA322 RNA323 RNA324 ^a RNA325 RNA326 RNA326 RNA327 RNA328 RNA329 ^a RNA330 RNA331 RNA331 RNA333	RF00504 x	x	C1_1462469R C1_1474436R C1_1533961R	1462286 1462681 1463011 1465615 1474297 1490830 1498654 1501432 1502035 1512856 1518964 1523665 1528429 152822 1533772 1536526 155317 1556576 1558153	1462472 1462807 1465699 147447 1491130 1498783 1501552 1502137 1512928 1519069 1523740 1528600 1528870 1533961 1536588 155368 1556638 1556638	186 126 66 84 150 300 129 120 102 72 105 75 111 48 189 62 51 62 99 219	IGR AS atu1466 IGR UTRS' rluC IGR UTRS' ripA UTRS' gyrA UTRS' gyrA UTRS' ssb IGR UTRS' gntZ UTRS' fixG IGR IGR IGR IGR IGR UTRS' fixG AS atu1566 AS atu1566 AS cyc UTRS' cysE UTRS' cysE UTRS' atu1573	+ + + + +
RNA315 ^a RNA316 RNA317 RNA318 RNA319 ^a RNA320 RNA321 RNA322 RNA323 RNA324 ^a RNA325 RNA326 RNA327 RNA326 RNA327 RNA328 RNA329 ^a RNA330 RNA331 RNA331 RNA333 RNA334 RNA335	RF00504	x	C1_1462469R C1_1474436R C1_1533961R	1462286 1462681 1463011 1465615 1474297 1490830 1498654 1502035 1512856 1518964 1523665 1528489 1528822 1533772 1536526 1553177 1556576 1558153 1559215 1561996	1462472 1463077 1465699 1474447 1491130 1498783 1501552 1502137 1512928 1519069 1523740 1528600 1528870 1533961 15336588 1553368 155638 155638 155638	186 126 66 84 150 300 129 120 102 72 105 75 111 48 189 62 51 62 99 219 39	IGR AS atu1466 IGR UTRS' rluC IGR UTRS' ripA UTRS' gyrA UTRS' gyrA UTRS' ssb IGR UTRS' gntZ UTRS' fixG IGR IGR IGR IGR IGR UTRS' fixG AS atu1546 AS atu1546 AS atu1546 AS atu1546 AS cyc UTRS' cysE UTRS' cysE UTRS' metC UTRS' metC UTRS' metC	+ + + + + -
RNA315 ^a RNA316 RNA317 RNA318 RNA319 ^a RNA320 RNA321 RNA322 RNA323 RNA324 ^a RNA325 RNA326 RNA326 RNA327 RNA328 RNA329 ^a RNA329 ^a RNA330 RNA331 RNA331 RNA333 RNA334 RNA335 RNA336	RF00504	x	C1_1462469R C1_1474436R C1_1533961R	1462286 1462681 1463011 1465615 1474297 1490830 1498654 1502035 1512856 1518964 1523665 1528489 1528422 1533772 1536526 1553317 1556576 1558153 1559215 1561996	1462472 1463077 1465699 1474447 1491130 1498783 1501552 1502137 1512928 151969 1523740 1528600 1528870 1533961 1533961 1553368 155638 155638 155638 155638	186 126 66 84 150 300 129 120 102 72 105 75 111 48 189 62 51 62 99 219 39 225	IGR AS atu1466 IGR UTRS' rluC IGR UTRS' ripA UTRS' gyrA UTRS' gyrA UTRS' ssb IGR UTRS' gntZ UTRS' fixG IGR IGR IGR IGR UTRS' fixG IGR UTRS' fixG AS atu1546 AS atu1546 AS atu1546 AS cyc UTRS' cysE UTRS' cysE UTRS' metC UTRS' metC UTRS' metC UTRS' metC	+ + + + - + - + -
RNA315 ^a RNA316 RNA317 RNA318 RNA319 ^a RNA320 RNA321 RNA322 RNA323 RNA324 ^a RNA325 RNA326 RNA326 RNA329 ^a RNA330 RNA331 RNA331 RNA333 RNA333 RNA334 RNA335 RNA336 RNA337	RF00504	x	C1_1462469R C1_1474436R C1_1533961R	1462286 1463011 1465615 1474297 1490830 1498654 1502035 1512856 1512856 1518964 1523657 1528489 1528422 1533772 1536526 1553317 1556576 1558153 1559215 1561996	1462472 1463077 1465699 1474447 1491130 1498783 1501552 1502137 1512928 1519069 1523740 1528600 1528870 1528870 1533961 1533961 1556388 155638 155638 155638 1559434 1562512 1563647	186 126 66 84 150 300 129 120 102 72 105 75 111 48 189 62 51 62 99 219 39 225 74	IGR AS atu1466 IGR UTRS' rluC IGR UTRS' ripA UTRS' gyrA UTRS' gyrA UTRS' gyrA UTRS' ghtZ UTRS' fixG IGR IGR IGR IGR IGR UTRS' fixG AS atu1566 AS cyc UTRS' cysE UTRS' cysE UTRS' atu1573 UTRS' metC UTRS' metC UTRS' atu1577 AS atu1578	+ + + + - + - + -
RNA315 ^a RNA316 RNA317 RNA318 RNA319 ^a RNA320 RNA321 RNA322 RNA323 RNA324 ^a RNA325 RNA326 RNA326 RNA327 RNA328 RNA329 ^a RNA329 ^a RNA330 RNA331 RNA331 RNA333 RNA334 RNA335 RNA336 RNA337 RNA338	RF00504	x	C1_1462469R C1_1474436R C1_1533961R	1462286 1462681 1463011 1465615 1474297 1490830 1498654 1501432 1502035 1512856 1518964 1523665 1528489 1528822 1533772 1536526 1553317 1556576 1558153 1559215 1561996 1562287 1563273	1462472 1463077 1465699 147447 1491130 1498783 1501552 1502137 1512928 1519069 1523740 1528600 152870 153368 155638 155638 155638 155638 155638 155638 1558434 1562312 1562512 1563647 1567495	186 126 66 84 150 300 129 120 102 72 105 75 111 48 189 62 51 62 99 219 39 225 74 75	IGR AS atu1466 IGR UTRS' rluC IGR UTRS' ripA UTRS' gyrA UTRS' gyrA UTRS' gyrA UTRS' gntZ UTRS' gntZ UTRS' fixG IGR IGR IGR IGR UTRS' fixG AS atu1566 AS cyc UTRS' cysE UTRS' cysE UTRS' cysE UTRS' atu1573 UTRS' metC UTRS' metC UTRS' atu1577 AS atu1578 IGR	+ + + + - + - + -
RNA315 ^a RNA316 RNA317 RNA318 RNA319 ^a RNA320 RNA321 RNA322 RNA323 RNA324 ^a RNA325 RNA326 RNA326 RNA327 RNA326 RNA327 RNA328 RNA329 ^a RNA330 RNA331 RNA331 RNA333 RNA334 RNA335 RNA336 RNA337 RNA338 RNA339	RF00504	x	C1_1462469R C1_1474436R C1_1533961R	1462286 1462681 1463011 1465615 1474297 1490830 1498654 1502035 1512856 1518964 1523665 1528489 1528822 1533772 1536576 155317 1556576 1558153 1559215 1561996 1562287 1563573 1564287	1462472 1463077 1465699 147447 1491130 1498783 1501552 1502137 1512928 1519069 1523740 1528700 152870 1528870 153368 155638 155638 155638 155638 1558252 1559434 1562312 1563647 1567495 1567495	186 126 66 84 150 300 129 120 102 72 105 75 111 48 189 62 51 62 99 219 39 225 74 75 117	IGR AS atu1466 IGR UTRS' rluC IGR UTRS' ripA UTRS' gyrA UTRS' gyrA UTRS' gyrA UTRS' gyrA UTRS' ghtZ UTRS' fixG IGR IGR IGR IGR IGR UTRS' fixG IGR UTRS' fixG AS atu1566 AS cyc UTRS' cysE UTRS' cysE UTRS' cysE UTRS' cysE UTRS' metC UTRS' atu1573 UTRS' metC UTRS' atu1577 AS atu1578 IGR	+ + + + - + - + -

RNA341				1573073	1573156	83	IGR	-
RNA342 ^ª		x		1580758	1580878	120	UTR5' fabF	-
RNA343				1581811	1581898	87	UTR5' acpXL	-
RNA344				1583431	1583599	168	UTR5' fnrN	+
RNA345				1585255	1585330	75	UTR5' phbC	+
RNA346				1606837	1607044	207	UTR5' gyrB	+
RNA347				1610614	1610686	72	atu1624	+
RNA348 ^a	х			1612690	1612825	135	IGR	+
RNA349 ^a	х			1617538	1617628	90	IGR UTR5'	-
RNA350				1620190	1620304	114	atu1632	-
RNA351				1623013	1623082	69	IGR	-
RNA352				1623662	1623688	26	AS atu1635	-
RNA353 ^a		x	C1 1625426F	1625422	1625562	140	IGR	+
RNA354 ^a	RE001068			1627108	1627180	72	IGR	+
RNA355	11 001000			1631620	1631746	126	UTR5' atu1647	
RNA356				163/211	163/205	84	AS atu1650	+
				1034211	1625046	62		т
	2500474			1055665	1055940	05		Ŧ
RNA358	RF00174		C1_1640604R	1640437	1640659	222	IGR	-
RNA359				1643371	1643503	132	IGR UTR5'	-
RNA360				1646187	1646301	114	<i>atu1660</i> UTR5'	+
RNA361				1649431	1649623	192	<i>atu1663</i> UTR5'	+
RNA362 ^a	x			1653691	1653955	264	atu8161	+
RNA363				1658041	1658119	78	IGR	+
RNA364				1666831	1666924	93	IGR	-
RNA365 ^a			C1_1667597F	1667620	1667767	147	AS ppiB	+
RNA366 [°]	x			1679311	1679407	96	UTR5' <i>kefB</i>	-
RNA367				1680105	1680183	78	UTR5' moaA	+
RNA368				1681285	1681423	138	IGR	+
RNA369				1686262	1686526	264	IGR	-
RNA370				1701307	1701376	69	IGR	+
RNA371				1703434	1703611	177	atu1716	-
RNA372				1708264	1708363	99	UTR5' vals	-
RNA373				1709398	1709764	366	IGR	-
		Y	C1 1712210E	1712214	1712202	70	AS atu1727	
		*	CI_1/13219F	1713214	1713232	78 60		т
RNA375				1714060	1/14120	00	IGR	-
RNA376		х		1716982	1717273	291	IGR	+
RNA377				1723966	1724062	96	AS <i>atu1738</i> UTR5'	+
RNA378				1725847	1725979	132	<i>atu1739</i> UTR5'	+
RNA379				1731361	1731433	72	atu1745	+
RNA380				1736921	1736975	54	UTR3' <i>folE</i> UTR5'	+
RNA381				1737547	1737640	93	atu1751	+
RNA382			C1 1745471R	1738564	1738660	96	IGR	-
RNA383 ^a		х	b	1745377	1745479	102	AS <i>atu1760</i> UTR5'	-
RNA384				1747087	1747255	168	atu8019	-
RNA385				1750798	1750888	90	IGR	+
RNA386				1756738	1756879	141	atu8165	+
RNA387 ^a				1757659	1757974	315	atu1773	-
RNA388				1758091	1758157	66	IGR	+

