

Université de Lille – Communauté d'Universités et d'Établissements Lille Nord de France

---

## Thèse de doctorat

soutenue le 28 septembre 2018

Discipline : Aspects Moléculaires et Cellulaires de la Biologie (DSPT 5)

Spécialité : Biochimie et Biologie Moléculaire

# Etude du transcriptome primaire codant et non-codant de *Bordetella pertussis*, caractérisation de l'impact des séquences d'insertion

Alexandre D'Halluin

Laboratoire de Transcriptomique et de Génomique Appliquée – CIIL – UMR8204  
U1019

Institut Pasteur de Lille - 1, rue du professeur Calmette 59000 Lille

### Composition du jury

<u>Rapporteurs :</u>	<b>Eliane HAJNSDORF</b>	DR CNRS, Université Paris Diderot 7
	<b>Florence HOMMAIS</b>	MCU, Université de Lyon 1
<u>Examineurs :</u>	<b>Brice FELDEN</b>	PR, Université de Rennes 1
	<b>Françoise JACOB- DUBUISSON</b>	DR CNRS, Université de Lille
	<b>Yves LEMOINE</b>	PR, Université de Lille
<u>Directeur de Thèse :</u>	<b>David HOT</b>	DR IPL, Université de Lille



## Avant-propos

Cette thèse a été financée entièrement par une bourse obtenue auprès de l'Université de Lille, sous la responsabilité de l'Ecole Doctorale Biologie Santé de Lille (« allocation de recherche »). Ce travail est le fruit d'une collaboration entre le laboratoire de Transcriptomique et de Génomique Appliquée de Lille, du laboratoire des Infections Respiratoires Bactérienne : Pertussis et Tuberculose de Lille, le laboratoire de Contrôle Post-Transcriptionnel de l'Expression des gènes de Prague, et du Groupe de Biochimie Théorique de Vienne.

Ce travail a été initié par le Dr. David Hot, travaillant sur la régulation transcriptomique de l'agent pathogène *Bordetella pertussis*. Le projet du transcriptome primaire a débuté peu de temps avant mon arrivée en laboratoire, à la suite d'une étude ayant permis la détection de 13 ARN régulateurs (Hot et al. 2011). J'ai initié mon travail au laboratoire durant mon année de Master 2, avec pour projet scientifique de détecter de nouveaux transcrits potentiellement régulateurs chez *Bordetella pertussis*.

La continuité de mon projet m'a donné la possibilité de bénéficier d'un échange scientifique entre mon laboratoire et l'équipe du Dr. Branislav Vecerek à Prague. J'ai ainsi pu aller pendant 2 mois à Prague afin d'apprendre des techniques de biologie moléculaire non maîtrisées dans mon équipe. Ce programme d'échange a été une expérience riche à la fois au niveau professionnel que personnel.

Les travaux décrits dans cette thèse ont été valorisés sous la forme d'un poster que j'ai été présenté au 5ème congrès international « Regulating with RNA in bacteria and Archaea » à Séville en mars 2018. Enfin mes travaux ont permis d'aboutir, au moment de la rédaction de cette thèse, à la publication d'un article dans le journal RNA Biology intitulé « Primary transcriptome analysis reveals importance of IS elements for the shaping of the transcriptional landscape of *Bordetella pertussis* » (RNA Biol. 2018 May 18:1-9. doi: 10.1080/15476286.2018.1462655) sur laquelle je suis en co-premier auteur avec Fabian Amman. Cet article de recherche correspond aux travaux du chapitre 1 et du chapitre 3 des résultats et discussions décrits dans cette thèse.

## Résumé

*Bordetella pertussis*, l'agent responsable de la coqueluche, provoque près de 200000 morts par an dans le monde. Malgré une forte couverture vaccinale, une réémergence de la maladie a été observée dans les pays développés, liée en partie à une adaptation à la pression vaccinale. Les souches capables d'échapper à la réponse immunitaire montrent des réarrangements génomiques importants, provoqués par des éléments génétiques mobiles (IS) présents en plus de 230 copies, qui pourraient impacter sur la transcription des gènes et ARN régulateurs du pathogène.

Pour élucider l'impact des IS sur le transcriptome global de *Bordetella pertussis*, nous avons d'abord déterminé le transcriptome codant et non-codant du pathogène par des approches de séquençage d'ARN couplées à des prédictions bioinformatiques. Cette étude a permis d'identifier les structures codantes mono- et polycistroniques, incluant des structures régulatrices telles que des riboswitches, un excludon et de longs 5' et 3'UTR chevauchants. Une cartographie de candidats ARN régulateurs non-codant a été édifiée à partir de nouveaux transcrits localisés en régions intergéniques (IGR) et de transcrits orientés en antisens de séquences codantes. Des prolongements de transcriptions des IS ont été observés, prenant leur origine de promoteurs internes à l'IS ou formés par insertion de celles-ci. Ces transcrits sont spécifiques d'une souche à l'autre du pathogène, et s'orientent en sens ou en antisens des gènes environnant, ou dans des IGR. Le potentiel caractère régulateur de ces transcrits a été étudié par la caractérisation et l'étude du mode d'action d'un ARN régulateur, BPnc264, orienté en antisens du gène de virulence *fim2*.

Mots-clés : *Bordetella pertussis*, Séquence d'insertion, Transcriptome primaire, ARN régulateurs, ARN non-codant

# Study of the coding and non-coding primary transcriptome of *Bordetella pertussis*, characterization of the impact of insertion sequences

## **Summary**

*Bordetella pertussis*, the causative agent of whooping cough, is responsible of more than 200000 deaths worldwide. Despite a high vaccine coverage in developed countries, a reemergence of the disease has been observed, which is in part linked to vaccine-pressure. Strains able to evade vaccine-induced immunity show high genome organization rearrangements, essentially due to mobile genetic elements (IS) present in more than 230 copies, which could impact on messenger and regulatory transcription of the pathogene.

To assess the impact of IS on the global transcriptome of *Bordetella pertussis*, we first determined the coding and non-coding primary transcriptome by a combination of different RNA-sequencing approaches and predictive bioinformatics analysis software packages. We identify mono- and polycistronic coding structures, including regulatory structures like riboswitches, excludon, and long overlapping 5' and 3'UTR. A list of candidates regulating transcripts (small RNA) has been mapped from new transcripts located in intergenic regions (IGR) and transcripts oriented in antisense of annotated coding sequences. Extended transcriptions emerging from IS elements have been observed, originating from internal promoters or newly formed promoters by insertion in a specific genomic region. Those transcripts can extend in sense or in antisense of the flanking gene, or in IGR. The regulatory function of those transcripts has been studied from the characterization and the mode of action of an extended regulatory RNA, BPnc264, oriented in antisense of the virulence gene *fim2*.

**Key-words** : *Bordetella pertussis*, Insertion Sequence, Primary transcriptome, small RNA, regulatory RNA

## Remerciements

Avant de commencer, je tiens à remercier les membres du jury pour avoir accepté de consacrer du temps à évaluer mon travail. J'ai pu lire certaines parties de vos travaux pour me mettre à niveau dans le monde des ARN régulateurs bactériens et de la coqueluche, et je suis très honoré de vous compter parmi mon jury de thèse.

Je tiens avant tout à remercier globalement tout mon entourage avant de rentrer dans les détails. Un énorme merci à vous tous, autant pour votre soutien, pour votre patience (surtout pour vous David ...) que pour vos conseils.

Un énorme merci à David, sans qui cela ne se serait jamais passé. Cela n'a pas du toujours être facile pour vous, mais sincèrement un grand merci pour vos enseignements, vos conseils, et votre dévouement. Je pense, et je le penserai toujours, avoir été formé par l'une des personnes les plus humaines et les plus ouvertes que j'aurai rencontré dans le domaine de la science. Un énorme merci pour votre enseignement, votre suivi, et votre dévouement. Cela m'a beaucoup apporté, et je vous en serai toujours reconnaissant.

Je tiens évidemment à remercier toute l'équipe, les nouveaux comme les anciens ! De l'équipe TAG, j'aimerais dire un grand merci à Peggy pour le soutien moral, les bonnes discussions et toute l'aide que tu m'as apporté pour les manip' ! Je pense bien évidemment aux cultures, aux qRT-PCR, et à toutes les fois où tu m'as grandement aidé. A toutes ces heures passées au P2 pour les cultures de biofilms et les tests de demi-vie, on s'en souviendra longtemps !

Ségolène, le prodige de la bioinformatique, c'était toujours un plaisir de manger avec toi le midi ! Je te serai toujours reconnaissant de tes conseils et de ton aide, et d'avoir partagé ton expérience. Tu as réussi à me redonner confiance en moi et à me donner envie d'aller de l'avant quand ça n'allait pas vraiment !

Un grand merci également à toutes les personnes du labo, notamment Florence (Merci de m'avoir initié à cutadapt) et Delphine (Merci pour les commandes), et également à deux personnes qui m'ont formé aux manipulations, Ludo et Steph. Je n'oublierai pas tous ces bons moments partagés avec vous. Ludo, merci pour tous ces cafés, ces conseils, et ces formations, j'espère toujours que je gagnerai un jour ce restau de sushis ! Et Steph, un énorme merci pour ta rigueur et ta bonne humeur, c'est grâce à toi que mes ARN sont intacts à chaque extraction. Un merci également aux deux nouveaux arrivants, Cécile et Franck, les personnes les plus motivées que je n'ai jamais vu pour aller faire du sport le midi (En plus de Ludo) !

J'aimerais également remercier toutes les personnes de Gène Diffusion de la plateforme Pégase, qui m'ont permis de faire passer le travail pour une partie de plaisir. Malgré toute attente, merci cher Christophe pour toutes ces années où, quelque part, tu as réussi à m'endurcir et à paraître un peu moins bête ! Un grand merci à Sophie de la table ronde pour ces années de déconnes, je sais que je te manquerai ... Tout comme la petite Sophie, vous allez embêter qui après ça ? Haha. Merci également à Gaël, qui

me fera toujours autant rire. J'ai du mal à parler de Gaël sans penser à Alex avec qui il faisait la paire, comme Laurel et Hardy. Merci à toi également, tu m'as bien fais rire et j'aimais beaucoup les midis au parc ou au fast-food que tu proposais. Merci à ton remplaçant, Pierre, qui sait résoudre n'importe quel problème d'informatique, et donner des conseils, notamment quand ton PC est sur le point de décéder quand tu rédiges ta thèse. Merci à toi à pour toutes ces discussions scientifiques et fantaisistes, et merci à toi aussi Tony pour les avoir enrichis. Tu resteras toujours l'homme le plus heureux sur Terre qui va mieux que tout le monde ... Haha.

Je tiens également à remercier tous les autres étudiants et post-doc que j'ai pu rencontrer, avec qui j'ai pu énormément discuter, boire des verres, ou autre, notamment Amandine, Léa, Alex, Olivia, etc ... Et à nos petits stagiaires qui m'ont fait rire ! Si je ne mets pas tous les noms, c'est que vous êtes trop nombreux, mais je me souviendrai de chacun d'entre vous !

J'ai également une grande pensée pour nos collaborateurs. J'aimerais beaucoup remercier Camille Locht, Rudy Antoine et Loïc Coutte pour nos réunions, vos conseils et votre aide pour nos projets. J'ai une pensée également pour chaque personne de votre laboratoire qui a pu m'aider pour me former à certaines cultures, à l'extraction protéique, Western Blot, et également pour toutes les fois où vous avez pu me dépanner, merci à vous tous.

Un énorme merci également à tous les tchèques que j'ai pu rencontrer ! Un grand merci à Branko, qui a pu m'accueillir pendant un mois et demi, qui m'a fait découvrir Prague et qui a pu me former a des techniques encore inconnues pour moi. Je compte bien reprendre un verre avec vous ! Un grand merci également à toute son équipe, Honza, Kuba, Denisa ... Je me suis bien amusé avec vous également.

Evidemment, je veux remercier tous mes amis pour m'avoir soutenu et supporté durant toutes ces années. A chaque fin de journée difficile, merci à tous ceux qui ont contribué à l'augmentation de mon alcoolémie. Un énorme merci à Vianney, Camel, Loïc, Athina, Manon, Bertrand, Amélie, Anne-Sophie, Justine, Vanessa, Thibaut, Rudy, Guillaume, Gaëtan, Corentin, etc .... Vous êtes très nombreux, mais je vous aime tous.

Un grand merci à ma famille également, pour les mêmes raisons, notamment à ma cousine Lynda et son copain Enzo, et aussi ma tante Brigitte. Merci de votre soutien et des bons moments passés avec vous.

Le meilleur pour la fin ? Un grand merci à celle qui a toujours été là pour moi, qui m'a toujours soutenu, qui m'a donné l'envie de réaliser mes rêves ... qui m'a donné l'opportunité de réaliser mes rêves contre toute attente. Je ne sais pas comment remercier ma mère, qui m'a toujours relevé quand j'étais au plus bas et qui m'encourageait quand je n'étais sûr de rien. Tu m'as laissé vivre la vie que je voulais, laissé faire mes propres choix tout en m'apportant toute l'attention qu'il fallait, et m'as poussé à concrétiser mes rêves et mes envies. Un énorme merci à toi, maman. C'est vraiment grâce à toi que j'en suis là aujourd'hui.

Merci à vous tous

## Abréviations

Abréviation	Signification	Abréviation	Signification
5' / 3' UTR	Région non traduite en 5' ou 3' d'ARN messenger	L / M	Marqueur de taille
5'RACE	Amplification rapide des extrémités 5' d'ADN complémentaire	LOS	Lipooligosaccharides
A / T / G / C	Adénine / Thymidine / Guanine / Cytosine	LPS	Lipopolysaccharides
ADN	Acide Désoxyribonucléique	M / mM	Molaire / millimolaire
ADNg / ADNc	ADN génomique / ADN complémentaire	Mg	Condition de culture liquide en présence de sulfate de magnésium
Am	Ampliciline	MgSO <sub>4</sub>	Sulfate de magnésium
ARNm	ARN messenger	NaOH	Hydroxyde de Sodium
ARNr	ARN ribosomique	NCES	Etude des encéphalopathies nationales de l'enfant
ARNt	ARN de transfert	Nt	Nucléotide
ARNtm	ARN transfert-messenger	OMS	Organisation Mondiale de la Santé
ATP / CTP / GTP / TTP	Adénine / Thymidine / Guanine / Cytosine TriPhosphate	PAMPs	Motifs moléculaires associés aux pathogènes
BG	Milieu de culture solide Bordet-Gengou	Pb/kb	Paire de bases / Kilopaires de bases
BPncXXX	Bordetella pertussis ARN non codant numéro XXX	PCI	Purification au Phénol-Chloroforme-Isoamyl Alcool
BpSM / SM	Bordetella pertussis Tohama I résistant à la streptomycine	PCR	Réaction en chaîne par polymérase
BPTXX	Bordetella pertussis ARNt numéro XX	P <sub>hyb</sub>	Promoteur hybride
Bvg+ / Bvgi / Bvg-	Phase de gène de virulence de Bordetella +/-	P <sub>in</sub>	Promoteur de l'ARN messenger de la transposase des IS
BvgA / BvgS / BvgR	Activateur / Senseur / Répresseur de gènes de virulence de Bordetella	P <sub>out</sub>	Promoteur de l'ARN régulateur des ARNm de la transposase des IS
BvgA-P / RisA-P	BvgA ou RisA phosphorylé	p-val	Probabilité d'hypothèse nulle
CDC	Centre de Contrôle des Maladies	qRT-PCR	Réaction de RT-PCR quantitative
c-di-GMP	Dicyclique GMP	RBS	Séquence de fixation du ribosome
Cm	Chloramphenicol	Rec	Domaine Receveur
Ct	Cycle Seuil de qRT-PCR	RisA / RisS / RisK	Activateur / Senseur / Kinase du Régulateur Intracellulaire de la réponse au Stress
DO	Densité Optique	RNA-IN	ARN messenger du gene de la transposase de l'IS10
DTaP	Vaccin acellulaire anti diphtérie-tétanos-coqueluche	RNA-OUT	ARN régulateur transcript en antisens du RNA-IN
DTwP	Vaccin à germe entier anti diphtérie-tétanos-coqueluche	rpm	Rotation par minute
EMSA	Retard sur gel	RT-PCR	PCR à partir de produits de transcription inverse
ETS	Espaceur transcriptionnel externe	SAH	S-Adénosylhomocystéine
E-Value	Valeur de confiance E	SAM	S-Adénosylméthionine
F/R (g/t/is/s/c)	Amorce Sens / Antisens (Gène, Transcrit, IS, Gène sens, Candidat)	Sm	Streptomycine
FHA	Hémagglutinine filamenteuse	SRP	Particule de Reconnaissance du Signal
FMN	Flavine MonoNucléotide	SS	Milieu de culture liquide Stainer-Scholte
Gm	Gentamycine	Tdap	Rappel du vaccin acellulaire DTaP
GMP	Guanine MonoPhosphate	TEX	5' Monophosphate Exonucléase
H	Heure	Th1 / Th2 / Th17	Réponse Immunitaire stimulant les lymphocytes T auxiliaires de type 1, 2 ou 17
Hfq	Facteur de l'Hôte de l'ARN du bactériophage Q $\beta$	TPP	Thiamine PyroPhosphate
HK	Domaine Histidine Kinase	TSS	Site de départ de transcription
Hpt	Domaine Histidine Phosphotransmetteur	U	Unité
IgG	Immunoglobuline G	UA	Unité Arbitraire
IS	Séquence d'insertion	vag	Gène activé en phase de virulence
ITS	Espaceur transcriptionnel interne	VFT	Domaine Vénus-Fly Trap
Kcal/mol	Kilocalorie par moles	vol	Volume
KDa	Kilodalton	vrg	Gène réprimé en phase de virulence
Km	Kanamycine	$\Delta\Delta$ Ct	Différence entre deux souches de la différence de Ct

## Contributions

La majorité des résultats présentés dans cette thèse sont issus d'expériences que j'ai pu réaliser moi-même au sein de mon laboratoire. Cependant, l'analyse du transcriptome primaire de *Bordetella pertussis* est un projet qui a été initié avant mon arrivée au laboratoire de Transcriptomique et de Génomique Appliquée par les Dr. David Hot, Branislav Vecerek et Fabian Amman. Certains travaux présentés dans cette thèse sont donc des résultats obtenus soit par des membres de mon équipe, soit par des collaborateurs.

Pour le premier chapitre des résultats, les ARN ont été séquencés en séquençage différentiel et profond par Ludovic Huot. Les données obtenues ont été analysées par notre collaborateur Fabian Amman de Vienne qui a développé le cheminement de traitement des données afin de prédire les structures mono- et polycistroniques (Chapitre 1) et les nouveaux candidats ARN régulateurs (Chapitre 2). Mon travail a donc été d'analyser les données générées, de valider ces prédictions par des expériences biologiques et de déterminer le caractère des transcrits prédits. Les TSS prédits chez *Bordetella pertussis* en phase Bvg<sup>+</sup> ont également été prédits par Fabian Amman, ainsi que la prédiction des motifs -10 et -35 des promoteurs. Les riboswitches prédits par Infernal ont été déterminés par Rudy Antoine et les transcrits dits « leaderless » ont été déterminés par David Hot. Les TSS ont été validés par 5'RACE en séquençage à Haut Débit par Peggy Bouquet.

Dans le deuxième chapitre des résultats, Fabian Amman a également déterminé la répartition des prédictions d'ARN régulateurs indépendant chez l'ensemble des Burkholdériales. Les logiciels SIPHT, NAPP et RNAz ont été appliqués sur le génome de *Bordetella pertussis* à la fois par notre laboratoire (David Hot), par notre collaborateur de Prague (Branislav Vecerek) et notre collaborateur de Vienne (Fabian Amman).

Enfin, concernant le troisième chapitre des résultats, la répartition des IS481 parmi 21 souches de *Bordetella pertussis* a été déterminée par notre collaborateur Rudy Antoine. La validation des départs de transcription par 5'RACE en séquençage à Haut Débit dans les chapitres 3 et 4 a été réalisée par ma collègue Peggy Bouquet.



## Table des matières

Introduction.....	15
I. Réémergence de la coqueluche : Des facteurs multiples.....	16
1. Description de la maladie.....	16
2. Programmes de vaccination.....	19
3. La recrudescence de la coqueluche.....	21
II. Les facteurs de virulence du pathogène.....	27
1. La régulation par les systèmes à deux composants.....	27
2. Les gènes de virulence de <i>Bordetella pertussis</i> .....	31
III. Les disparités génomiques inter-souche.....	34
1. Caractéristiques génomiques de <i>Bordetella pertussis</i> .....	34
2. Les séquences d'insertion, un facteur d'adaptation.....	38
3. La régulation des IS et l'expression de gènes environnant.....	40
IV. Les objectifs de thèse.....	43
Résultats et Discussions.....	47
I. Annotation du transcriptome primaire de <i>Bordetella pertussis</i> .....	48
1. Annotation du transcriptome en condition Bvg <sup>+</sup> .....	48
2. Identification de l'architecture et des structures ARN.....	51
3. Discussion : Architecture et régulateurs non-indépendants.....	60
II. Mise en évidence des ARN régulateurs indépendant.....	64
1. Edification d'une liste de prédictions d'ARN régulateurs non codants.....	64
2. Sélection et validation des prédictions d'ARN régulateurs indépendant.....	67
3. Discussion : Un complexe de régulation encore inexploré.....	75
III. Impact des IS481 sur la transcriptome de <i>Bordetella pertussis</i> .....	78
1. Les prolongements de transcription issus des séquences d'insertion.....	78
2. Les prolongements d'IS et les diversités transcriptomiques inter-souches.....	82
3. Discussion : Une régulation transcriptomique par les IS481 ?.....	86
IV. Caractérisation de BPnc264, ARN régulateur orienté en antisens de BP1118 et de <i>fim2</i> .....	89
1. Identification de BPnc264.....	89
2. Double interaction et régulation de BP1118.....	92
3. Régulation de l'ARN messenger de <i>fim2</i> dans la cellule.....	94
4. Discussion : Les IS481, un réservoir de double ARN régulateurs potentiels.....	100
Discussion et perspectives.....	103
Matériel et Méthodes.....	106
Bibliographie.....	122

<b>Figures Supplémentaires</b> .....	138
<b>Figure Supplémentaire S1</b> : .....	139
<b>Figure Supplémentaire S2</b> : .....	147
<b>Figure Supplémentaire S3</b> : .....	150
<b>Figure Supplémentaire S4</b> : .....	151
<b>Figure Supplémentaire S5</b> : .....	153
<b>Figure supplémentaire S6</b> : .....	155
<b>Figure Supplémentaire S7</b> : .....	156
<b>Figure Supplémentaire S8</b> : .....	157
<b>Figure Supplémentaire S9</b> : .....	158



# Introduction

## I. Réémergence de la coqueluche : Des facteurs multiples

Selon l'Organisation Mondiale de la Santé (OMS), les maladies transmissibles sont l'une des causes majeures de mortalité. Parmi les 57 millions de décès recensés en 2015, l'OMS a comptabilisé 12 millions de cas mortels de maladie transmissible et dont 3,2 millions dus à des maladies infectieuses respiratoires, les classant en tant que maladies transmissibles les plus mortelles (WHO, 2017). Elles comprennent de graves maladies comme la tuberculose, la rubéole, la rougeole, ou encore la coqueluche. Cette dernière, appelée « pertussis » en anglais, est une maladie infantile, causée par une bactérie pathogène du nom de *Bordetella pertussis*, responsable de violentes toux pouvant persister plusieurs semaines (Kuchar et al., 2016). Elle infecte les cellules épithéliales ciliées du tractus respiratoire inférieur, ce qui entraîne une sécrétion de mucus abondante, et cause des toux spasmodiques prononcées (Figure 1) (Soane et al., 2000). Les individus en bas âge, encore en apprentissage de la déglutition, peuvent dans certains cas s'étouffer par obstruction du système respiratoire, entraînant la mort (Paddock et al., 2008).

### 1. Description de la maladie

#### i. Découverte de *Bordetella pertussis*

La première description de la coqueluche date de 1578, par Guillaume de Baillou, qui établit une première description clinique de quintes de toux virulentes après une épidémie d'un violent rhume sévissant à Paris (Guiso, 2006). Il appelle cette maladie *tussis quinta* ou encore *tussis quintana*, et la relie à une maladie affectant préférentiellement les individus en bas âge. En 1679, Thomas Sydenham donne le premier nom latin de l'agent pathogène, *infantum pertussis*, signifiant violente (*per*) toux (*tussis*) de l'enfant (*infantum*) (Cherry, 1996). Le caractère épidémique de la maladie a ensuite été décrit trois années plus tard par Thomas Willis, qui la renomme, après observation d'épisodes de convulsion tussif chez des patients atteints, en *Tussis puerorum convulsiva* (Guiso, 2006). Le développement industriel et le commerce entre pays lointains propagent la maladie à partir du XIX<sup>ème</sup> siècle, notamment en Suède et aux Etats-Unis, qualifiant la coqueluche comme la maladie la plus mortelle, avec la rougeole (Guiso, 2006). Il est estimé qu'à cette époque près d'un enfant sur mille mourrait de la coqueluche. En 1816, Adalbert Friedrich Marcus définit les différentes dénominations de la coqueluche dans plusieurs pays du monde, encore appelé chant du coq en français (Marcus, 1821). La



**Figure 1:** Capture de microscopie électronique modifiée de *Bordetella pertussis* (cellules jaunes) infectant une cellule épithéliale ciliée (cellule rouge). La bactérie s'agglutine au niveau des cils tapissant le tractus respiratoire de l'hôte.

Copyright © 2004 Pearson Education, publié par Benjamin Cummings

maladie tirerait son nom d'une croyance de l'époque selon lui : « On croyait autrefois qu'elle dépendait d'humeurs qui découlaient de la tête, et qu'on pouvait y remédier en tenant la tête chaudement, au moyen d'un bonnet pareil à un capuchon de moine, qu'on appelait coqueluche ».

Il faudra attendre en 1906 la publication de Jules Bordet et d'Octave Gengou, qui décrit la composition d'un milieu de culture portant leur nom, permettant d'isoler le pathogène (Bordet and Gengou, 1906). Les bactéries ont été isolées sur le milieu Bordet-Gengou (BG) supplémenté en sang à partir d'expectorations d'un nourrisson atteint de la maladie. Etant donné son caractère hémolytique et sa morphologie, ils renomment la bactérie *Haemophilus pertussis* ou encore *Bacillus pertussis*. C'est en 1954 que Moreno Lopez donne la dénomination actuelle *Bordetella pertussis*, en hommage à la découverte de Jules Bordet, en décrivant le genre *Bordetelleae* (Moreno López, 1990).

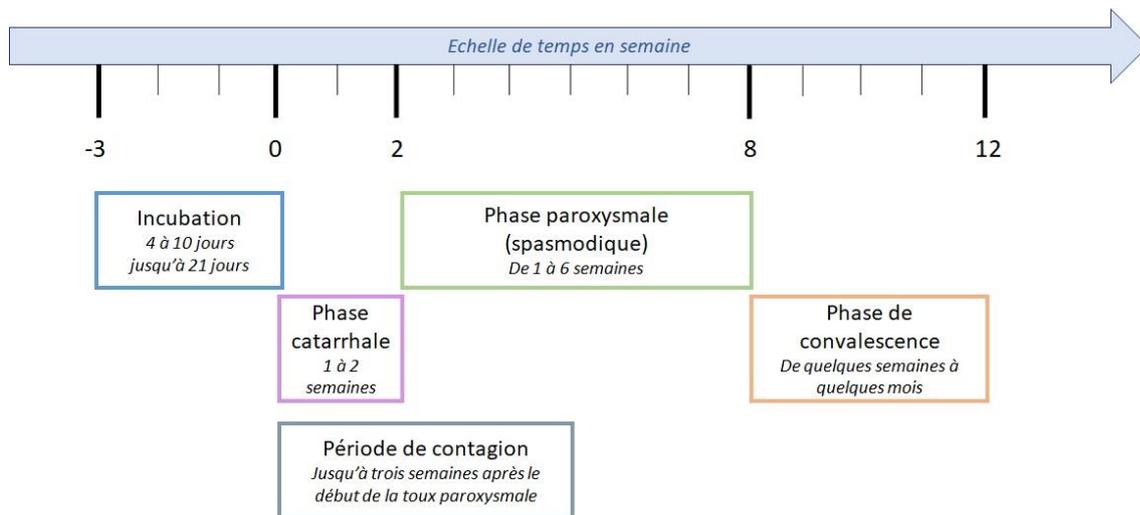
## ii. Description du pathogène

*Bordetella pertussis* est ainsi décrite comme étant un coccobacille à Gram négatif, doté d'une double paroi lipidique séparée par un périplasme (Bordet and Gengou, 1906). Sa taille moyenne est de 0,8 micron sur 0,4 micron. C'est une bactérie aérobique stricte, ne produisant pas de spores, et qui survit dans la nature à 37°C. De plus, c'est une bactérie qui, malgré la présence de flagelles, n'est pas motile (Parkhill et al., 2003). Ces caractéristiques laissent supposer que *Bordetella pertussis* ne serait pas capable de survivre en dehors de son hôte. A l'heure actuelle, aucune autre niche écologique, autre que l'être humain, n'a pu être identifiée. Malgré ces contraintes, la coqueluche serait responsable de 200000 à 400000 morts par an dans le monde, principalement des nourrissons, pour environ 45 million de cas (Crowcroft et al., 2003). En 2015, l'OMS a pu estimer et recenser près de 65000 décès dus à la coqueluche (WHO, 2017).

*Bordetella pertussis* est une bêta-protéobactérie de l'ordre des *Burkholderiales* et de la famille des *Alcaligenaceae* (Gerlach et al., 2001). Le genre *Bordetelleae* comporte 9 espèces distinctes, dont la quasi-totalité du genre est pathogène, principalement par infection du tractus respiratoire. *Bordetella bronchiseptica* et *Bordetella parapertussis* sont considérées, avec *Bordetella pertussis*, comme des sous-espèces de par leur proximité génomique (Parkhill et al., 2003). Ces deux autres sous-espèces peuvent également infecter l'homme, mais sont principalement pathogènes de plusieurs mammifères, notamment le chien, le porc et le mouton. D'autres, comme *B. avium* et *B. hinzii* infectent les oiseaux. La dernière espèce découverte récemment, *B. petrii*, une bactérie environnementale colonisant les bioréacteurs, a été associée à un cas clinique d'ostéomyélite mandibulaire (Fry et al., 2005).

## iii. Les phases de la maladie

Il a été montré que la maladie se transmettait préférentiellement par les microgouttelettes en aérosols dégagées lors de quintes de toux (Fernandez, 2012; Warfel et al., 2012). Cette transmission du pathogène d'un hôte vers un autre se fait de préférence de l'adulte vers l'enfant, plutôt que de l'enfant vers l'enfant, ou de l'adulte



**Figure 2 :** Représentation schématique de l'évolution clinique de la coqueluche. Les échelles de temps représentent un cas classique de patient atteint de la maladie. Schéma modifié du Centre de Contrôle des Maladies (CDC) (<https://www.cdc.gov/pertussis>)

vers l'adulte (Bisgard et al., 2004; Skoff et al., 2015; Wendelboe et al., 2007). L'infection touche le tractus respiratoire supérieur, la trachée et les bronches où s'agglomèrent les bactéries (van den Berg et al., 1999). Le pathogène adhère aux cellules épithéliales et, dans un premier temps, se multiplie activement, engendrant le syndrome infectieux. Dans un second temps, le pathogène sécrète des toxines pour engendrer la mort cellulaire de l'épithélium, où sont relâchés les métabolites nécessaires à sa survie, et dérive vers le syndrome toxinique. La maladie se développe en 10 à 12 semaines environ après infection, et est divisée en trois différentes phases (Figure 2) (Melvin et al., 2014).

La première phase est la phase catarrhale. Elle survient après un temps d'incubation, soit 4 à 10 jours après l'infection généralement, et persiste entre 1 et 2 semaines. A cette étape, les bactéries ayant adhérées à l'épithélium cilié vont survivre au dégagement de mucus sécrété par l'hôte, et contourner les défenses du système immunitaire. Les symptômes se confondent généralement avec des maladies respiratoires communes, comme le rhume ou la rhinite (Cornia et al., 2010; Wirsing Von König, 2014). Ils comprennent alors des rhinorrhées, une légère toux, et des éternuements, parfois accompagnés de fièvre. Ces symptômes cliniques rendent difficile le diagnostic de la coqueluche par les médecins généralistes. C'est pourtant durant cette phase que la maladie reste la plus contagieuse et donc, la plus à risque pour les nourrissons et les personnes asthmatiques.

S'ensuit la phase paroxysmale, également appelée phase spasmodique. C'est l'étape la plus longue de la maladie, elle peut durer 1 semaine au minimum, et jusqu'à 10 semaines. Elle se caractérise par des périodes d'alternance entre des respirations claires et normales, et des toux spasmodiques, souvent très violentes. Cette toux lui a valu le nom de « chant du coq » de par les quintes très caractéristiques et les inspirations très bruyantes de la respiration. Souvent, ces toux répétées ont lieu la nuit, empêchant le malade de dormir et altérant le repos (Tozzi et al., 2005). Elle peut s'aggraver dans

certains cas et s'accompagner de vomissements. La sécrétion de mucus étant de plus en plus importante, les individus en bas âge et les personnes asthmatiques, ayant des difficultés à déglutir de façon répétée, peuvent s'étouffer et bloquer leur respiration, causant dans certains cas la mort par apnée et cyanose (Paddock et al., 2008). Le traitement lors de cette phase nécessite l'intervention de kinésithérapeutes, qui aident le patient à se dégager le tractus respiratoire de l'excès de mucus. Cette phase est souvent associée à de sévères lymphocytoses et à de l'hypertension pulmonaire, causant des pneumonies et parfois même des convulsions et encéphalopathies (Paddock et al., 2008).