							UTR5'	
RNA389				1766266	1766353	87	atu1779	-
RNA390				1768838	1768973	135	AS atu1782	+
RNA391				1779946	1780087	141	IGR	-
RNA392				1784062	1784137	75	IGR	-
RNA393				1785115	1785334	219	IGR	-
RNA394				1788406	1788523	117	UTR5' cspA	+
RNA395				1789648	1789897	249	UTR5' atu1805	_
RNA396				1789990	1790272	282	UTR5' atu8166	+
RNA397				1794769	1794877	108	IGR	_
				1904501	1004700	207	LITP5' cufP	
				1906012	1004700	105		-
RNA400				1814515	1814608	93	UTR5' atu1832	+
			C1 101024CD	1014313	1010040	99		
KNA401	x	x	CI_1818340K	1818247	1818340	99	UTR5'	
RINA4U2	X			1819399	1819600	201	ulu1837	+
RNA403	х			1840279	1840519	240	UTR5' rpsD	-
RNA404				1840618	1840660	42	IGR	+
RNA405				1848223	1848319	96	UTR5' icdA	+
RNA406 ^a	х		C1_1849636R	1849546	1849633	87	IGR	-
RNA407				1853647	1853758	111	UTR5' alaS	-
RNA408				1854811	1855065	254	UTR5' recA	-
RNA409				1862658	1862743	85	IGR UTR5'	+
RNA410				1863034	1863175	141	atu1883	+
RNA411				1864345	1864393	48	AS ssuA	+
RNA412 ^a	RF00050			1872196	1872364	168	IGR	+
RNA413				1883464	1883518	54	IGR	+
RNA414				1883569	1883638	69	IGR	-
RNA415 ^ª			C1 1885370F	1885350	1885416	66	IGR	+
RNA416			_	1897339	1897564	225	IGR	-
RNA417				1902148	1902469	321	IGR	-
RNA418				1903309	1903495	186	IGR	_
RNA419				1907032	1907209	177	UTR5' rnsF	_
RNA420				1915684	1915900	216	AS rnsl	_
RNA420				1917169	1917307	138	LITR5' tufB	_
				1020264	1020586	222		
				1021407	1021520	122		-
RNA423	X			1931407	1931529	122	IGR	-
RNA425				1935139	1935164	25	AS atu1964	+
RNA426				1937522	1937609	87	UTR5' <i>tufB</i> UTR5'	-
KNA427				1938813	1938894	81	atu8023	+
RNA428				1952356	1952437	81	IGR	+
RNA429°				1956501	1956550	49	IGR	-
RNA430				1956872	1956956	84	AS <i>atu1992</i> UTR5'	+
RNA431				1962033	1962225	192	atu1999	-
RNA432 ^a	х			1977186	1977303	117	IGR UTR5'	-
RNA433				1978839	1978926	87	atu2016	+
RNA434				1985536	1985628	92	UTR5' <i>ndh</i> UTR5'	-
RNA435				1986498	1986699	201	atu2025	+
RNA436 ^a			C1_1990387R	1990342	1990414	72	AS atu2030	-
RNA437 ^a	х	x (C8)		1995912	1995978	66	IGR	-
RNA438				2005188	2005330	142	UTR5' <i>hflK</i>	-
RNA439				2006601	2006679	78	UTR5' thyA	-

							UTR5'	
RNA440				2006939	2007102	163	atu8170	-
	DE019E0	Atu2049;	Atu2049;	2007220	2007560	220		
RNA441	KF01850	UTIKNA	UTIKNA	2007230	2007560	330	IGR	+
RINA442				2007591	2007819	228	IGR	+
KINA443				2014533	2014077	144	UTR5'	-
RNA444				2023860	2023944	84	atu2059	-
RNA445				2030760	2030823	63	AS atu2070	+
RNA446				2032764	2032970	206	AS atu2071	+
RNA447 ^ª	х	x	C1 2032975F	2032971	2033097	126	IGR	+
RNA448			-	2034640	2034747	107	UTR5' <i>sco1</i>	-
RNA449				2045430	2045739	309	atu2084	_
RNA450				2046896	2046953	57	IGR	_
$RNA/51^{a}$	RE001383			2048685	20/18022	227	LITR5' ftc72	_
	RF001383			2048085	2040922	237		-
				2052054	2052524	270		-
KINA455		Atu2109:	Atu2109:	2008914	2009190	202		-
RNA454 ^a	x RF00010	RnaseP	RnaseP	2073546	2073939	393	IGR	-
RNA455				2080290	2080506	216	IGR	+
RNA456 ^a	х	x (C9)	C1 2087200F	2087199	2087358	159	IGR	+
		()					UTR5'	
RNA457				2089107	2089347	240	atu2124	-
RNA458				2095233	2095329	96	IGR	+
RNA459				2097719	2097863	144	AS atu2132	+
RNA460				2111415	2111523	108	AS atu2146	+
RNA461				2123816	2123928	112	UTR5' metH	-
RNA462				2125377	2125473	96	IGR	-
RNA463				2126775	2126883	108	IGR	-
RNA464				2127009	2127207	198	UTR5' omp	+
RNA465 [°]	х			2129571	2129634	63	IGR	-
RNA466				2129994	2130033	39	IGR	-
RNA467				2130606	2130756	150	IGR	+
RNA468			C1 2139202F	2136652	2136916	264	IGR	-
RNA469 ^ª	х	х	b	2139177	2139315	138	IGR	+
RNA470				2141303	2141408	105	UTR3' carA	+
RNA471				2142543	2142609	66	AS atu2171	-
RNA472			C1_2143068R	2143005	2143077	72	AS atu2171	-
RNA473				2148948	2149053	105	atu2175	-
RNA474				2150580	2150622	42	IGR	+
RNA475 ^a		x		2158746	2158848	102	UTR5' areA	+
RNA476				2159295	2159367	72	IGR	_
RNA477				2159299	2159370	71	IGR	-
RNA478				2162582	2162663	81	AS atu2186	_
	x; RF00519; SmrC16.SmrC15.							
RNA479 ^a	sra41	x (AbcR1)	C1_2163256F	2163252	2163354	102	IGR	+
RNA480				2167503	2167704	201	AS atu2191	-
RNA481				2168232	2168358	126	AS atu2194	+
RNA482				2173698	2173941	243	IGR	+
RNA483				2175072	2175135	63	UTR5' atu2203	+
RNA/8/ a		×		2180270	2180/12	42	IGR	_
DU14407		A		2100370	2100412	-12	UTR5'	-
KNA485				2180790	2180991	201	atu8273	-
KNA486				2186206	2186299	93	IGR	+
RNA487 d	х	х		2193513	2193589	76	IGR	-
RNA488 ^ª	х			2195061	2195151	90	UTR5' aldA	-

RNA489				2199501	2199570	69	IGR	+
RNA490				2210538	2210688	150	IGR	-
RNA491				2217744	2217846	102	IGR	+
RNA492				2224212	2224296	84	IGR	+
RNA493				2233698	2233767	69	IGR UTR5'	-
RNA494				2233905	2233959	54	atu2261	+
RNA495				2243589	2243703	114	UTR5' <i>clpP3</i>	-
RNA496				2244990	2245107	117	AS <i>atu2272</i> UTR3'	-
RNA497				2245296	2245371	75	atu2272	+
RNA498 ^a	х			2248458	2248614	156	IGR	-
RNA499				2250342	2250450	108	UTR5' <i>nolR</i> UTR5'	-
RNA500				2253999	2254107	108	atu2281	-
RNA501 ^a	х	х		2261976	2262039	63	IGR	+
RNA502				2270898	2270979	81	UTR5' <i>kdgK</i>	+
		X	C1_2273315F	2222206	222261	70		
		x	61 22000120	2275260	2275504	78		Ŧ
RNA504		X	C1_2290013R	2289969	2290014	45	UTR5'	-
RNA505				2298396	2298489	93	atu2321 UTR5'	+
RNA506				2301960	2302029	69	atu2326 UTR5'	-
RNA507				2310123	2310213	90	atu2332	-
RNA508 °			C1_2310423F	2310417	2310516	99	IGR	+
RNA509				2328827	2328917	90	IGR UTR3'	-
RNA510				2329066	2329426	360	<i>atu2354</i> UTR5'	-
RNA511				2329542	2329671	129	atu2354	-
RNA512				2331753	2331807	54	UTR5' exoQ	-
RNA513				2345460	2345571	111	UTR5' <i>xynA</i> UTR5'	-
RNA514				2348826	2348907	81	atu2375	-
RNA515				2349057	2349165	108	AS atu2376	+
RNA516				2350776	2350872	96	IGR	-
RNA517				2350929	2351073	144	AS atu2378	-
RNA518				2367582	2367792	210	UTR5' sinR	+
RNA519				2369817	2369940	123	UTR5' ureG	-
RNA520				2370024	2370160	136	IGR	-
RNA521				2372202	2372244	42	IGR	-
RNA522				2377026	2377177	151	IGR	-
RNA523 [°]		x		2379211	2379264	53	IGR	+
RNA524				2386401	2386539	138	UTR5' glnA	+
RNA525				2392959	2393163	204	IGR	-
RNA526 ^ª			C1_2393362R	2393264	2393366	102	AS <i>atu2420</i> UTR3'	-
RNA527				2395665	2395767	102	atu8177	-
RNA528				2399169	2399214	45	AS atu2426	+
RNA529				2400429	2400555	126	UTR5' pssB	+
RNA530				2401284	2401386	102	UTR3' <i>pssB</i> UTR3'	+
RNA531				2403603	2403750	147	<i>atu2431</i> UTR3'	+
RNA532				2403744	2403804	60	atu2432	-
RNA533				2404845	2404926	81	IGR	+
RNA534				2405049	2405262	213	UTR5' ctrA	+
RNA535 [°]				2409060	2409165	105	IGR UTR3'	+
RNA536				2409768	2409897	129	atu2439	+

							UTR3'	
RNA537				2414976	2415033	57	atu2444	+
RNA538 ^a			C1_2416295F	2416260	2416359	99	IGR	+
RNA539				2420130	2420286	156	UTR5' <i>purA</i> UTR5'	-
RNA540				2422305	2422431	126	atu2450	-
RNA541 ^a			C1 2437621F	2437596	2437728	132	IGR	+
RNA542			-	2442426	2442477	51	IGR	+
RNA543 ^a		x	C1_2446240R	2446158	2446242	84	AS atu2478	-
							UTR5'	
RNA544				2447700	2447817	117	atu2481	+
RNA545				2448483	2448663	180	IGR	+
RNA546				2450911	2450929	18	AS mexF	-
RNA547				2460435	2460525	90	IGR	+
RNA548				2462568	2462661	93	UTR5' pntA	+
RNA549 ^a	х			2465703	2465745	42	IGR	-
RNA550				2468313	2468337	24	AS atu2499	+
RNA551				2470251	2470311	60	IGR	+
RNA552 ^a			C1_2479274R	2478999	2479302	303	AS atu2510	-
RNA553				2480487	2480547	60	AS atu2512	+
RNA554				2520846	2520930	84	IGR	+
RNA555 [°]			C1 2523992R	2523917	2524010	93	AS atu2551	-
RNA556				2525370	2525433	63	IGR UTR5'	+
RNA557 ^a		х		2530158	2530557	399	atu2557	+
RNA558				2534670	2534766	96	IGR	+
a			C1_2541934R					
RNA559°	x; RF00059		5	2541837	2541933	96	IGR	-
RNA560				2543553	2543656	103	UTR3' acd	+
RNA561				2552736	2552820	84	IGR	+
RNA562				2553117	2553159	42	AS atu2578	-
RNA563				2553285	2553417	132	IGR	+
RNA564				2561166	2561265	99	IGR	+
RNA565				2583145	2583247	102	AS malF	+
RNA566 ^a		х		2593317	2593491	174	IGR	-
RNA567				2593584	2593700	116	IGR UTR5'	+
RNA568				2594112	2594223	111	atu2615	+
RNA569				2600355	2600514	159	UTR5' cheW	+
RNA570				2600937	2600985	48	UTR3' cheW	+
RNA571				2603850	2603970	120	UTR5' fdhD	-
RNA572 ^a	Sm270			2609195	2609461	266	UTR5' atpH	-
RNA573				2622396	2622537	141	UTR5' sucA	-
RNA574				2625654	2625822	168	UTR5' mdh	-
RNA575				2631370	2631665	295	UTR5' sdhC	-
RNA576 ^ª				2659185	2659455	270	IGR UTR5'	-
RNA577				2659871	2659932	61	atu2675	-
RNA578 [°]	RF00520		C1_2667196F	2665545	2665713	168	IGR	+
RNA579 ^a		x (C10)	D	2667189	2667282	93	IGR	+
RNA580				2667258	2667300	42	IGR	-
RNA581				2670969	2671095	126	IGR	-
RNA582				2671137	2671203	66	UTR5' ccmA	+
RNA583				2686611	2686854	243	UTR5' <i>rplS</i> UTR3'	+
RNA584 [°]	х			2691852	2691963	111	atu2708	-
RNA585 ^a	RF00059			2700240	2700342	102	IGR	-
RNA586 ^a	RF00521			2703456	2703528	72	IGR	+

RNA587 ^a			C1_2707755F	2707740	2707821	81	AS ompA	+
RNA588				2725239	2725302	63	AS <i>chvB</i> UTR5'	+
RNA589				2735538	2735685	147	atu2739	+
RNA590				2737047	2737179	132	IGR	+
RNA591				2737986	2738055	69	IGR	+
RNA592				2744409	2744517	108	IGR UTR5'	+
RNA593				2745867	2745939	72	<i>atu2747</i> UTR5'	-
RNA594				2755992	2756109	117	<i>atu2755</i> UTR5'	-
RNA595				2761945	2762022	77	atu2760	+
RNA596				2764827	2764989	162	UTR5' mtrA	-
RNA597 ^ª		х		2766237	2766463	226	IGR UTR5'	-
RNA598				2773434	2773692	258	atu2768	-
RNA599 ^a	х	х		2776245	2776329	84	IGR	+
RNA600 ^ª	x			2788551	2788623	72	IGR UTR5'	-
RNA601				2791230	2791368	138	atu2789	+
RNA602 ^a	RF00174			2809917	2810175	258	IGR	-
RNA603				2813247	2813325	78	AS cobD	+
RNA604				2832177	2832294	117	UTR5' htpX	-

Linear

RNA605				12214	12370	156	IGR	-
RNA606				20228	20330	102	AS <i>atu3025</i> UTR3'	-
RNA607				37924	38005	81	atu3041	+
RNA608				64822	64924	102	IGR	+
RNA609				86290	86407	117	UTR5' <i>chrA</i> UTR3'	-
RNA610				98230	98325	95	<i>atu3091</i> UTR5'	+
RNA611				102144	102272	128	atu3095	+
RNA612				102716	102779	63	IGR	+
RNA613 ^ª			C2_105462R	105438	105500	62	AS <i>ugpE</i> UTR3'	-
RNA614				114327	114391	64	atu3106	-
RNA615				117590	117635	45	AS atu3108	-
RNA616				130538	130819	281	AS atu3120	+
RNA617 ^ª	х			131066	131484	418	UTR5' cspA	+
RNA618 ^a	х			132342	132453	111	IGR	+
RNA619 ^a	x; SmrB35		C2_132595F	132592	132715	123	IGR	+
RNA620 ^ª		х	C2_133474F	133444	133670	226	IGR	+
RNA621 ^ª			C2_140230F	140336	140435	99	AS atu3132	+
RNA622				150787	150832	45	IGR	+
RNA623 ^a	х			170848	170972	124	IGR	-
RNA624 ^a			C2_175078F	175065	175140	75	AS atu3165	+
RNA625				187275	187351	76	IGR	+
RNA626 ^a		х		194502	194544	42	IGR	+
RNA627				198879	198948	69	AS atu3189	+
RNA628 ^ª	х			199845	199896	51	IGR UTR5'	-
RNA629				201193	201334	141	atu3191	-
RNA630 ^ª	х	х	C2_204149R	204048	204151	103	IGR	-
RNA631 ^a	x; RF00518; SmrC45		C2_205780R	205570	205785	215	IGR	-
RNA632				207806	207909	103	IGR	-

							UTR5'	
RNA633				210287	210581	294	atu3200	-
RNA634				211316	211430	114	IGR UTR5'	-
RNA635				218556	218663	107	<i>atu3205</i> UTR5'	-
RNA636				219670	219800	130	<i>atu3206</i> UTR5'	-
RNA637				219964	220145	181	atu3207	+
RNA638				229519	229639	120	AS atu3214	-
RNA639 ^ª			C2 232866F	232858	233116	258	AS atu3218	+
RNA640			01_1010000	237297	237342	45	AS atu3223	_
RNA641				261677	261759	82	UTR5' atu3244	+
RNA642				268987	269284	297	UTR5' exuR	+
RNA643				269604	269992	388	IGR	+
RNA644				280174	280255	81	UTR5' atu3261	_
			C2 2015625	201561	201685	124		+
			C2_291303P	291301	201046	24		т
	5500050		62 2427705 ^b	291922	291940	24	AS pssiv	-
RNA647	RF000059		C2_312778F	312//5	312880	105	IGR	+
RNA648				317692	317767	75	UTR5' dadB	+
RNA649				329041	329305	264	UTR5' fixR	-
RNA650				340424	340489	65	UTR5' celA	-
RNA651				348860	348904	44	IGR	+
RNA652				349446	349632	186	AS atu3318	+
RNA653 ^a			C2_352528R	352422	352532	110	AS atu3321	-
RNA654				353520	353615	95	AS atu3321	-
RNA655 ^ª	х			361668	361872	204	UTR5' exoY	-
RNA656				368334	368469	135	IGR	-
RNA657				369588	369728	140	IGR	-
RNA658				369727	369818	91	IGR	+
RNA659				370027	370132	105	AS atu3335	-
RNA660 ^ª		х		370683	370770	87	IGR	-
RNA661				370955	371090	135	atu3336	-
RNA662				372383	372464	81	IGR	-
RNA663				403441	403526	85	atu3364	+
RNA664 ^ª	х			405458	405656	198	atu8094	-
RNA665 ^a	х			405853	405911	58	IGR	+
RNA666 ^ª	х	х	C2_408068F	408064	408140	76	IGR	+
RNA667				413456	413537	81	atu3372	-
RNA668 ^ª		x		427553	427613	60	IGR	+
RNA669 ^a	RE00174			441107	441347	240	IGR	_
	11 00174			441107	441547	109	AS atu 2200	
			C2_444544N	444574	444572	198	AS 0105599	-
RNA671	x			453785	453841	56	IGK	-
RNA672			C2_465314F	465275	465452	177	AS atu3417	+
RNA673				487988	488033	45	AS atu3440	-
RNA674				510944	511004	60	IGR UTR3'	-
KNA675 °	х			513164	513242	/8	atu8032	+
RNA676				515702	515966	264	AS <i>atu3468</i> UTR5'	-
RNA677				523732	523889	157	<i>atu3475</i> UTR5'	-
RNA678				544764	544835	71	<i>atu3493</i> UTR5'	-
RNA679				545769	545910	141	atu3494	-
RNA680				546668	546706	38	AS atu3495	+

RNA681				548406	548528	122	IGR	-
RNA682 ^ª	х			550518	550646	128	IGR UTR5'	-
RNA683				556776	556878	102	atu3505	-
RNA684				562751	562886	135	AS <i>pilA</i> UTR5'	+
RNA685				563369	563477	108	atu3515	-
RNA686				574625	574769	144	AS <i>atu8188</i> UTR5'	-
RNA687				575338	575807	469	atu3525	-
RNA688				581492	581642	150	IGR	-
RNA689				583547	583623	76	IGR	-
RNA690				596817	596967	150	AS atu3546	-
RNA691 ^ª			C2_615336R	615233	615367	134	IGR	-
RNA692				616742	616877	135	UTR5' greB	-
RNA693				617003	617093	90	IGR	+
RNA694				618217	618538	321	UTR3' <i>exsH</i> UTR5'	+
RNA695				621059	621122	63	atu3568	+
RNA696				630723	630971	248	UTR5' xylF	-
RNA697				636013	636103	90	AS mutA	+
RNA698				653102	653188	86	AS argH1 UTR5'	-
RNA699				656188	656305	117	atu3602	+
RNA700 ^ª	Х			659050	659151	101	IGR UTR5'	+
RNA701				659769	659915	146	atu3603	-
RNA702 ^a			C2_661206F	661340	661391	51	IGR	+
RNA703				661412	661512	100	UTR5' <i>ftsE</i> UTR5'	+
RNA704				663194	663285	91	atu3608	+
RNA705				671647	671671	24	AS atu3616	+
RNA706				671743	671857	114	IGR	+
RNA707				671888	672031	143	UTR5' <i>rpmB</i> UTR5'	+
RNA708				673123	673216	93	atu3619	-
RNA709				681595	681640	45	AS atu3682	-
RNA710				681736	681928	192	UTR5' <i>accA</i> UTR5'	+
RNA711				683094	683325	231	atu3631	+
RNA712				684697	684736	39	IGR	-
RNA713				688690	688961	271	IGR	+
RNA714				699165	699438	273	AS atu3645	-
RNA715 [°]	х			708876	708975	99	IGR UTR5'	-
RNA716				716532	716625	93	atu3668	-
RNA717 ^a			C2_717107F	716905	717184	279	AS atu3669	+
RNA718				729336	729579	243	AS atu3676	-
RNA719 ^ª			C2_735151R	735096	735168	72	AS atu3683	-
RNA720				750076	750181	105	AS atu3685	-
RNA721				759368	759449	81	AS atu3686	-
RNA722				759800	759881	81	IGR	+
RNA723 ^ª	Х			760781	760936	155	UTR5 ' <i>fecA</i> UTR5'	+
RNA724				779249	779317	68	atu3699	-
RNA725				779480	779561	81	UTR5' <i>atu3700</i>	+
RNA726 [°]	Х	x	C2_780641F	780636	780688	52	IGR UTR3'	+
RNA727				782372	782557	185	<i>atu3704</i> UTR5'	-
RNA728				783349	783685	336	atu3704	-
RNA729 ^a	RF00517			788173	788258	85	IGR	-