La dernière phase de la maladie est la phase de convalescence. Sa durée est plus variable, souvent de 2 à 3 semaines, mais peut perdurer jusqu'à plusieurs mois. Les symptômes s'amenuisent au court du temps, les quintes de toux sont de plus en plus espacées, le malade parvient à déglutir plus facilement. Lors de cette phase, la toux, qui est alors devenue plus sèche, est propice à la transmission de l'adulte vers l'enfant sans traitement approprié (Tozzi et al., 2005). En général, les antibiotiques efficaces, comme l'érythromycine, sont prescrits au cours de la phase de convalescence pour éviter cette transmission. La toux étant provoquée principalement par l'action des toxines sécrétées par le pathogène, les antibiotiques ne peuvent amener à la réduction des symptômes de la maladie.

## **2. Programmes de vaccination**

Après la découverte du pathogène par Jules Bordet et Octave Gengou, plusieurs scientifiques se sont intéressés à la production d'un vaccin efficace pour éradiquer la coqueluche. Les premières tentatives se sont concentrées sur le développement d'un vaccin à partir de cellules mortes, bien que celui-ci ait été testé pour soigner la coqueluche chez les personnes atteintes plutôt que pour prévenir la propagation du pathogène dans la population humaine (Cherry, 1996).

### **i. La production d'un vaccin composé de cellules mortes**

Jules Bordet et Octave Gengou de l'Institut Pasteur de Bruxelles, à la suite de leur découverte, proposèrent un vaccin à partir de cellules mortes en 1912 (Howson et al., 1991). S'ensuivirent Charles Nicolle de l'Institut Pasteur de Tunis en 1913 et Thorvald Madsen du Danish Serum State Institut en 1914. C'est cette année-là que le premier vaccin constitué de cellules mortes de *Bordetella pertussis* a été enregistré à l'Association Médicale Américaine, alors que les effets du vaccin n'étaient pas encore caractérisés (Howson et al., 1991).

Il faudra ensuite attendre 1925 pour que Thorvald Madsen décrive la première utilité d'un vaccin composé de cellules vivantes conservées dans une solution saline contenant du phénol (Howson et al., 1991). Par ses pratiques, de 1925 à 1933, il réussit à instaurer un vaccin capable d'induire une protection immunitaire efficace et de prévenir l'infection de *Bordetella pertussis*. Il aurait même permis de prévenir deux épidémies de coqueluche aux Iles Féroées. Cependant, en 1933, il reporte que 2 enfants en bas âge sont décédés dans les 48h après injection du vaccin.

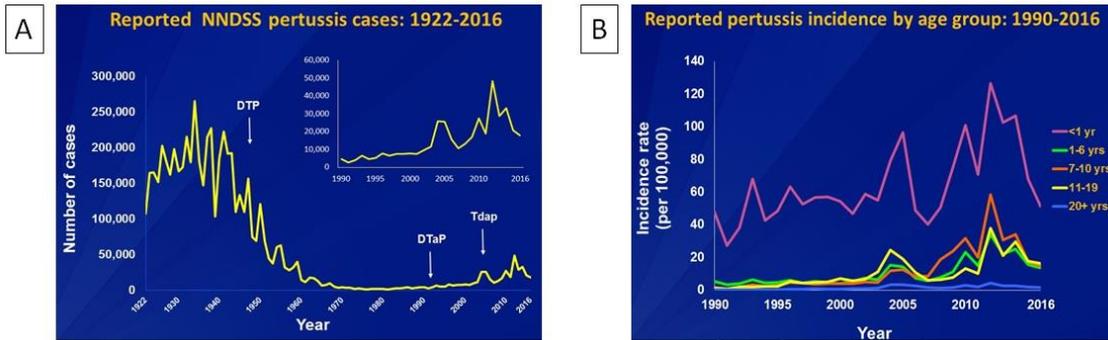
Dans les années 1930 jusqu'en 1943, Pearl Kendrick et Grace Eldering concevaient un nouveau vaccin anti-coqueluche après avoir étudié l'immunité induite par le pathogène (Kendrick and Eldering, 1936; Kendrick, 1942). Elles proposèrent un renouveau, en combinant le vaccin anti-coqueluche composé également de cellules mortes, avec les toxoïdes du vaccin déjà existant anti-diphtérique et anti-tétanique, combinaison encore utilisée aujourd'hui. Le vaccin, ayant une efficacité estimée de 89%, a ensuite été produit et commercialisé dans plusieurs pays du monde, résultant en une diminution drastique du nombre de cas de coqueluche à cette époque (Wadman and You, 2017). Ce vaccin, le DTwP, est administré trois fois dans la vie de l'enfant pour l'immuniser contre la coqueluche.

## ii. Transition vers le vaccin acellulaire : Des effets secondaires ?

En 1974, une étude de chercheurs anglais sous-entend un lien de causalité entre le vaccin à cellule entière et des complications neurologiques apparus chez 36 nourrissons 24h après les trois administrations du DTwP (Kulenkampff et al., 1974). Les scientifiques conseillent, alors, de ne pas administrer ce vaccin s'il existe des antécédents familiaux reliés à l'autisme ou autre désordre neurologique. Cette étude a été médiatisée et relayée dans la presse, résultant en la formation de mouvements anti-vaccin et d'associations voulant abolir le DTwP.

Dans les années qui suivirent, le Japon et la Suède ordonnèrent l'interdiction d'administrer le vaccin anti-coqueluche (Gangarosa et al., 1998). En Angleterre naît le NCES (National Childhood Encephalopathy Study) pour appuyer ces études sur les désordres neurologiques causés chez les enfants, et soutient la publication de Kulenkampff (Miller et al., 1993). Ces conclusions conduisirent à une augmentation spectaculaire du nombre de cas de coqueluche au Japon, en Suède, en Angleterre, en Irlande, en Australie et en Italie. La diminution de la couverture vaccinale résulta en l'apparition de deux épidémies de coqueluche en Angleterre entre 1978 et 1982, et également en Suède en 1983 et 1985 (Kuchar et al., 2016).

Pour faire face à la recrudescence de coqueluche dans plusieurs pays du globe, des équipes de recherche se sont intéressées à développer un vaccin qui ne serait être relié à des effets secondaires graves. Dans les années 1980 apparaissent alors des vaccins composés d'antigènes de *Bordetella pertussis* (Klein, 2014). Ils comprennent généralement 1 à 5 protéines inactives, majoritairement composés d'adhésines et de toxines. Ce vaccin a été accepté plus facilement par la population et a résulté en une diminution de cas de coqueluche à travers le monde. Ce vaccin acellulaire est également combiné avec des toxoïdes de Diphtérie et de Tétanos (DTaP). Le premier a été développé en 1981 au Japon, et contenait deux antigènes inactivés au formaldéhyde et au formalin (Sato and Sato, 1999). Quelques années plus tard en 2006, un booster fut créé pour une administration entre 11 et 18 ans chez les adolescents, composé de toxoïdes tétaniques, de toxoïdes diphtériques réduites et du vaccin acellulaire anti-coqueluche (Tdap) (Kline et al., 2013).



**Figure 3:** Incidence de cas de coqueluche aux Etats-Unis.

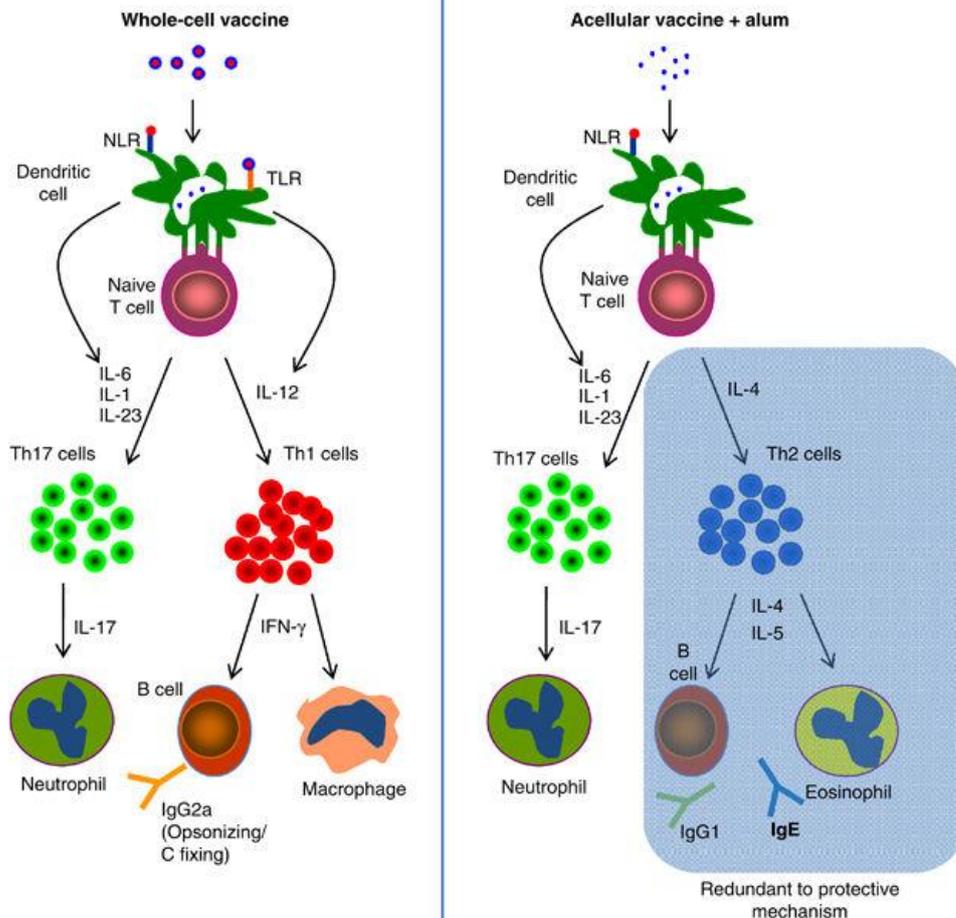
(A) : Nombre de cas de coqueluche recensé aux Etats-Unis entre 1922 et 2016. DTP : Vaccin composé de cellules mortes. DTaP : Vaccin acellulaire. Tdap : Booster du vaccin acellulaire. (B) : Incidence de cas de coqueluche par tranche d'âge pour 100000 individus entre 1990 et 2016. Figure tiré du CDC (<https://www.cdc.gov/pertussis>)

### 3. La recrudescence de la coqueluche

#### i. 1<sup>ère</sup> cause : Le déclin de la réponse immunitaire

Les vaccins à cellules mortes et acellulaires ont résulté en une diminution drastique de cas de coqueluche dans le monde. Cependant, depuis le début des années 2000, une recrudescence du nombre de cas et de mortalité a été recensée malgré une forte couverture vaccinale (Figure 3A) (Kline et al., 2013; Poolman, 2014; Sealey et al., 2016). Elle touche principalement les individus en bas âge, chez qui la maladie est mortelle, et les adolescents (Figure 3B). Très peu de cas ont été recensés chez les adultes. Pour comprendre ce phénomène, des études épidémiologiques et immunologiques ont été entreprises dans les pays développés. L'étude de cas cliniques, d'infections *in vivo* chez la souris et chez le babouin ont soulevé l'hypothèse de la diminution de la durabilité de la protection immunitaire (« waning »). Les cellules immunitaires permettent de contrôler l'infection, tandis que l'éradication de la maladie nécessite l'intervention des cellules T auxiliaires (« T-helper ») 1 et 17 (Higgs et al., 2012).

Des études d'infection de cellules immunitaires humaines et de souris *in vivo* ont permis de comprendre les mécanismes d'action des deux vaccins, et les réponses immunitaires impliquées dans l'éradication de la maladie (Park et al., 2012). Les toxines et les adhésines du pathogène lui permettent de contrer les défenses immunitaires et de coloniser et lyser les cellules épithéliales ciliées (de Gouw et al., 2011). La reconnaissance de ces cibles par le système immunitaire va induire différentes réponses en fonction des récepteurs impliqués et interleukines produites (Higgs et al., 2012). Les toxines et adhésines majeures sont reconnues par les récepteurs Nod-like, qui tend vers une réponse Th17, impliquée dans le recrutement des neutrophiles et l'inflammation (Figure 4). Les lipopolysaccharides et les autres PAMPs sont reconnus par les récepteurs Toll-like des cellules dendritiques et les macrophages qui produisent alors l'interleukin-12 (Mahon et al., 1996). Cette cytokine pro-inflammatoire induit une différenciation des lymphocytes T vers des cellules Th1, pour privilégier le recrutement de macrophages supplémentaires et l'activation de neutrophiles, nécessaires à l'élimination des



**Figure 4:** Réponses immunitaires stimulées suite à l'injection du DTwP ou du DTaP. Le vaccin à germe entier est reconnu par des Toll-Like Receptor stimulant la production d'IL-12, pour activer une réponse Th1. Le vaccin acellulaire n'étant pas reconnu par les TLR, la production de l'IL-4 par les cellules T entraîne une réponse de type Th2. Les deux composants des vaccins sont reconnus par les Nod-Like Receptor stimulant une réponse de type Th17. Figure tirée de Higgs *et al.*, 2012

bactéries pathogènes. Les cellules Th1 vont également stimuler la production d'anticorps IgG2a, un anticorps opsonisant, pour marquer le pathogène reconnu par les phagocytes. Ces réponses sont retrouvées après injection du vaccin à germes entier chez des souris *in vivo* et des préadolescents. Le vaccin acellulaire contient des antigènes reconnus par ces systèmes également. Cependant, le vaccin acellulaire ne stimule aucune production d'interleukine-12, mais celle de l'interleukine-4 qui tend vers une différenciation des lymphocytes T vers une réponse Th2. Elle stimule l'activation des éosinophiles et la production d'anticorps IgG1, IgG3 et IgE impliqués dans la reconnaissance de toxines et des adhésines pour empêcher une infection et une adhésion de bactéries pathogènes. Cette réponse correspond à un mécanisme de protection, cependant les lipopolysaccharides présents dans les vaccins acellulaires ne sont pas reconnus par les récepteurs Toll-like (Zaitsev *et al.*, 2014). L'inactivation du LPS par le formol change la nature de la molécule qui n'est pas reconnue.

Ces études sous-entendent en globalité que la protection immunitaire induite par le vaccin acellulaire serait aussi efficace que celle induite par le vaccin à germes entiers, mais que celle-ci perdurerait moins longtemps. Ce phénomène est encore appelé « Waning » en anglais, et serait, à l'heure actuelle, l'une des causes principales de la réémergence du pathogène. Ces études se confirment avec des études cliniques et épidémiologiques, notamment avec une épidémie qui s'est déclarée aux Etats-Unis en 2010 (Latasa et al., 2018; Sheridan et al., 2014; Witt et al., 2012). La prévention de la maladie par un vaccin acellulaire est de ce fait remise en cause, notamment après observation de l'augmentation de cas de coqueluche chez les individus âgés de 8 à 12 ans.

De nos jours, des appels à la communauté scientifique à développer de nouveaux vaccins plus efficaces se font entendre. Une troisième génération de vaccin serait inéluctable, mais nécessiterait potentiellement un retour vers un vaccin à germes entiers, ou à un changement des adjuvants du vaccin acellulaire (Higgs et al., 2012; Locht, 2016; Locht and Mielcarek, 2014; Poolman, 2014; Warfel and Edwards, 2015). Différentes solutions ont été proposées et sont encore en phase de test. Les derniers développements incluent des vaccins (i) à cellules mortes comportant moins d'endotoxines, (ii) des vésicules membranaires, (iii) acellulaires avec de nouveaux adjuvants, (iv) acellulaires avec de nouveaux antigènes, et le plus prometteur, (v) des germes atténués vivants (Locht, 2017). Le vaccin le plus avancé est BPZE1, composé de germes vivants atténués, et est un des premiers du genre, ayant prouvé son efficacité chez le babouin (Locht et al., 2017). Une autre nouveauté concernerait son administration. Le vaccin est à appliquer par voie nasale, pour imiter une infection contre la coqueluche et maximiser l'induction de la réponse Th17 et Th1 chez l'homme (Mielcarek et al., 2006). Cependant, malgré ce vaccin prometteur, l'évolution constante du pathogène pourrait devenir un frein à la prévention d'épidémies au sein des pays développés et à l'efficacité vaccinale.

## ii. 2<sup>ème</sup> cause : Des souches asymptomatiques

Jusqu'à récemment, les études de la coqueluche se concentraient sur le caractère épidémique de la maladie, sa transmission et la répartition en classe d'âge. De par l'aspect évolutif du pathogène, la question de porteurs sains de la coqueluche s'est posée au sein de la communauté scientifique depuis 2014. En comparant l'infection de *Bordetella pertussis* chez des enfants babouins après trois injections du vaccin à germes entiers ou du vaccin acellulaire, il a été montré que ce dernier ne prévenait pas la colonisation du tractus respiratoire, ni sa transmission vers un autre individu (Warfel et al., 2014). Cette conclusion sous-entend que la bactérie peut coloniser le tractus respiratoire d'un hôte et se transmettre vers un autre et ce malgré l'absence de la maladie, démontrant l'existence de porteurs asymptomatiques. Malgré la vaccination de la population, chaque individu pourrait être porteur du pathogène sans en déclencher les symptômes, et transmettre la maladie asymptomatique à leur entourage, chez des personnes vaccinées ou non. Cette deuxième cause de la recrudescence de la

coqueluche sous-entend également que le nombre de porteurs serait largement sous-estimé, malgré les tests statistiques et épidémiologiques.

Un modèle mathématique, à l'échelle de la population humaine, basé sur les cas de coqueluche aux Etats-Unis et en Angleterre, les pays les plus représentés, confirme ces données biologiques (Althouse and Scarpino, 2015). Selon ce modèle, l'augmentation du nombre de cas observée sur ces dernières années corrèlerait avec l'existence de porteurs sains et ne serait être expliquée uniquement par le « waning » du vaccin acellulaire. Ceci remettrait bien évidemment en cause l'efficacité vaccinale, notamment lors des phases de test et de développement de nouveaux vaccins si les porteurs asymptomatiques ne sont pas pris en considération. De plus, cette étude remet en cause l'efficacité de la couverture vaccinale. Une personne vaccinée porteuse pourrait malgré tout transmettre la maladie chez des nourrissons non-vaccinés, conclusion similaire aux études biologiques chez le babouin. L'apparition des souches asymptomatiques seraient cependant liées à la transition du vaccin à germes entiers vers le vaccin acellulaire (Sealey et al., 2016). De plus, des études dénombrant la séroprévalence de personnes ayant été en contact avec le micro-organisme montrent que le nombre de personnes séropositives est supérieur au nombre de cas recensés de coqueluche. Cette cause de la réémergence soulève une troisième hypothèse, qui concernerait l'adaptation du pathogène lui-même à la vaccination.

### iii. 3<sup>ème</sup> cause : Adaptabilité du pathogène

Les vaccins utilisés permettent d'induire une protection immunitaire efficace. Cependant, pour contrer ces défenses, certaines souches de *Bordetella pertussis* se sont adaptées à la reconnaissance des antigènes par l'hôte. Ainsi ont pu être recensées des souches cliniques n'exprimant plus certaines adhésines ou toxines, que l'on croyait jusqu'alors nécessaire à la virulence du pathogène.

Le génome de *Bordetella pertussis* est très plastique et sujet à un changement constant. Ce phénomène est provoqué par la présence de séquences d'insertion (IS), des éléments génétiques mobiles responsables de remaniements génomiques conférant un avantage évolutif et adaptatif (Vandecraen et al., 2017). L'adaptation de pathogènes à leur environnement donne naissance à l'apparition de souches distinctes, pouvant devenir plus ou moins virulente ou lui conférant un avantage lors de l'infection de l'hôte, notamment sur la pression vaccinale. Il a été montré que le génome de *Bordetella pertussis*, contenant plus de 230 copies de l'IS481, pouvait s'adapter à la pression vaccinale induite dans les pays développés (Parkhill et al., 2003; Xu et al., 2015). Plusieurs souches du pathogène provenant de Finlande, de Chine et des Pays-Bas, où les programmes de vaccination ont différés et évolués au fur et à mesure des années, ont été séquencées et analysées, datant de prélèvements des années 1950 à aujourd'hui. Parmi ces souches, des variants génétiques, des pertes de gènes et / ou des remaniements importants du génome ont été mis en corrélation avec la couverture vaccinale de chaque pays. L'apparition de ces changements génétiques majeurs correspondrait aux différents changements de politique vaccinale dans chaque pays. Une de ces souches présenterait même 116 variants génétiques, pouvant



Récemment, une nouvelle souche de *Bordetella pertussis* a été recensée aux Etats-Unis, déficiente en hémagglutinine filamenteuse, protéine que l'on considérait comme majeure et essentielle jusqu'à maintenant pour l'infection d'une coqueluche (Weigand et al., 2018).

Enfin, le pathogène s'adapte à la pression vaccinale également en présentant différents allèles des protéines de virulence, reconnues par l'immunité induite par le vaccin acellulaire. Ainsi, en Angleterre, les souches cliniques prédominantes entre l'aire prévaccinale et l'aire du vaccin acellulaire montrent des variants alléliques différents pour des gènes codants pour la toxine pertussique, la pertactine, et les fimbriae, des adhésines également nécessaires à la colonisation du tractus respiratoire (Sealey et al., 2015). Ces variants alléliques ne sont pas nécessairement reconnus par le système immunitaire selon l'allèle qui a été utilisé pour la composition du vaccin.

Cette adaptation du pathogène à la pression vaccinale rend la prévention de la coqueluche encore plus délicate. Malgré la couverture vaccinale, la coqueluche est la maladie infantile à prévention vaccinale prédominante. De par son pouvoir adaptatif, il est encore nécessaire, à l'heure actuelle, d'étudier les mécanismes de virulence et évolutifs du pathogène dans le but d'identifier de nouvelles cibles thérapeutiques.

## **II. Les facteurs de virulence du pathogène**

La virulence de *Bordetella pertussis* est basée sur la sécrétion de toxines et la présence de nombreuses adhésines le long de la paroi extracellulaire. Ces facteurs interagissent avec diverses protéines des cellules épithéliales du tractus respiratoire pour contrer les défenses du système immunitaire, résister aux dégagements de mucus et ainsi permettre l'infection (Melvin et al., 2014). De ce fait, la sécrétion de ces facteurs prolongent la durée des symptômes de la coqueluche et ce, même après traitement antibiotique et lyses bactériennes. Plusieurs facteurs de virulence et régulateurs ont ainsi pu être identifiés, et leur impact lors de l'infection étudié.

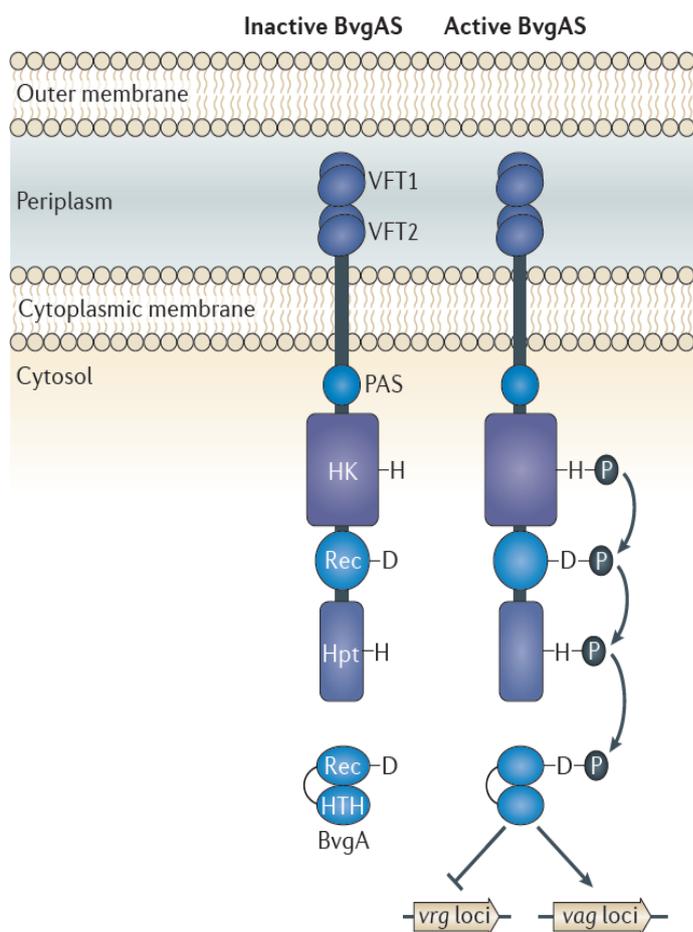
### **1. La régulation par les systèmes à deux composants**

Comme de nombreux procaryotes, la régulation de la virulence parmi le genre *Bordetellae* est régie par des systèmes à deux composants, encore appelés phosphorelais, dont le plus connu se nomme BvgAS (*Bordetella* virulence genes Activator / Sensor) (Scarlato and Rappuoli, 1991). Ces systèmes de régulation bactériens sont composés de deux protéines, un senseur ancré à la membrane et un activateur transcriptionnel qui permet à la cellule de percevoir les variations de l'environnement pour s'adapter aux stress éventuels (Zschiedrich et al., 2016). Ils sont généralement impliqués dans la régulation de voies métaboliques, du chimiotactisme, de la réponse aux stress environnementaux, ou encore de la virulence. La protéine senseur est capable de lier un ligand, une molécule présente dans l'environnement, provoquant l'autophosphorylation du senseur. Le groupement phosphate est alors transmis vers l'activateur transcriptionnel, par une cascade de phosphorylation, induisant ainsi sa capacité à se lier à l'ADN pour activer ou inhiber l'expression de gènes.

#### **i. Le système à deux composants BvgAS**

La virulence de *Bordetella pertussis* est régulée par le système à deux composants BvgAS. Le phosphorelai est actif en condition classique de culture à 37°C, et peut être retrouvé sous une forme inactive à 20°C (Dupré et al., 2015). Encore à l'heure actuelle, aucun ligand n'a pu être mis en évidence. En laboratoire, il est possible d'inactiver également le phosphorelai en supplémentant le milieu en sulfate de magnésium ou en acide nicotinique (Dupré et al., 2013; Scarlato and Rappuoli, 1991). La modulation du signal perçu par le domaine transmembranaire entraîne un changement de fonction du senseur, BvgS, d'une activité kinase fournie par le domaine Histidine-Kinase, vers une activité phosphatase fournie par le domaine receveur (Uhl and Miller, 1996) (Figure 6). Ces modulateurs agissent sur la dynamique structurale des hélices reliant les différents domaines enzymatiques du senseur (Lesne et al., 2018, 2016). Les changements de structure, potentiellement induits par des variations de température, des « linker » régulent directement l'activité de kinase ou de phosphatase par une régulation mécanique (Abriata et al., 2017).

Lorsque les dimères de BvgS sont actifs à 37°C, le domaine Histidine-Kinase induit une autophosphorylation du phosphorelai (Figure 6) (Uhl and Miller, 1994). Le



**Figure 6:** Le système à deux composants BvgAS régulateur de la virulence de *Bordetella pertussis*.

Le senseur BvgS est ancré à la membrane cytoplasmique, ayant un segment périplasmique pouvant percevoir les modulateurs et variations du milieu par le domaine Venus-Fly-Trap (VFT), et un segment cytoplasmique comprenant les activités kinases et phosphatases fournies par les domaines Histidine-Kinase (HK) et Recepteur (Rec). L'activation du senseur entraîne une cascade de phosphorylation relayée par les phosphotransmetteurs Rec et Histidine-phosphotransmetteur (Hpt) jusqu'à l'activateur transcriptionnel BvgA. BvgA phosphorylé va permettre la transcription des gènes activés lors de la virulence (*vag*) et inhiber les gènes réprimés lors de la virulence (*vrg*). PAS : Domaine Per-Arnt-Sim. Figure de Melvin *et al.* 2014

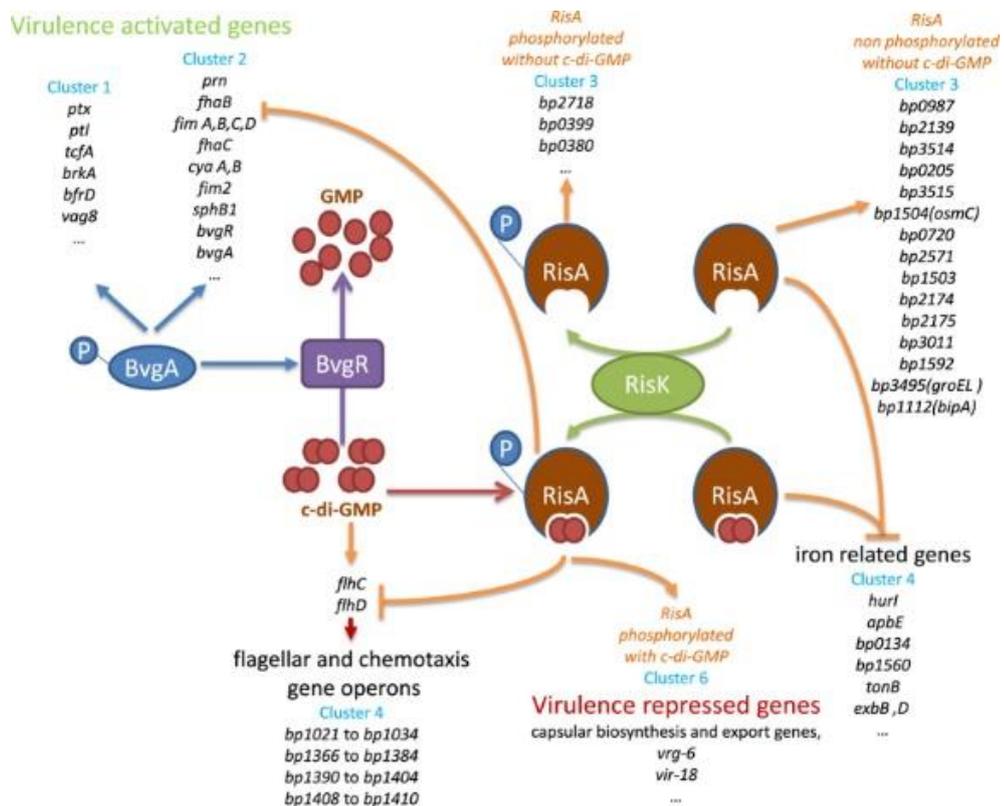
groupement phosphate est alors transmis vers le domaine receveur sur un résidu acide aspartique, exerçant alors une activité de phosphotransmetteur, jusqu'au domaine histidine-phosphotransmetteur (Uhl and Miller, 1996). Ce dernier domaine va transmettre le groupement vers la protéine activatrice, BvgA, un facteur de transcription. Le transfert du groupement phosphate rend possible la dimérisation de BvgA phosphorylé (BvgA-P), et lui permet d'exercer son rôle d'activateur transcriptionnel via son domaine « Helix-Turn-Helix » (Decker *et al.*, 2012).

BvgA-P se lie à la séquence ADN consensus (T/A)TTC(C/T)TA (Decker *et al.*, 2012). Elle est retrouvée en amont des séquences promotrices de gènes de virulence, comprenant principalement des toxines et adhésines (Boucher and Stibitz, 1995; Boulanger *et al.*, 2015; Deora, 2002; Karimova and Ullmann, 1997; Kinnear *et al.*, 1999; Merkel *et al.*, 2003). De ce fait, la phosphorylation de BvgA active la transcription de gènes impliqués dans la virulence, appelés *vag* (Virulence Activated Genes) (Melvin *et al.*, 2014). D'autres gènes ont été montrés comme étant réprimés lors de l'activation du phosphorelai, et ont été nommés *vrg* (Virulence Repressed Genes). La majorité de ces gènes codent pour des protéines impliquées dans le métabolisme, les flagelles ou

encore le chimiotactisme (Hot et al., 2003). Ils sont réprimés par l'intermédiaire d'une autre protéine, BvgR (*Bordetella* virulence gene Repressor), dépendant de l'activation du système à deux composants BvgAS et longtemps considérée comme étant le répresseur de ces gènes (Merkel et al., 2003). Ce système de régulation de la virulence est retrouvé parmi toutes les espèces du genre *Bordetellae* et est nécessaire à l'expression de leurs toxines et adhésines respectives (Linz et al., 2016).

## ii. Le système RisA / RisK

Des études plus récentes ont pu mettre en évidence l'existence d'un autre système régulateur de la virulence chez les *Bordetellae*. Le système à deux composants, appelé RisAS (Regulator of Intracellular Stress response Activator / Sensor), est essentiel pour l'expression de gènes exprimés en réponse au stress oxydatif chez *Bordetella bronchiseptica*, et nécessaire à la persistance de l'infection chez l'animal (Jungnitz et al., 1998). RisA peut se fixer en amont des promoteurs de *vrg* pour induire leur expression sur la séquence consensus AAAT(G/T)TA, et est de ce fait également impliqué dans la régulation de la virulence des *Bordetellae* (Croinin et al., 2005).