							UTR5'	
RNA730				795407	795530	123	atu3712	-
RNA731				799855	800214	359	UTR5' tolO	-
RNA732				804380	804636	256	UTR5' ruvC	-
RNA733				809283	809422	139	UTR5' atu3728	+
RNA734 ^ª	x; RF00013; SmrC22, sra56	Atu8108; ssrS	Atu8108; ssrS	813850	814009	159	IGR	-
DN14725				044600	04 4000	424	UTR5'	
RNA735				814688	814809	121	atu8192	-
KNA736			h	823078	823154	76	IGK	-
RNA737°	x; Sm84		C2_824767F [°]	824764	824848	84	IGR UTR5'	+
RNA738				824952	825141	189	atu3742	+
RNA739 [°]	х			825524	825608	84	IGR UTR5'	-
RNA740				826049	826217	168	atu3744	+
RNA741				830651	830702	51	IGR UTR5'	-
RNA742				831827	831899	72	atu3747	-
RNA743				834715	834859	144	UTR5' thiE	-
RNA744 ^a	х			835767	836007	240	IGR	-
RNA745				838208	838325	117	UTR3' <i>rpmJ</i> UTR5'	+
RNA746				839161	839393	232	atu3758	-
RNA747				844560	844642	82	IGR UTR5'	-
RNA748				850256	850391	135	atu3768	+
RNA749				851809	851935	126	UTR5' <i>omp10</i>	+
RNA750 ^a	v			852391	852529	138	IGR	+
	X			052351	052525	130		÷
	X			052594	052525	129		Ŧ
RNA752				801001	801/5/	90		-
RNA753				861826	861943	117	IGR	+
RNA754				863105	863242	137	IGR	+
RNA755				876820	876883	63	AS kdpD	+
RNA756 °	х			883112	883148	36	IGR	-
RNA757				887419	887458	39	AS <i>atu3794</i> UTR5'	+
RNA758				909242	909314	72	atu3813	+
RNA759				919217	919258	41	IGR	-
RNA760 [°]	RF000174			921716	921896	180	IGR UTR5'	+
RNA761				932039	932138	99	<i>atu3835</i> UTR5'	-
RNA762				978415	978568	153	<i>atu3883</i> UTR5'	-
RNA763				978625	978703	78	atu3884	+
RNA764 ^ª	x	x		981672	981762	90	IGR UTR3'	-
RNA765				986039	986160	121	<i>atu3891</i> UTR5'	-
RNA766 ^a	x; RF00717			987680	987980	300	atu3891	-
RNA767				993683	993842	159	UTR5' qlpD	-
RNA768				1012715	1012798	83	AS atu3912	-
RNA769 ^ª			C2 1022833R	1022236	1022483	247	AS atu3920	-
	PE00400		01_10110000	102/110	1024226	117		
RNA771 ^a	ctRNA_p42d	Atu8080; ctRNAp42d	Atu8080; ctRNAp42d	1024113	1024230	117	IGR	-
	cumA_p42u	cuninap420	cumAp420	1029070	1020045	-+J CC		т
				1041036	1041705	157		
				1041028	1042224	721		+
KNA//4				1043288	1043384	90	IGK	+
RNA775 [°]	х			1126808	1126892	84	UTR3' exsl	-
RNA776 ^a		х		1133774	1133933	159	IGR	-

							UTR5'	
RNA777				1141616	1141958	342	atu4026	+
RNA778 ^ª		x		1148986	1149064	78	AS atu8035	+
RNA779 [°]				1168085	1168226	141	IGR	-
RNA780				1174226	1174397	171	UTR5' exoH	-
RNA781				1179329	1179515	186	UTR5' exoU	-
RNA782 ^a	х	x (L1)	C2_1179748F	1179754	1179818	64	AS exoX	+
RNA783				1183216	1183236	20	AS soxB	+
RNA784				1189091	1189217	126	UTR5' nodN	-
RNA785				1198166	1198184	18	AS glgP	+
RNA786 ^a		х		1200582	1200692	110	AS <i>atu4079</i> UTR5'	-
RNA787				1202242	1202317	75	atu4081	-
RNA788				1211403	1211526	123	IGR UTR5'	+
RNA789				1215260	1215452	192	atu4095	-
RNA790				1231866	1232073	207	IGR	+
RNA791 °			C2_1261393R	1261270	1261396	126	AS atu4144 UTR5'	-
				1270054	1270329	295	atu4155	+
KINA793			C2 1201001D	1280109	1280310	147	AS ULU4102	+
RNA794	X	x	CZ_1281891K	1281792	1281903	93	UTR5' atu4165	+
RNA796				1286436	1286783	347	UTR5' atu4168	_
RNA797				1293089	1293341	252	UTR5' ptsl	+
RNA798 ^a		x		1297490	1297634	144	IGR UTR5'	+
RNA799				1297706	1297886	180	atu4176	+
RNA800				1301918	1301960	42	AS atu4178	+
RNA801 ^a		х	C2_1303985F	1303693	1303741	48	IGR	-
RNA802 ^a			b	1303985	1304131	146	AS atu4179	+
RNA803			C2_1305779F	1304207	1304278	71	IGR	+
RNA804 °			5	1305788	1305869	81	IGR	+
RNA805				1305970	1306057	87	IGR	+
RNA806				1306895	1306958	63	AS atu4183	+
RNA807 °				1310255	1310367	112	IGR UTR5'	+
RINAOUO				1313307	1313520	219	0104191	-
			C2 12202C0D	1220204	1220275	70	AS 0104192	-
			C2_1320309K	1320294	1220575	20	AS 0104190	-
				1321077	1321107	30	AS 0104197	-
			62 12027705	1334077	1334170	99	AS 0104210	-
RNA813 RNA814 ^ª			C2_1383778F C2_1386463F	1383776	1383858 1386557	99	IGR	+
RNA815				1397031	1397217	186	01R5 [°] atu4271	-
RNA816				1414733	1414788	55	atu4291	-
RNA817 ^a	х		C2 1418097R	1418048	1418183	135	IGR	_
RNA818				1433742	1433955	213	UTR5' atu4308	+
RNA819				1434582	1434753	171	01K3 atu4308	+
RNA820 ^a			C2 1456982P	1456675	1456763	88	AS atu4327	_
RNA821			22_2.00002N	1456897	1456981	84	AS atu4327	_
RNA822				1457193	1457256	63	AS atu4328	+
RNA823				1458269	1458449	180	IGR	+

RNA825				1477219	1477330	111	UTR5' <i>clpB</i>	+
RNA826 ^a	х	х	C2_1492506R	1492471	1492542	71	IGR	-
RNA827				1494107	1494152	45	AS Atu4357	+
RNA828				1518008	1518196	188	UTR5' nnrO	_
RNA829				1524959	1525044	85	IGR	+
RNA830				1527720	1527807	87	LITR5' norC	_
				1525720	1525/007	192	LITES' fon A	+
		(1.2)		1555250	1555421	105		
RNA832	X	X (L2)		1563972	1564059	87	IGR	+
RNA833 °	Х			1570254	1570349	95	IGR	-
RNA834 ^a				1583219	1583390	171	IGR UTR5'	+
RNA835				1587840	1588054	214	atu4444	+
RNA836				1594812	1594872	60	AS atu4449 UTR5'	-
RNA837				1601322	1601580	258	atu4454	-
RNA838				1602120	1602163	43	AS atu4456	-
RNA839				1603748	1603859	111	IGR	+
RNA840				1606587	1606730	143	UTR5' <i>tal</i> UTR5'	+
RNA841				1613564	1613717	153	atu4472	+
RNA842 ^a			C2_1623126R	1623789	1623861	72	AS emrB	-
RNA843 ^a			C2_1630159F	1630155	1630305	150	AS atu4484	+
RNA844 ^a		х	C2 1636311R	1636128	1636323	195	AS atu4490	-
RNA845			_	1638270	1638339	69	IGR	+
RNA846 ^a				1647094	1647233	139	IGR	+
				1649470	1649633	153		
				1650840	1650000	153	UTR5'	-
				1050649	1050990	141	ulu4304	
RNA849				1652640	1652/2/	87	IGK	+
RNA850				16/9215	16/9322	107	AS aroQ2	+
RNA851				1703034	1703100	66	AS atu4557	+
RNA852				1703553	1703640	87	AS atu4557	+
RNA853				1725074	1725158	84	IGR	-
RNA854				1725322	1725407	85	IGR	+
RNA855				1735556	1735637	81	IGR UTR5'	+
RNA856				1747147	1747723	576	atu4600	-
RNA857 ^a			C2_1753644F	1753619	1753842	223	AS nodX	+
RNA858 ^a	х	x (L3)		1758585	1758795	210	IGR	-
RNA859				1760937	1761054	117	AS atu4613	+
RNA860 ^ª			C2_1763538F	1763709	1763787	78	AS rfbD	+
RNA861 ^ª			C2 1764441F	1764491	1764593	102	AS rffB	+
RNA862				1767827	1767988	161	LITR5' UVrD	_
			C2 17027E0D	1702607	1707500	07	AS atu4622	
			C2_1785758K	1705097	1705704	07	AS 0104052	-
RNA864		х	C2_1/91414R	1/9131/	1/91488	1/1	AS atu4639	-
RNA865				1794151	1794224	73	IGR UTR5'	+
				1/90004	T/30/30	132	u104043	-
RNA867		х		1809342	1809464	122	IGR	-
RNA868 ^a				1825065	1825233	168	01K3 atu4666	+
				1020152	1920202	1/1		
NINAOUS		(1.4)		1020122	10207233	141		Ŧ
KNA870 °		x (L4)	C2_1830949F	1830947	1831112	165	AS atu4669	+
RNA871 ^ª	RF001746			1831432	1831675	243	IGR	-
RNA872 ^ª	х	x (L5)	C2_1831446F	1831444	1831584	140	IGR UTR5'	+
RNA873				1831738	1832123	385	atu4671	-
RNA875				1848323	1848380	57	AS atu4690	-
RNA876				1867244	1867355	111	UTR5' <i>nuoE</i>	+

RNA877				1886314	1886396	82	AS <i>atu4723</i> UTR5'	-
RNA878				1888204	1888443	239	<i>atu4724</i> UTR3'	-
RNA879 ^a	х			1890698	1891015	317	atu4727	+
RNA880 ^ª		х	C2_1891119F	1891117	1891172	55	IGR	+
RNA881				1899824	1899965	141	UTR5' acnB	+
RNA882 ^ª	х		C2_1902921F	1902890	1902979	89	IGR	+
RNA883				1913042	1913156	114	IGR	+
RNA884 ^a			C2_1929391F	1929395	1929495	100	AS atu4760	+
RNA885 ^a	x			1938545	1938644	99	IGR	-
RNA886				1947153	1947253	100	UTR5' fdhF	-
RNA887				1949566	1949725	159	IGR	+
RNA888 ^ª			C2_1952079F	1952248	1952435	187	AS <i>atu4781</i> UTR5'	+
RNA889				1953731	1953857	126	atu4782	-
RNA890				1954026	1954180	154	UTR5' afuA2	+
RNA891				1983499	1983547	48	AS pssF	+
RNA892				1986619	1986719	100	AS <i>amsJ</i> UTR5'	+
RNA893				2009079	2009207	128	atu4827	-
RNA894				2011986	2012069	83	AS atu4832	-
RNA895				2039771	2040014	243	IGR UTR5'	+
RNA896				2040106	2040364	258	atu4856	+
RNA897				2041247	2041384	137	UTR5' mobC	+
RNA898				2043370	2043497	127	IGR	+

At Plasmid

RNA899				28	220	192	UTR5' repA	+
RNA900 ^a		x		2470	2512	42	IGR	-
RNA901 ^a	RF00490			2502	2759	257	UTR5' repC	+
RNA902				8840	8933	93	IGR UTR5'	-
RNA903				10285	10530	245	atu8044	+
RNA904 ^a			pAt_30777R	30571	30781	210	IGR UTR5'	-
RNA905				34533	34764	231	atu5033	-
RNA906				34993	35067	74	IGR	+
RNA907 ^a			pAt_37435R	37343	37463	120	AS atu5037	-
RNA908 ^a		x	pAt_43691F	43686	43779	93	AS atu5042	+
RNA909 ^a			pAt_45638F	45618	45892	274	AS atu5045	+
RNA910				51235	51279	44	IGR	+
RNA911				51727	51826	99	UTR3' hspAT1	+
RNA912 ^a		x (At1)	pAt_55271R	55158	55277	119	IGR UTR5'	-
RNA913				55383	55497	114	atu5056	+
RNA914				58296	58396	100	AS upgA2	-
RNA915				63183	63319	136	AS atu5064	+
RNA916				65495	65668	173	AS dapA	+
RNA917 ^ª	Х			71555	71699	144	AS <i>atu8046</i> UTR5'	-
RNA918				71869	72043	174	atu5072	+
RNA919				77793	77851	58	IGR	-
RNA920 ^ª			pAt_85273R	85154	85274	120	AS atu5083	-
RNA921				88295	88451	156	IGR	-
RNA922				90850	90870	20	AS atu5087	-
RNA923 ^a		x		91654	91790	136	IGR	+

RNA924 ^a		pAt_92195R	92180	92221	41	AS atu5089	-
RNA925			94457	94582	125	AS rcdA	+
RNA926			95417	95537	120	AS rcdB	+
RNA927			96357	96477	120	UTR5' rcdB	-
RNA928 ^a	х	pAt 98054F	98023	98202	179	IGR	+
RNA929		• _	102339	102494	155	IGR	+
RNA930 ^a		nΔt 111447R	111229	111450	221	$\Delta S atu 5104$	-
RNA931		p/((_11111)))	115778	116018	240	LITR5' traC2	_
			110054	116456	402		
KINA952	X		110054	110450	402	UTKS LTUAZ	Ŧ
RNA933 °	X		124915	125158	243	IGR UTR5'	-
RNA934			126049	126334	285	utrs118	-
KNA935			12/321	127534	213	UTR5'	+
RNA936			128305	128424	119	atu5121	-
RNA937 °		pAt_142309R	142162	142366	204	IGR	-
RNA938			143058	143211	153	UTR5' blcR	-
RNA939			150124	150189	65	IGR	+
RNA940			150741	150885	144	AS attS	-
RNA941			151155	151297	142	AS attT	+
RNA942			152747	152978	231	UTR5' attW	-
RNA943			154444	154570	126	IGR	+
RNA944 ^ª		pAt_157836F	157834	157972	138	AS atsD	+
RNA945			158164	158275	111	AS atu5158	+
RNA946			159050	159209	159	UTR5' atsE	-
RNA947			159845	159935	90	UTR5' rctA	_
		nAt 162266P	162222	162295	62	AS aub PA	
RNA948		pAt_102200K	102222	102205	207	AS avhB6	-
			105079	170022	207	AS UVIIDO	-
RINA950			170773	170932	159	IGR	+
RNA951			173740	173800	60		-
RINA952			174370	174460	90	ASTITUCA	-
RNA953		pAt_177141R	177052	177109	57	IGR	-
RNA954°	х		183988	184078	90	IGR	-
RNA955 ^a		pAt_185242R	185173	185253	80	AS atu5189	-
RNA956			199660	199882	222	AS xylB	-
RNA957			202238	202355	117	AS <i>gbsA</i> UTR5'	-
RNA958			205042	205220	178	atu5206	-
RNA959			213433	213575	142	IGR	+
RNA960 ^a			214071	214308	237	IGR	+
RNA961			215441	215629	188	AS <i>proW2</i> UTR5'	-
RNA962			217600	217804	204	<i>atu5221</i> UTR5'	+
RNA963			234539	234641	102	atu5239	+
RNA964			235964	236161	197	IGR	+
RNA965 ^a	x		238435	238546	111	IGR	+
RNA966			239789	239941	152	IGR	-
RNA967			249989	250133	144	IGR	+
RNA968			251740	251863	123	UTR5' dppB2	+
RNA969			255148	255364	216	AS dppF2	-
RNA970 ^ª	x		257760	257894	134	IGR	+
RNA971			264908	265007	99	IGR	_
RNA072			265012	265076	63	UTR5'	+
RNA072			203013	203070	87	$\Delta S dc+M$	т _
DNA074			20040/	2003/1	04	AS abrog	т
111/13/4			213104	213231	55	~2 UIII V2	Ŧ
						UTR3'	
----------------------	---	-------------	--------	--------	-----	----------------------------	---
RNA975 ^a	x		275394	275636	242	atu5276	-
RNA976 [°]	х		277256	277461	205	UTR5' <i>yciF</i>	+
RNA977			278675	278730	55	IGR UTR5'	+
RNA978			279495	279744	249	<i>atu5281</i> UTR5'	-
RNA979			286914	287016	102	atu5289	-
RNA980			294734	294756	22	IGR	+
RNA981			296625	296758	133	UTR5' dgdA	-
RNA982 ^a	x		298473	298846	373	AS atu5305	+
RNA983			303205	303379	174	UTR5' fecR2	+
RNA984			305580	305854	274	AS fhuA2	-
RNA985			307652	307792	140	AS <i>fepB</i> UTR5'	-
RNA986			319726	319777	51	<i>atu5328</i> UTR3'	+
RNA987			320636	320900	264	atu5328	+
RNA988			323186	323256	70	AS <i>atu5332</i> UTR5'	-
RNA989			329247	329428	181	atu5338	-
RNA990			330659	330782	123	IGR	+
RNA991			340772	340824	52	IGR	+
RNA992			352310	352425	115	IGR UTR5'	+
RNA993			352470	352527	57	<i>atu5358</i> UTR3'	+
RNA994 ^a	x		352912	353246	334	atu5358	+
RNA995			356620	356705	85	AS atu5364	+
RNA996			360659	360860	201	AS <i>atu5368</i> UTR5'	+
RNA997			365036	365150	114	atu5374	-
RNA998			365455	365788	333	AS exsG	-
RNA999 [°]	х		370683	370918	235	IGR	-
RNA1000			375268	375298	30	AS atu5382	+
RNA1001			378625	378784	159	AS <i>atu5384</i> UTR5'	+
RNA1002			385582	385744	162	<i>atu5392</i> UTR5'	+
RNA1003			387769	387851	82	atu5393	-
RNA1004			391861	392114	253	AS dapA	+
RNA1005			397956	397977	21	AS <i>atu5403</i> UTR5'	+
RNA1006			399259	399343	84	<i>atu5404</i> UTR5'	-
RNA1007			401455	401554	99	atu5406	-
RNA1008			403323	403446	123	AS atu5408	+
RNA1009			407503	407597	94	AS atu5412	-
RNA1010 °		pAt_414717R	414568	414754	186	AS atu5419	-
RNA1011			423741	423960	219	IGR UTR5'	-
RNA1012			424344	424483	139	<i>atu5430</i> UTR5'	-
RNA1013			426925	427099	174	atu5433	-
RNA1014			429502	429560	58	IGR UTR5'	-
RNA1015			432682	432835	153	<i>atu5439</i> UTR5'	-
RNA1016			437113	437244	131	atu5443	-
RNA1017			442805	442980	175	UTR3'	+
RNA1018 ^a		pAt_444075R	443789	444080	291	AS attE2	-
RNA1019			460640	460739	99	AS kup2	+
RNA1020			462086	462173	87	AS dinB2	-
RNA1021			463936	464041	105	IGR	-

RNA1022				476937	477033	96	IGR UTR5'	-
RNA1023				477063	477219	156	<i>atu5485</i> UTR5'	+
RNA1024				479022	479322	300	atu5489	+
RNA1025 ^a			pAt_481533F	481521	481609	88	AS katE	+
RNA1026				484664	484903	239	IGR UTR3'	+
RNA1027				491916	492189	273	atu5500	+
RNA1028 ^a				492222	492368	146	IGR UTR5'	-
RNA1029				492996	493141	145	atu5501	-
RNA1030 ^a			pAt_497789R	497710	497818	108	AS atu5506	-
RNA1031 ^a			pAt_500617R	500560	500710	150	AS gst	-
RNA1032				504868	504959	91	IGR	-
RNA1033 ^a	х	х		522656	522807	151	IGR	+
RNA1034				523107	523323	216	AS amiC	-
RNA1035 ^a	х	х		524741	524865	124	IGR	+
RNA1036				534048	534109	61	IGR	-
RNA1037				535729	535804	75	AS <i>atu5541</i> UTR5'	+
RNA1038				537745	537902	157	atu5543	+

Ti plasmid

							UTR5'	
RNA1039				2347	2494	147	atu6001	+
RNA1040 ^a	х		pTi_5440F ^b	5419	5673	254	IGR	+
RNA1041				6300	6496	196	AS rolB	-
RNA1042				9661	9763	102	AS atu6005	-
RNA1043				10123	10288	165	IGR	-
RNA1044				15058	15235	177	AS tms2	+
RNA1045				16968	17058	90	AS tms1	-
RNA1046				20752	20836	84	AS atu6013	-
RNA1047 ^a		х		22530	22689	159	AS atu8062	+
RNA1048 ^a		х	pTi_23992F	23988	24238	250	AS nos	+
RNA1049 ^a	х			24808	24928	120	IGR	+
RNA1050				26965	27043	78	AS odh	+
RNA1051				31156	31339	183	IGR	-
RNA1052 ^a	x			31479	31567	88	IGR	-
RNA1053				36970	37033	63	AS nocM	+
RNA1054				41032	41110	78	AS nocR	-
RNA1055				41193	41346	153	AS atu6030	+
RNA1056				52183	52305	122	UTR5' repA	+
RNA1057 ^a		х	pTi_54770R	54396	54450	54	AS repB	-
RNA1058 ^a	х			54629	54767	138	IGR	+
RNA1059 ^a	RF00490	х		54789	54942	153	IGR UTR5'	+
RNA1060				57406	57497	91	atu6047	+
RNA1061				59358	59473	115	AS atu6048	+
RNA1062				68865	69025	160	AS ssuC	-
RNA1063				70320	70489	169	AS ssuB	-
RNA1064 ^a		х		76719	76852	133	AS dhaA	+
RNA1065				77578	77626	48	AS atu6065	+
RNA1066 ^a			pTi_79525F	79555	79740	185	AS atu6067	+
RNA1067				80836	80905	69	AS rbsC	+
RNA1068 ^a			pTi_84241F	84206	84567	361	AS aiiB	+
RNA1069				86430	86716	286	AS atu6072	+