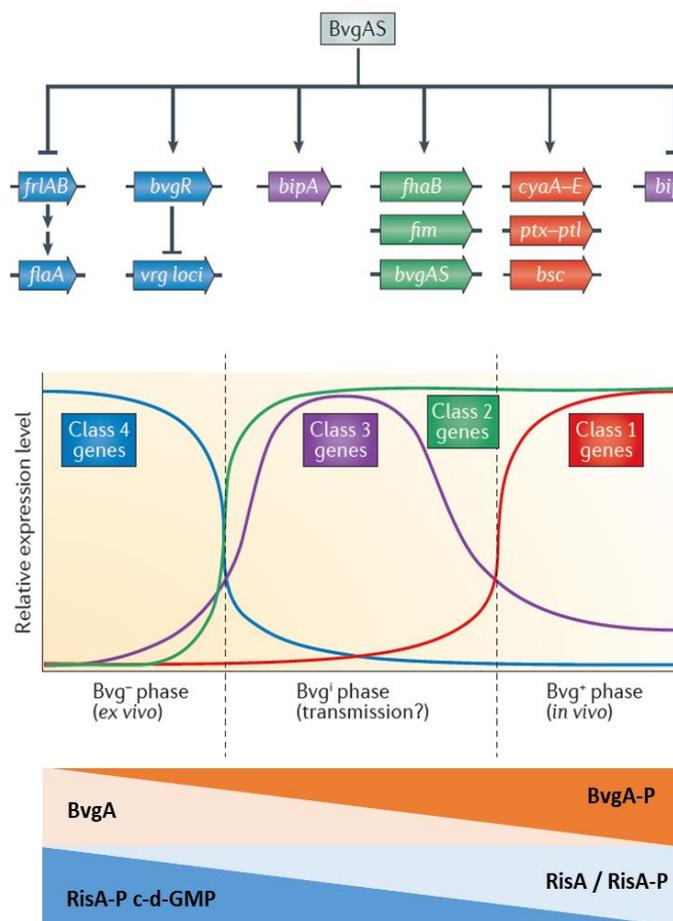


**Figure 7:** Le système à deux composant RisAK chez *Bordetella pertussis*. RisA régule différentes catégories de gènes en fonction de son état de phosphorylation, et de sa liaison au cyclique-di-GMP (c-di-GMP). RisA est phosphorylé par la protéine RisK et, lorsque le pathogène est dans un état non-virulent, lie le c-di-GMP non-métabolisé par la protéine BvgR. Dans cet état, la fixation de RisA sur le site consensus de fixation à l'ADN permet l'expression des *vrg*, une inhibition de l'expression des *vag* et de *flhC* et *flhD*, impliqués dans la régulation de la motilité et du chimiotactisme. Figure de Coutte et al., 2016

Chez *Bordetella pertussis*, le gène *risS* est un pseudogène codant pour une protéine tronquée (Stenson et al., 2005). Il a été montré récemment que la protéine RisA du pathogène peut être phosphorylée par la protéine Risk, dont le gène n'est pas transcrit en opéron avec l'activateur transcriptionnel (Chen et al., 2017). Selon son état de phosphorylation, RisA régule l'expression de différents gènes, encore peu étudiés à ce jour (Coutte et al., 2016) (Figure 7). Cependant, RisA peut lier le cyclique di-GMP produit et accumulé dans la cellule en condition de non-virulence. Cette molécule est métabolisée par la protéine BvgR, dont l'expression est dépendante du phosphorelai BvgAS. Lorsque RisA est phosphorylé et lié au cyclique di-GMP, il entraîne l'activation des *vrg* et une inhibition de certains *vag*. La régulation de la virulence par ces deux systèmes à deux composants implique une régulation fine de chaque gène de virulence, nécessaire à l'infection du tractus respiratoire de l'hôte en fonction de la localisation du pathogène et des facteurs environnant.

### iii. La cinétique de la virulence : Une régulation fine

Le degré d'activation des systèmes à deux composants montre une régulation fine de chaque gène de virulence, reflétant diverses catégories de gènes. Par exemple, le taux de BvgA phosphorylé va permettre l'expression de certains gènes en fonction du nombre de sites de fixation localisés en amont de leur séquence promotrice (Decker et



**Figure 8:** Cinétiques de virulence de *Bordetella pertussis*.

La proportion de BvgA et de RisA activés ou non dans la cellule permet de réguler à différents taux d'expression les gènes de virulence. L'activation de BvgAS inhibe l'expression des gènes de classe 4, les *vrg*, et à fort taux, des gènes de classe 3, dont *bipA* est le seul connu. Un faible degré d'activation suffit à l'expression des gènes de classe 2, comportant les adhésines du pathogène et *bvgAS*. Lorsqu'il est pleinement actif, les gènes de classe 1 comportant les toxines et systèmes de sécrétion du pathogène sont exprimés. Ces cinétiques d'expression des classes de gènes reflètent les trois phases de virulence du pathogène (Bvg-/i/+). Les dégradés représentent les proportions de BvgA et RisA actifs. Figure tirée de Melvin et al., 2014

al., 2012). Une faible activation du phosphorelai suffit à l'expression des gènes présentant un unique site de fixation de BvgA-P, appelés « early *vag* ». Si plusieurs sites de fixation sont présents, le gène sera exprimé uniquement lorsque les phosphorelais sont totalement activés. Ces gènes sont appelés « late *vag* ». De ce fait, chaque gène de virulence est exprimé à une étape ou condition donnée en fonction du nombre de site de fixation des facteurs de transcription et de leur affinité. Actuellement, quatre classes de gènes ont pu être identifiées, exprimés à des conditions et à des temps différents dans la cinétique de virulence (Figure 8) (Melvin et al., 2014). L'expression de ces gènes reflète les trois phases de virulence de *Bordetella pertussis*, la phase avirulente Bvg<sup>-</sup>, la phase virulente Bvg<sup>+</sup> et une phase intermédiaire Bvg<sup>i</sup> correspondant potentiellement à la transmission du pathogène d'un hôte vers un autre.

## 2. Les gènes de virulence de *Bordetella pertussis*

La cinétique de la virulence de *Bordetella pertussis* permet au pathogène de s'adapter à chaque étape de l'infection en fonction de leur environnement. Chaque classe de gène a donc une ou des fonctions caractéristiques dépendant de l'état de virulence de la bactérie.

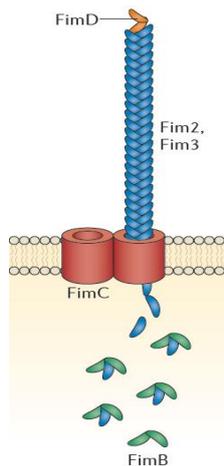
### i. La phase avirulente Bvg<sup>-</sup>

En phase Bvg<sup>-</sup> sont exprimés les gènes de classe 4, correspondant aux *vrg* de *Bordetella pertussis* dont l'expression est dépendante de RisA phosphorylé lié au cyclique di-GMP (Coutte et al., 2016). Chez *Bordetella bronchiseptica*, les gènes exprimés dans cette phase sont impliqués dans la motilité (Akerley et al., 1992; Akerley and Miller, 1993). Ces gènes sont également présents chez *Bordetella pertussis*, cependant l'opéron codant pour les flagelles est interrompu par des éléments transposables, formant des pseudogènes (Parkhill et al., 2003). Le pathogène n'est pas retrouvé dans cette phase *in vivo*, mais il est cependant possible de moduler l'expression du système à deux composants BvgAS, comme indiqué précédemment (Byrd et al., 2013; Chen et al., 2017; Veal-Carr and Stibitz, 2005).

Les gènes de classe 4 correspondent à différentes fonctions qui permettent au pathogène de survivre dans des conditions de stress *in vitro*. Ils regroupent des gènes impliqués dans la réponse au choc thermique, dans le métabolisme des acides gras et des lipides, dans la synthèse du lipopolysaccharide et enfin dans la synthèse de la capsule polysaccharidique (Hoo et al., 2014; Moon et al., 2017).

### ii. La phase de transmission Bvg<sup>i</sup>

En phase Bvg<sup>i</sup> sont exprimés les gènes de classe 3 et les gènes de classe 2. Le seul gène de classe 3 répertorié chez *Bordetella pertussis* est *bipA*, codant pour une protéine impliquée dans la formation de biofilm et la prévention de l'autoagglutination chez les *Bordetellae* (Hiramatsu et al., 2016). Ces gènes pourraient avoir un rôle dans la transmission du pathogène par les aérosols et seraient réprimés en excès de BvgA-P par fixation du facteur de transcription sur la séquence codante du gène (Decker et al., 2012; Stockbauer et al., 2001).



**Figure 9:** Composition d'une fimbriae de *Bordetella pertussis*.

Les fimbriae sont des pilis bactériens permettant d'adhérer sur les surfaces et cellules de l'hôte. Elles sont constituées d'un corps composé de protéine de deux différents sérotypes (Fim2 ou Fim3), amenées à la membrane par une protéine chaperonne FimB. L'extrémité distale de la fimbriae est composée de la protéine FimD, suivie par les protéines Fim2 ou Fim3 du pathogène. Les fimbriae sont maintenues à la membrane par le transporteur membranaire FimC.

Figure de Melvin *et al.*, 2014

Les gènes de classe 2 sont des « *early-vag* » et correspondent pour la plupart d'entre eux aux adhésines de *Bordetella pertussis*, impliqués dans la colonisation du tractus respiratoire et à la résistance au dégageement de mucus (Melvin *et al.*, 2014). Ces adhésines sont très régulièrement retrouvées dans la composition des vaccins acellulaires de par leur nécessité pour le pathogène (Cherry, 1996; Gorringer and Vaughan, 2014; Sato and Sato, 1999). La première adhésine majeure est l'hémagglutinine filamenteuse (FHA), impliquée dans l'adhésion du pathogène aux cellules épithéliales, la formation de biofilms et la suppression de l'inflammation (Hoffman *et al.*, 2017; Scheller *et al.*, 2015; van den Berg *et al.*, 1999). FHA est une protéine adressée à la membrane extracellulaire de *Bordetella pertussis*, et parfois relâchée dans le milieu extracellulaire (Melvin *et al.*, 2014). Elle a une affinité pour les polysaccharides sulfatés comme l'héparine (Menozzi *et al.*, 1994). Récemment, certaines souches virulentes et déficientes en cette adhésine par insertion d'IS481 dans la séquence codante ont été répertoriées, adhésine que l'on pensait jusqu'alors nécessaire à l'infection du pathogène (Weigand *et al.*, 2018).

Une autre adhésine majeure est la fimbriae, également utilisée dans la composition des vaccins acellulaire (Gorringer and Vaughan, 2014). Cette adhésine, également adressée à la membrane extracellulaire, coopère avec l'hémagglutinine filamenteuse pour renforcer l'adhérence du pathogène aux cellules épithéliales ciliées du tractus respiratoire de l'homme et prévenir des dégageements de mucus induits par les cellules inflammatoires (Scheller *et al.*, 2015). Ces pilis de type I sont impliqués également dans la formation de biofilm chez *Bordetella bronchiseptica*, et détermineraient la spécificité d'hôte chez ce pathogène (Burns *et al.*, 1993; Cattelan *et al.*, 2016). Le pili est composé de protéines agglomérées en polymères dont deux sous-types ont été répertoriés chez *Bordetella pertussis*, le sérotype 2 Fim2 et le sérotype 3 Fim3 (Figure 9) (Savelkoul *et al.*, 1996). Les gènes codant pour ces protéines, *fim2* et *fim3*, sont localisés à différents endroits du génome et permettent au pathogène de s'adapter à la pression vaccinale et au système immunitaire par des variations de sérotype (Vaughan *et al.*, 2014). La séquence promotrice de *fim2* et de *fim3* comprend une séquence poly-C entre les boîtes -10 et -35 dont le nombre de résidus cytosines permet à la polymérase de transcrire ou non le gène (Otsuka *et al.*, 2017; Willems *et al.*, 1990). De cette manière, les souches de *Bordetella pertussis* peuvent présenter une fimbriae (sérotypes Fim2+ Fim3- ou Fim2-

Fim3+), les deux fimbriae (sérotypage Fim2+ Fim3+), ou aucune fimbriae (sérotypage Fim2- Fim3-) (Vaughan et al., 2014).

La pertactine est une adhésine également présentée à la surface membranaire et dont le rôle est encore controversé. Il a pu être montré que cette protéine était nécessaire à l'adhésion aux cellules épithéliales ciliées *in vitro*, mais pas *in vivo* (Khelef et al., 1994). Cependant, chez *Bordetella bronchiseptica*, elle lie les intégrines et serait nécessaire à la résistance aux neutrophiles (Inatsuka et al., 2010). Cette adhésine est également couramment utilisée dans le vaccin acellulaire, et a fait l'objet de beaucoup de controverses depuis quelques années dû à la mise en évidence de souches déficientes en pertactine, dont la plupart présentent une IS481 insérée dans la séquence codante du gène ou une délétion du gène médiée par ces éléments génétiques mobiles (Otsuka et al., 2012; Queenan et al., 2013; Williams et al., 2016).

### iii. La phase de virulence Bvg<sup>+</sup>

Lors de la phase de virulence Bvg<sup>+</sup>, la bactérie exprime les gènes de classe 2 correspondant principalement aux adhésines et les gènes de classe 1 qui regroupent la majorité des toxines du pathogène et le système de sécrétion de type III (Melvin et al., 2014). Actuellement, quatre toxines du pathogène ont pu être répertoriées : la cytotoxine trachéale, la toxine dermonécrotique, l'adénylate cyclase hémolytique et la toxine pertussique. Ces toxines ont pour fonction d'inhiber le recrutement des cellules immunitaires et de lyser ces cellules et les cellules épithéliales de l'hôte (Melvin et al., 2014). La plus étudiée est la toxine pertussique, spécifique de *Bordetella pertussis*, et présente également dans le vaccin acellulaire (Sato and Sato, 1999). Elle est responsable d'une lymphocytose sévère chez les animaux et inhibe le recrutement de cellules immunitaires. Des souches déficientes en cette toxine ont également pu être répertoriées en liaison avec des délétions induites par les IS481 (Williams et al., 2016).

L'apparition de souches cliniques de *Bordetella pertussis* déficientes en hémagglutinine filamenteuse, pertactine, et en toxine pertussique témoigne de la complexité de la virulence du pathogène. La perte de ces antigènes, utilisés dans le vaccin acellulaire, confère un avantage sélectif et évolutif du pathogène, médié par les IS481, et ce malgré l'importance de leur rôle dans l'infection. L'acquisition de ces éléments génétiques mobiles procure une plasticité génomique importante, reflétant une grande diversité inter-souche en termes d'expression de gènes et de facteurs de virulence (Xu et al., 2015).

### III. Les disparités génomiques inter-souche

Pour comprendre l'adaptation de *Bordetella pertussis* à la pression vaccinale et comprendre son évolution en pathogène spécifique de l'être humain, des études de comparaison génomique et protéomique au sein du genre *Bordetellae* ont été menées (Park et al., 2015; Parkhill et al., 2003; Williamson et al., 2015). Les facteurs de virulence sont retrouvés en grande majorité chez toutes les *Bordetellae*. La différence majeure réside en l'apparition de la toxine pertussique, que l'on pensait jusqu'alors comme toxine indispensable et responsable des symptômes de la coqueluche (Williams et al., 2016). Une autre différence majeure réside en l'apparition de la cytotoxine trachéale, une toxine dérivée des monomères de peptidoglycane (Cookson et al., 1989; Goldman and Cookson, 1988). *Bordetella pertussis* n'est pas capable de recycler les monomères de sucres durant le remodelage de la paroi cellulaire et les relâche dans le milieu extracellulaire. La dernière différence notable concerne la perte de l'antigène-O (Preston et al., 1999). Contrairement aux autres *Bordetellae*, *Bordetella pertussis* ne présente pas d'antigène-O à la surface. Elle possède une couche de lipo-oligosaccharide (LOS), alors que les autres espèces du genre ont une couche de lipopolysaccharide (LPS) complète (Melvin et al., 2014). La perte du LPS confère également un avantage au pathogène, celui-ci étant reconnu par les cellules du système immunitaire pour combattre l'infection, sa disparition pourrait donc être en lien avec une diminution de l'inflammation (Mann et al., 2005).

#### 1. Caractéristiques génomiques de *Bordetella pertussis*

Ces différences protéomiques ne sont pas suffisantes pour expliquer la spécificité d'hôte de *Bordetella pertussis* et adaptation à la pression vaccinale. L'apparition du séquençage à haut débit a permis d'étudier en profondeur les différences génomiques parmi le genre *Bordetellae* pour identifier des facteurs spécifiques au pathogène (Parkhill et al., 2003).

##### i. Comparaison des génomes des *Bordetellae*

*Bordetella pertussis* aurait évolué indépendamment de *Bordetella parapertussis* à partir d'un ancêtre commun se rapprochant d'une ancienne *Bordetella bronchiseptica* (Diavatopoulos et al., 2005). Une comparaison des trois génomes respectifs des souches modèles a pu montrer que les génomes de *Bordetella pertussis* et de *Bordetella parapertussis* sont réduits en comparaison du génome de *Bordetella bronchiseptica* (Tableau 1). La diminution de la taille du génome a conduit à une perte de gènes et s'accompagne d'une augmentation du nombre de pseudogènes. Cependant, les autres caractéristiques du génome restent similaires entre ces sous-espèces : le nombre d'ARN ribosomiques, le nombre d'ARN de transfert, le pourcentage en Guanine / Cytosine, le pourcentage de portions codantes du génome et la taille moyenne de ces loci restent pratiquement inchangés.

L'une des plus grandes différences réside en l'apparition d'éléments génétiques mobiles en grand nombre, appelés séquences d'insertions (IS) (Parkhill et al., 2003),

chez *Bordetella pertussis* et *Bordetella parapertussis*. Le génome de *Bordetella parapertussis* transporte 22 IS1001 et 90 IS1002, alors que le génome de *Bordetella pertussis* transporte 6 IS1002, 17 IS1663 et 238 IS481. Les IS481 peuvent être encore retrouvées chez certaines souches de *Bordetella bronchiseptica* de classe IV, alors que les IS1001 et IS1002 sont principalement retrouvées chez quelques souches de classe I (Diavatopoulos et al., 2005). Ce trait renforce le fait que *Bordetella pertussis* serait issue d'une lignée de *Bordetella bronchiseptica* qui aurait évolué indépendamment d'une autre lignée ayant elle évoluée en *Bordetella parapertussis*. L'acquisition des IS serait responsable de remaniements génomiques majeurs observés chez le pathogène humain (Figure 10) (Parkhill et al., 2003). Le génome de *Bordetella pertussis* présente de grandes similarités avec celui de *Bordetella bronchiseptica*, mais l'organisation génomique est très différente contrairement à celle de *Bordetella parapertussis*.

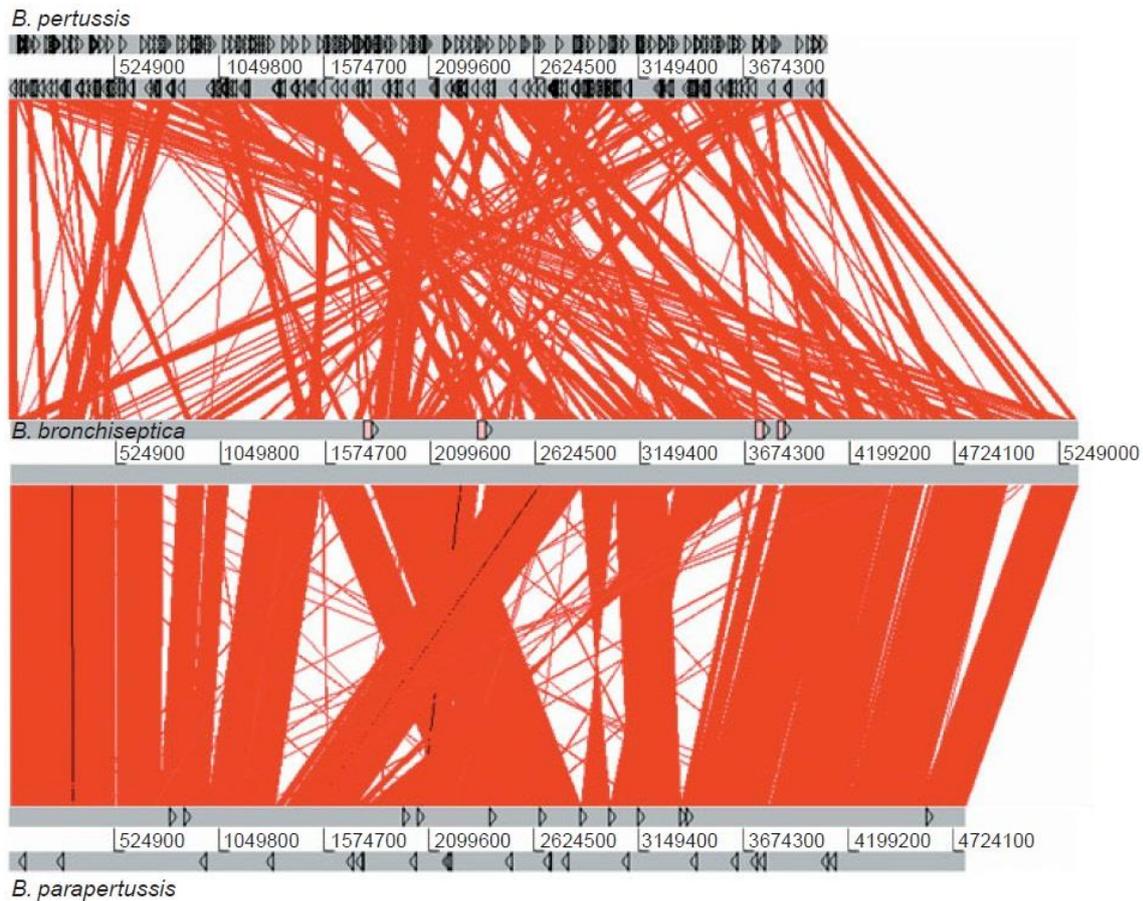
## ii. Avantages des séquences d'insertion IS481

Les IS481 seraient responsables de la diminution de la taille du génome de *Bordetella pertussis*, qui serait elle-même corrélée avec la spécificité d'hôte chez les *Bordetellae* (Cummings et al., 2004). Les différences génomiques affectent principalement les locus codant pour des protéines de surfaces et des antigènes, dont l'antigène-O et la fimbriae, également liée à la spécificité d'hôte chez les *Bordetellae* (Burns et al., 1993; Cummings et al., 2004; Linz et al., 2016; Preston et al., 1999). La réduction de la taille du génome favorisée par l'acquisition des IS est un événement majeur ayant contribué à l'évolution de *Bordetella pertussis* et à sa spécificité d'hôte (Siguier et al., 2014).

Outre le fait d'être impliqué dans l'évolution du pathogène, les IS481 sont responsables de remaniements génétiques majeurs responsables de grandes diversités génomiques entre les souches de *Bordetella pertussis* (Weigand et al., 2017).

	<i>B. pertussis</i>	<i>B. parapertussis</i>	<i>B. bronchiseptica</i>
Size (bp)	4,086,186	4,773,551	5,338,400
G+C content (%)	67.72	68.10	68.07
Coding sequences	3,816	4,404	5,007
Pseudogenes	358 (9.4%)	220 (5.0%)	18 (0.4%)
Coding density (intact genes)	82.9%	86.6%	91.4%
Coding density (all genes)	91.6%	92.2%	92.0%
Average gene size (bp)	978	987	978
rRNA operons	3	3	3
tRNA	51	53	55
IS481	238	0	0
IS1001	0	22	0
IS1002	6	90	0
IS1663	17	0	0

**Tableau 1:** Comparaison des caractéristiques génomiques de *Bordetella pertussis*, *Bordetella parapertussis* et *Bordetella bronchiseptica*. Tableau de Parkhill et al., 2003



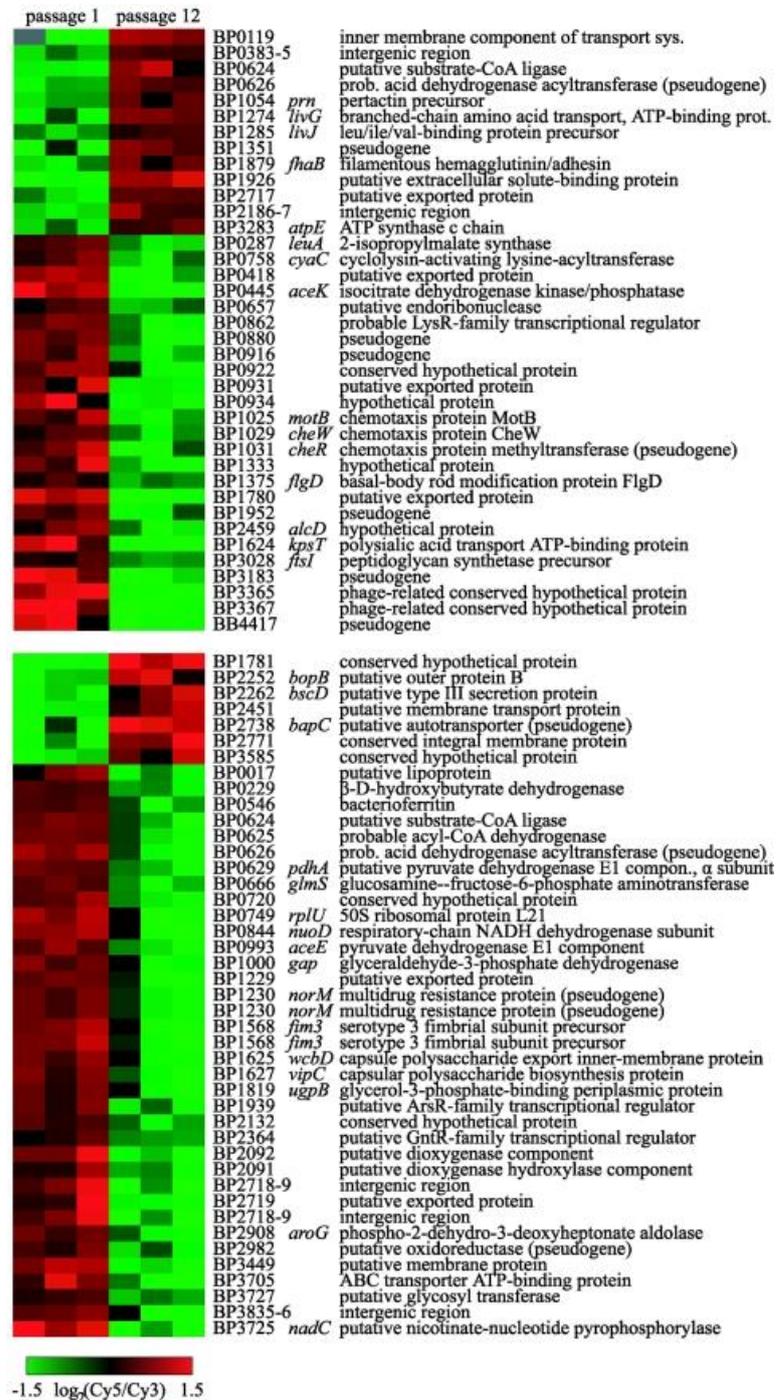
**Figure 10:** Alignement des séquences génomiques respectives de *Bordetella pertussis*, *Bordetella bronchiseptica* et *Bordetella parapertussis*.

En gris sont représentés les brins positifs (Haut) et négatifs (Bas) de chaque sous-espèce. Les similarités génomiques en rouge ont été détectées par BLASTN. Les triangles noirs représentent les éléments génétiques mobiles et les carrés roses les prophages. Figure de Parkhill *et al.*, 2003

L'organisation génomique est changée principalement par des événements de double recombinaison homologue et par l'activation des IS. Le génome du pathogène varie très peu en terme de contenu entre les souches, cependant l'ordre des gènes et l'organisation génomique influe directement sur l'expression de gènes, et notamment de gènes de virulence (Brinig *et al.*, 2006). Des variations d'expression de gènes de virulence ont pu être observées entre différentes souches cliniques et vaccinales malgré la présence des *loci* génétiques chez toutes ces souches. De plus, l'expression des gènes de virulence des souches cliniques peut être altérée après quelques cultures en laboratoire. Par exemple, les gènes *fhaB*, *prn*, *cyaC* et *fim3* sont exprimés différemment chez la souche Bpe280 après uniquement 12 passages sur boîte de pétri, en condition Bvg<sup>+</sup> et Bvg<sup>-</sup> (Figure 11) (Brinig *et al.*, 2006).

Ces remaniements génomiques seraient corrélés à la pression vaccinale exercée sur les souches circulantes du pathogène (Xu *et al.*, 2015). La variation de l'ordre des gènes conférerait un avantage sélectif potentiel agissant sur l'expression des gènes de virulence (Brinig *et al.*, 2006). les antigènes présents dans le vaccin acellulaire seraient

inactivés par les IS481 de *Bordetella pertussis* soit directement par insertion dans la séquence du gène, soit indirectement par remaniement du chromosome (Martin et al., 2015; Weigand et al., 2018; Williams et al., 2016).



**Figure 11:** Variation d'expression de gènes après 12 passages sur milieu de culture de la souche Bpe280 observée par puces à ADN. Chaque colonne représente un réplica de culture en condition Bvg<sup>+</sup> (Haut) et Bvg<sup>-</sup> (Bas). L'expression des gènes a été représentée par une carte thermique en Log<sub>2</sub> par comparaison avec la souche Tohamia I, modèle de laboratoire. Figure de Brinig et al., 2006

## 2. Les séquences d'insertion, un facteur d'adaptation

La majorité des procaryotes portent des éléments génétiques mobiles au sein de son génome. Les séquences d'insertions rendent le génome de leur hôte très plastique, et induisent des remaniements génomiques importants en cas de stress environnemental dans le but de sélectionner des souches pouvant s'y adapter capables de coloniser de nouvelles niches (Vandecraen et al., 2017).

### i. Caractéristiques des séquences d'insertion

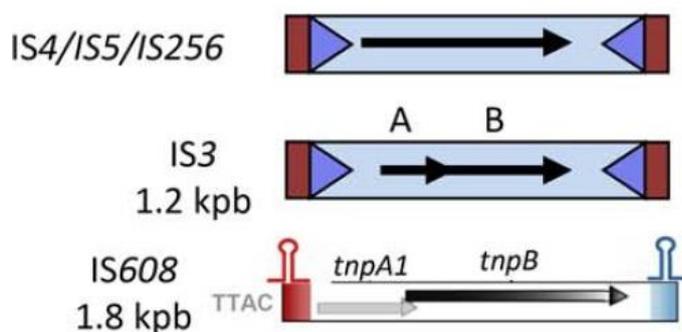
Les IS sont considérées comme des îlots génétiques parasites, incapables de se répliquer par eux-mêmes, infectant un hôte pour se répliquer et lui conférer un avantage sélectif (Siguier et al., 2015). Les IS sont des éléments génétiques mobiles d'une taille comprise entre 0,7 et 2,5kb codant pour une transposase. Elles contiennent généralement un ou deux cadres ouverts de lecture, et arborent à leurs extrémités 5' et 3' des séquences répétées (Figure 12).

Les séquences gauches et droites sont généralement des répétitions inversées imparfaites qui permettent à l'IS de s'intégrer au génome. La transposase, lorsqu'elle est exprimée, est capable de catalyser le clivage du premier brin de l'ADN où elle se trouve (Siguier et al., 2015). Selon le motif de la transposase, l'IS est ensuite soit « copiée », auquel cas une copie de l'IS est générée et s'insère à un autre endroit du génome, doublant le nombre de copie, soit « coupée » pour s'insérer ailleurs dans le génome. Dans les deux cas, l'ADN excisé se circularise dans la majorité des cas et s'insère dans le génome par reconnaissance d'une séquence consensus qui encadrera l'IS.

Les IS sont impliquées dans divers mécanismes pouvant conférer un avantage à la cellule hôte. Elles peuvent être impliquées par exemple dans la résistance aux antibiotiques, à la réponse à un stress environnemental et aux remaniements génomiques comme chez *Bordetella pertussis* (Depardieu et al., 2007; Humayun et al., 2017; Schneider and Lenski, 2004; Vandecraen et al., 2017).

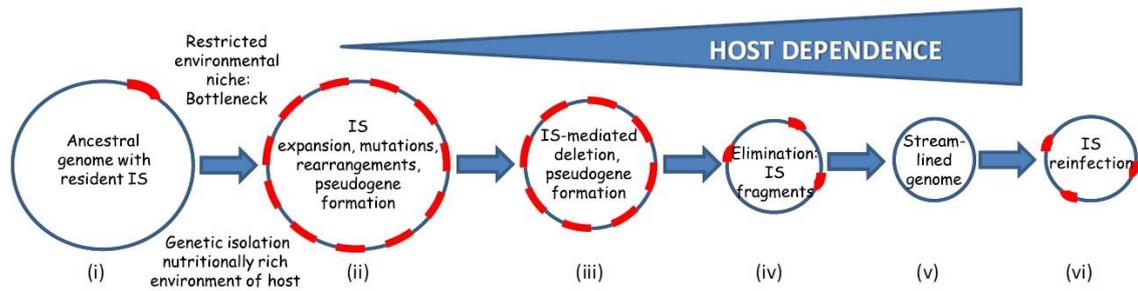
### ii. Les caractéristiques des IS481 de *Bordetella pertussis*

Les IS sont regroupées en différents groupes selon le motif catalytique de la transposase qu'elles portent, reflétant le mécanisme utilisé pour la mobilité (Siguier et al.,



**Figure 12:** Schéma de l'organisation de séquences d'insertion.

Les IS contiennent un (IS4, IS5, IS256) ou deux (IS3, IS608) cadres ouverts de lecture (Flèche noire et grise) et sont encadrées par des séquences répétées (Rouge) inversées (Triangle bleu). Les tiges et boucles représentent les terminateurs de transcription. Image de Siguier *et al.*, 2015



**Figure 13:** Théorie sur l'expansion des IS, élimination des IS et linéarisation des génomes en lien avec la dépendance à l'hôte.