NALAD70'PTL 87172FST357ST367SA Sd pAASd pAARNA1072'xPTL 91643FS158891658S70Sd galoARNA1072'FPTL 916587R020280GUG082ARNA1073'FPTL 1665581066151020SG mu6088ARNA1074'FPTL 16655816161116111611161ARNA1074'FPTL 1625871161311601300GR AARNA1074'FF11613116131160100081AARNA1074'FF13008130141300GR AAARNA1074'FF130081301413011301130AAARNA1074'FF130031301413011301130AAAARNA1084'KF1301413011301130AAA<								dfpA	
NALDY1xpT_91043F918339163475A S d/pm4RNALDY2FT_9701RPT_9701R<	RNA1070 ^a			pTi_87172F	87157	87337	180	AS dfpA	+
NALADZ*PT_9C11RPT_9C11RPT_9C12	RNA1071 ^a		х	pTi_91643F	91583	91658	75	AS dfpB	+
RNA1073NALO73NAU074NAU074NAU076NAU077NAU0	RNA1072 ^a			pTi_97611R	97420	97471	51	AS <i>atu6081</i> UTR5'	-
RNA1074Infra <t< td=""><td>RNA1073</td><td></td><td></td><td></td><td>97743</td><td>98029</td><td>286</td><td>atu6082</td><td>+</td></t<>	RNA1073				97743	98029	286	atu6082	+
RNA1075 ¹ pTI_066598 106613 10614 102 AS <i>abu0288</i> - RNA1077 pTI_1152658 116163 116140 22 IGR - RNA1077 pTI_1152658 115050 19000 19000 10000 - RNA1079 I 118887 19000 10000 A <i>abu0010</i> - RNA1081 x I 13030 128140 99 A <i>sau60101</i> - RNA1082 x I 13030 13023 13020 1600 A <i>sau60111</i> - RNA1083 x I 13030 13023 13020 A <i>sau60111</i> - RNA1084 x pTI_48378 148278 148278 1000 A <i>sau61111</i> - RNA1085 x pTI_40378 148278 148278 1600 A <i>sau61111</i> - RNA1085 x pTI_40378 14939 14937 16070 15010 15020	RNA1074				101653	101720	67	IGR	+
RNA1076Infaile <t< td=""><td>RNA1075^a</td><td></td><td></td><td>pTi_106595R</td><td>106513</td><td>106615</td><td>102</td><td>AS atu6088</td><td>-</td></t<>	RNA1075 ^a			pTi_106595R	106513	106615	102	AS atu6088	-
RNA1007 , pT_116265R 116163 116302 139 , GR UTRS' UTRS' - RNA1078 , 124939 15114 1950 A5 atu6101 - RNA1087 , 244939 12314 1950 A5 atu6101 - RNA1080 , 244939 12314 1950 A5 atu6101 - RNA1081 x , 24393 12340 106R - RNA1082 , 3 13922 1230 IGR - RNA1083 x (Ti1) pTi_148374R 148278 148378 100 A5 atu6111 - RNA1085* x (Ti1) pTi_14938F 14993 19933 14993 40 A5 atu6123 + RNA1085* x (Ti1) pTi_14938F 15020 15070 IS070 A5 atu6135 + RNA1093 x (Ti1) pTi_140938F 15021 15040 A5 atu6147 + RNA1094 x (Ti1) pTi_16074R 16067F 16129 33 A5 atu6147 + RNA1095 x (Ti1) pTi_160609R 16529 17529 <t< td=""><td>RNA1076</td><td></td><td></td><td></td><td>116118</td><td>116140</td><td>22</td><td>IGR</td><td>-</td></t<>	RNA1076				116118	116140	22	IGR	-
NNA10078IIISBATII19006IIIIIC008IIIC008IIIIIIIINNA108012439312514195AS atu6105.RNA1081 ¹ x13231712841599AS atu6110.RNA10821390713923130013001300NGR.RNA1083 ¹ xTIII1390719223100AS traß.RNA1085 ¹ x (TT1)pTI_148374148278148278100AS traß.RNA1085 ¹ x (TT1)pTI_149384149331997300UTRS'tra<	RNA1077 ^a			pTi_116265R	116163	116302	139	IGR UTR5'	-
RNA10671248301251.34195AS atubi10-RNA1081x12831712841694AS atubi10-RNA1083x13006130429123IGR-RNA1083xTT113107131101310979AS atubi11-RNA1083x (TT1)PT1_1433748148278148378100AS tra8-RNA1084x (TT1)PT1_14938148148278148070100AS tra8-RNA1085x (TT1)PT1_149381481500015020150UTR5' tra0+RNA1086x (TT1)PT1_1607481502015020149AS atubi15+RNA1090x153356153467111AS tra0+RNA1091xPT1_16074815052149AS atubi15+RNA1092xPT1_16074815052167469AS atubi16+RNA1093xPT1_16074816649715129133IGR+RNA1094xPT1_1600748166129233AS atubi14+RNA1095xPT1_1600748167421751IGR+RNA1095xPT1_180481860701684+RNA1095xPT1_180481860701684+RNA1096xTFT1_180481860701684+RNA1097xT16745186170168-+R	RNA1078				118887	119006	119	atu6098	-
RNA1080 x 128317 128416 99 AS artb6105 - RNA1081 [*] x 130306 130429 123 IGR - RNA1083 [*] x 130073 139223 150 IGR - RNA1083 [*] x (Ti1) ρTi_149378 148278 148378 149973 40 IGR + RNA1085 [*] x (Ti1) ρTi_149387 148278 149973 400 IGR + RNA1086 x (Ti1) ρTi_149387 149973 149973 400 IGR + RNA1088 x (Ti1) ρTi_149387 149273 15512 149 AS artb6135 + RNA1089 x Ti5356 153467 111 AS traC + RNA1091 x FTi_1607418 160687 160749 53 AS artb6147 + RNA1092 [*] FTi_1606088 166759 16279 87 GGR + RNA1095 [*] x FTi_1606088 16679 16751 157 IGR + RNA1095 [*] <t< td=""><td>RNA1079</td><td></td><td></td><td></td><td>124939</td><td>125134</td><td>195</td><td>AS atu6101</td><td>-</td></t<>	RNA1079				124939	125134	195	AS atu6101	-
NAND081* NAND082*X130300130420123016GR-NAND084*x (TI1)pTI_148374R148278100AS arab(11)-RNA1085*x (TI1)pTI_14938F1993314993001GR+RNA1085*x (TI1)pTI_14938F1993314993001GR+RNA1085*x (TI1)pTI_14938F150200150200150AS mapA+RNA1086x (TI1)FT15020115020150AS mapA+RNA1087x (TT1)FT153356153467111AS trad+RNA1089xFT154973155122149AS aracA-RNA1090*xFT_16074IR1606781674457AS aracA-RNA1091xFT_16074IR16267116276493AS arab(147)-RNA1093rPTI_16074IR166871674293AS arab(147)-RNA1094*rFT_160608R1652916219133IGR+RNA1095xrFT_160608R16752750IGR-RNA1096xrFT_16074IR166291362054AS virA-RNA1097xFT_118044R1879913620167IGRRNA1098xFT_1190638R190571626193AS virB4RNA1104*xFT_190638R19057138AS virB4<	RNA1080				128317	128416	99	AS atu6105	-
NA108213311113319079A S attb211-RNA1083 *r139073139223150IGR-RNA1083 *x (Ti1)FT_1483741482781482714837100AS traß-RNA1085 *x (Ti1)PT_14938F14993514997340IGR+RNA1086.150201507015700IS700AS mcpA+RNA10881505115072015070AS mcpA+RNA1090 *x.159381593653AS actb135+RNA1091159281593853AS accb-RNA1092 *162671607453AS accb-RNA109315122149AS accb-RNA1094 *162671607453AS accb-RNA1095 *17266175493AS accb-RNA1095 *172661751IGR-RNA1095 *16762175IGR-RNA1096 *16762175IGR-RNA1097 *RNA1098 * <td< td=""><td>RNA1081^a</td><td>х</td><td></td><td></td><td>130306</td><td>130429</td><td>123</td><td>IGR</td><td>-</td></td<>	RNA1081 ^a	х			130306	130429	123	IGR	-
NA1083* x (T1) pT_148374 148273 148178 100 AS traB - NA1085* x (T1) pT_149338 14973 14973 40 GR + NN1085* x (T1) pT_149338 149933 149973 40 GR + NN1086 x pT_149338 149933 15021 206 UTRS' traM + NN1087 x 150561 150720 159 AS mcpA + NN1098 x 154973 15123 63 AS accB - NN1092* x pT1_160741 16087 16074 57 AS accB - NN1093 x pT1_160741 16087 16074 53 87 AS accD - NN1094* x pT1_160609 16529 16329 133 IGR + NN1095 x pT1_160901 16589 166129 230 IGR + NN1097 x pT1_190683 170651 17011 90 IGR + NN1097	RNA1082				133111	133190	79	AS atu6111	-
NA1084* x (T1) pTI_143374 148378 148378 100 AS traB - NNA1085* x pTI_149338 149933 14973 40 IGR + NNA1085 x pTI_149338 149033 14973 40 UTRS' traM + NNA1086 . . 15004 15020 159 AS mcA + RNA1089 153356 153467 111 AS traB + RNA1090* x RNA1091 x .	RNA1083 ^a				139073	139223	150	IGR	-
NA1085°xpT_149938F14993814997340IGR+NA1086I15020205UTRS' traM+NA1088I1502015070151A5 mcpA+NA1087I153367110A5 mcpA+RNA1090xI51375155127149A5 mcbA-RNA1091xI159281593653A5 accA-RNA1092FT_16009R1667416074470A5 accA-RNA1093FT_16009R16589166129233A5 accA-RNA1094FI674521751987IGR+RNA1095xFT_16009R1658916129133IGR-RNA1096xF1720617219133IGR+RNA1096xF176221771190IGR-RNA1097FI6769217219133IGR+RNA1098FI7118204R1807018070168A5 virA-RNA1008FI7119180701807054A5 virA-RNA1010FF193319044111A5 virA-RNA103X (TI3)FI_19068R193747265UTRS' virA-RNA1103X (TI4)F194331944111A5 virA-RNA1104X (TI4)F194391516A5 virA-RNA11	RNA1084 ^a		x (Ti1)	pTi_148374R	148278	148378	100	AS traB	-
NA1086ISO004150210206UTRS' traM+RNA1088150720159A5 mcpA+RNA10891536615376711A5 traR+RNA1090"x159365346711A5 traR+RNA1091159385336A5 actub135++RNA1092 *159381593853A5 actub135+RNA1093 *pT1_16074R16068716074993A5 actub-RNA1094 *pT1_16009R16526166129233A5 actub147-RNA1095 *xPT1_16009R1725913316R-RNA1096 *xT1706161721913316R-RNA1097 *r1706161721913316R-RNA1098 *xpT1_18204R18050715018A5 virA-RNA1098 *xpT1_18204R18050918084894UTRS' virA+RNA1098 *xpT1_18204R1805071801400A5 virA-RNA1098 *xpT1_18204R18050718084094UTRS' virA+RNA1098 *xpT1_18204R18050718014016RRNA109 *xpT1_190683819070518084094UTRS' virA+RNA103 *x (Ti3)pT1_20052819347265UTRS' virA-RNA104 *xpT1_20052819349161A5 virC4 <t< td=""><td>RNA1085^a</td><td></td><td>х</td><td>pTi_149938F</td><td>149933</td><td>149973</td><td>40</td><td>IGR</td><td>+</td></t<>	RNA1085 ^a		х	pTi_149938F	149933	149973	40	IGR	+
NA1088ISOT20159AS mcpA+RNA1089x15346111AS trach+RNA1090x1531215122149AS atu6135+RNA1091x159721597315512149AS atu6137+RNA1092x1607411607457AS accA-RNA1092167471627616276233AS atu6147-RNA1093xTPT_166009R1658916129233AS atu6147-RNA1095xTPT_166009R1658916129133IGR-RNA1096xTT1720617219133IGR-RNA1096xTT1720617219133IGR-RNA1096xTT1720617219133IGR-RNA1096xTT1701100IGR-RNA1097xTPT_18204R181961820554AS virA-RNA1008rT1701100IGRRNA1010rrPT_19063871903319044111AS virB4-RNA102rT1906387190705138AS virD4-RNA103x (Ti3)PT_1906387190705138AS virD4-RNA103x (Ti4)F1933719444111AS virD4-RNA1104r	RNA1086				150004	150210	206	UTR5' <i>traM</i>	+
RNA1089 x 153356 15346 111 AS traR + RNA1090 ^a x 154973 155122 149 AS atu6135 + RNA1091 ^a x 15928 15938 53 AS arcA - RNA1092 ^a FTI_16074R 16074 16074 57 AS arcA - RNA1091 ^a FTI_16600R 16276 16276 33 AS arcA - RNA1094 ^a FTI_16600R 16528 16612 233 AS arcA - RNA1095 x FTI_16600R 16528 16752 75 IGR - RNA1096 ^a x FTI_16600R 17219 133 IGR - RNA1097 x FTI_18204R 17616 17619 10 ITS'virA + RNA100 ^a x FTI_18204R 18196 18205 54 AS virA - RNA100 ^a x FTI_190683R 19343 19444 111 AS virA - RNA1102 ^a x (Ti3) FTI_190683R 19343 19444<	RNA1088				150561	150720	159	AS mcpA	+
RNA1090 ^a x 15497 155122 149 AS atu6135 + RNA1091 5338 53 AS accA - RNA1092 ^a FTI_16074R 160687 160744 57 AS accB - RNA1093 ^a FTI_16074R 160687 160740 93 AS atu6147 - RNA1094 ^a FTI_16007R 165896 16589 17519 87 IGR + RNA1095 FT 17646 17519 87 IGR + RNA1096 ^a x FT 17646 17519 133 IGR + RNA1097 FT 17646 17519 163 H + + RNA1098 F FT 17604 18079 IGR + + RNA100 ^a F FT 18070 18084 94 UTR5'virA + RNA1100 ^a FT I90333 19044 111 AS virB - RNA1102 ^a X (Ti3) FT 190383 19044 111 AS virC + <td>RNA1089</td> <td></td> <td></td> <td></td> <td>153356</td> <td>153467</td> <td>111</td> <td>AS traR</td> <td>+</td>	RNA1089				153356	153467	111	AS traR	+