Les cercles bleus représentent le génome d'un procaryote. Les rectangles rouges symbolisent les séquences d'insertion. Entre parenthèses figurent le numéro des étapes constituant la théorie. Figure de Siguier *et al.*, 2014

2015). Les IS481 font partie de la famille qui porte leur nom, les IS481, et dont les IS1002 présentent chez *Bordetella pertussis* et *Bordetella parapertussis* font également parties. Elles comportent uniquement un cadre ouvert de lecture et ont une taille approximative de 1050pb, dont la séquence codante est de 950pb (Parkhill *et al.*, 2003).

Les IS481 et les IS1002 reconnaissent un motif de 6pb, retrouvé de façon commune au sein du génome du pathogène, NCTAGN (Stibitz, 1998). Ce motif est présent sur plusieurs gènes de virulence, et notamment dans le gène de l'opéron *bvgAS*. Les motifs protéiques prédits de la transposase laissent supposer que son mode de fonctionnement serait de copier la séquence de l'IS pour doubler le nombre d'IS481 (Siguier *et al.*, 2015). Le nombre d'IS481 présentes dans le génome de *Bordetella pertussis* laisse supposer qu'il y aurait une probabilité d'apparition à chaque 15kb, renforçant l'hypothèse de la réplication de la séquence (Parkhill *et al.*, 2003). Elles sont retrouvées quasiment exclusivement chez *Bordetella pertussis*, mais sont présentes également chez une lignée de *Bordetella bronchiseptica* de façon moins proéminente (Diavatopoulos *et al.*, 2005). Certaines de ces IS481 sont également retrouvées chez *Bordetella holmesii*, une espèce capable en de rares occasions d'infecter l'homme (Diavatopoulos *et al.*, 2006).

### iii. Les IS, des médiateurs de l'évolution ?

Le nombre de copies d'IS481 chez *Bordetella pertussis* peut se comparer à d'autres espèces procaryotes pathogènes infectées également par certaines IS, faisant parties d'une ou de plusieurs familles différentes (Siguier *et al.*, 2014). Par exemple, *Shigella flexneri* souche 2457T comporte plus de 100 copies de l'isoforme IS1X1 dans son génome, et *Yersinia pestis* CO92 qui comporte plus 120 IS au total (Parkhill *et al.*, 2001; Zaghoul *et al.*, 2007). Cependant, le nombre d'IS chez *Yersinia pestis* varie d'une souche à l'autre (Chain *et al.*, 2006). Par exemple, la souche Antiqua possède 75 IS1 parmi les 176 IS présentes dans son génome, alors que la souche KIM possède 34 IS1 parmi les 116 IS présentes dans son génome. La variation du nombre d'IS au sein d'une même espèce et des espèces proches mettent en évidence une théorie reliant l'expansion des

IS puis leur élimination avec l'augmentation de la dépendance à l'hôte (Figure 13) (Siguier et al., 2014).

L'acquisition d'une IS chez un génome ancestral constitue le point de départ de la théorie, *Bordetella bronchiseptica* se situerait à ce stade (i) (Soumana et al., 2017)(Siguier et al., 2014). Lorsque l'organisme se retrouve en condition de stress, dans une niche environnementale restreinte, l'activation de l'IS engendre une multiplication de copies de l'IS en question (ii). Ce phénomène est appelé expansion des IS, et engendre des réarrangements chromosomiques par des phénomènes de double recombinaisons homologues, la formation de pseudogènes et des mutations. La divergence des deux lignées de *Bordetella bronchiseptica* classes I et classes IV aurait eu lieu entre ces deux phases, *Bordetella parapertussis* en serait encore au stade (ii). La répétition des événements de recombinaison engendre de nombreux remaniements génomiques responsables de la réduction de la taille du génome au sein de la niche environnementale. *Bordetella pertussis* se situerait actuellement au stade (iii) de cette théorie. Les éléments génétiques mobiles sont éliminés au fur et à mesure des recombinaisons et de leur activation (iv), jusqu'à une linéarisation complète du génome sans IS (v). Eventuellement, une réinfection peut avoir lieu (vi). La réduction de la taille du génome et la perte de séquences codantes renforcerait la dépendance d'un pathogène à un seul hôte. Ce serait actuellement le cas de *Bordetella pertussis*, mais également de *Shigella flexneri* et de *Yersinia pestis* (Chain et al., 2006; Diavatopoulos et al., 2005; Zaghloul et al., 2007).

L'expansion des IS chez *Bordetella pertussis* lui aurait permis de devenir un pathogène spécifique de l'homme. Les IS481 de par leur mobilité, permettent d'enrichir les souches de *Bordetella pertussis* déficientes en production de facteur de virulence. Cette théorie indique que la pression vaccinale exercée sur les souches circulantes augmenterait la probabilité de sélectionner une souche déficiente en un antigène, suite à l'insertion d'une IS481 ou d'une délétion provoquée par ces éléments génétiques mobiles.

### **3. La régulation des IS et l'expression de gènes environnant**

Les études de variabilité génomique en termes d'ordre des gènes ont montré que les réarrangements génomiques influent sur le taux d'expression des gènes. Les transposases sont transcrites par un promoteur interne, localisé dans l'IS, et sont dans la majorité des cas régulées négativement par un ARN antisens, transcrit par un promoteur interne localisé dans la séquence codante de l'IS.

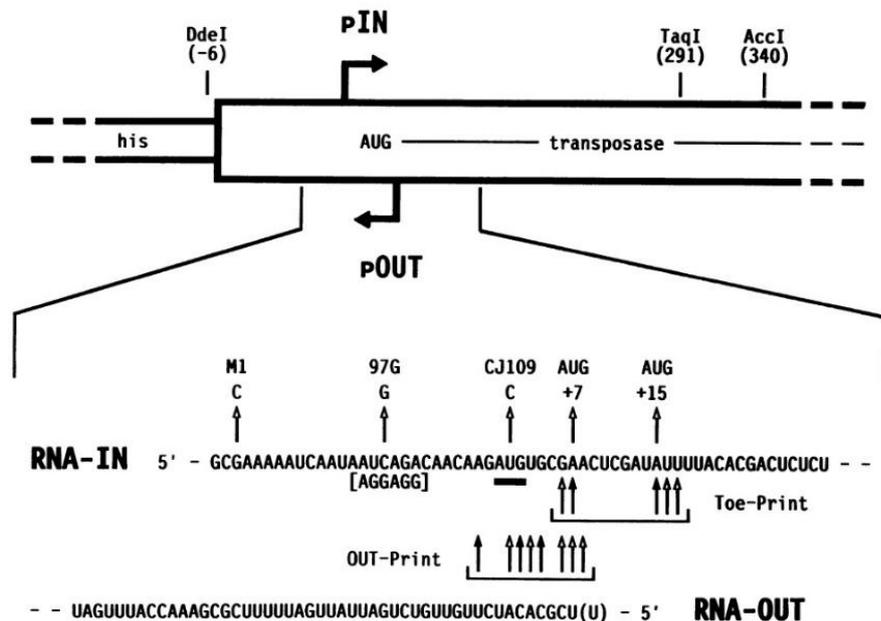
#### **i. Régulation et prolongements de transcription des IS**

Chez l'IS10, l'ARNm de la transposase est appelé RNA-IN, et est transcrit au départ d'un promoteur appelé  $P_{in}$  (Case et al., 1988). Cet ARN est régulé négativement par un ARN, appelé RNA-OUT, transcrit par un promoteur orienté en antisens du gène de la transposase appelé  $P_{out}$  (Figure 14). Le RNA-OUT s'hybride sur le site de fixation du ribosome et le codon d'initiation de la traduction du RNA-IN, bloquant la traduction de l'ARNm (Kittle et al., 1989; Ma and Simons, 1990; Simons and Kleckner, 1983). De plus,

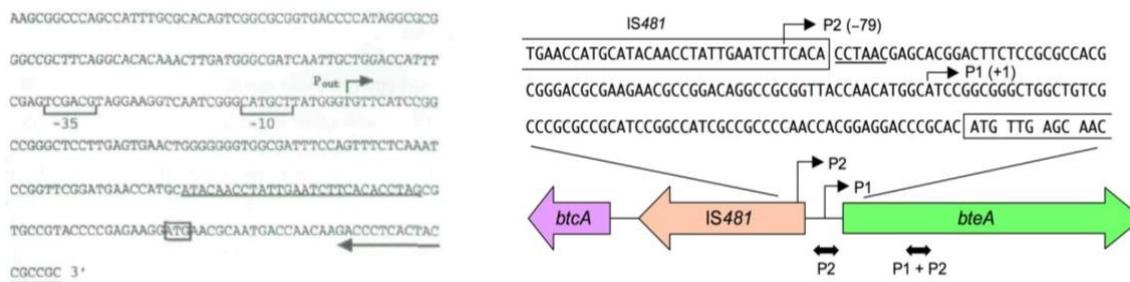
l'appariement des deux transcrits déstabilise l'ARNm de la transposase par recrutement de RNases clivant les ARN double-brins (Case et al., 1990). Bien que non décrites, d'autres IS comprennent ces deux promoteurs P<sub>in</sub> et P<sub>out</sub> (Safi et al., 2004).

Il a été observé pour plusieurs IS des prolongements de transcription orientés en sens de gènes, transcrits à partir des promoteurs P<sub>out</sub>. Selon la localisation de ces IS, elles peuvent pour certaines d'entre-elles transcrire un ARN au-delà de la séquence correspondant à l'IS jusqu'à la région génomique voisine, pour donner un ARNm supplémentaire, appelé « escaping the IS ». Ce genre de cas a pu être observé par exemple pour l'IS10, l'IS5 et l'ISAbal (Ciampi et al., 1982; Kamruzzaman et al., 2015; Wang and Roth, 1988).

Les IS peuvent activer l'expression des gènes des régions génomiques avoisinantes par la formation d'un promoteur hybride, P<sub>hyb</sub>, par insertion près d'une boîte promotrice -10 localisée dans la région génomique (Prentki et al., 1986). Les régions répétées gauches et droites possèdent des boîtes promotrices -35, et la région répétée gauche possède une boîte -10 (Chandler et al., 2015). Ces séquences cachées dans les régions répétées ont pour but de former un promoteur fort lors de la circularisation de l'IS après copie ou excision. De cette manière, la transposase est très fortement exprimée pour promouvoir l'insertion de l'élément mobile. L'insertion de l'IS près d'une boîte -10 dans la région génomique peut former un promoteur fort à l'aide des boîtes -35 situées dans les régions répétées de l'IS et transcrire le gène en aval (Prentki et al., 1986).



**Figure 14:** Organisation des séquences promotrices de l'IS10 chez *Escherichia coli*. Le rectangle noir correspond à la représentation schématique de l'IS10 sur le génome. La transposase est transcrite par le promoteur P<sub>in</sub> (Flèche noire) à 85pb de l'extrémité 5' de l'IS et est régulée par un ARN antisens, le RNA-OUT, transcrit à partir d'un promoteur P<sub>out</sub> interne à l'IS10 et orienté en antisens du gène à 115pb de l'extrémité 5'. L'ARN antisens s'hybride sur le codon d'initiation de la traduction (Souligné) et sur la séquence de fixation du ribosome (entre crochets) de l'ARNm. Figure de Ma & Simons, 1990



**Figure 15:** Activation de l'expression de gènes par les IS481.

A gauche, séquence du P<sub>out</sub> identifié chez l'IS481 BP3851. L'ARNm du gène *katA* est transcrit uniquement par le P<sub>out</sub> de l'IS481 située en amont du gène. Les boîtes -10 et -35 sont soulignées avec des crochets et le départ de transcription est indiqué par une flèche. Le codon d'initiation de la traduction de *katA* est encadré, la région répétée gauche de l'IS481 est soulignée. Figure de DeShazer *et al.*, 1994. A droite, activation de la transcription du gène *bteA* par son propre promoteur P1 ou par le P<sub>out</sub> de l'IS481 en amont P2. Les gènes sont indiqués par des flèches de couleur sur le génome. Figure de Han *et al.*, 2011

## ii. Les prolongements des IS481

Chez *Bordetella pertussis*, il a été montré que les IS481 pouvaient parfois activer l'expression des gènes avoisinant. La première mise en évidence concerne le gène *katA*, codant pour une catalase nécessaire à la protection contre le peroxyde d'hydrogène (DeShazer *et al.*, 1994). L'ARNm de *katA* est exclusivement transcrit à partir du P<sub>out</sub> de l'IS481 localisé en amont du gène (Figure 15). Le deuxième exemple de la littérature scientifique concerne le gène *bteA*, un gène codant pour une protéine effectrice du système de sécrétion de type III (T3SS), *vag* impliqué dans la virulence de *Bordetella pertussis* (Han *et al.*, 2011). Chez certaines souches, le gène *bteA* peut être transcrit par son propre promoteur et également par le P<sub>out</sub> d'une IS481 localisée en amont du gène (Figure 15). Cette configuration génomique est spécifique à certaines souches vaccinales, et n'est pas présente dans les souches cliniques de l'étude.

De ce fait, les IS481 ont un rôle dans l'adaptation du pathogène à la pression vaccinale induite dans les pays développés par le vaccin acellulaire. L'expansion de ces IS au sein du génome de *Bordetella pertussis* lui confère une capacité d'adaptation par des remaniements chromosomiques fréquents et des inactivations des antigènes utilisés dans la composition du vaccin actuel. Une autre particularité de ces IS est d'activer l'expression de gènes localisés dans les régions génomiques proches, permettant la transcription d'un ARNm, expliquant des différences de niveau d'expression de gènes impliqués dans la défense aux peroxydes et dans la sécrétion de facteurs de virulence (DeShazer *et al.*, 1994; Han *et al.*, 2011). Les différences d'expression de gènes observés lors des remaniements chromosomiques pourraient être expliquées par l'existence de ces prolongements de transcription (Brinig *et al.*, 2006). L'impact des prolongements de transcription de ces IS sur le transcriptome global de *Bordetella pertussis* pourrait apporter d'autres indices sur l'adaptation du pathogène à la pression vaccinale.

#### **IV. Les objectifs de thèse**

Les IS481 ont un rôle dans l'adaptation du pathogène à la pression vaccinale. Les études actuelles se sont concentrées essentiellement sur les remaniements génomiques, l'aspect évolutif et l'aspect adaptatif de *Bordetella pertussis* à son environnement (Diavatopoulos et al., 2005; Parkhill et al., 2003; Weigand et al., 2017; Xu et al., 2015). Malgré l'existence d'études mettant en évidence des prolongements de transcription des IS481, il n'existe à l'heure actuelle aucune étude montrant l'impact global des éléments génétiques mobiles sur le transcriptome du pathogène (DeShazer et al., 1994; Han et al., 2011). Une seule étude a pu relier les différences d'expression global de gènes avec l'ordre de leur séquence sur le génome, provoquées par les remaniements génomiques (Brinig et al., 2006). L'impact potentiel des IS sur le transcriptome total des procaryotes n'est pas non plus documenté chez d'autres espèces bactériennes, seules des études transcriptomiques sur l'adaptation des pathogènes aux antibiotiques et à différents stress par les éléments génétiques mobiles ont été mis en évidence (Depardieu et al., 2007).

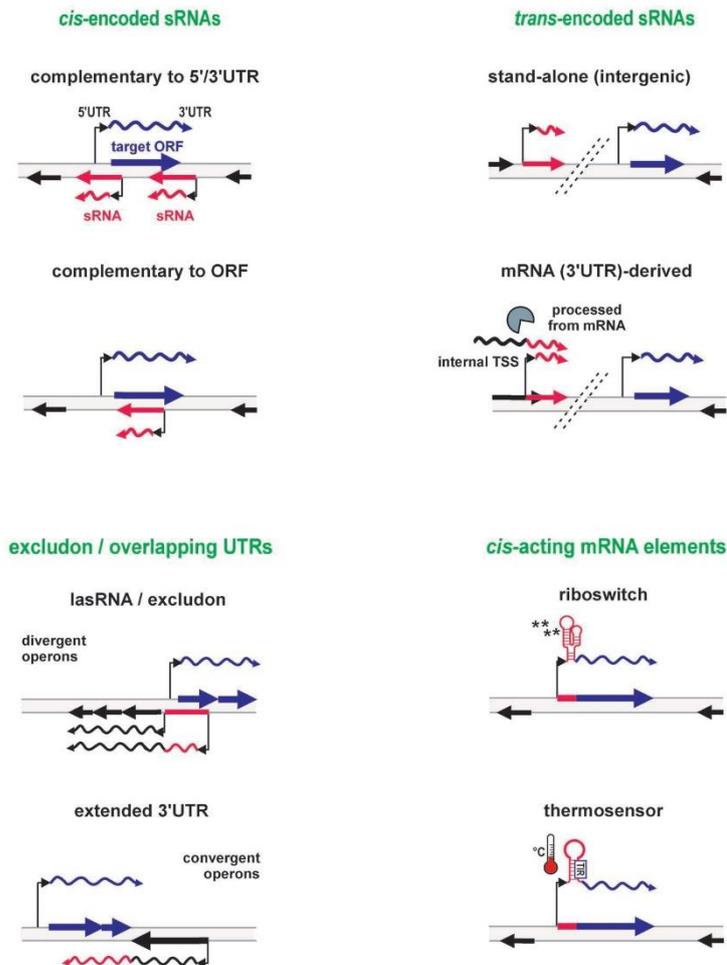
L'objectif de ma thèse a donc été d'étudier l'impact des IS481, IS1002 et IS1663 sur l'architecture globale du transcriptome du pathogène. Le transcriptome de *Bordetella pertussis* n'étant pas documenté dans la littérature scientifique, le premier objectif a été de déterminer le transcriptome primaire du pathogène et d'en déduire l'architecture de chaque transcrit détecté. Pour ce faire, nous avons utilisés dans un premier temps des approches de séquençage à ARN (RNA-seq) combinées à des analyses de prédictions bioinformatiques. Le cumul de ces approches nous a permis de mettre en évidence l'architecture globale du transcriptome de la souche de référence Tohama I en mettant en évidence les départs et fin de transcription de chaque ARN messenger, et de prédire les structures particulières des ARN mono- et polycistroniques et des transcrits sans régions non codantes en 5'. L'étude du transcriptome primaire a également permis de mettre en évidence des structures ARN potentiellement régulatrices et non codantes, tel que des 5' et 3' UTR chevauchants, des riboswitches et des excludons (Figure 16).

Il a été montré également que la transcription des ARN régulateurs non codant constituait l'une des divergences majeures entre les souches d'une même espèce, en plus des remaniements génomiques provoqués par les éléments génétiques mobiles (Kopf et al., 2015; Siguier et al., 2014). Ces ARN ont pour rôle de favoriser une adaptation et une réponse rapide à un stress environnemental ou à une condition donnée permettant à la cellule de survivre (Svensson and Sharma, 2016; Wagner and Romby, 2015). Il existe différents types d'ARN régulateurs, transcrits indépendamment ou non des ARN messagers (Figure 16). Pour déterminer l'impact des séquences d'insertion sur l'architecture globale du transcriptome codant et non-codant de *Bordetella pertussis*, une cartographie de candidats ARN régulateurs a été édifiée en parallèle des transcrits messagers et de leurs prolongements de transcription. Actuellement, peu de structures transcriptomiques ont été étudiées chez les *Bordetellae*. Seuls 2 riboswitches liant la glycine et la flavine mononucléotide, 13 ARN potentiellement non codants et régulateurs, dont un ARN antisens de l'opéron *bvgAS* et un ARN différenciellement

exprimé en condition de virulence, ont été validés chez *Bordetella pertussis* (Hot et al., 2011). Cette liste de candidats ARN régulateurs a donc été prédite en cumulant des prédictions issues des séquençages à ARN et des prédictions issues d'analyses bioinformatiques.

A partir de ces données transcriptomiques, les prolongements de transcription issus des séquences d'insertion ont été annotés et leur impact sur le transcriptome global étudié. Une liste de ces prolongements a été comparée à d'autres souches du pathogène. Le rôle de ces transcrits a été également investigué à partir des exemples étudiés dans la littérature scientifique et d'un ARN candidat identifié comme étant un régulateur potentiel d'un gène majeur de virulence transcrit à partir d'une IS481. L'impact de ces transcrits dépend de leur orientation par rapport aux gènes environnants, et serait spécifique de chaque souche du pathogène.

Cette thèse de doctorat a donc pour but de (i) déterminer le transcriptome primaire de *Bordetella pertussis* afin de cartographier les structures codantes mono- et polycistroniques des ARN messenger en condition classique de virulence, (ii) de mettre en évidence le transcriptome non-codant du pathogène en identifiant des candidats ARN régulateurs transcrits en *cis* et en *trans*, (iii) d'étudier les prolongements de transcription émanant des séquences d'insertion et de les comparer chez différentes souches du pathogène, et (iv) d'étudier la fonction potentielle de ces transcrits sur les gènes environnants à partir d'un ARN orienté en antisens d'un gène majeur de virulence, *fim2*. Cette étude soulève l'hypothèse que les IS, en plus d'inactivation par insertion dans des séquences codantes, peuvent aussi agir constamment sur l'expression des gènes environnant par transcription d'ARN messagers et/ou d'ARN régulateurs. Ce phénomène pourrait apporter de nouveaux éléments d'explication de l'adaptation du pathogène à la pression vaccinale.



**Figure 16:** Les différents types d'ARN régulateurs bactériens.

Il existe plusieurs types d'ARN régulateurs. Les ARN non codant (small RNA) sont soit transcrits en antisens de leur cible en cis, auxquels cas ils peuvent être complémentaires des régions non traduites (UTR) ou de la séquence codante. Ils peuvent également être transcrits à un autre endroit du génome en trans, possédant leur propre locus de transcription, enfin ils peuvent être transcrits ou issus du clivage d'un 3'UTR de gène. Les ARN régulateurs peuvent également être des prolongements de transcription d'ARNm. Il existe des longs 5' UTR qui peuvent se chevaucher, formant ce que l'on appelle des excludons ou des longs ARN antisens. De même, de longs 3'UTR peuvent se chevaucher. Les gènes issus de ces structures codent généralement pour des fonctions opposées. Enfin, il existe des ARN régulateurs très structurés en 5'UTR de gènes régulant directement le niveau de transcription ou de traduction de l'ARNm qui en est issu. Les riboswitches sont capables de lier une molécule appelée ligand pour réguler l'expression du gène en aval, impliqué dans le métabolisme ou les réactions enzymatiques de celle-ci. Enfin, les thermosenseurs sont des structures sensibles à la température. La transcription ou la traduction du gène dépend de ce fait de la température de culture. Ces derniers sont généralement utiles aux pathogènes, la température de l'hôte étant différente des températures du milieu extérieur.

Barre grise : Génome bactérien. Flèches bleues : Gène cible/régulé. Flèche noire : Gène. Flèches rouges : Locus d'ARN régulateur. Lignes ondulantes bleues : ARNm. Lignes ondulantes rouges : ARN régulateur. Flèche noire fine : Départ de transcription / promoteur. Figure de Svesson & Sharma, 2016



# Résultats et Discussions

# I. Annotation du transcriptome primaire de *Bordetella pertussis*

Pour déterminer l'impact des IS481 sur le transcriptome de *Bordetella pertussis*, la première nécessité était de cartographier tous les types d'ARN du pathogène. Pour cela, nous avons déterminé le transcriptome primaire de la souche de laboratoire BpSM, un mutant dérivé de la souche vaccinale Tohama I résistant à la streptomycine (Antoine and Loch, 1990).

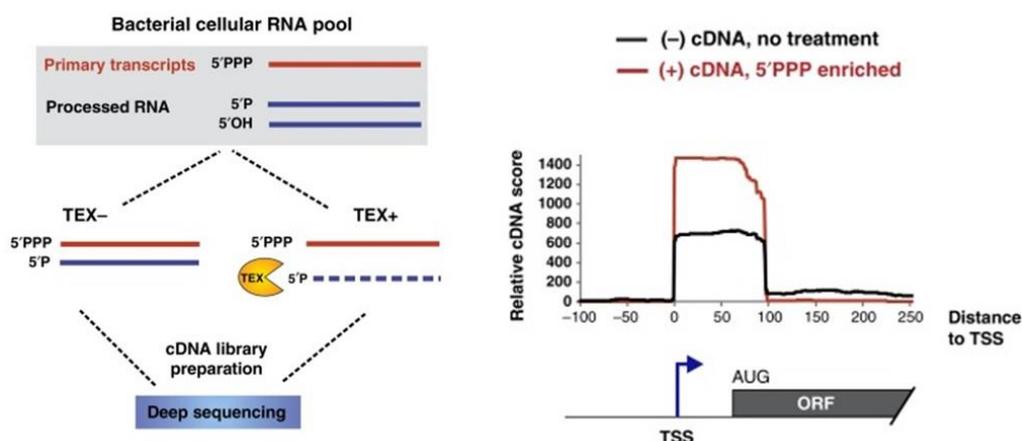
## 1. Annotation du transcriptome en condition Bvg<sup>+</sup>

*Bordetella pertussis* est un pathogène qui est retrouvé essentiellement virulent *in vivo* (Byrd et al., 2013; Veal-Carr and Stibitz, 2005). De ce fait, le transcriptome primaire a été déterminé en condition classique de culture pour maintenir le pathogène en phase Bvg<sup>+</sup> (Matériel et Méthodes 1).

### i. Les séquençages ARN

Pour déterminer l'architecture globale des transcrits de *Bordetella pertussis*, nous avons choisi de séquencer les ARN du pathogène en séquençage profond (« Deep RNA-seq ») par une approche de séquençage Haut-Débit. Les bibliothèques d'ADN complémentaires ont été construites à partir des ARN totaux extraits du pathogène et épuisés en ARN ribosomiques (Matériel et Méthode 3 & 4). Les ADN complémentaires ont été séquencés et orientés sur le génome de façon à distinguer les ARN transcrits du brin positif des ARN transcrits du brin négatif.

La seconde approche de séquençage choisie permet de prédire de nouveaux départs de transcription, appelée séquençage différentiel d'ARN (« differential RNA-seq »)



**Figure 17 :** Prédiction de départs de transcription par séquençage différentiel d'ARN. Cette technique consiste à séquencer une bibliothèque dont les ARN ont été traités à la 5' monophosphate exonucléase (TEX) avec une bibliothèque dont les ARN n'ont pas été traités (Gauche). Le traitement enrichit la bibliothèque en transcrits primaires, présentant un triphosphate sur leur extrémité 5'. Les enrichissements en départs de séquences de la bibliothèque TEX+ permettent de prédire de nouveaux départs de transcription (Droite). Figure de Sharma & Vogel, 2014

(Matériel et Méthode 5) (Figure 17) (Sharma and Vogel, 2014). Ce séquençage est basé sur la comparaison du séquençage de deux librairies dont les ARN ont été traités (TEX+) ou non (TEX-) à la 5' monophosphate exonucléase. Cette enzyme dégrade les ARN non primaires, issus de maturation ou de dégradation pour enrichir la librairie en transcrits primaires. Cet enrichissement est obtenu grâce à la spécificité de la TEX pour les extrémités 3' mono-phosphate des ARN non primaires. Cette technique permet de prédire de nouveaux départs de transcription par comparaison de ces enrichissements après alignement des séquences sur le génome. Quatre répliques biologiques ont été séquencés par séquençage différentiel, et un total de 36,7 millions de séquences a pu être aligné sur le génome de *Bordetella pertussis*.

## ii. Prédiction de départs et annotation du transcriptome

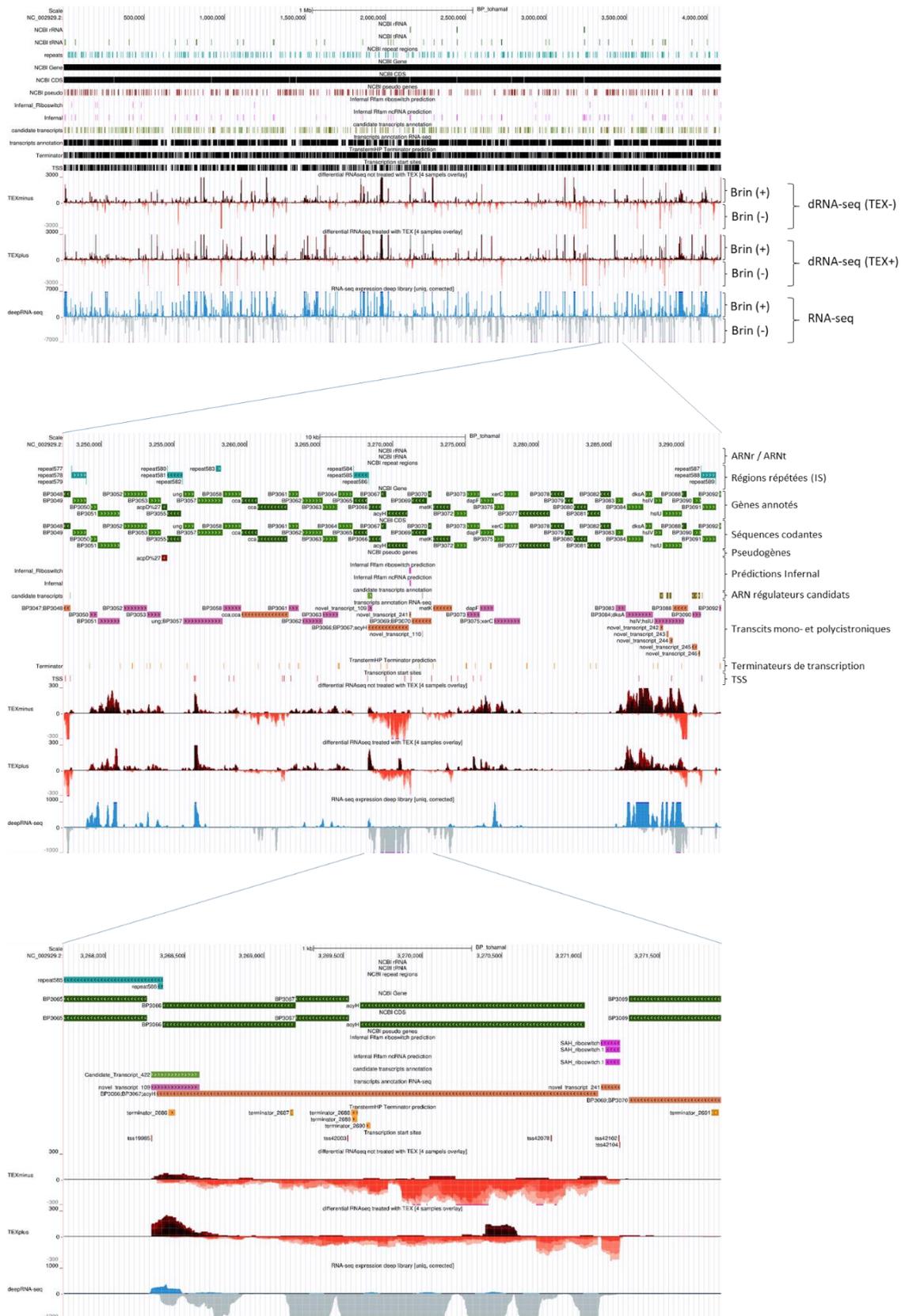
La combinaison des deux approches de séquençage du transcriptome nous a permis de déterminer à la fois l'architecture de chaque transcrit, de prédire de nouveaux ARN et de quantifier leur abondance dans la cellule en phase de virulence. Ces prédictions ont été obtenues par diverses analyses bioinformatiques permettant de déterminer les départs de transcription (TSS), les structures opéroniques, les terminateurs de transcription (Matériel et Méthode 17).

Les séquences obtenues ont été alignées sur le génome de *Bordetella pertussis* Tohama I (NC\_002929.2). Les données transcriptomiques ont été annotées à partir de la fusion des données des répliques de séquençage TEX- issues du séquençage différentiel. L'alignement des séquences sur les brins spécifiques entraîne entre 0,5 et 10% d'alignements non-spécifiques des séquences en antisens les unes par rapport aux autres, engendrant des biais dans la prédiction de nouveaux transcrits (Levin et al., 2010). Pour pallier à ce problème, une approche de correction a été développée par notre collaborateur Fabian Amman pour détecter et corriger l'effet des séquences erronées en antisens (Matériel et méthode 17). Le pourcentage d'erreur associé à l'annotation *de novo* du transcriptome de *Bordetella pertussis* est de 1,4% et a été pris en compte dans la prédiction de l'architecture des ARN.

Les départs de transcription ont été prédits par le logiciel TSSAR par comparaison des deux séquençages différentiels (Amman et al., 2014). Les départs de transcription ont été sélectionnés avec un seuil d'enrichissement d'un facteur 2 minimum, et les départs consécutifs séparés par moins de 10 paires de bases ont été fusionnés et ajustés à l'enrichissement le plus important. Au total, 2722 TSS ont été prédits en condition Bvg<sup>+</sup>. L'architecture des transcrits a été prédite en prenant compte des séquences en antisens erronées, les prédictions de départ de transcription, l'annotation du génome et les données de séquençage (Matériel et méthode 17).

## iii. Visualisation des données

Les résultats de séquençage et les prédictions bioinformatiques ont été déposés dans les banques de données (BioProject ID: PRJNA430365) et téléchargés sur le centre d'assemblage UCSC Genome Browser visible à l'adresse : <http://bit.ly/2Euo5HZ>.



**Figure 18:** Visualisation des données de prédictions sur UCSC Genome Browser. Les résultats de séquençage sont alignés sur le génome de *Bordetella pertussis* Tohama I annoté et représenté sur un histogramme de profondeur. Les correspondances sont annotées à droite des captures d'écran. RNA-seq : Séquençage profond des ARN. dRNA-seq : Séquençage différentiel des ARN

Type de transcrit testé	Nom du transcrit	TSS prédit	TSS publié	TSS validé par 5'RACE
ARNm	<i>fim2</i>	-	1 176 514 (Chen et al. 2010)	1 176 514
	<i>bipA</i>	1 164 469	1 164 472 (Deora et al. 2001)	1 164 471
		1 164 588		
<i>ptxA</i>	3 988 233	3 988 233 (Nicosia A & Rappuoli R. 1987)	3 988 233	
Riboswitches prédits	<i>SAM <math>\alpha</math></i>	48 079	-	48 080
	<i>Glycine</i>	197 081	-	197 082
		197 142		
		197 190		
	<i>SAH</i>	197 284	-	3 271 247
3 217 242				
	3 271 247			
ARN régulateurs non codant prédits	<i>BPnc180</i>	776 799	-	776 799
	<i>BPnc468</i>	1 968 972	-	1 968 972
	<i>BPnc962</i>	3 956 913	-	3 956 960
		3 956 936		
		3 956 951		
	3 956 960			

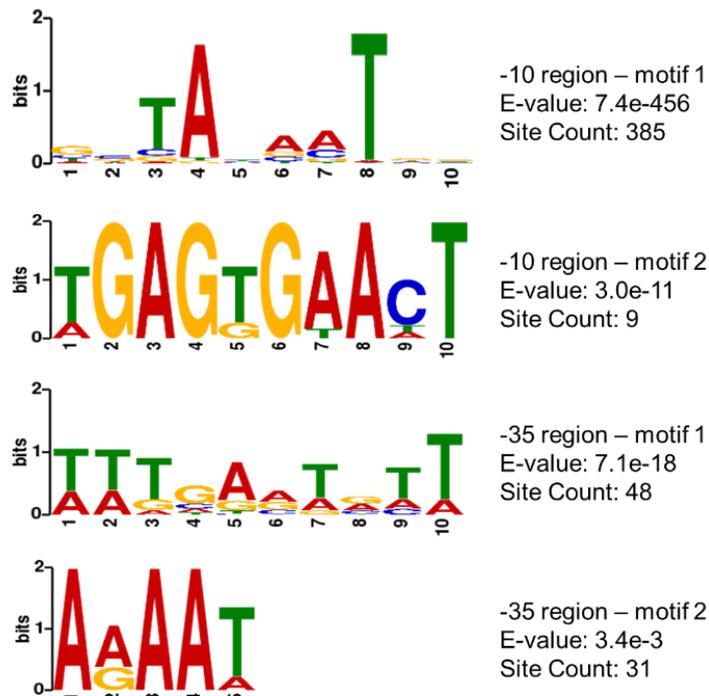
**Tableau 2:** Position des TSS prédits et des TSS validés par 5'RACE sur séquençage à Haut-Débit des transcrits candidats.

Les TSS choisis ont été testés pour des ARN messagers, des riboswitches prédits (Résultats Chapitre 1, Partie 2, Sous-partie iii) et des ARN régulateurs potentiellement non-codants prédits (Résultats Chapitre 2, Partie 1, Sous-partie 1)

La représentation visuelle regroupe les données de séquençage à ARN (TEX+, TEX-, séquençage profond), alignées sur l'annotation du génome de *Bordetella pertussis* Tohama I NC\_002929.2 regroupant les séquences codantes, les pseudogènes, les séquences d'insertion et les ARN de transfert et ribosomiques (Figure 18). Les résultats de séquençage sont représentés en histogramme de profondeur pour chaque brin d'ADN. Les corrections des séquences erronées sont également visibles dans la piste « pooled RNAseq librairies ». Les prédictions de transcrits sont également alignées sur le génome en parallèle des séquençages ARN.

## 2. Identification de l'architecture et des structures ARN

L'analyse des séquençages à ARN couplés à différents logiciels bioinformatiques nous ont permis d'identifier la structure des ARN messagers, et d'identifier des candidats ARN régulateurs couplés à leur transcription. Les départs de transcription prédits ont été classés en fonction de leur orientation et de leur localisation sur le génome en regard de l'annotation. Les TSS positionnés à moins de 250 pb en sens des gènes annotés ont été catégorisés comme TSS primaires pour le plus éloigné, et TSS secondaires pour les autres. Certains TSS ont été catégorisés comme étant internes, ils sont situés dans la séquence codante en sens du gène. Enfin, des TSS orientés en antisens de gène et des TSS orphelins situés à plus de 250 pb des annotations seront détaillés dans la seconde partie de résultats.



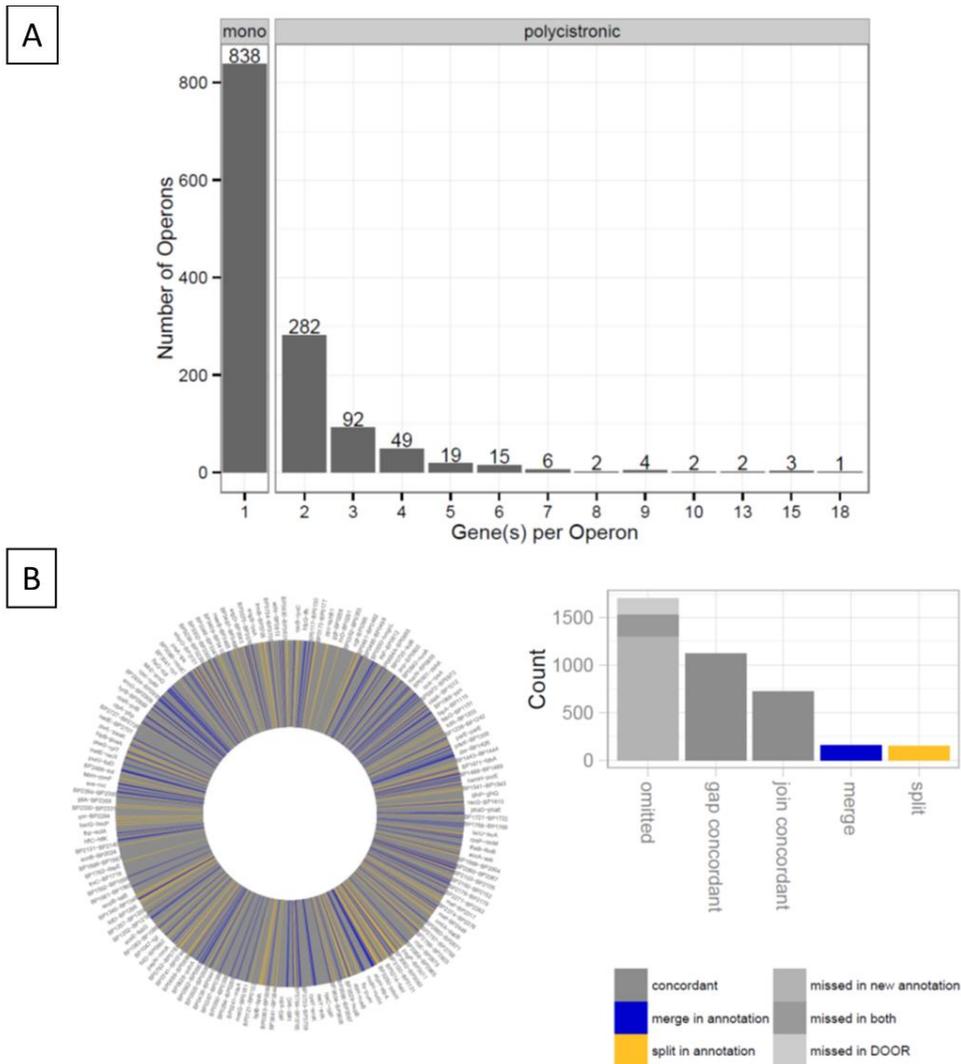
**Figure 19:** Motifs des séquences promotrices prédites par le logiciel MEME. Les deux motifs prédominant des boîtes -10 et -35 sont représentés dans l'ordre. La probabilité d'apparition aléatoire (E-value) et le nombre d'apparition du motif au sein du génome de la souche Tohama I sont indiqués. La probabilité d'apparition aléatoire des bases est donnée en bit-score.

### i. Les ARN messagers mono- et polycistroniques

Sur les 2722 TSS prédits en condition Bvg<sup>+</sup>, 614 TSS ont pu être identifiés comme TSS primaires d'un ARN messenger mono- ou polycistronique, dont 99 d'entre eux présentent au moins un TSS secondaire. Pour valider la solidité des prédictions de départ de transcription, certains de ces TSS ont pu être validés par 5'RACE en séquençage à haut-débit en suivant un protocole publié (Beauregard et al., 2013) (Matériel et Méthode 6) (Tableau 2). Des TSS d'ARN messenger et de prédictions d'ARN régulateurs ont été ainsi validés. La position des TSS déterminée par 5'RACE est identique à celle des TSS prédits par séquençage différentiel pour quasiment tous les candidats testés, seuls 3 TSS prédits s'éloignent de 1 ou 2 bases de la position déterminée par 5'RACE. De plus, les TSS testés des ARNm de *fim2*, *bipA* et *ptxA* par 5'RACE sont conformes aux TSS publiés (Chen et al., 2010; Deora et al., 2001; Nicosia and Rappuoli, 1987).

Les séquences retrouvées en amont des TSS prédits ont été utilisées pour identifier une éventuelle séquence consensus pour les boîtes -10 et -35 à l'aide du logiciel MEME (Matériel et Méthode 17) (Bailey and Elkan, 1994) (Figure 19). La séquence consensus de la boîte -10 en amont des TSS, TANAAT, se rapproche de séquences retrouvées chez d'autres procaryotes et est retrouvée plus de 385 fois sur les 614 TSS primaires (TANNNT) (Narlikar, 2014). Cependant, aucune séquence consensus de la boîte -35 n'a pu être retrouvée chez *Bordetella pertussis*.

Les données de séquençage à ARN ont été exploitées pour caractériser l'architecture transcriptionnelle des gènes codants annotés (Matériel et Méthode 17). L'analyse a permis de prédire la structure des transcrits de 2262 gènes sur les 3871 gènes annotés du génome de la souche Tohama I. La structure des gènes manquant correspond pour certains d'entre eux à des gènes exprimés en phases Bvg<sup>-</sup> ou Bvg<sup>i</sup>, et ne sont pas exprimés dans les conditions étudiées. Au total, 1315 structures ont été prédites, 838

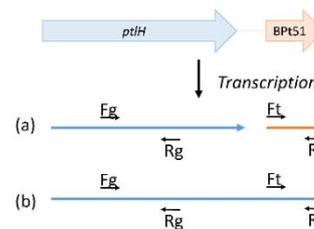
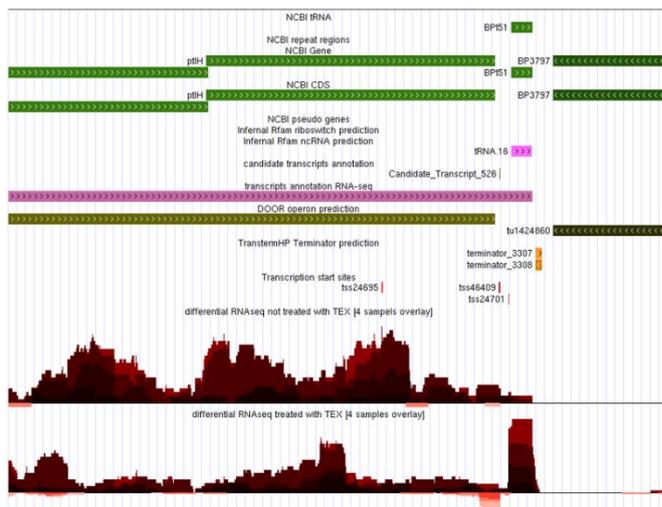


**Figure 20:** Prédiction des structures mono- et polycistroniques du transcriptome de la souche Tohama I.

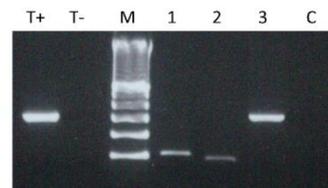
(A): Quantification du nombre de gènes transcrits par cistrons. (B) : Comparaison des prédictions opéroniques entre l’annotation développée et les prédictions de DOOR 2.0.

gènes sont transcrits en monocistron et 477 sont transcrits en opéron polycistronique (Figure 20A). La majorité des gènes polycistroniques correspondent à deux gènes cotranscrits et un polycistron regroupant jusqu’à 18 séquences annotées a pu être détecté. Les gènes de cet opéron prédit (BP3626 à BP3643) correspondent à des gènes impliqués dans la traduction des ARN messagers.

Les gènes mono- et polycistroniques prédits par notre analyse ont été comparés aux prédictions du logiciel DOOR 2.0 (Mao et al., 2014) (Figure 20B) (Matériel et Méthode 17). Les extrémités des transcrits de 2151 gènes ont été analysées par les deux méthodes. Les deux méthodes ont prédits une organisation architecturale des transcrits similaires pour 1840 de ces extrémités, co-transcrites ou non. Notre annotation prédit une rupture opéronique dans 152 cas et une continuité de transcription dans 159 cas qui ne sont prédites par DOOR 2.0. Cependant, les deux méthodes prédisent une rupture dans 1121 cas, et une continuité de transcription dans 719 cas. L’architecture



Cas (a) : Unité de transcription indépendante, PCR négative  
 Cas (b) : Unité de transcription identique, PCR positive



**Figure 21:** Validation de la co-transcription de l'ARNt BpT51 avec l'opéron de la toxine pertussique.

Les prédictions de séquençage sont montrées sur la capture d'écran de gauche représentant nos données de séquençage. L'indépendance de transcription de BpT51 avec l'opéron *ptx-ptl* a été testée par RT-PCR en utilisant deux couples d'amorces différents. Les ADN complémentaires ont été générés avec les amorces Rg (amorce antisens du gène *ptlH*) et Rt (amorce antisens de l'ARNt) et l'expression de chaque transcrite a été validée indépendamment avec les amorces Fg/Rg (amorces sens et antisens de *ptlH*) (Ligne 1) et Ft/Rt (amorces sens et antisens de BpT51) (Ligne 2). La co-transcription de BpT51 et de *ptlH* a été testée à partir des ADNc générés avec l'amorce Rt, et amplifiés avec les amorces Fg/Rt (Ligne 3). T+ : Témoins positif de PCR, test des amorces Fg/Rt sur de l'ADN génomique. T- : Témoin négatif de PCR en absence de matrice ADN. M : Marqueur de taille. C : Contrôle négatif d'absence d'ADN génomique : transcription inverse faite en absence d'enzyme.

transcriptomique des séquences codantes est retrouvée en majorité par les deux méthodes.

L'opéron de la toxine pertussique a été prédit à 15 cistrons, incluant les gènes de la toxine, les gènes du système de sécrétion de la toxine et un ARN de transfert. L'opéron de la toxine pertussique est prédit à 14 cistrons avec le logiciel DOOR 2.0, et n'inclue pas l'ARN de transfert à asparagine BpT51 (Figure 21). L'indépendance de transcription de cet ARN de transfert a été testée par RT-PCR pour déterminer si celui-ci est co-transcrit ou non avec l'opéron de la toxine pertussique (Matériel et Méthode 9). L'amplification de l'ADN complémentaire a montré une jonction de transcription entre les deux transcrits, ils seraient donc issus d'une même unité de transcription. La validation du cistron renforce la qualité des prédictions issues de notre annotation du transcriptome primaire.

## ii. Les transcrits sans régions non codantes en 5' : « leaderless »

L'architecture de certains ARN messagers montre un départ de transcription aligné sur le codon d'initiation de la traduction de la séquence codante du gène. Sur les 614 TSS primaires, 35 d'entre eux ne présentent aucune séquence non codante en aval de leur séquence codante (Tableau 3A).

A

ID <sup>a</sup>	Gene Name	Predicted Function	Functional categories	Other start codon <sup>b</sup>
BP0109	rfbB	dTDP-glucose 4,6-dehydratase	Surface proteins	no
BP0209		putative GntR-family transcriptional regulator	Regulators	no
BP0570		putative exported protein	Surface proteins	yes: 8 codons further
BP0690		conserved hypothetical protein	Conserved/Hypothetical protein	no
BP0701		conserved hypothetical protein	Conserved/Hypothetical protein	no
BP0744		putative C-terminal region of dehydrogenase(Pseudogene)	Regulators	no
BP1079		conserved hypothetical protein	Conserved/Hypothetical protein	no
BP1088		putative membrane protein	Surface proteins	yes: 27 codons further
BP1089		putative membrane protein	Surface proteins	no
BP1116		conserved hypothetical protein	Conserved/Hypothetical protein	yes: 21 codons further
BP1227	radA	DNA repair protein	Information transfer	no
BP1477		conserved hypothetical protein	Conserved/Hypothetical protein	yes: 25 codons further
BP1556		conserved hypothetical protein	Conserved/Hypothetical protein	no
BP1983		putative extracellular solute-binding protein	Surface proteins	yes: 4 and 19 codons further
BP2131		conserved hypothetical protein	Conserved/Hypothetical protein	no
BP2145		putative ArsR-family transcriptional regulator	Regulators	yes: 24 codons further
BP2286		putative transcriptional regulator (MarR family)	Regulators	no
BP2312		conserved hypothetical protein	Conserved/Hypothetical protein	no
BP2379		putative chaperonin	Pathogenicity/Adaptation/Chaperones	no
BP2467		hypothetical protein	Unknown	no
BP2571		putative MerR-family transcriptional regulator	Regulators	no
BP2720		putative TetR-family transcriptional regulator	Regulators	no
BP2905	glcD	glycolate oxidase subunit	Central/Intermediary/misc metabolism	no
BP3003		putative membrane protein	Surface proteins	no
BP3070		conserved hypothetical protein	Conserved/Hypothetical protein	no
BP3222	ompR	transcriptional regulatory protein OmpR	Regulators	no
BP3381		phage-related hypothetical protein		no
BP3466		putative ArsR family transcriptional regulator	Regulators	no
BP3491	ndh	putative NADH dehydrogenase	Energy metabolism	no
BP3534	basR	response regulator protein	Regulators	no
BP3656	aroK	shikimate kinase I	Central/Intermediary/misc metabolism	no
BP3733		conserved hypothetical protein	Conserved/Hypothetical protein	no
BP3745		conserved hypothetical protein	Conserved/Hypothetical protein	no
BP3748	rpoH	RNA polymerase sigma-32 factor	Information transfer	yes: 13 codons further
BP3768	hisD	putative histidinol dehydrogenase	Central/Intermediary/misc metabolism	no

a: ID: gene identification

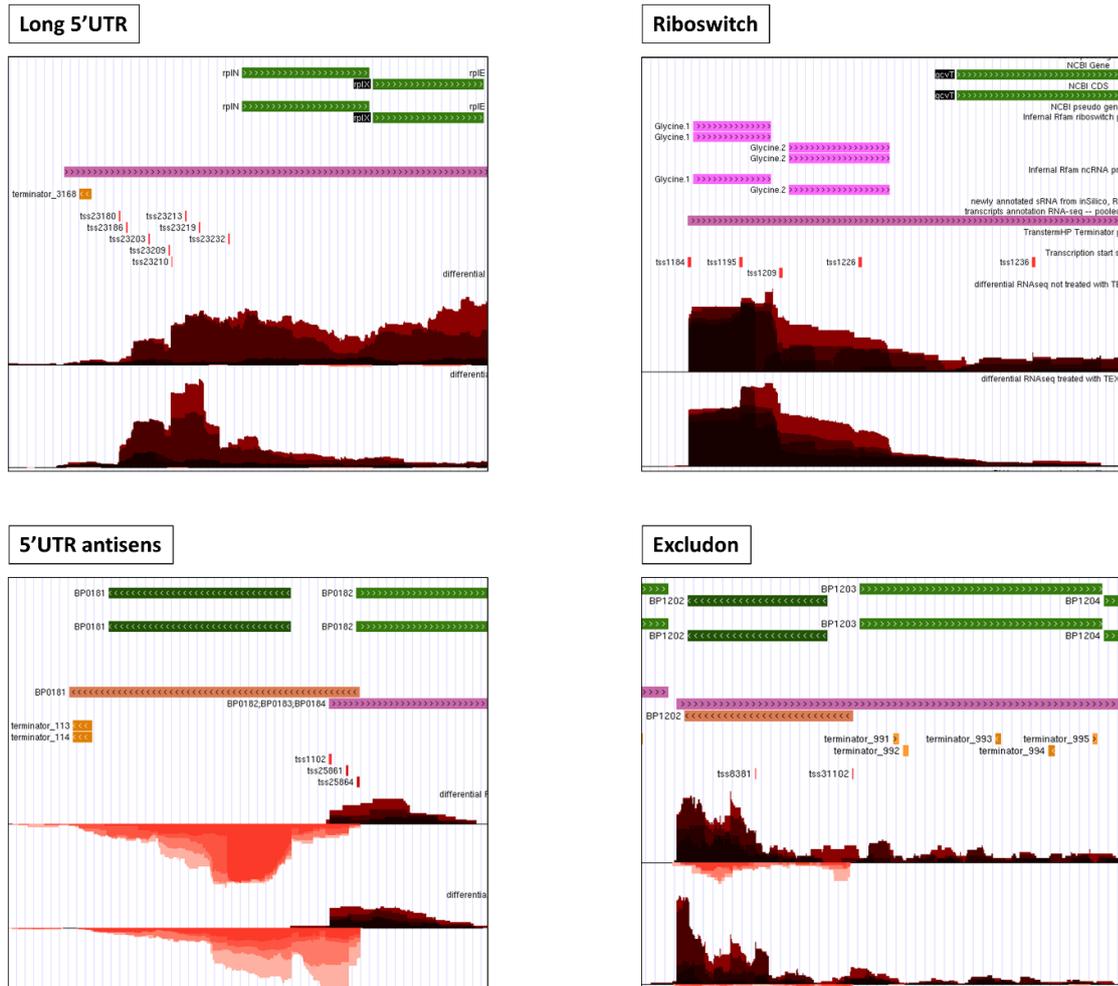
b: Other start codon: Gene ORF were scanned for any Methionine residue within the first 30 annotated amino acids. The presence of a methionine could indicate a wrong translation start annotation.

B

Functional categories	Nombre per Functional categories	%	Pval Hypergeo test
Central/Intermediary/misc metabolism	3	8.8	n.s.
Conserved/Hypothetical protein	11	32.3	n.s.
Energy metabolism	1	2.9	n.s.
Information transfer	2	5.8	n.s.
Pathogenicity/Adaptation/Chaperones	1	2.9	n.s.
Regulators	9	26.5	0.0014
Surface proteins	6	17.6	n.s.
Unknown	1	2.9	n.s.

**Tableau 3:** Prédications de gènes "leaderless".

(A) : Liste des gènes prédits sans région non codantes en 5'. ID : Identifiant du gène. (B) : Enrichissement des catégories de fonctions des gènes évalué par calculs hypergéométriques. Pval hypergeo test : P-value des proportions hypergéométriques. N.s. : Non-significatif. Tests calculés en utilisant le site : <https://www.geneprof.org/GeneProf/tools/hypergeometric.jsp>



**Figure 22:** Configuration des prédictions des longues régions non traduites en 5' de gènes (5'UTR).

Les analyses ont permis de prédire des long 5'UTR de plus de 100 nucléotides, des riboswitches, des 5'UTR potentiellement régulateurs en antisens de gènes, et un excludon recouvrant la totalité d'un gène en sens opposé et en aval de la séquence codante du gène auquel il appartient. Un exemple de chaque cas est illustré ci-dessus.

Pour 7 de ces gènes, un second codon d'initiation potentiel a été trouvé et pourraient résulter d'une annotation erronée des séquences codantes. La majorité des gènes « leaderless » codent pour un régulateur transcriptionnel (26%) (Tableau 3B). Les autres codent en majorité pour des protéines hypothétiques de fonction encore inconnue. La représentation de la classe 'régulateur transcriptionnel' est significative ( $p$ -value = 0,0014), elle inclue le régulateur transcriptionnel *ompR*, prédit comme étant co-transcrit avec le gène *risK* par notre annotation. Le facteur sigma 32, codé par le gène *rpoH* impliqué de la réponse aux chocs thermiques, est également prédit « leaderless » (Grossman et al., 1987). Ces transcrits représentent 0,90% de la totalité des séquences annotés, proportion similaire à celle retrouvée chez d'autres espèces procaryotes (0,46% pour *Helicobacter pilory*, 0,48% pour *Salmonella typhimurium* et 1,41% pour *Klebsiella pneumonia*) (Kroger et al., 2012; Seo et al., 2012; Sharma et al., 2010).

### iii. Les extrémités non traduites en 5' des ARN messagers

A l'instar des transcrits « leaderless », 120 transcrits annotés possèdent des régions non traduites en 5' d'une taille supérieure à 100 nucléotides. Les longues régions non codantes en 5' sont souvent des structures complexes ayant un rôle de régulateur de la transcription ou de la traduction du gène en aval (Oliva et al., 2015; Svensson and Sharma, 2016). L'analyse du transcriptome de *Bordetella pertussis* en phase Bvg<sup>+</sup> a permis d'identifier et de prédire différentes structures potentiellement régulatrices (Figure 22).

Parmi les long 5'UTR, certains d'entre eux se retrouvent en antisens des transcrits des gènes en amont. Les structures prédites par notre annotation montrent qu'il existe 37 longs 5'UTR chevauchant entre deux gènes. Cette organisation remarquable est le siège de régulations potentielles. Un de ces chevauchements, donné par le promoteur de BP1203, recouvre la séquence codante entière du gène opposé en aval, BP1202. Cette structure correspond à la définition d'un excludon, déjà montré chez d'autres pathogènes, notamment chez *Listeria monocytogenes* (Sesto et al., 2013).

Certains de ces longs 5'UTR ont été prédits comme étant des riboswitches *in silico* à l'aide du logiciel Infernal (Matériel et Méthode 17) (Tableau 4A) (Nawrocki, 2014). Les riboswitches sont des 5'UTR de gènes ayant une structure tertiaire complexe, capable

A	Cible	Rfam ID	Localisation	Départ	Fin	Brin	E-value
	SAM_alpha	RF00521	BP0047_promoter_47985_48208	48078	48158	+	3.7e-08
	Glycine	RF00504	BP0195_promoter_197008_197462	197088	197180	+	2.8e-11
	Glycine	RF00504	BP0195_promoter_197008_197462	197202	197321	+	7.3e-12
	TPP	RF00059	BP0206_promoter_209453_209831	209596	209707	+	2.4e-07
	FMN	RF00050	BP0471_promoter_481772_482205	481881	482034	+	3.1e-29
	SAH_riboswitch	RF01057	BP3068_promoter_3270966_3271305	3271127	3271247	-	4.1e-09
	YybP-ykoY (manganese)	RF00080	BP3410_promoter_3619131_3619507	3619234	3619427	-	9.6e-15
	Cobalamin	RF00174	BP3595_promoter_3811442_3811838	3811547	3811706	+	1.1e-11

B	Symbole	Aliases	Description	Départ de prédiction	Fin de prédiction	Brin	TSS en amont (>100nt)	Prédiction d'ARN non codant
	glyQ	BP0033, glyS(a)	glycine--tRNA ligase sous-unité alpha	37013	37924	+	-	-
	glyS	BP0034, glyS(b)	glycine--tRNA ligase sous-unité bêta	37926	40064	+	-	-
	argS	BP0115	arginine--tRNA ligase	111937	113724	-	Oui	-
	metG	BP0180	methionine--tRNA ligase	178634	180712	-	-	-
	aspS	BP0709, tIs	aspartate--tRNA ligase	723420	725210	+	-	-
	lysS	BP1102, asuD, herC	lysine--tRNA ligase	1151365	1152885	-	-	-
	trpS	BP1143	tryptophan--tRNA ligase	1206090	1207445	+	-	-
	thrS	BP1497	threonine--tRNA ligase	1574769	1576718	+	Oui	BPnc372
	ileS	BP1753, ilvS	isoleucine--tRNA ligase	1837609	1840470	-	-	-
	alaS	BP1836, lovB	alanine--tRNA ligase	1927646	1930270	+	-	BPnc458
	cysS	BP1908	cysteine--tRNA ligase	2010516	2011976	+	Oui	BPnc480
	leuS	BP2044	leucine--tRNA ligase	2161527	2164184	-	-	-
	hisS	BP2198	histidine--tRNA ligase	2319344	2320648	-	-	BPnc552
	valS	BP2203	valine--tRNA ligase	2324418	2327300	+	-	BPnc552
	BP2470	BP2470	serine--tRNA ligase	2612694	2614049	-	-	-
	pheT	BP2573	phenylalanine--tRNA ligase sous-unité bêta	2724449	2726866	-	-	BPnc652
	pheS	BP2574	phenylalanine--tRNA ligase sous-unité alpha	2726879	2727901	-	-	BPnc652
	tyrS	BP2954	tyrosine--tRNA ligase	3141499	3142728	-	Oui	-
	gltX	BP2978	glutamate--tRNA ligase	3166143	3167525	+	Oui	-
	glnS	BP3226	glutamine--tRNA ligase	3438453	3440216	+	-	-
	proS	BP3348, drpA	proline--tRNA ligase	3569728	3571458	-	-	BPnc848

**Tableau 4:** Prédiction de riboswitches sur le génome de la souche Tohama I.

(A) : Prédiction issues du logiciel Infernal. (B) : Liste des tRNA-ligases de *Bordetella pertussis* Tohama I et des riboswitches potentiels. Les gènes ayant un long 5'UTR et/ou une prédiction d'ARN non codant régulateur sont susceptibles d'avoir un riboswitch sur leur extrémité 5'.

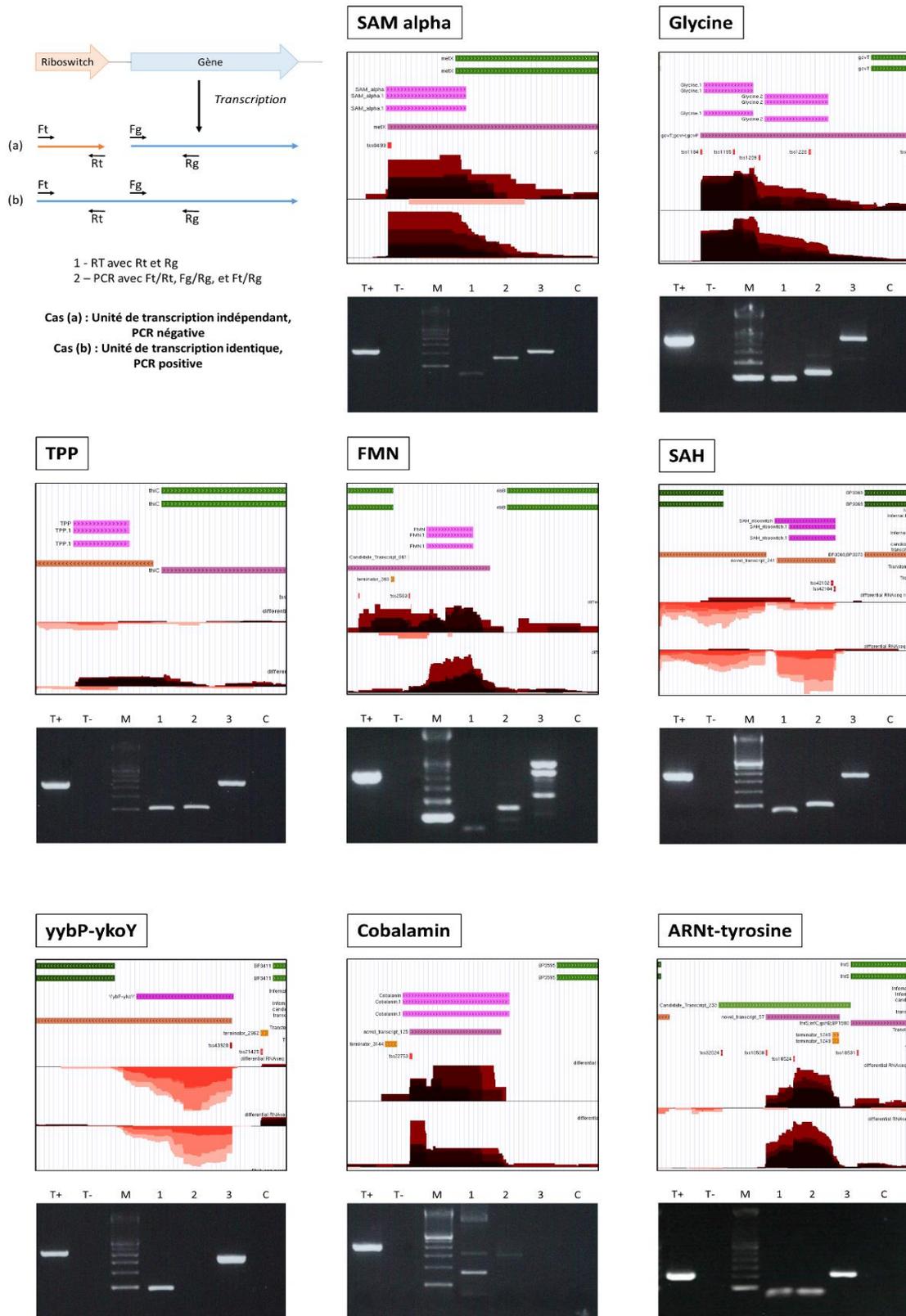
de masquer la séquence de fixation du ribosome pour empêcher la traduction de l'ARN messager, ou de former un terminateur de transcription pour détacher l'ARN polymérase de la matrice ADN (De Lay and Garsin, 2016; Oliva et al., 2015; Stazic and Voß, 2016; Svensson and Sharma, 2016). L'activation ou l'inhibition de la transcription ou de la traduction dépend de la liaison d'un ligand, une molécule impliquée dans la fonction du gène en aval, déstructurant le riboswitch. Cette déstructuration peut masquer ou libérer la séquence de fixation du ribosome, ou former ou déstabiliser un terminateur de transcription.