RNA1091ISP330S3AS accAIRNA1092 aPTI_160741R160687160740S7AS accBIRNA1093 aPTI_1600781627016276493AS acuB IIRNA1094 aPTI_16009R165896166129233AS atu6147IRNA1095 aPTI_16009R16745216753987IGRIRNA1096 aXII17206617219133IGRIRNA1097 aII176211771190IGRIIRNA1098 aIII </td <td>RNA1090^a</td> <td></td> <td>х</td> <td></td> <td>154973</td> <td>155122</td> <td>149</td> <td>AS atu6135</td> <td>+</td>	RNA1090 ^a		х		154973	155122	149	AS atu6135	+
RNA1092 a pTi_160741 160744 57 AS accB - RNA1093 - 162671 16276 93 AS accD - RNA1094 a - 162671 16276 93 AS acuB147 - RNA1095 a x - 166098 166129 233 AS acuB147 - RNA1096 a x - 17606 17219 133 IGR - RNA1097 a x - 17664 17611 90 IGR - RNA1098 a x - 17692 17011 90 IGR - RNA1090 a - - 16074 180750 54 AS virA - RNA1091 a - - 17626 17011 90 IGR - RNA1092 a - - 17092 17011 90 AS virA - RNA1010 a - - 16074 18070 96 AS virB4 - RNA1010 a x (Ti3) PTi_190683R 190545 19075 1808<	RNA1091				159283	159336	53	AS accA	-
RNA1093 RNA1094 ^a RNA1094 ^a RNA1095 RNA1096 ^a x .	RNA1092 ^a			pTi_160741R	160687	160744	57	AS accB	-
RNA1094 ^a pTi_166009R 165896 166129 233 AS atu6147 - RNA1095 167452 167539 87 IGR + RNA1096 ^a x 172066 172199 133 IGR + RNA1097 x 176446 176521 75 IGR + RNA1098 x 176921 17011 90 IGR + RNA1099 0 IGR - + - + RNA1099 0 IGR - <t< td=""><td>RNA1093</td><td></td><td></td><td></td><td>162671</td><td>162764</td><td>93</td><td>AS accD</td><td>-</td></t<>	RNA1093				162671	162764	93	AS accD	-
RNA1095 x 167452 167539 87 IGR + RNA1096 ⁸ x 172066 17219 133 IGR - RNA1097 176446 176521 75 IGR + RNA1098 176446 176521 75 IGR + RNA1098 176921 177011 90 IGR - RNA1099 176921 177011 90 IGR - RNA1099 180790 180884 94 UTR5'virA + RNA1101 pTi_182044R 181966 182050 54 AS virB4 - RNA1101 pTi_190683R 180674 186170 96 AS virB4 - RNA1102 ⁸ x (Ti3) pTi_190683R 190363 190444 111 AS virB4 - RNA1104 x (Ti4) pTi_190683R 190375 180444 111 AS virC4 + RNA1104 x (Ti4) pTi_190683R 190375 190755 138 AS virC4 + RNA1106 x (Ti4) pTi_200229R	RNA1094 ^a			pTi_166009R	165896	166129	233	AS atu6147	-
RNA1096 ³ x I7206 17219 133 IGR - RNA1097 I7646 17621 75 IGR + RNA1098 I7701 90 IGR - RNA1099 IT 176921 17701 90 IGR - RNA1099 IT 180790 18084 94 UTRS' virA + RNA100 ³ IT IT 18070 18080 54 AS virA - RNA101 IT It <t< td=""><td>RNA1095</td><td></td><td></td><td></td><td>167452</td><td>167539</td><td>87</td><td>IGR</td><td>+</td></t<>	RNA1095				167452	167539	87	IGR	+
RNA1097 IGR + RNA1098 17692 177011 90 IGR - RNA1099 18079 180884 94 UTR5' virA + RNA100 ^a pTi_18204R 18196 182050 54 AS virB4 - RNA1101 186074 186170 96 AS virB4 - RNA1101 pTi_190683R 19044 111 AS virB9 - RNA1102 ^a x (Ti3) pTi_190683R 190567 190705 138 AS virB9 - RNA1104 rtTi3) pTi_190683R 190567 190705 138 AS virC2 + RNA1104 rtTi3) pTi_190683R 190567 190705 138 AS virC2 + RNA1104 rtTi4) rtTi4 194397 194913 516 AS virC1 + RNA1106 rtTi4 rtTi4 194397 195274 195 AS virD4 - RNA1106 rtTi4 rtTi4 rtTi4 rtTi4 200506 200617 111 AS virD4 -	RNA1096 ^a	х			172066	172199	133	IGR	-
RNA1098 Image: sector of the sector of t	RNA1097				176446	176521	75	IGR	+
RNA1099 UTRS' virA + RNA1100 ^a pTi_18204R 181996 182050 54 AS virA - RNA1101 pTi_190683R 186074 186170 96 AS virB4 - RNA1102 ^a r 190333 190444 111 AS virB9 - RNA1102 ^a r(Ti3) pTi_190683R 190567 190705 138 AS virB9 - RNA1104 pTi_190683R 190567 190705 138 AS virG1 + RNA1104 pTi_190683R 190567 190705 138 AS virG2 + RNA1104 r r 193482 19377 265 UTRS' virG + RNA1106 r 195079 195274 195 AS virC1 + RNA1107 ^a x (Ti4) pTi_200229R 200140 200245 105 AS virD4 - RNA1108 r pTi_2014638 201017 201617 111 AS virD4 - RNA1110 ^a r r 201307 201475 168 AS virD4	RNA1098				176921	177011	90	IGR	-
RNA1100 ^a pTi_182044R 181996 182050 54 AS virA - RNA1101 - - 186074 186170 96 AS virB4 - RNA1102 ^a - - </td <td>RNA1099</td> <td></td> <td></td> <td></td> <td>180790</td> <td>180884</td> <td>94</td> <td>UTR5' virA</td> <td>+</td>	RNA1099				180790	180884	94	UTR5' virA	+
RNA1101 Image: prime prima prime prima prime prime prime prima prime prima prima	RNA1100 ^a			pTi_182044R	181996	182050	54	AS virA	-
RNA1102 a pTi_190683R RNA1102 a x (Ti3) pTi_190683R 190567 190705 138 AS virB9 - RNA1104 193482 193747 265 UTR5' virG + RNA1105 a x (Ti4) 194397 194913 516 AS virC2 + RNA1106 x (Ti4) 195079 195274 195 AS virD4 - RNA1107 a x (Ti4) 195079 195274 195 AS virD4 - RNA1106 Fri_200229R 200140 200245 105 AS virD4 - RNA1108 Fri_200229R 200130 20163 150 AS virD4 - RNA1109 Fri_201463R 201013 201163 150 AS virD4 - RNA1110 a Fri_201463R 205561 106 IGR - RNA1111 S X PTi_206590R 206608 123 UTR5' virE1 + RNA11113 a X PTi_206590R 206607 206000 143 AS virE2 - RNA1114 a X <t< td=""><td>RNA1101</td><td></td><td></td><td></td><td>186074</td><td>186170</td><td>96</td><td>AS virB4</td><td>-</td></t<>	RNA1101				186074	186170	96	AS virB4	-
RNA1102 ° 190333 190444 111 AS virB9 - RNA1103 ° x (Ti3) pTi_190683R 190567 190705 138 AS virB9 - RNA1104 193482 193747 265 UTR5' virG + RNA1105 ° x (Ti4) 194397 194913 516 AS virC2 + RNA1106 195079 195074 195 AS virC1 + RNA1107 ° x pTi_200229R 200140 200245 105 AS virD4 - RNA1108 x pTi_200229R 200140 200245 105 AS virD4 - RNA1108 x pTi_200249R 200506 200617 111 AS virD4 - RNA1109 x pTi_201463R 201013 201163 150 AS virD4 - RNA1110 ° x pTi_201463R 201307 201475 168 AS virD4 - RNA1111 x x pTi_20559R 205561 106 IGR - RNA11112 x pTi_20559R 2064	2			pTi_190683R					
RNA1103 a x (Ti3) pTi_190683R 190567 190705 138 AS virB9 - RNA1104 193482 193747 265 UTR5' virG + RNA1105 a x (Ti4) 194397 194913 516 AS virC2 + RNA1106 195079 195274 195 AS virC1 + RNA1107 a x pTi_200229R 200140 200245 105 AS virD4 - RNA1108 x pTi_200229R 200140 200245 105 AS virD4 - RNA1109 x pTi_201463R 201013 201163 150 AS virD4 - RNA1110 a x pTi_201463R 205455 205561 106 IGR - RNA1111 x pTi_206590R 206457 20600 143 AS virE2 - RNA1113 a x pTi_205590R 206457 206600 143 AS virE2 - RNA1114 a x pTi_205590R 206457 20600 143 AS virE2 - RNA1114 a	RNA1102 °			b	190333	190444	111	AS virB9	-
RNA1104 193482 193747 265 UTR5' virG + RNA1105 a x (Ti4) 194397 194913 516 AS virC2 + RNA1106 195079 195274 195 AS virC1 + RNA1107 a x pTi_200229R 200140 200245 105 AS virD4 - RNA1108 x pTi_200229R 200140 200245 105 AS virD4 - RNA1109 x pTi_201463R 201013 201163 150 AS virD4 - RNA1110 a x pTi_201463R 201307 201475 168 AS virD4 - RNA1110 a x pTi_201463R 205455 205561 106 IGR - RNA1111 x pTi_201463R 205455 205561 106 IGR - RNA1111 x pTi_206590R 206457 206009 123 UTR5' virE1 + RNA1113 a x pTi_206590R 206457 206600 143 AS virE2 - RNA1114 a x <td>RNA1103^a</td> <td></td> <td>x (Ti3)</td> <td>pTi_190683R</td> <td>190567</td> <td>190705</td> <td>138</td> <td>AS virB9</td> <td>-</td>	RNA1103 ^a		x (Ti3)	pTi_190683R	190567	190705	138	AS virB9	-
RNA1105° x (Ti4) 194397 194913 516 AS virC2 + RNA1106 195079 195274 195 AS virC1 + RNA1107° x pTi_200229R 200140 200245 105 AS virD4 - RNA1108 x pTi_200229R 200140 200245 105 AS virD4 - RNA1109 pTi_201463R 201013 201475 168 AS virD4 - RNA1110° pTi_201463R 201307 201475 168 AS virD4 - RNA1111° 205455 205561 106 IGR - RNA1111 x pTi_206590R 206457 20600 143 AS virE2 - RNA1113° x pTi_20590R 206457 206600 143 AS virE2 - RNA1114° x pTi_205590R 206457 206600 143 AS virE2 -	RNA1104				193482	193747	265	UTR5' virG	+
RNA1106 195079 195274 195 AS virC1 + RNA1107 ^a x pTi_200229R 200140 200245 105 AS virD4 - RNA1108 200506 200617 111 AS virD4 - RNA1109 pTi_201463R 201013 201475 168 AS virD4 - RNA1110 ^a pTi_201463R 201307 201475 168 AS virD4 - RNA1110 ^a 205455 205561 106 IGR - RNA1112 pTi_206590R 206457 20600 143 AS virE2 - RNA1113 ^a x pTi_20559R 206457 20600 143 AS virE2 -	RNA1105 [°]		x (Ti4)		194397	194913	516	AS virC2	+
RNA1107 ^a x pTi_200229R 200140 200245 105 AS virD4 - RNA1108 200506 200617 111 AS virD4 - RNA1109 pTi_201463R 201013 201163 150 AS virD4 - RNA1110 ^a pTi_201463R 201307 201475 168 AS virD4 - RNA1111 205455 20561 106 IGR - RNA1112 pTi_206590R 206457 20600 143 AS virE2 - RNA1113 ^a x pTi_20590R 206457 20610 243 IGR -	RNA1106				195079	195274	195	AS virC1	+
RNA1108 200506 200617 111 AS virD4 - RNA1109 201013 201163 150 AS virD4 - pTi_201463R 201307 201475 168 AS virD4 - RNA1110 a 201307 201475 168 AS virD4 - RNA1111 205455 20561 106 IGR - RNA1112 205966 206089 123 UTR5' virE1 + RNA1113 a x pTi_206590R 206457 206600 143 AS virE2 - RNA1114 a x 210367 210610 243 IGR +	RNA1107 ^a		х	pTi_200229R	200140	200245	105	AS virD4	-
RNA1109 201013 201163 150 AS virD4 - pTi_201463R b 201307 201475 168 AS virD4 - RNA1110 ° 205455 205561 106 IGR - RNA1112 205966 20609 123 UTR5' virE1 + RNA1113 ° x pTi_206590R 206457 20600 143 AS virE2 - RNA1114 ° x 210367 210610 243 IGR +	RNA1108				200506	200617	111	AS virD4	-
pTi_201463R RNA1110 ° 201307 201475 168 AS virD4 - RNA1111 205455 205561 106 IGR - RNA1112 205966 206089 123 UTR5' virE1 + RNA1113 ° x pTi_206590R 206457 20600 143 AS virE2 - RNA1114 ° x 210367 210610 243 IGR +	RNA1109				201013	201163	150	AS virD4	-
RNA1110 201307 201475 168 AS VIED4 - RNA1111 205455 205561 106 IGR - RNA1112 205966 206089 123 UTR5' vieE1 + RNA1113 ^a x pTi_206590R 206457 206600 143 AS vieE2 - RNA1114 ^a x 210367 210610 243 IGR +				pTi_201463R	201207	201475	160	AS wirD4	
RNA1112 205455 205501 100 IGK - RNA1112 205966 206089 123 UTR5' virE1 + RNA1113 ^a x pTi_206590R 206457 206600 143 AS virE2 - RNA1114 ^a x 210367 210610 243 IGR +					201307	2014/5	106	AS VII D4	-
RNA1113 ° x pTi_206590R 206457 206600 143 AS virE2 RNA1114 ° x 210367 210610 243 IGR +	RNA1111 RNA1112				205455	205501	123	UTR5' virE1	+
RNA1114 ^a x 210367 210610 2/3 IGR ±	RNA1113 ^a		x	pTi 206590R	206457	206600	143	AS virE2	-
· · · · · · · · · · · · · · · · · · ·	RNΔ111/1 ^a	v		F2000000	210367	210610	243	IGR	+