Les riboswitches et leur ligand prédits sont listés dans le tableau 4A. Les TSS prédits de 3 de ces riboswitches ont été validés par 5'RACE : pour le SAM alpha riboswitch, le SAH riboswitch et le riboswitch liant la glycine (Tableau 2). D'autres riboswitches peuvent lier des ARN de transfert pour moduler l'expression des ARN-t ligases (Condon et al., 1996; Grundy and Henkin, 1993). Ce type de riboswitch n'est pas prédit par le logiciel Infernal, cependant certains de ces gènes codant pour des ARNt-ligase ont été prédits avec de longs 5'UTR ou des prédictions d'ARN non codant potentiels détaillés dans le deuxième chapitre des résultats (Tableau 4B). Sur les 21 gènes d'ARNt-ligases, des TSS de plus de 100 pb en amont ont été prédits pour 5 de ces gènes, et des prédictions d'ARN non codant en 5' ont été prédites pour 8 de ces gènes. Au total, 11 de ces gènes pourraient présenter un riboswitch potentiel sur leur extrémité 5'.

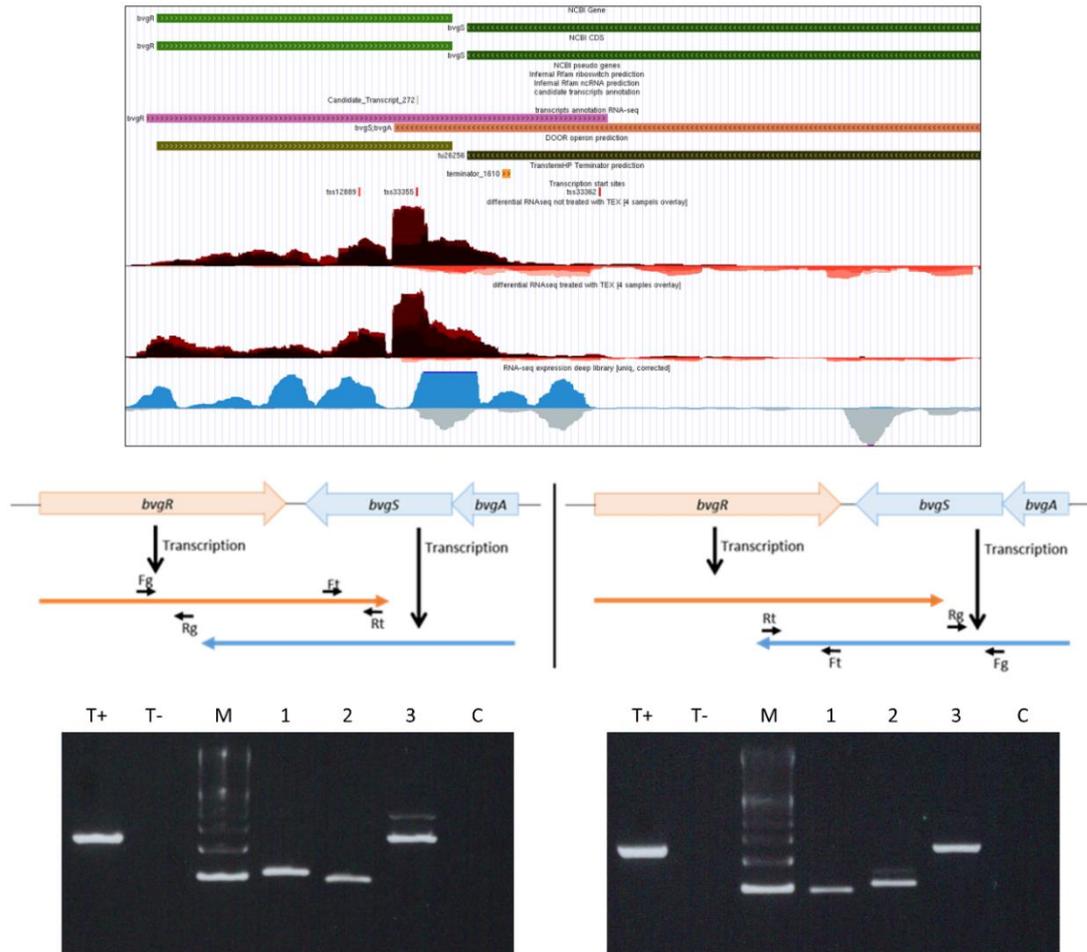
Les extrémités prédites comme riboswitch ont été validées par RT-PCR (Figure 23). Les prédictions et les transcrits messagers ont une jonction commune pour le riboswitch SAM, Glycine, TPP, SAH et ARNt-tyrosine. Les riboswitches FMN, *yybP-ykoY* et Cobalamine ont été détectés par RT-PCR (Ligne 1), cependant l'ARN messager ne semble pas transcrit (ligne 2) et expliquerait l'absence du signal de la jonction commune en ligne 3. Ces riboswitches sont prédits pour inhiber la transcription de l'ARN messager par la formation d'un terminateur de transcription (Dambach et al., 2015; Pedrolli et al., 2015; Polaski et al., 2016). L'absence d'ion manganèse, la présence d'adénosylcobalamine et de flavine mononucléotide dans le milieu de culture inhiberait donc l'expression des gènes BP3410, BP3595 et *ribB*.

#### iv. Les extrémités non traduites en 3' chevauchantes

L'annotation du transcriptome primaire a également permis de détecter de nombreux chevauchements antisens d'ARN messagers en 3'UTR. Au total, 89 transcrits chevauchants ont pu être comptabilisés, dont 66 avec un chevauchement de plus de 100 nucléotides. Un de ces exemples concerne les gènes du système à deux composants *bvgAS*, et du gène codant le catalyseur de GMP *bvgR*. Le transcrit de *bvgR* s'étend sur plus de 400 nucléotides en antisens du transcrit de *bvgAS*, et le transcrit de *bvgAS* s'étend sur plus de 170 nucléotides en antisens du transcrit de *bvgR* (Figure 24). De ce fait, les deux transcrits se chevauchent de 630 nucléotides, comprenant la région intergénique entre les deux annotations. Ces prolongements de transcription ont pu être validés par RT-PCR en utilisant deux couples d'amorces spécifiques de chaque transcrits.



**Figure 23:** Validation des longs 5'UTR prédits comme riboswitch par RT-PCR. Le plan de RT-PCR est détaillé précédemment, avec un test de prédiction en 5' (Figure 21). Les ADNc ont été générés avec les amorces Rt (antisens du riboswitch prédit) et Rg (antisens du gène). Les transcrits ont été détectés avec les amorces Fg/Rg (ARNm, ligne 2), Ft/Rt (Riboswitch, ligne 1) et Ft/Rg (Test de jonction, ligne 3).



**Figure 24:** Validation du chevauchement des transcrits de *bvgAS* et de *bvgR* par RT-PCR. L'annotation du transcriptome primaire a prédit un chevauchement des extrémités 3' des deux transcrits (Haut). La validation du chevauchement par RT-PCR suit le même plan qu'en figure 21. Rapidement, les transcriptions inverses ont été faites avec les amorces Rt (3'UTR) et Rg (Gène annoté) et amplifiés avec les amorces Fg/Rg (Gène annoté, ligne 1), Ft/Rt (3'UTR, ligne 2) et Fg/Rt (Test de prolongement, ligne 3).

D'autres transcrits impliqués dans la virulence montrent des chevauchements de leur extrémité 3'. Par exemple, l'opéron du système de sécrétion des adhésines, codé par *fimB*, *fimC*, *fimD* et *fhaC*, s'étend en antisens d'un autre opéron, comprenant les gènes BP1885, BP1886 et BP1887, codant pour une protéine membranaire et une protéine exportée dont la fonction reste encore inconnue, sur 215 nucléotides. Un autre exemple, le transcrit du gène *fhaL* s'étend sur plus de 500 nucléotides en antisens de l'opéron *aroG*, BP2909 et BP2910 impliqués potentiellement dans la biosynthèse des acides aminés.

### 3. Discussion : Architecture et régulateurs non-indépendants

Le couplage du séquençage à ARN profond et du séquençage à ARN différentiel a permis d'établir l'architecture globale du transcriptome primaire de *Bordetella pertussis* en condition Bvg<sup>+</sup>. Le nombre de TSS et de gènes annotés révèle une architecture dans laquelle la majorité des gènes sont exprimés, incluant certaines séquences d'insertion,

et des structures régulatrices potentielles qui n'étaient pas encore détaillées chez le pathogène de la coqueluche. La mise en évidence de 2722 TSS, dont certains ont été validés par 5'RACE, et de 1315 structures validés par RT-PCR pour nombre d'entre elles ont permis de recouvrir l'architecture de 2262 gènes annotés sur 3871. La co-transcription de l'opéron de la toxine pertussique avec l'ARN de transfert BPt51, possédant l'anticodon GTT, nécessaire à l'incorporation de Valine, pourrait définir un autre niveau de régulation. Chez d'autres bactéries, l'usage des codons régule la quantité de protéines, leur sélection et leur exportation dans le milieu extracellulaire (Korkmaz et al., 2014; Ran et al., 2014). L'ARN de transfert valine possédant l'anticodon GTT serait surexprimé en phase Bvg<sup>+</sup>, la toxine pertussique étant un gène de classe I (Melvin et al., 2014). Ce codon pourrait être utilisé principalement pour les gènes tardifs de virulence de classe I, co-exprimés avec la toxine pertussique.

La majorité des ARN mono- et polycistroniques présentent des structures régulatrices prédites en 5' ou en 3' de ces transcrits. Ces éléments de régulation pourraient également être affectés par les remaniements chromosomiques induits par les IS481, comme cela l'a déjà été montré pour l'expression des gènes (Brinig et al., 2006). La description du transcriptome primaire a révélé un réseau de régulation potentiel très complexe comme cela a déjà été observé chez d'autres procaryotes. Cette complexité du réseau de régulation pourrait être en lien avec la pathogénicité et l'efficacité de ces pathogènes (Sharma et al., 2010; Stazic and Voß, 2016). Certains ARN messagers sont transcrits au départ du codon d'initiation de la traduction, appelés ARN messagers « leaderless », représentant 0,90% de la totalité des gènes annotés. Ces gènes correspondent en majorité à des régulateurs transcriptionnels chez *Bordetella pertussis*, dont nombre d'entre eux codent pour des protéines de type HU, capables de lier l'ADN. La surreprésentation de cette catégorie de gènes a déjà été observée chez *Deinococcus deserti* et *Deinococcus radiodurans*, ils sont essentiels pour la survie de ces espèces en environnement extrême (Bouthier de la Tour et al., 2015). L'absence de site de fixation du ribosome remet en question notre paradigme du mécanisme traductionnel des ARN messagers, ceux-ci sont traduits à partir d'extrémités 5' AUG phosphatées, reconnues directement par le ribosome chez *Escherichia coli* (Brock et al., 2008; Giliberti et al., 2012). Il a été montré que ces transcrits peuvent être régulés, transcrits et traduits en réponse à un stress, ou en fonction de la disponibilité des protéines impliquées dans la machinerie traductionnelle (Moll et al., 2002; Vesper et al., 2011). La majorité de ces ARN codant pour des régulateurs transcriptionnels, il se pourrait que ces ARN messagers « leaderless » soient traduits peut être moins efficacement que les ARN messagers canoniques, mais de façon plus fidèle pour répondre efficacement à un stress environnemental (Kozak, 1991; Malys and McCarthy, 2011; Moll et al., 2002; O'Donnell and Janssen, 2001).

A l'opposé des ARN messagers « leaderless », plusieurs transcrits mono- et polycistroniques présentent des 5'UTR supérieurs à 100 nucléotides. Certaines de ces structures correspondent à des prédictions de riboswitches, validés pour nombre d'entre eux par RT-PCR. Certains de ces riboswitches peuvent lier des co-facteurs enzymatiques (SAM, SAH, FMN, TPP et Cobalamine), permettant à la cellule de s'adapter

rapidement à la production où la quantité de ces facteurs présents dans l'environnement (Oliva et al., 2015; Pedrolli et al., 2015; Polaski et al., 2016). D'autres de ces riboswitches peuvent lier des acides aminés, comme la glycine, ou encore les ARN de transfert, pour réguler les composants de la machinerie traductionnelle (Grundy and Henkin, 1993; Hot et al., 2011). Le riboswitch prédit du gène *thrS* a déjà été mis en évidence chez *Bacillus subtilis*, permettant de limiter la production d'ARNt ligase en absence de tyrosine. Enfin, un autre riboswitch, *yybP-ykoY*, peut lier les ions métalliques, comme le manganèse (Dambach et al., 2015). Les ions sont considérés comme un des éléments essentiels pour l'infection chez les pathogènes, et notamment chez *Bordetella pertussis* (Beall, 1998; de Gouw et al., 2011). L'adaptation rapide à la quantité de métaux et de cofacteurs peut être un élément de régulation supplémentaire, encore non documenté, nécessaire à différentes étapes de l'infection du pathogène. La caractérisation du transcriptome primaire n'a, cependant, pas permis de prédire des thermosenseurs, structures activant ou inhibant la transcription ou la traduction de gènes en aval en fonction des changements de température, un attrait important dans la virulence du pathogène (Dupré et al., 2015, 2013). La détection de ces thermosenseurs nécessiterait des études transcriptomiques approfondies sur des cellules cultivées à différentes températures. Cependant, parmi la détection des 120 longs 5'UTR supérieurs à 100 nucléotides à 37°C, plus 10 d'entre eux ne sont pas prédits comme des structures antisens ou des riboswitches. Il est possible que ceux-ci correspondent à d'autres structures régulatrices, notamment à des thermosenseurs potentiellement impliqués dans les étapes de l'infection de l'hôte.

Une large proportion des ARN mono- et polycistroniques se prolongent en antisens des gènes environnant, par leur extrémité 5' ou leur extrémité 3'. Les 5'UTR chevauchant en antisens de deux gènes résultent en une compétition d'expression. Le degré d'activation de ces gènes va permettre, en fonction des changements environnementaux, d'exprimer uniquement l'un ou l'autre de ces gènes. Cette définition s'applique également à l'excludon BP1203, recouvrant entièrement la séquence de BP1202, un régulateur transcriptionnel de la famille TetR. Ces deux gènes auraient, selon cette définition, des fonctions opposées, phénomène déjà observé pour la régulation des flagelles chez *Listeria monocytogenes* (Sesto et al., 2013; Wurtzel et al., 2012). Le 5'UTR du gène codant pour MogR, le répresseur de la motilité, est transcrit en antisens des deux premiers gènes de l'opéron codant pour le flagelle. Différentes familles de gènes présentent des 5'UTR orientés en antisens des gènes environnant chez *Bordetella pertussis*, reflétant un réseau de régulation complexifié par des compétitions de niveau d'expression de gènes de fonctions opposées.

La majorité des transcrits antisens détectés provient des 3'UTR issus des ARN mono- et polycistroniques, se prolongeant jusqu'à plus de 400 nucléotides en aval de certains d'entre eux. Ce système de régulation serait également un système de répression entre deux gènes de fonctions potentiellement opposées (Brantl, 2007; Lasa et al., 2012; Thomason and Storz, 2010). Plusieurs prolongements de transcrits en 3' ont été mis en évidence en déterminant le transcriptome primaire, certains sont impliqués dans la virulence de *Bordetella pertussis*. L'un des exemples les plus frappants concerne le

principal régulateur de la virulence, le système à deux composants BvgAS. Le prolongement de transcription des gènes *bvgAS* et *bvgR* pourrait ajouter un nouveau niveau de régulation de l'expression du phosphorelai et de la virulence du pathogène. L'expression de BvgR étant modulée par BvgA phosphorylé pour catalyser le cyclique di-GMP en GMP, la phosphorylation de l'activateur transcriptionnel activerait la transcription de *bvgR*, se prolongeant en antisens, entraînant une boucle de rétrocontrôle potentielle de l'expression de *bvgAS* dépendant de l'activation de facteurs sigma. Les transcrits antisens pourraient de ce fait jouer un rôle fondamental lors des différentes étapes de l'infection.

L'analyse du transcriptome primaire de *Bordetella pertussis* en condition Bvg<sup>+</sup> a permis d'identifier de nouvelles structures régulatrices non-indépendantes issues de prolongements des ARN messagers sur leurs extrémités 5' et/ou 3'. Le remaniement du génome du pathogène induit par les IS481 pourrait altérer l'expression de ces structures régulatrices et potentiellement affecter la régulation des gènes de virulence et altérer l'expression des antigènes utilisés dans le vaccin acellulaire. Les diversités génomiques inter-souches se reflètent également dans la diversité du transcriptome non codant régulateur (Kopf et al., 2015). Elle concerne principalement une catégorie majeure de transcrits, les ARN régulateurs indépendants, majoritairement non codants (Svensson and Sharma, 2016). Le transcriptome primaire de *Bordetella pertussis* a donc été analysé également dans le but d'identifier des ARN indépendant et potentiellement régulateurs.

## II. Mise en évidence des ARN régulateurs indépendant

Les ARN messagers transcrits en conditions Bvg<sup>+</sup> ainsi que leurs structures particulières, comprenant pour certains des éléments régulateurs potentiels, ont été annotés. Cependant, plusieurs transcrits révélés par notre étude ne correspondent à aucune annotation actuelle des séquences codantes, et ont pour la plupart été notés comme des candidats ARN régulateurs indépendant potentielles.

### 1. Edification d'une liste de prédictions d'ARN régulateurs non codants

Une liste de prédiction d'ARN régulateurs potentiellement non codant a été générée à partir des prédictions issues de l'annotation du transcriptome du pathogène en condition Bvg<sup>+</sup>, couplée à des prédictions de différents logiciels bioinformatiques. Ces deux approches prédictives nous ont permis d'édifier une liste exhaustive de prédictions de candidats ARN régulateurs, disponible en tableau supplémentaire S1 (<https://tinyurl.com/y7el869z>). Les prédictions ont été annotées dans l'ordre de coordonnées sur le génome et nommées « BPncXXX » pour « *Bordetella pertussis* non coding RNA prediction number XXX ».

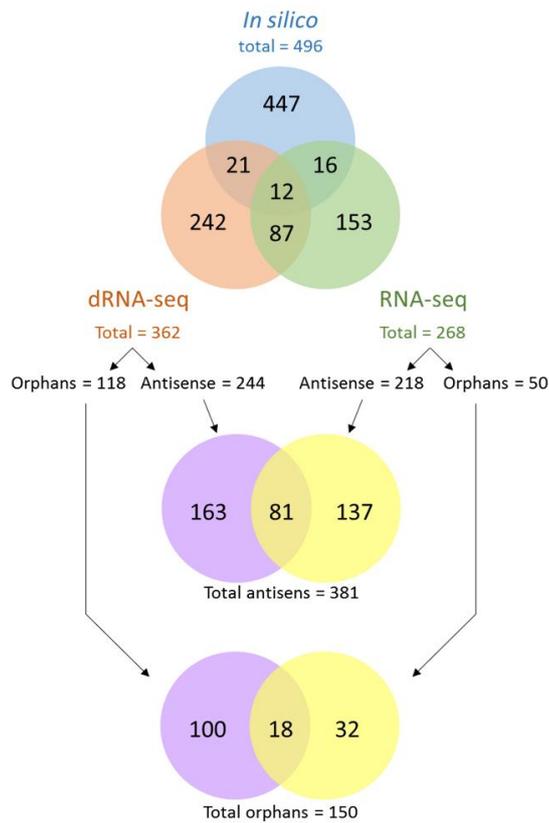
#### i. Prédictions issues des séquençages à ARN

Sur les 2722 TSS prédits issus de l'étude du transcriptome primaire de *Bordetella pertussis*, 244 ont été définis comme étant orientés en antisens des gènes annotés, correspondant potentiellement à de nouveaux ARN régulateurs non codant. De plus, 118 TSS ont été identifiés dans les régions intergéniques. Ces derniers se retrouvent à plus de 250 nucléotides en amont ou en aval des transcrits annotés orientés dans le même sens, et ont donc été qualifiés de TSS orphelins. Au total, 362 ARN régulateurs indépendant ont été prédits à partir du séquençage à ARN différentiel. Les TSS de trois ARN régulateurs prédits ont été validés par 5'RACE pour les prédictions BPnc180, BPnc468 et BPnc962 (Résultats chapitre 1, Tableau 2) (Matériel et Méthode 6).

Le séquençage profond du transcriptome de *Bordetella pertussis* a permis de détecter 268 structures ARN, alignées soit en antisens des séquences codantes annotées, soit en région intergénique sans gène annoté. Ces prédictions de transcrits absent de l'annotation actuelle du génome de la souche Tohama I ont été catégorisées comme candidats ARN régulateur non codant.

#### ii. Prédictions *in silico*

Pour compléter la liste de prédictions de candidats ARN régulateurs non codant, des prédictions issues de l'analyse du génome de la souche Tohama I par des logiciels bioinformatiques ont été ajoutées. Trois algorithmes différents, basés sur différents critères de prédiction, ont été utilisés dans le but d'étoffer la liste (Matériel et Méthode 17). L'algorithme SIPHT (sRNA Identification Protocole using High-Throughput Technologies) se base sur la conservation de séquences localisées en amont des terminateurs de transcription Rho-dépendant prédits ici par TransTerm HP (Livny, 2012). Le second algorithme, NAPP (Nucleic Acid Phylogenetic Profiling), se base sur la



**Figure 25:** Diagramme de Venn comparant le nombre de prédictions d'ARN régulateurs potentiellement non codant retrouvés par chaque approche.

*In silico* : Cumul des prédictions des algorithmes SIPHT, NAPP et RNAz. dRNA-seq : Prédictions issues du séquençage à ARN différentiel. RNA-seq : Prédictions issues du séquençage profond des ARN totaux. Orphans : Prédictions localisées en régions intergéniques. Antisense : Prédictions orientées en antisens des gènes annotés.

comparaison de séquences en mosaïques localisées dans les régions intergéniques (Ott et al., 2012). Enfin, le dernier algorithme, RNAz, se base sur la prédiction de structures ARN stables et sur la conservation de ces structures au sein des espèces phylogénétiquement proches (Gruber et al., 2010). Cette dernière, déjà utilisée auparavant, a été étendue à une liste de 18 espèces proches appartenant au genre *Burkholderiale* dont le génome est annoté. Au total, les approches *in silico* ont permis de détecter 496 prédictions de candidats ARN non codant régulateurs. Ces approches sont très complémentaires, uniquement 16 prédictions sont retrouvées par les trois algorithmes, et uniquement 80 retrouvées par deux des trois méthodes.

### iii. Comparaison des approches prédictives

La combinaison des trois approches a permis de prédire 978 candidats ARN régulateurs non codant (Tableau supplémentaire S1). Le cumul de ces approches permet d'élargir la liste des prédictions, avec une grande complémentarité, uniquement 12 prédictions sont retrouvées par les trois approches (Figure 25). Cette cartographie des candidats ARN régulateurs non codant comprend un total de 381 antisens de gènes annotés, et 597 nouveaux transcrits prédits et localisés dans des régions intergéniques.

Nom du candidat	Début de prédiction	Fin de prédiction	Prédiction <i>in silico</i>	Prédiction par dRNA-seq	Prédiction par RNA-seq	Test de Détection par Northern Blot	Orientation	Cette étude	Hot et al., 2011
BPnc006	10396	10512	✓	-	-	Non détecté	-	✓	✓
BPnc008	11752	11752	-	✓	✓	Validé par NB	<	✓	
BPnc035	168998	169072	✓	-	-	Non détecté	-		✓
BPnc082	364288	364389	✓	-	-	Validé par NB	<	✓	
BPnc109	488484	488614	✓	-	-	Validé par NB	<		✓
BPnc180	776391	776820	-	✓	✓	Validé par NB	<	✓	
BPnc181	784618	784619	-	✓	-	Non détecté	-	✓	
BPnc199	868916	869224	✓	✓	-	Non détecté	-		✓
BPnc218	978089	978089	-	✓	-	Validé par NB	>	✓	
BPnc233	1014574	1014783	✓	-	-	Validé par NB	<	✓	
BPnc264	1175905	1175906	-	✓	-	Validé par NB	>	✓	
BPnc344	1483831	1483831	-	✓	✓	Validé par NB	>	✓	
BPnc355	1494111	1494207	✓	-	-	Validé par NB	>		✓
BPnc363	1510669	1510822	✓	✓	✓	Validé par NB	>	✓	
BPnc370	1564905	1565002	✓	-	-	Non détecté	-	✓	
BPnc375	1581003	1581151	✓	✓	✓	Validé par NB	>	✓	
BPnc384	1626733	1627267	✓	✓	✓	Validé par NB	<	✓	
BPnc392	1647741	1647741	-	✓	-	Validé par NB	<	✓	
BPnc425	1819837	1819911	✓	-	-	Non détecté	-	✓	
BPnc451	1888983	1889132	✓	✓	-	Validé par NB	>	✓	
BPnc452	1894765	1894865	✓	-	-	Non détecté	-	✓	
BPnc465	1941726	1941817	✓	-	-	Validé par NB	<	✓	
BPnc468	1968769	1968976	-	✓	✓	Validé par NB	<	✓	
BPnc469	1980441	1980592	-	-	✓	Validé par NB	-	✓	
BPnc494	2127981	2128089	✓	-	-	Validé par NB	<	✓	
BPnc519	2208284	2208636	✓	-	-	Non détecté	-	✓	
BPnc560	2375218	2375428	-	✓	✓	Non détecté	-	✓	
BPnc588	2503141	2503239	✓	-	-	Non détecté	-		✓
BPnc627	2612002	2612166	✓	-	-	Non détecté	-	✓	
BPnc629	2623936	2624241	✓	✓	✓	Validé par NB	<		✓
BPnc640	2699393	2699489	✓	-	-	Validé par NB	>		✓
BPnc642	2703571	2703802	✓	-	✓	Non détecté	-	✓	
BPnc651	2723562	2723632	✓	-	-	Non détecté	-	✓	
BPnc656	2739741	2740028	-	-	✓	Validé par NB	<	✓	
BPnc670	2835490	2835490	-	✓	-	Non détecté	-	✓	
BPnc741	3099552	3099726	✓	✓	-	Validé par NB	<		✓
BPnc748	3138544	3138544	-	✓	-	Validé par NB	<	✓	
BPnc755	3173584	3173723	✓	-	-	Validé par NB	<		✓
BPnc758	3178089	3178339	✓	-	-	Validé par NB	>		✓
BPnc783	3263690	3263815	✓	-	-	Validé par NB	<		✓
BPnc807	3362907	3363416	✓	✓	✓	Validé par NB	<	✓	
BPnc811	3378749	3378863	✓	-	-	Non détecté	-	✓	
BPnc861	3597094	3597159	✓	-	-	Non détecté	-	✓	
BPnc863	3597269	3597434	-	-	✓	Non détecté	-	✓	
BPnc870	3605201	3605397	✓	-	-	Validé par NB	>		✓
BPnc875	3617998	3618072	✓	✓	-	Validé par NB	<	✓	
BPnc878	3651305	3651533	-	-	✓	Validé par NB	<	✓	
BPnc923	3825740	3825895	✓	-	-	Non détecté	-	✓	
BPnc947	3896305	3896451	✓	-	-	Validé par NB	> et <		✓
BPnc962	3956825	3957191	✓	✓	✓	Validé par NB	<		✓
BPnc968	3992002	3992388	-	✓	✓	Non détecté	-	✓	
BPnc969	3999760	3999761	-	✓	-	Non détecté	-	✓	
BPnc972	4041724	4041777	✓	-	-	Validé par NB	<	✓	
BPnc977	4077015	4077111	✓	-	-	Non détecté	-		✓

**Tableau 5 :** Validation des prédictions d'ARN régulateurs non codant par Northern Blot. Chaque prédictions est indiquée avec ses coordonnées sur le génome, la ou les approche(s) ayant permis leur prédiction, et l'orientation détectée par Northern Blot. L'étude ayant testé chaque prédiction est également indiquée (Cette étude : étude actuelle ; Hot *et al.*, 2011 : études précédentes). *In silico* : Prédictions bioinformatiques. dRNA-seq : Prédictions issues du séquençage à ARN différentiel. RNA-seq : Prédictions issues du séquençage profond des ARN totaux. NB : Northern Blot.

Pour évaluer la fiabilité de cette liste de prédiction, certaines de ces prédictions ont été testées en Northern Blot (Matériel et Méthode 7). Les candidats ont été testés en condition classique de culture c'est-à-dire en phase Bvg<sup>+</sup> (Tableau 5) (Figures Supplémentaires 1 et 2). La précision et la sensibilité de chaque approche a été calculée en cumulant les candidats testés dans cette étude (38) et de ceux testés précédemment dans la publication Hot *et al.*, 2011 (16). Les prédictions *in silico* ont la précision la plus faible avec 59%, alors que le séquençage différentiel atteint une précision de 74% et le séquençage profond une précision de 76%. En considérant deux des trois approches une précision de 86% est atteinte et si les prédictions par les trois approches sont cumulées, une précision de 100% est atteinte. Cependant, la sensibilité de prédiction de chaque méthode séparée reste faible. Le séquençage profond a une sensibilité de 40%, le séquençage différentiel une sensibilité de 51% et les approches bioinformatiques de 67%. La faible sensibilité de chaque méthode prise individuellement illustre l'intérêt de combiner plusieurs approches prédictives.

## **2. Sélection et validation des prédictions d'ARN régulateurs indépendant**

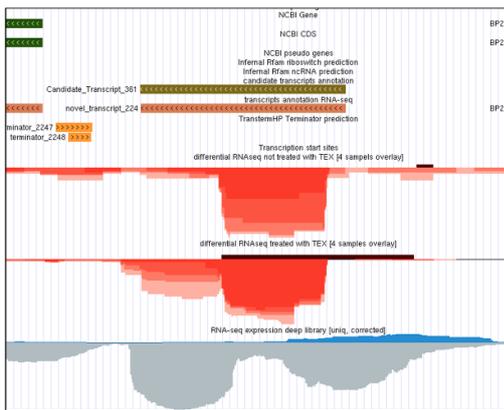
Les trois approches prédictives ont permis de prédire l'existence d'ARN régulateurs soit orientés en antisens de gènes, soit localisés dans des régions intergéniques (Figure 26). Cependant, l'analyse de certaines de ces prédictions a permis de distinguer des structures particulières, probablement non-indépendantes de transcrits de gènes ou de séquences annotées, indiquées dans la ligne « commentaire » du tableau des prédictions (Tableau supplémentaire S1).

### **i. Prédictions d'ARN non indépendant**

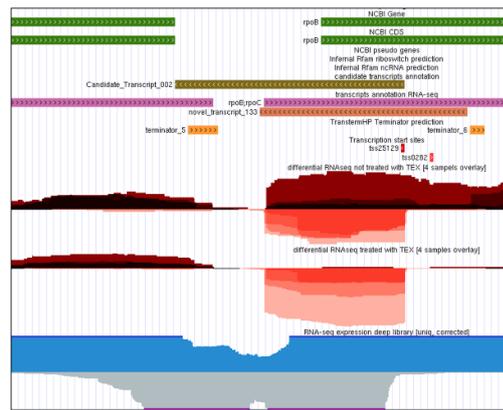
Parmi la liste de prédictions, 29 d'entre elles correspondent à des espaceurs transcriptionnels internes ou externes (ITS et ETS), des structures ARN non traduites localisées aux extrémités 5' ou 3', ou encore entre deux séquences codantes ou deux séquences d'ARNt (Lalaouna et al., 2015). Les 11 ITS ont été prédits par les analyses *in silico* et comprennent 2 ITS d'ARNt et 9 ITS d'ARNr. Les 18 ETS ont été majoritairement prédits par les analyses *in silico* également, seules trois d'entre elles ont été prédites par les séquençages. Les ETS prédits se situent aux extrémités 3' des ARNt pour 15 prédictions.

Certaines prédictions sont localisées à moins de 150nt et orientées dans le même sens que des séquences codantes annotées. Au total, 144 prédictions de candidats ARN régulateurs non codant pourraient correspondre à des 3'UTR, et 175 prédictions pourraient correspondre à des 5'UTR. L'indépendance de transcription d'ARN validés en Northern Blot et localisés en 5'UTR ou en 3'UTR de gène a été testée par RT-PCR (Figure 27) (Figure Supplémentaire S1) (Matériel et Méthode 9). BPnc363 et BPnc375 sont transcrits indépendamment de *ppsA* et de *ptsI* respectivement. Cependant, BPnc513 est cotranscrit avec l'ARN messager de BP2055, et serait un long 5'UTR. BPnc451 est également cotranscrit avec l'ARN messager de *ptpA*, et serait le 3'UTR de ce gène.

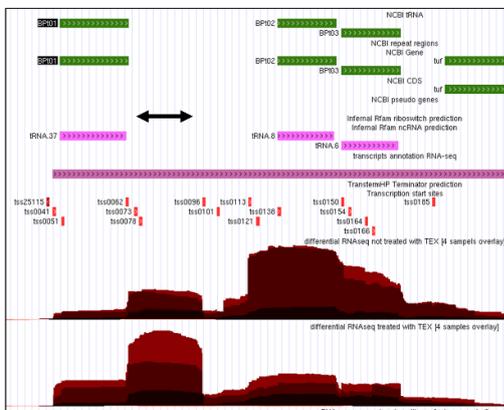
### ARN régulateur prédit en *trans*



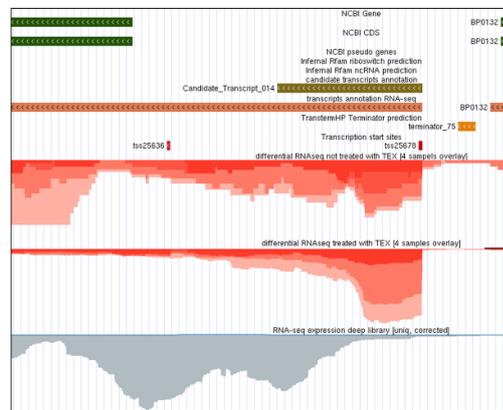
### ARN régulateur prédit en *cis*



### ETS/ITS



### Longs 5' et 3'UTR



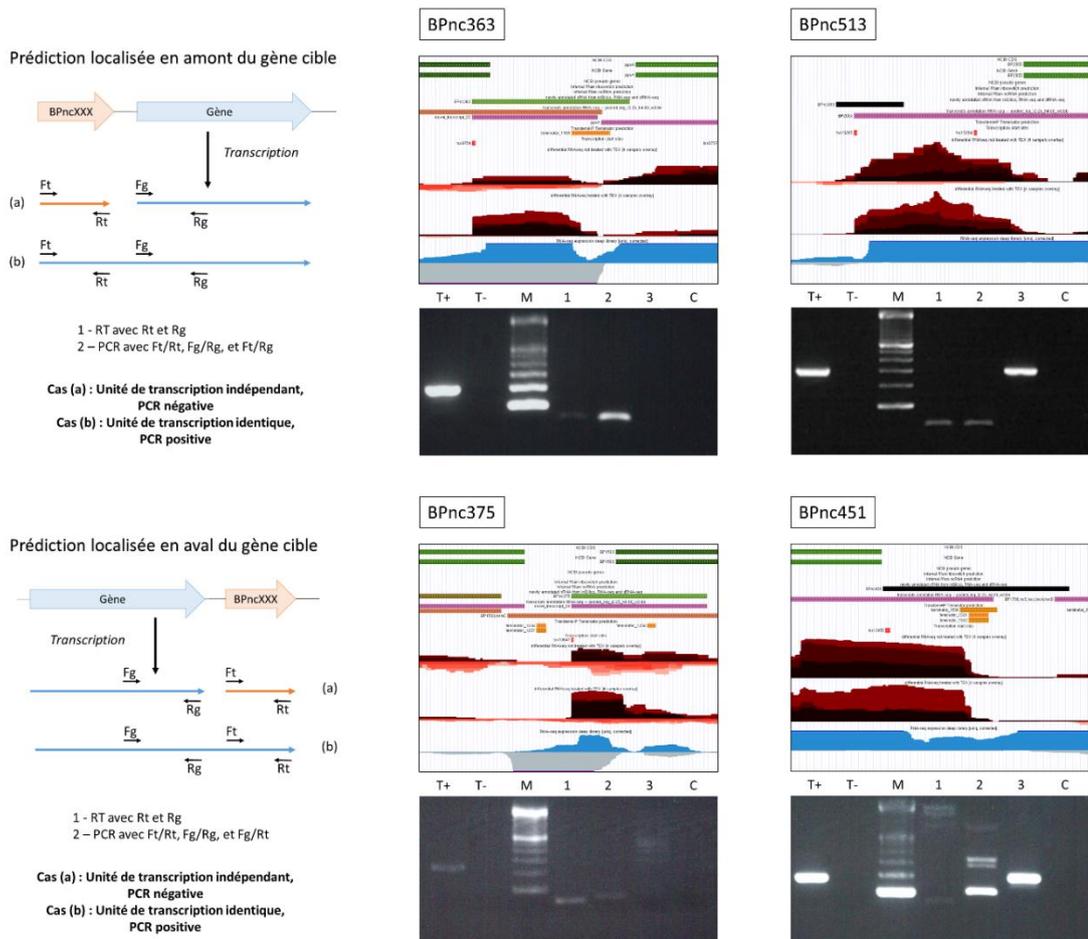
**Figure 26:** Configurations des prédictions d'ARN régulateurs non codant.

Les prédictions d'ARN régulateurs sont localisées dans des régions intergéniques, donnant des potentiels régulateurs en *trans*, ou en antisens de gènes, donnant des potentiels régulateurs en *cis*. Certaines prédictions correspondent à des espaceurs transcriptionnel internes ou externes (ITS, ETS, flèche noire), ou encore à des longs 5' ou 3'UTR. Chaque exemple est illustré ci-dessus.

Cependant, les prédictions localisées en 3'UTR de gènes pourraient correspondre à des ARN régulateurs indépendants, co-transcrits ou non avec l'ARN messager en amont (Chao et al., 2012; Miyakoshi et al., 2015a; Ren et al., 2017). Leur caractéristique d'ARN régulateur potentiel a été testée dans la partie suivante de ce chapitre.

## ii. Validation des ARN régulateurs potentiels

Les prédictions testées par Northern Blot ont permis de valider 22 nouveaux transcrits (Figure Supplémentaire S1 et S2). Par ailleurs, le logiciel Infernal a prédit la position de trois ARN identifiés chez d'autres procaryotes, le motif *sucA*, un élément pouvant correspondre à un riboswitch chez les *Burkholdérialés* (BPnc255), l'ARN 4,5S SRP (« Single Recognition Particule ») (BPnc384), et l'ARN toxique des bêtaprotéobactéries (BPnc807) (Kimelman et al., 2012; Miralles, 2010; Peterson and

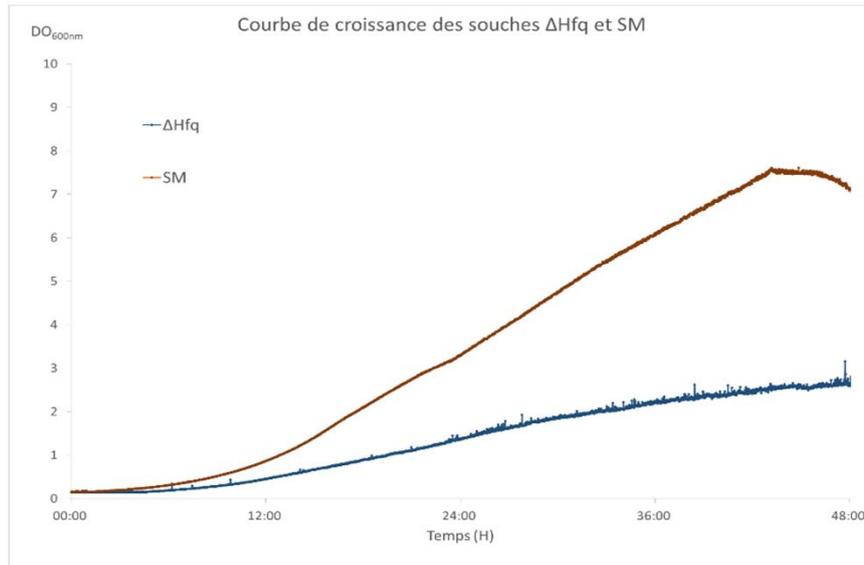
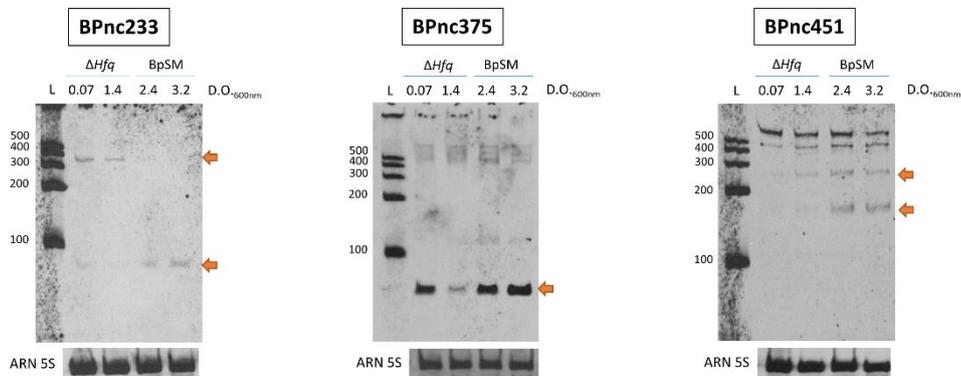


**Figure 27:** Test d'indépendance de transcription des ARN candidats.

Le plan de RT-PCR est détaillé en figure 21 (Résultats, Chapitre 1, Partie 2). Les ADNc ont été générés avec les amorces Rt (antisens du candidat prédit) et Rg (antisens du gène). Les transcrits ont été détectés avec les amorces Fg/Rg (ARNm, ligne 2), Ft/Rt (BPncXXX, ligne 1) et Ft/Rg ou Fg/Rt (Test de jonction, ligne 3)

Phillips, 2008; Weinberg et al., 2007). Parmi les 22 transcrits, l'ARN 4,5S SRP et l'ARN toxique des bêta-protéobactéries ont été validés.

Sur les 22 nouveaux transcrits validés par Northern Blot, 10 d'entre eux sont orientés en antisens de gènes annotés, 7 d'entre eux sont localisés en régions intergéniques, et 4 d'entre eux sont localisés en 3'UTR de gènes annotés. Les transcrits en antisens de gènes sont localisés dans la région 5' des gènes annotés dans 6 cas, et en 3' des gènes annotés dans 5 autres cas. Ainsi ont pu être mis en évidence un antisens du gène *rpoB* codant pour la sous-unité Bêta de l'ARN polymérase (BPnc008), un antisens du gène *flaG* la protéine flagellaire (BPnc344), un antisens du gène BP1435 impliqué dans la synthèse du phosphoénol-pyruvate (BPnc363) et un antisens de BP1503, codant pour une protéine hypothétique (BPnc375) (Figure S1). La majorité des antisens validés par Northern Blot sont localisés en antisens de gènes de virulence (Figure S2). Deux ARN sont orientés en antisens des gènes de toxines *cyxA* (BPnc180) et *dnt* (BPnc878), et quatre ARN sont orientés en antisens des gènes codant pour des adhésines, *fhaB*

**A****B**

**Figure 28:** Stabilité des ARN validés par Northern Blot chez le mutant  $\Delta$ Hfq.

(A) : Courbe de croissance des souches BpSM et  $\Delta$ Hfq en condition classique de culture. SM : Souche de référence BpSM.  $DO_{600nm}$  : Densité Optique à 600nm. (B) : Northern Blot de BPnc233, BPnc375 et BPnc451 chez les souches BpSM et  $\Delta$ Hfq. Les flèches oranges indiquent les transcrits modulés entre les deux souches. L : Marqueur de taille.

(BPnc468), l'opéron *fimBCD-fhaC* (BPnc469), et les gènes codant pour le corps de la fimbriae *fim3* (BPnc392) et *fim2* (BPnc264). Les prédictions d'ARN orientés en antisens des gènes de virulence *cyd* (BPnc181), *bcrH2* (BPnc560), *ptIC* (BPnc968) et *ptIH* (BPnc969) n'ont pas été validés par Northern Blot.

La conservation des prédictions d'ARN régulateurs en région intergénique et en 3'UTR a été testée à la fois parmi les Burkholdériales et différentes souches de *Bordetella pertussis* à l'aide du logiciel RNAlien (Eggenhofer et al., 2016). Les conservations phylogénétiques sont indiquées dans la liste de prédiction d'ARN régulateurs du Tableau S1. La majorité des prédictions d'ARN régulateur est très peu conservé au sein des Burkholdériales, seulement cinq séquences sont retrouvées dans plus de 150 génomes sur les 185 comparés. Ces séquences sont très conservées chez les différentes souches de *Bordetella pertussis*, cependant 10 prédictions d'ARN régulateurs

sont spécifiques du pathogène, et ne sont pas retrouvées chez les autres espèces du genre *Bordetellae*.

Les ARN non codant transcrits en région intergénique ou en 3'UTR de gène sont très régulièrement stabilisés par un hexamère protéique appelé Hfq, une protéine chaperonne capable de lier un ARN non codant régulateur et un ARN messager cible à l'aide de ses faces proximales et distales (Vogel and Luisi, 2011). L'expression de certains transcrits validés par Northern Blot a été testée entre la souche de laboratoire BpSM et un mutant de délétion  $\Delta hfq$ . Hfq étant nécessaire à l'expression de facteurs de virulence chez le pathogène et potentiellement à d'autres protéines impliquées dans la réponse aux stress, la croissance du mutant a été mesurée pour comparer l'expression des ARN non codant entre les deux souches (Figure 28A) (Matériel et Méthode 1) (Bibova et al., 2013, 2015). Le mutant ayant une croissance altérée, la stabilité des transcrits a été testée à une densité optique à 600nm de 0,07 et de 1,4 pour le mutant  $\Delta hfq$ , et de 2,4 et 3,2 pour la souche BpSM. La stabilité de 3 transcrits sur 6 testés, localisés en 3'UTR de gène ou en région intergénique, est altérée chez le mutant comparé à la souche de référence (Figure 28B). La stabilité de BPnc451, un transcrit localisé en 3'UTR de gène, est réduite chez le mutant  $\Delta hfq$ , renforçant sa nature d'ARN régulateur potentiel. Les plus petits transcrits de BPnc233 et BPnc375, localisés en régions intergéniques, sont déstabilisés également chez le mutant, alors que le transcrit supérieur de BPnc233 est plus stable. Le polymère Hfq pourrait donc potentiellement lier ces trois transcrits, probablement pour les appairer avec des ARN messagers cibles. La déstabilisation du transcrit 350nt de BPnc233 uniquement pour la souche BpSM indiquerait que seulement le transcrit de petite taille serait lié au polymère, et qu'il pourrait soit résulter d'une maturation du transcrit supérieur, soit être transcrit par un second promoteur.

### iii. Modulation des prédictions en phase Bvg<sup>+</sup> et Bvg<sup>-</sup>

Pour identifier des ARN régulateurs potentiellement impliqués dans la régulation de la virulence de *Bordetella pertussis*, le pathogène a été cultivé en présence de 50mM de sulfate de magnésium. L'ion sulfate a pour conséquence d'inactiver la phosphorylation du senseur BvgS, et de rendre la bactérie non virulente en phase Bvg<sup>-</sup> (Dupré et al., 2015, 2013; Scarlato and Rappuoli, 1991). Le passage de la souche BpSM à l'état Bvg<sup>-</sup> a été vérifié par puce à ADN en comparaison avec une culture classique en phase Bvg<sup>+</sup> (Figure S3) (Matériel et Méthode 11). Les toxines et les adhésines du pathogène sont sous-exprimées chez la bactérie cultivée en présence de sulfate de magnésium, tout comme les gènes du système à deux composants *bvgAS*. A l'inverse, les gènes de la capsule polysaccharidiques sont exprimés, affirmant que les cellules sont en phase Bvg<sup>-</sup>. Le transcriptome de la souche non virulente a été séquencé en séquençage différentiel et les départs de transcription prédits ont été comparés aux départs prédits de la souche virulente à l'aide du logiciel TSSAR (Matériel et Méthode 5 et 17).

Le séquençage différentiel de la bactérie en phase Bvg<sup>-</sup> a permis de prédire 1808 TSS, parmi lesquels 183 sont orientés en antisens de gènes, et 304 sont localisés dans des régions intergéniques.

Nom du candidat	Ancienne dénomination	TSS prédit en phase Bvg+	TSS prédit en phase Bvg-	Orientation	Prédiction du phénotype	Test par Northern Blot	Commentaire
BPnc006	-	-	10306	>	vrg	Non détecté	
BPnc008	-	11752	11762	<	-	Validé par NB	
BPnc035	-	-	-	-	-	Non détecté	
BPnc082	-	-	-	<	-	Validé par NB	
BPnc109	BprA	-	-	<	-	Validé par NB	
BPnc180	-	776799	-	<	vag	Validé par NB	
BPnc181	-	784619	-	<	vag	Non détecté	
BPnc199	-	868941	-	>	vag	Non détecté	
BPnc218	-	978089	978204	>	-	Validé par NB	Transcrit par un P <sub>out</sub> d'une IS481
BPnc233	-	-	-	<	-	Validé par NB	
BPnc264	-	1175906	1175774	>	-	Validé par NB	Transcrit par un P <sub>out</sub> d'une IS481
BPnc344	-	1483831	-	>	vag	Validé par NB	
BPnc355	BprB	-	-	>	-	Validé par NB	
BPnc363	-	1510649	-	>	vag	Validé par NB	
BPnc370	-	-	-	-	-	Non détecté	
BPnc375	-	1581072	1581062	>	-	Validé par NB	
BPnc384	-	1627167	1627174	>	-	Validé par NB	
BPnc392	-	1647742	-	<	vag	Validé par NB	
BPnc425	-	-	1819957	>	vrg	Non détecté	3'UTR
BPnc451	-	-	-	<	-	Validé par NB	
BPnc452	-	-	-	-	-	Non détecté	
BPnc465	-	-	-	<	-	Validé par NB	
-	BprC	-	-	<	-	Validé par NB	
BPnc468	-	1968972	-	<	vag	Validé par NB	
BPnc469	-	-	-	<	-	Validé par NB	
BPnc494	-	-	-	<	-	Validé par NB	
BPnc519	-	-	-	-	-	Non détecté	
BPnc560	-	2375219	-	>	vag	Non détecté	
BPnc588	-	-	-	-	-	Non détecté	
BPnc627	-	-	2612176	<	vrg	Non détecté	3'UTR
BPnc629	BprD	2624141	2624146	-	-	Validé par NB	
BPnc640	BprE	-	-	>	-	Validé par NB	
BPnc642	-	-	-	-	-	Non détecté	
BPnc651	-	-	2723662	<	vrg	Non détecté	3'UTR
BPnc656	-	-	-	<	-	Validé par NB	
BPnc670	-	2835490	-	<	vag	Non détecté	
BPncX01	-	-	2956996	<	vrg	Non détecté	3'UTR
BPnc741	BprF	3099723	3099732	<	-	Validé par NB	
BPnc748	-	3138544	-	<	vag	Validé par NB	
BPnc755	BprG	-	-	<	-	Validé par NB	
BPnc758	BprH	-	-	>	-	Validé par NB	
BPnc783	BprI	-	-	<	-	Validé par NB	
BPnc807	-	3363134	3363139	<	-	Validé par NB	
BPnc811	-	-	3378738	>	vrg	Non détecté	
BPnc861	-	-	-	-	-	Non détecté	
BPnc863	-	-	-	>	-	Non détecté	
BPnc870	BprJ	-	-	>	-	Validé par NB	
BPnc875	-	3618066	3618077	<	-	Validé par NB	
BPnc878	-	-	-	<	-	Validé par NB	
BPnc923	-	-	-	-	-	Non détecté	
BPnc947	BprM	-	-	> et <	-	Validé par NB	
BPnc962	BprN	3956960	3956973	<	-	Validé par NB	
BPnc968	-	3992316	-	<	vag	Non détecté	
BPnc969	-	3999761	-	<	vag	Non détecté	
BPnc972	-	-	-	<	-	Validé par NB	
BPnc977	-	-	-	-	-	Non détecté	

**Tableau 6:** Prédictions des départs de transcription des ARN validés par Northern Blot en phase Bvg<sup>+</sup> et en phase Bvg<sup>-</sup>.

Chaque transcrit est indiqué avec son indicatif de prédiction (BPncXXX), et l'indicatif de l'étude précédente de Hot *et al.*, 2011 (BprX). Les candidats validés ou non en Northern Blot sont également indiqués. La position des TSS est donnée pour chaque prédiction en phase Bvg<sup>+</sup> et en phase Bvg<sup>-</sup>, ainsi que l'orientation prédite des transcrits candidats. *Vag* : Prédiction d'un transcrit à profil *vag*, activé en phase de virulence. *Vrg* : Prédiction d'un transcrit à profil *vrg*, réprimé en phase de virulence. La colonne commentaire regroupe les prédictions particulières des candidats. 3'UTR : Candidat potentiellement transcrit en 3'UTR de gène. Transcript from P<sub>out</sub> of IS481 : Prédiction d'un ARN régulateur issu de prolongement de transcription des promoteurs antisens des IS481 (Voir Résultats Chapitre 3 Partie 1).

Parmi ces TSS, de nouvelles prédictions d'ARN régulateurs sont apparues. Les TSS de tous les transcrits testés en Northern Blot ont été analysés dans les deux phases de virulence afin d'identifier les prédictions d'ARN régulateurs potentiellement modulées en fonction de l'état de virulence du pathogène (Tableau 6). Douze prédictions d'ARN régulateurs testés en Northern Blot ont été prédites avec un départ de transcription uniquement en phase  $Bvg^+$ . Ils ont donc un profil de modulation similaire aux *vag* (virulence activated genes), des ARN exprimés uniquement en phase de virulence et potentiellement impliqués dans sa régulation. Sur ces 12 prédictions, 6 d'entre elles n'ont pas été validées par Northern Blot en phase  $Bvg^+$ . Les prédictions ayant un profil *vag* comprennent 8 prédictions de transcrits orientés en antisens de gènes de virulence. Seuls les transcrits antisens du gène *fim2* (BPnc264) et du gène *dnt* (BPnc878) ne sont pas prédits comme étant modulés en fonction de l'état de virulence du pathogène.

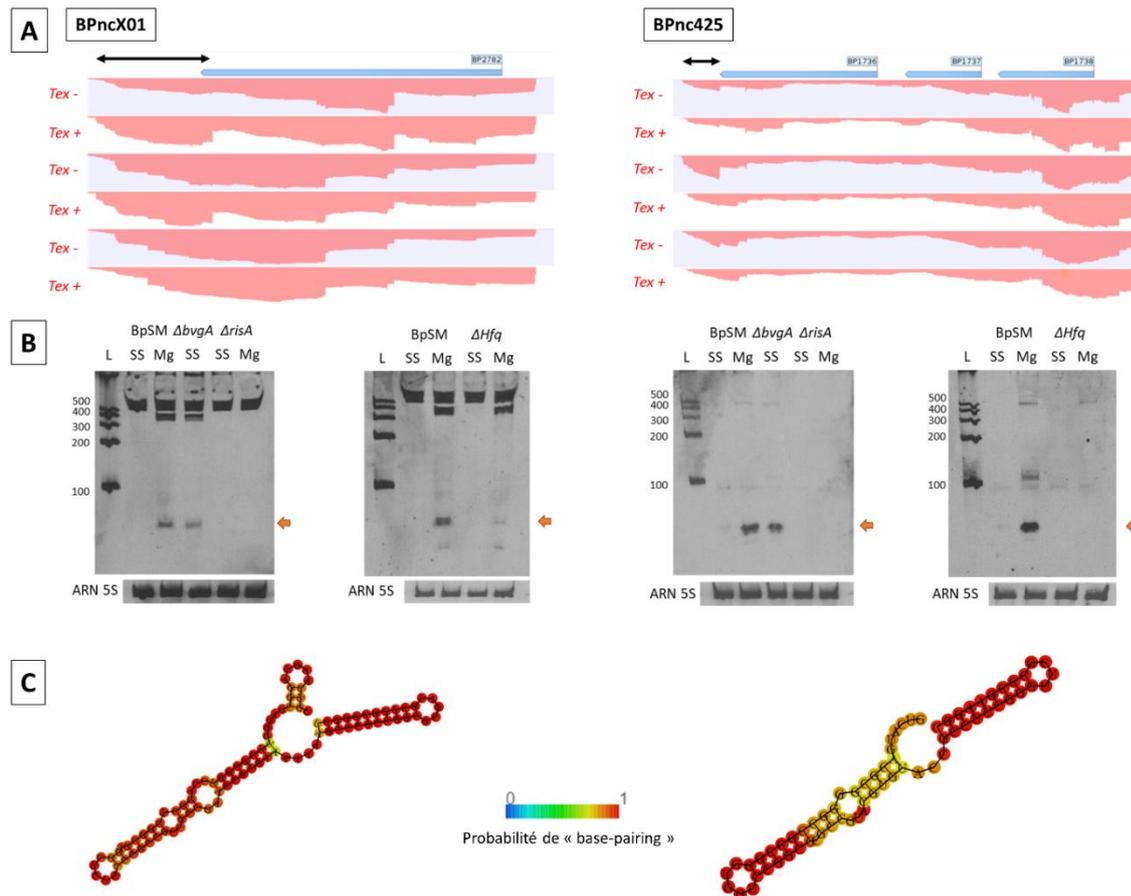
A l'inverse, 6 transcrits candidats ont été prédits avec un TSS en phase  $Bvg^-$ , et absent en phase  $Bvg^+$ , et peuvent donc être considérés comme un profil d'expression similaire aux *vrg* (virulence repressed genes), des transcrits réprimés en phase de virulence. Aucun de ces transcrits n'a été validé par Northern Blot en phase  $Bvg^+$ , mais pourraient cependant être transcrits en phase  $Bvg^-$ . Quatre de ces transcrits sont localisés en 3'UTR de gène, dont une nouvelle prédiction, appelée BPncX01, prédit en 3'UTR du gène BP2782, un *vrg* majeur de *Bordetella pertussis* dont la fonction est encore inconnue (Coutte et al., 2016).

#### iv. Etude de deux ARNs : BPncX01 et BPnc425

Parmi la liste de prédictions de transcrits modulés en fonction de l'état de virulence du pathogène, deux transcrits, BPnc425 et BPncX01, sont localisés en 3'UTR de gènes exprimés uniquement en phase  $Bvg^-$  (Figure 29A). Un TSS a été prédit pour chacun de ces candidats à l'extrémité 3' des gènes BP2782 pour BPncX01, codant pour une lipoprotéine dont la fonction est inconnue, et à l'extrémité de l'opéron BP1736-BP1737-BP1738 pour BPnc425. Cet opéron pourrait potentiellement être impliqué dans la réponse au stress osmotique, BP1736 ayant des homologies avec le gène *osmY* d'*Escherichia coli* (Yim and Villarejo, 1992).

Ces deux transcrits sont exprimés uniquement en phase  $Bvg^-$ . Il a déjà été montré par puces à ADN, en condition classique de culture, que BP2782 et l'opéron B1736-BP1737-BP1738 ne sont pas exprimés (Coutte et al., 2016). Les deux transcrits BPncX01 et BPnc425 ont un profil similaire aux gènes localisés en amont (Figure 29B). Leur expression est induite par ajout de sulfate de magnésium dans le milieu de culture ou chez un mutant  $\Delta bvgA$ . De plus, ces deux transcrits ne sont pas exprimés chez un mutant  $\Delta risA$  avec ou sans ajout de sulfate de magnésium. Ils ont donc un profil de *vrg*, dont l'expression est dépendante de l'activation RisA lié au cyclique di-GMP et l'inactivation du système à deux composants BvgAS. De plus, BPncX01 et BPnc425 ne sont pas exprimés chez un mutant  $\Delta hfq$  avec ou sans ajout de sulfate de magnésium dans le milieu de culture.

La prédiction du TSS et des terminateurs de transcription a permis de déterminer les séquences des deux ARN régulateurs potentiels dans le but de prédire leur structure (Figure 29C). La probabilité de l'appariement des bases indique que la structure de BPncX01 et de BPnc425 seraient stables. Ces résultats cumulés avec la perte de stabilité chez un mutant  $\Delta hfq$  renforcent la probabilité que ces deux transcrits puissent avoir une fonction d'ARN régulateur, agissant potentiellement en *trans*.



**Figure 29:** Caractérisations préliminaires de BPncX01 et de BPnc425.

(A): Résultats en triplicatas des séquençages différentiels d'ARN en phase Bvg-. Les séquences ont été alignées avec le logiciel CLC Genomics Workbench. Les deux ARN régulateurs potentiels, représentés par une flèche noire, sont transcrits en 3'UTR des gènes BP2782 pour BPncX01 (Gauche), et de l'opéron BP1736, BP1737 et BP1738 pour BPnc425 (Droite). Flèche jaune : Cadre ouvert de lecture. Flèche bleue : Annotation du gène. Tex- : Séquençage des ARN totaux. Tex+ : Séquençage enrichi en ARN primaires. (B) : Validation de l'expression de BPncX01 et de BPnc425 chez différents mutants par Northern Blot. Les flèches orange indiquent les transcrits BPncX01 (Gauche) et BPnc425 (Droite). L : Marqueur de taille. SS : Culture classique. Mg : Culture à 50mM sulfate de magnésium. BpSM : Souche de laboratoire.  $\Delta bvgA$  : Mutant de délétion du gène *bvgA*.  $\Delta risA$  : Mutant de délétion du gène *risA*.  $\Delta hfq$  : Mutant de délétion du gène *hfq*. (C) : Prédiction de structure des transcrits. Les structures ont été prédites avec le logiciel RNAfold disponible à l'adresse : <http://rna.tbi.univie.ac.at/cgi-bin/RNAWebSuite/RNAfold.cgi>

### 3. Discussion : Un complexe de régulation encore inexploré

L'étude du transcriptome primaire de *Bordetella pertussis* en phase Bvg<sup>+</sup> a permis d'établir une liste exhaustive d'ARN régulateurs potentiels. Un total de 978 prédictions d'ARN régulateur a été prédit par le cumul de trois approches prédictives, se basant sur le séquençage du transcriptome et sur l'identification de séquences stables et non annotées, et contiennent un taux de 70% de vrais-positifs. La combinaison des approches biologiques et bioinformatiques recouvrent un nombre de prédictions similaire à celles retrouvées chez d'autres pathogènes, notamment chez *Clostridium difficile*, chez qui 511 ARN régulateurs ont été prédits par des approches bioinformatiques, et 251 retrouvés par séquençage à ARN différentiel (Beauregard et al., 2013; Papenfort et al., 2015; Soutourina et al., 2013; Toledo-Arana et al., 2009).

Dans la liste d'ARN candidats issus de mon travail, 175 pourraient être de longs 5'UTR de gènes codant, correspondant potentiellement à des riboswitches ou des thermosenseurs discutés dans le premier chapitre de cette étude, dont un nouveau long 5'UTR a été validé par RT-PCR. Cependant, d'autres prédictions d'ARN régulateurs sont localisées au niveau des prolongements de transcription issus d'ARN ribosomiques et d'ARN de transfert. Ces prolongements correspondent à des espaceurs, clivés et rapidement dégradés pendant le processus de maturation, mais il a été montré très récemment que certains de ces espaceurs externes et internes peuvent agir comme une éponge à ARN (Lalaouna et al., 2015). Le 3'ETS du précurseur polycistronique *glyW-cysT-leuZ* chez *Escherichia coli* peut s'hybrider avec les ARN RyhB et RybB pour inhiber une action trop importante de ces régulateurs. Chez *Bordetella pertussis* également, les prédictions correspondant aux espaceurs internes et externes d'ARN de transfert pourraient agir comme des éponges à ARN régulateurs pour limiter leur action.

De la même façon, 144 prédictions sont localisées au niveau des 3'UTR d'ARN messagers, dont 4 ont été validés par Northern Blot. Récemment, il a été montré que ce genre d'ARN régulateurs était soit issu du clivage de l'ARN messager, classé comme ARN régulateur en 3'UTR de type I, soit transcrit par un promoteur localisé dans la séquence codante du gène en amont, classé alors comme ARN régulateur en 3'UTR de type II (Chao et al., 2012; Miyakoshi et al., 2015a; Ren et al., 2017; Svensson and Sharma, 2016). Ces ARN régulateurs agissent en *trans* et sont stabilisés par l'hexamère Hfq, ce qui serait le cas de BPnc425, BPnc451 et BPncX01 (Guo et al., 2014; Miyakoshi et al., 2015b). Ces trois prédictions, validées par Northern Blot, sont déstabilisées en absence du polymère Hfq. L'absence de TSS pour BPnc451 indique qu'il serait un transcrit de type I, issu du clivage de l'ARNm du gène *ptpA* codant pour une tyrosine-phosphatase. Ces protéines sont régulièrement impliquées dans la transduction du groupement phosphate des systèmes à deux composants, et également impliquées dans la réponse aux stress et dans la virulence de certains pathogènes (Caselli et al., 2016; Musumeci et al., 2005). A l'inverse, des TSS ont été prédits pour BPnc425 et BPncX01, deux ARN ayant un profil *vrg*, transcrits en 3'UTR de type II. Ces deux transcrits sont localisés en 3'UTR des gènes les plus fortement exprimés en phase Bvg<sup>-</sup> (Coutte et al., 2016). Leurs localisations pourraient suggérer une implication dans la réponse aux stress, comme l'ARN régulateur

MicL chez *Escherichia coli* régulant négativement la synthèse de lipoprotéines pour éviter les tensions de l'enveloppe et qui est transcrit en 3'UTR de *cutC*, impliqué dans la réponse aux excès de cuivre (Guo et al., 2014; Miyakoshi et al., 2015a). L'opéron BP1736-BP1737-BP1738 étant potentiellement impliqué dans la réponse au stress osmotique, BPnc425 pourrait avoir une fonction parallèle à celle de l'opéron.