Implication des ARN non codant dans la virulence du phytopathogène Agrobacterium fabrum C58

L'une des caractéristiques majeures des microorganismes, et donc des bactéries, est qu'ils sont en contact direct avec l'environnement et doivent donc percevoir et répondre rapidement à ses variations. Pour cela, plusieurs niveaux de contrôle existent, et récemment le rôle des ARN non codants régulateurs, ou riborégulateurs, a été mis en lumière comme un mécanisme de contrôle peu couteux et rapide pour la cellule. Chez le phytopathogène *Agrobacterium fabrum* (anciennement appelé *Agrobacterium tumefaciens*), la virulence est principalement régulée au niveau transcriptionnel par le système à deux composants VirA/VirG. L'implication des riborégulateurs dans la virulence d'*A. fabrum* est encore mal connue et ces travaux de thèse ont eu pour objectif de déterminer l'implication de riborégulateurs dans le cycle infectieux de cette bactérie.

Pour cela, nous avons identifié l'ensemble des transcrits d'*A. fabrum* C58 en combinant l'utilisation de plusieurs méthodes d'analyses globales et nous avons étudié la fonction de différents candidats transcrits à partir du plasmide Ti (plasmide de virulence). Des souches modifiées dans la production des riborégulateurs candidats ont été construites, leurs ARNm cibles ont été prédits et validés, et des tests phénotypiques, en particulier des tests de virulence, ont été réalisés.

Ainsi, le séquençage des transcrits de petite taille a permis d'identifier plus d'un millier de riborégulateurs potentiels dont plusieurs sont exprimés à partir de régions en relation avec le cycle infectieux. Nous avons validé 4 de ces petits transcrits comme étant des riborégulateurs puisqu'ils sont de petite taille, non traduits en protéine et fortement structurés (RNA1111, RNA1083, RNA1059 et RNA1051). Plus particulièrement, nous avons montré que le riborégulateur RNA1111 était nécessaire pour la virulence d'*A. fabrum* C58, et que son action semblait se faire au travers du contrôle post-transcriptionnel de gènes impliqués dans les fonctions de virulence et de transfert du plasmide Ti. Un rôle plus modéré du riborégulateur RNA1083 dans le contrôle du cycle infectieux a également été observé, potentiellement au travers de la modulation de la mobilité et du transfert conjugatif du plasmide Ti. D'autre part, nous avons mis en évidence deux autres riborégulateurs, RNA1059 et RNA1051, qui sont impliqués dans le contrôle du maintien du plasmide Ti via une implication dans la réplication du plasmide (RNA1059) et via une implication dans un nouveau system de toxine-antitoxine présent sur le plasmide Ti (RNA1051).

Ainsi à partir d'une analyse globale nous avons mis en évidence le rôle des riboregulateurs dans les systèmes de mise en place de l'infection bactérienne, soit via le contrôle de facteurs de virulence, soit via le contrôle de la persistance du plasmide responsable de la virulence.

Mots-clés : Agrobacterium, virulence, plasmide Ti, ARN non codant, riborégulateurs, régulation post-transcriptionnelle, transcriptome.

Implication of non coding RNA in the virulence of the phytopathogene *Agrobacterium fabrum* C58

One of the main characteristics of microorganisms, including bacteria, is their direct interaction with their environment. They thus need to perceive and quickly answer to its variations. Several steps of control exist, and recently the role of regulatory non-coding RNA, or riboregulator, was highlighted as a fast and economic mechanism of regulation. In the phytopathogen *Agrobacterium fabrum* (previously named *Agrobacterium tumefaciens*), the virulence is mainly controlled transcriptionally by the two components system VirA/VirG. The implication of riboregulators in the virulence of this bacterium is still unknown. The objectives of this thesis were to identify *A. fabrum* riboregulators and to determine their involvement in the infectious cycle of the bacteria.

To this end, we identified small transcripts of *A. fabrum* C58 strain by combining several global analyses, and we studied the function of different candidates transcribed from the Ti plasmid (the virulence plasmid). Strains modified in the production of these candidates were constructed, their mRNA targets were predicted and validated, and phenotypic analyses - especially virulence tests- were realized.

Thereby, small transcript deep-sequencing allowed the identification of a thousand potential riboregulators, some of them being transcribed from regions related to the infectious cycle. We validated 4 of these transcripts as riboregulators according to their small size, their strong secondary structure and their non-translation into protein (RNA1051, RNA1059, RNA1083 and RNA1111). In particular, we showed that RNA1111 was necessary for the virulence of *A. fabrum* C58, and that it seems to act through the posttranscriptional control of genes implicated in virulence functions and in Ti plasmid conjugation. A more moderated role of RNA1083 was also observed, potentially by the modulation of the bacterial mobility and of the plasmid conjugation. Furthermore, we highlighted two riboregulators, RNA1059 and RNA1051, involved in the control of the Ti plasmid persistence, through their implication in the replication of the plasmid (RNA1059) and in a toxinantitoxin system present on the Ti plasmid (RNA1051).

Thus, from a global analysis, we brought out the role of riboregulators in the control of several steps of the infectious cycle of *A. fabrum* C58, through the control of virulence factors, or through the control of the persistence of the main actor of the virulence, the Ti plasmid.

Key words: Agrobacterium, virulence, Ti plasmid, small non-coding RNA, riboregulation, posttranscriptional regulation, transcriptome.