Un grand nombre d'ARN régulateurs prédits sont localisés dans des régions intergéniques, agissant potentiellement en « trans ». Les ARN régulateurs transcrits dans ces régions sont en général stabilisés par l'hexamère Hfq, facilitant également l'appariement avec leur cible (Brennan and Link, 2007; Kim et al., 2015; Vogel and Luisi, 2011). Chez *Bordetella pertussis*, il a déjà été montré que ce polymère était impliqué dans la virulence du pathogène, et plus particulièrement dans la régulation du système de sécrétion de type III, un facteur de virulence majeur nécessaire à la persistance de l'infection chez la souris (Bibova et al., 2013, 2015; Melvin et al., 2014). La perte de virulence en absence de l'hexamère Hfq en phase Bvg<sup>+</sup> pourrait impliquer l'action d'ARN régulateurs prédits dans notre liste de candidats. Ces ARN potentiellement impliqués dans la régulation de la virulence du pathogène pourrait être également modulés en fonction de l'état d'activation du phosphorelai BvgAS. Les différences de prédictions identifiées en fonction de l'état de virulence du pathogène pourrait amener à la découverte d'ARN impliqués dans la régulation de la virulence du pathogène. La comparaison des départs de transcription en phase Bvg<sup>+</sup> et Bvg<sup>-</sup> a ainsi permis d'identifier 12 prédictions à profil *vag*, et 6 prédictions à profil *vrg*.

Le transcriptome bactérien est, en général, constitué de nombreux ARN antisens naturels (Brantl, 2007; Inouye, 1988; Lasa et al., 2012). L'étude du transcriptome primaire de *Bordetella pertussis* a permis d'identifier 381 prédictions d'ARN régulateurs orientés en antisens des gènes annotés. Parmi ces ARN, 6 antisens de gènes majeurs de virulence ont été validés par Northern Blot. Les ARN transcrits en antisens de gènes ont généralement un rôle d'inhibiteur, soit en altérant la stabilité de l'ARN messenger par recrutement de RNases, soit par un blocage de la séquence de fixation du ribosome, ou soit par la formation d'un terminateur précoce de transcription (Brantl, 2007; Chang et al., 2015; Georg and Hess, 2011; Inouye, 1988; Pinel-Marie et al., 2014). BPnc180, antisens du gène *cyaA*, codant pour la toxine hémolytique du pathogène, est localisé en 5' de l'ARN messenger et est exprimé uniquement en phase Bvg<sup>+</sup>. La séquence s'appariant avec la séquence codante du gène, il pourrait avoir un rôle de déstabilisateur. L'antisens de *dnt*, BPnc838, localisé à l'extrémité 3', pourrait jouer un rôle similaire, proche des systèmes de toxine-antitoxine connus (Brantl and Jahn, 2015; Pinel-Marie et al., 2014). Enfin, les antisens des adhésines *fhaB*, *fimB* et *fim3* sont également localisés en 5' de la séquence codante. Ces ARN régulateurs potentiels ont été prédits comme étant modulés en fonction de l'état de virulence du pathogène. Ils seraient donc régulés par le système à deux composants BvgAS, et souligneraient un nouveau système de régulation de l'expression des toxines et adhésines du pathogène.

Les ARN régulateurs permettent à la cellule de s'adapter rapidement aux changements environnementaux (Svensson and Sharma, 2016). L'induction d'un stress

causé par un changement environnemental serait en lien avec des changements d'expression d'ARN régulateurs (Kopf et al., 2015). De ce fait, l'expression de ces transcrits pourrait varier d'une souche à une autre au sein de la même espèce. Les remaniements génomiques causés par la pression vaccinale chez *Bordetella pertussis* pourraient également altérer l'expression de certains de ces ARN régulateurs (Brinig et al., 2006; Sealey et al., 2016).

Parmi la liste de candidats régulateurs, 72 prédictions seraient issues des prolongements de transcription provenant des IS481. L'impact de ces prolongements de transcription des IS481 sur le transcriptome primaire de *Bordetella pertussis* a été analysé. Un de ces candidats ARN régulateur, BPnc264, orienté en antisens du 3' de la séquence codante de *fim2*, prendrait son origine dans l'IS481 BP1118 localisée en aval du gène. L'étude de l'impact des prolongements de transcription des IS481 et l'étude de BPnc264 font l'objet des prochains chapitres de cette thèse.

### **III. Impact des IS481 sur la transcriptome de *Bordetella pertussis***

L'analyse du transcriptome primaire de *Bordetella pertussis* a permis d'identifier plusieurs candidats ARN régulateurs prenant leur origine au sein des séquences d'insertion IS481, IS1002 et IS1663. Pour déterminer l'impact des éléments génétiques mobiles sur le transcriptome du pathogène, des analyses d'expression des transcrits de toutes les IS ont été entreprises.

#### **1. Les prolongements de transcription issus des séquences d'insertion**

L'architecture des transcrits issus des IS481, IS1002 et IS1663 a été déterminée à partir de notre annotation du transcriptome primaire. Des prolongements de transcription prenant leur origine dans les éléments génétiques mobiles ont été observés pour la plupart des séquences d'insertion (Figure 30A et Figure 30B). Les prolongements observés à partir de l'extrémité 5' sont orientés en antisens de la transposase de l'IS481 annotée, alors que les prolongements de transcription observés à partir de l'extrémité 3' se prolongent dans le même sens que l'ARN messager de la transposase. Ces prolongements sont répertoriés dans le tableau supplémentaire S2 (<https://tinyurl.com/y7el869z>)

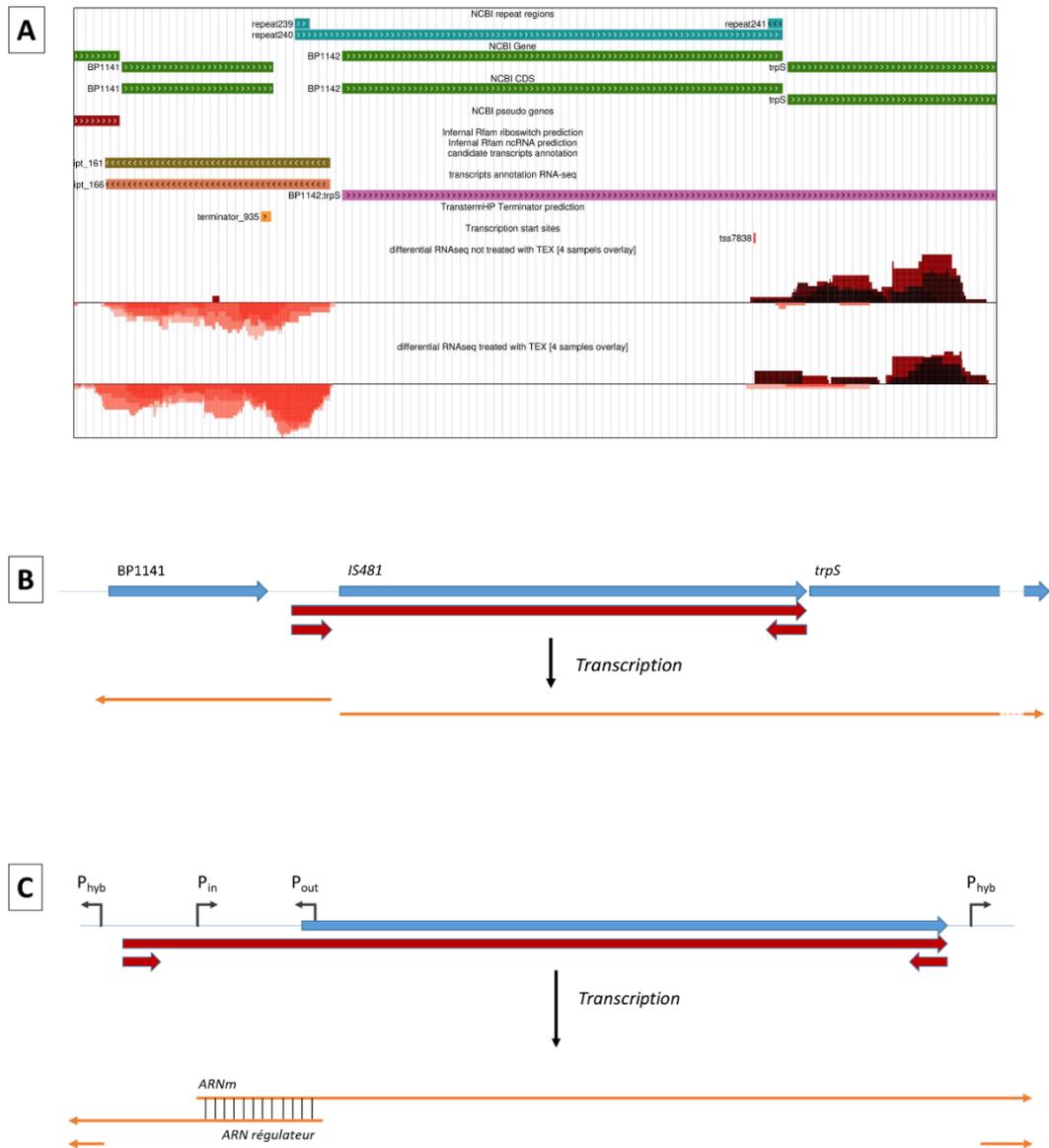
##### **i. Les promoteurs des IS481, IS1002 et IS1663**

Les séquences d'insertion contiennent un promoteur interne pour transcrire l'ARN messager du gène de la transposase, appelé  $P_{in}$ , retrouvé dans la région non codante en 5' (Case et al., 1988; Siguier et al., 2015) (Figure 30C). La majorité des éléments génétiques mobiles portent un promoteur interne, orienté en antisens du promoteur  $P_{in}$ , appelé  $P_{out}$ , localisé en aval du codon d'initiation de la traduction de l'ARN messager (Kittle et al., 1989; Ma and Simons, 1990; Simons and Kleckner, 1983). Pour l'IS10, cet ARN antisens, appelé RNA-OUT, s'hybride au niveau du codon d'initiation de la traduction de l'ARN messager et du site de fixation du ribosome pour inhiber la traduction de la transposase, limitant les remaniements génomiques. Enfin, certaines IS peuvent transcrire un ARN messager d'un gène flanquant par formation d'un promoteur fort hybride, appelé  $P_{hyb}$  (Prentki et al., 1986). Les IS portent des motifs -35 dans les régions répétées gauches et droites afin de transcrire abondamment le gène de la transposase sous forme circulaire (Figure supplémentaire S4). De ce fait, l'insertion d'une IS près d'une boîte -10 peut former un promoteur fort.

Chez *Bordetella pertussis*, les séquences des boîtes -10 et -35 du  $P_{out}$  ont déjà été mises en évidence par extension d'amorces (DeShazer et al., 1994; Han et al., 2011). Pour valider les prolongements de transcription des deux promoteurs  $P_{in}$  et  $P_{out}$ , et la formation de promoteurs forts  $P_{hyb}$ , plusieurs transcrits ont été testés par 5'RACE (Matériel et Méthode 6), (Figure 31).

Le départ de transcription du gène de la transposase a été déterminé avec une amorce spécifique de la région codante de toutes les IS481. Les séquences ont été alignées sur le gène BP1118, sélectionné pour représenter les IS481 de manière

générale. La plus grande partie des séquences alignées (40%) s'étendent au-delà de la région répétée gauche de BP1118 et proviendraient de prolongements de transcrits



**Figure 30:** Architecture commune des transcrits issus des IS481.

(A) : Capture d'écran des résultats de séquençage centrée sur l'IS481 BP1141. (B) Schématisation des transcrits se prolongeant en dehors de l'IS481 BP1141. Flèche bleue : Cadres de lecture ouverts des gènes annotés. Flèche rouge : Annotation des IS481. Flèche orange : Transcrits observés par séquençage à ARN. (C) : Identification des promoteurs des séquences d'insertion décrits dans la littérature scientifique. P<sub>in</sub> : Promoteur du gène codant pour la transposase. P<sub>out</sub> : Promoteur interne orienté en antisens du P<sub>in</sub>, transcrit un ARN régulateur en antisens de l'ARN messager de la transposase. P<sub>hyb</sub> : Promoteur hybride formé par des motifs -35 retrouvés dans les régions répétées gauches et droites des IS (Voir figure supplémentaire S4) et un motif -10 dans la région génomique environnante. La formation du promoteur dépend de l'insertion de l'IS à une distance presque parfaite de la boîte -10. Lignes noires : Appariement documenté entre l'ARN messager de la transposase et l'ARN régulateur orienté en antisens.

issus des régions génomiques localisées en amont des IS481, dépendant de leur localisation génomique. Un TSS majoritaire a été détecté à 12 bases (+/- 1 base) en amont du codon d'initiation de la traduction, représentant 30% des séquences alignées. Un second TSS, représentant 20% des séquences alignées, a été détecté sur le codon d'initiation de la traduction, indiquant un potentiel transcrit « leaderless ».



**B**

Origine du transcrit	Nom du transcrit / gène	Prédiction du TSS	TSS validé par 5'RACE
TSS de l'ARN messenger de la transposase originant du P <sub>in</sub>	<a href="#">BP1118</a>	-	1 175 754
			1 175 766
Transcrit originant d'un P <sub>out</sub> d'une IS481 orienté en sens de gène	<a href="#">bfr</a>	174 307	174 165
			174 306
			868 908
			869 138
Transcrit originant d'un P <sub>out</sub> d'une IS481 orienté en antisens de gène	<a href="#">BP0840</a>	868 907	3 543 852
			<a href="#">BP3322</a>
			3 543 853
			<a href="#">katA</a>
			4061786 (DeShazer <i>et al.</i> , 1994)
Transcrit originant d'un P <sub>out</sub> d'une IS481 orienté en antisens de gène	<a href="#">BPnc266</a>	1 205 136	1 205 051
			1 205 135
			<a href="#">AS-talB</a>
			1 524 716
			1 524 717
Transcrit originant d'un P <sub>out</sub> d'une IS481 orienté en antisens de gène	<a href="#">BPnc385</a>	1 630 661	1 630 584
			1 630 660
			1 781 569
			1 781 568
Transcrit originant d'un P <sub>out</sub> d'une IS481 se prolongeant en région intergénique	<a href="#">BPnc418</a>	Promoteur hybride : 1 781 435	1 781 435
			<a href="#">BPnc576</a>
			2 448 013
Transcrit originant d'un P <sub>out</sub> d'une IS481 se prolongeant en région intergénique	<a href="#">BPnc917</a>	3 824 592	2 447 936
			2 448 012
			3 824 593

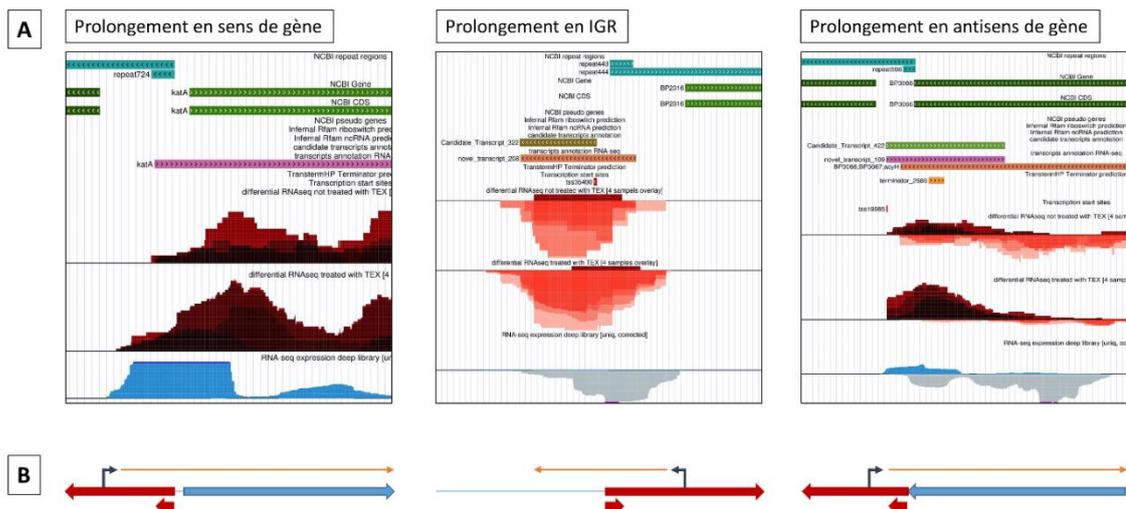
**Figure 31:** Validation des départs de transcription issus des IS481 par 5'RACE. (A) : Capture d'écran des alignements de séquence sur CLC Genomics Workbench v10. Flèche rouge : Annotation des IS481. Flèche bleue : Séquences codantes annotées. Flèche noire : Départ de transcription des promoteurs. (B) : Position des TSS prédits et validés par 5'RACE.

Les départs de transcription issus du  $P_{out}$  ont été déterminés à partir d'une amorce spécifique des transcrits se prolongeant dans la région génomique en amont de l'IS481. Dix TSS issus de  $P_{out}$  ont été validés par 5'RACE (Figure 31). Les motifs -10 et -35 décrits par DeShazer *et al.* ont été identifiés pour chaque transcrit, mais la position du TSS varie d'un nucléotide par rapport à la position prédite par ces auteurs. Le TSS issu d'un  $P_{hyb}$  a également pu être validé par 5'RACE et la séquence du promoteur a pu être identifiée (Figure supplémentaire S4).

## ii. Orientation des transcrits issus des IS

Sur les 259 IS annotés du génome de la souche Tohama I, uniquement 9 d'entre elles ne montrent aucun prolongement de transcription en 5' et en 3' (Tableau Supplémentaire S2). L'annotation du transcriptome primaire a permis d'identifier 6 TSS formés par des  $P_{hyb}$  entre une boîte -35 de régions répétées gauches et droites des IS481 et une boîte -10 des régions génomiques voisines. Des prolongements de transcription de l'ARN messager des transposases transcrit à partir du  $P_{in}$  ont été observés pour 101 séquences d'insertion, et des prolongements de transcription à partir du  $P_{out}$  ont été observés pour 243 séquences d'insertion.

Les prolongements de transcription ont été classés en fonction de leur orientation par rapport à l'annotation du génome (Figure 32). Pour 70 IS (66 IS481, 2 IS1002 et 2 IS1663), le prolongement de transcription issu du  $P_{out}$  est orienté dans le même sens que le gène en aval, donnant ainsi un ARN messager pour ce gène. Cet ARN vient s'ajouter en tant que ARN messager supplémentaire si le gène a son propre promoteur, il peut aussi faire office de seul messager si le gène n'a pas de promoteur propre. Ces configurations se retrouvent pour les gènes *kata* et *bteA* (DeShazer et al., 1994; Han et al., 2011). De la même façon, 23 prolongements de transcription issus du  $P_{out}$  s'étendent



**Figure 32:** Orientation des prolongements de transcription types issus des  $P_{out}$  des IS481. (A) : Captures d'écran des prolongements de transcription orientés en sens de gène, en antisens de gène, et en région intergénique. (B) : Schématisation des captures d'écran. Flèche rouge : Annotation des IS481. Flèche bleue : Séquence codante de gène. Flèche noire : Promoteurs  $P_{out}$ . Flèche orange : Transcrits des  $P_{out}$

en région intergénique, se retrouvant dans la liste de candidats ARN régulateurs non codant (Tableau Supplémentaire S1 et S2). Enfin, 150 prolongements de transcription issus du P<sub>out</sub> (139 IS481, 2 IS1002 et 9 IS1663) s'étendent en antisens des ARN messagers des gènes annotés. Ces derniers sont les plus représentés, et ont été annotés pour certains d'entre eux comme candidats ARN régulateurs non codant. L'orientation de ces transcrits et le départ de transcription de certains d'entre eux a été validé par 5'RACE également (Figure 31).

De la même manière, l'ARN messenger du gène de la transposase issus du P<sub>in</sub> des IS se prolonge dans la même orientation que le gène en amont dans 46 cas, en antisens du gène en amont dans 38 cas, et dans une région intergénique dans 17 cas. De plus, certains gènes semblent transcrits uniquement par les promoteurs des IS. Par exemple, comme évoqué plus haut, le gène *kata*, transcrit par le P<sub>out</sub> de l'IS481 en aval, ne possède pas son propre promoteur et est retrouvé dans la même situation par notre annotation (DeShazer et al., 1994). La région intergénique entre l'IS481 BP1142 et le gène *trpS* n'est que de 9pb, l'annotation du transcriptome primaire montre que ce gène serait uniquement co-transcrit avec l'ARN messenger de l'IS481 en amont (Figure 30).

## **2. Les prolongements d'IS et les diversités transcriptomiques inter-souches**

La diversité des génomes entre les différentes souches de *Bordetella pertussis* serait en lien avec les remaniements des génomes provoqués par les IS, en particulier par les IS481 présentes en grand nombre (Brinig et al., 2006; Heikkinen et al., 2007). Un exemple marquant est la comparaison des génomes des souches de *Bordetella pertussis* avant et après l'ère vaccinale dans les pays développés (Xu et al., 2015). Les prolongements de transcription étant observés pour presque la totalité des IS, la formation d'ARN messagers et de candidats ARN régulateurs issus de ces IS pourrait, elle aussi, varier entre différentes souches du pathogène en fonction de l'organisation génomique induite par les remaniements.

### **i. La diversité de localisation génomique des IS481**

Pour déterminer l'impact global des IS481 sur l'architecture du transcriptome, la position et le nombre d'IS481 a été comparé entre 21 souches de *Bordetella pertussis* entièrement séquencées et annotées (Matériel et Méthode 17). Les IS481 sont retrouvées en plus de 230 copies pour chaque souche analysée (Tableau 7A). Cependant, ce nombre varie d'une souche à une autre, allant de 236 copies pour la souche 18323 à 260 copies pour la souche VA-190.

Les différences de localisation génomique entre les 21 souches ont été déterminées en comparant les séquences du départ de transcription des P<sub>out</sub> jusqu'à 200 nucléotides en aval de l'extrémité de l'IS (Matériel et Méthode 17). Certaines localisations d'IS481 sur le génome sont spécifiques de certaines souches (Tableau 7B). Par exemple, il existe 9 positions d'IS481 spécifiques sur le génome de la souche de référence Tohama I (Colonne) par rapport au génome de la souche D420 (Ligne), souche représentative des souches circulantes. A l'inverse, il existe 15 positions d'IS481 spécifiques sur le génome de la souche D420 (Colonne) par rapport au génome de la souche Tohama I (Ligne).

A

Souche	Taille du génome	Nombre d'IS481
Tohamal	4.102.978	246
D420	4.106.389	252
B1917	4.102.186	248
B1920	4.114.630	254
VA-190	4.127.222	260
CS	4.133.777	241
B202	4.128.979	251
B203	4.134.643	247
18323	4.043.846	236
137	4.134.593	246
1838	4.108.472	253
1865	4.105.363	251
3405	4.109.988	257
3582	4.104.305	250
3585	4.106.388	252
3621	4.100.705	251
3629	4.101.147	248
3640	4.110.990	257
3658	4.103.254	249
3913	4.109.548	255
3921	4.111.557	257

B

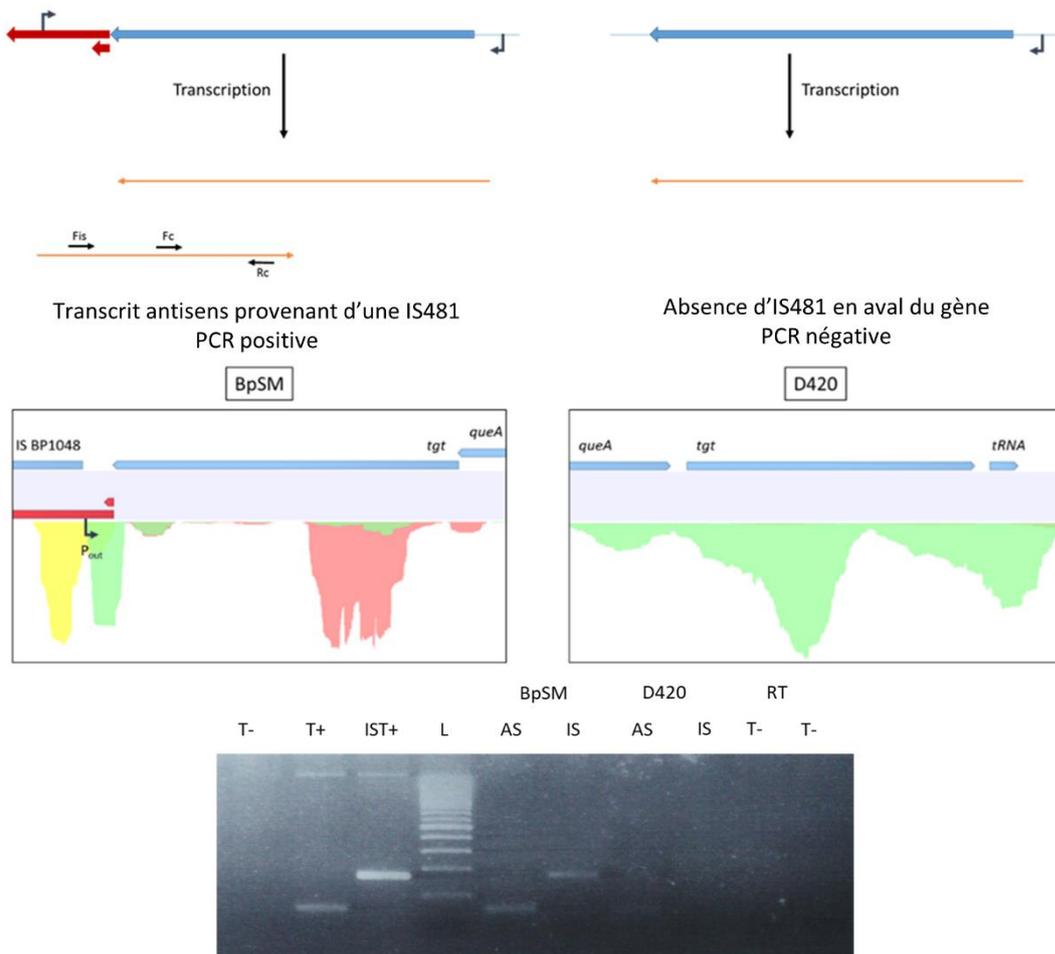
	Tohl (246)	VA-190 (260)	18323 (236)	B3629 (248)	B1917 (248)	D420 (252)	B3913 (255)	CS (241)	B3405 (257)	B202 (251)	B203 (247)	137 (246)	B3621 (251)	B3921 (257)	B3658 (249)	B1920 (254)	B3582 (250)	B3585 (252)	B1865 (251)	B3640 (257)	B1838 (253)
Tohl	0	19	74	17	14	15	14	6	16	15	21	20	14	16	14	14	16	13	14	16	15
VA-190	5	0	71	3	0	1	1	6	3	5	18	17	1	3	1	3	2	0	0	3	1
18323	84	95	0	83	83	87	91	76	93	87	82	81	86	93	85	90	85	88	86	93	88
B3629	15	15	71	0	3	7	11	8	13	9	26	25	7	13	5	11	4	8	6	13	9
B1917	12	12	71	3	0	4	8	6	10	6	23	22	4	10	2	8	2	5	3	10	6
D420	9	9	71	3	0	0	5	6	7	5	20	19	1	7	1	5	2	2	0	7	3
B3913	5	6	72	4	1	2	0	6	2	5	18	17	2	2	1	3	3	0	1	3	2
CS	11	25	71	15	13	17	20	0	22	15	25	24	16	22	14	19	15	17	16	22	19
B3405	5	6	72	4	1	2	0	6	0	5	18	17	2	1	1	3	3	0	1	3	2
B202	10	14	72	6	3	6	9	5	11	0	21	20	6	11	3	8	5	6	5	11	7
B203	20	31	71	27	24	25	26	19	28	25	0	0	24	28	24	25	26	23	24	28	25
137	20	31	71	27	24	25	26	19	28	25	1	0	24	28	24	25	26	23	24	28	25
B3621	9	10	71	4	1	2	6	6	8	6	20	19	0	8	2	6	3	3	1	8	4
B3921	5	6	72	4	1	2	0	6	1	5	18	17	2	0	1	2	3	0	1	3	2
B3658	11	12	72	4	1	4	7	6	9	5	22	21	4	9	0	7	3	4	3	9	6
B1920	6	9	72	5	2	3	4	6	6	5	18	17	3	5	2	0	4	1	2	6	3
B3582	12	12	71	2	0	4	8	6	10	6	23	22	4	10	2	8	0	5	3	10	6
B3585	7	8	72	4	1	2	3	6	5	5	18	17	2	5	1	3	3	0	1	5	2
B1865	9	9	71	3	0	1	5	6	7	5	20	19	1	7	1	5	2	2	0	7	3
B3640	5	6	72	4	1	2	1	6	3	5	18	17	2	3	1	3	3	0	1	0	2
B1838	8	8	71	4	1	2	4	7	6	5	19	18	2	6	2	4	3	1	1	6	0

**Tableau 7:** Répartition des IS481 chez 21 souches séquencées de *Bordetella pertussis*. (A) : Nombre d'IS481 retrouvées dans chaque souche séquencée. (B) : Nombre de localisations d'IS481 spécifiques entre les souches. Les nombres correspondent aux positions spécifiques retrouvées dans la souche nommée en colonne par rapport à la souche nommée en ligne. Voir exemple de lecture dans le texte.

Certaines souches montrent très peu de disparités en terme de localisation d'IS481, comme les souches D420 et B1917, deux souches circulantes, où seulement 4 localisations d'IS481 sont spécifiques de la souche D420. Cependant, d'autres souches montrent des disparités plus importantes, comme la souche circulante 18323 qui

présente plus de 70 localisations d'IS481 spécifiques en comparaison des 20 autres souches.

Les différences de position et de nombre d'IS481 sont le reflet de la diversité génomique des souches de *Bordetella pertussis*. Les divergences génomiques de la souche Tohama I avec les souches cliniques B1917, D420, CS et 18323 ont été étudiées à l'aide du logiciel Mauve (Figure Supplémentaire S5) (Darling et al., 2004). Malgré un génome plus proche de celui de la souche CS que de D420 et B1917, le nombre et les localisations spécifiques des IS481 dans Tohama I est comparable pour les 3 souches. Le génome de la souche Tohama I semble cependant très éloigné de celui de la souche



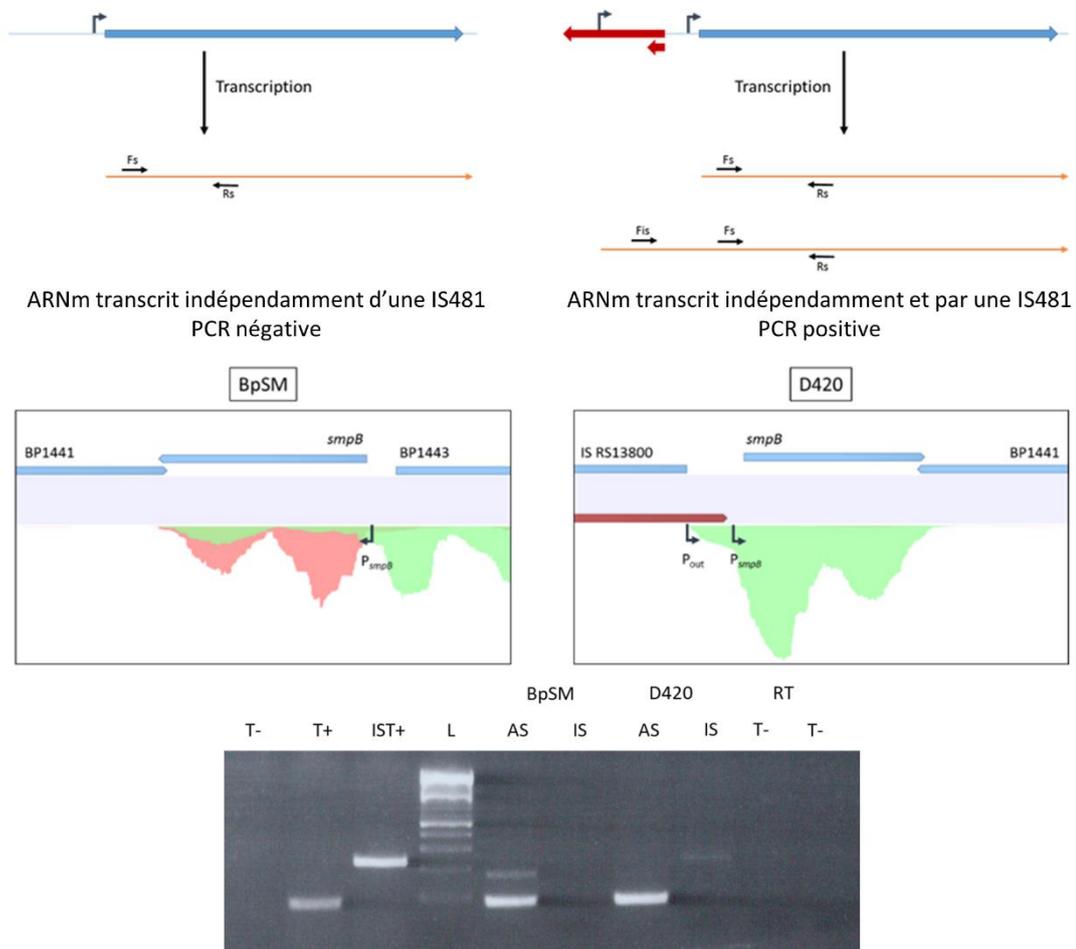
**Figure 33 :** Validation des variations de prolongements de transcription issues de  $P_{out}$  orientés en antisens des gènes en aval entre les souches BpSM et D420.

Les séquences ont été alignées sur les génomes des souches BpSM et D420 avec le logiciel CLC Genomics Workbench v10. La profondeur de séquençage en rouge de séquences alignées sur le brin négatif est en rouge, sur le brin positif en vert, et sur les régions répétées en jaune. Pour les RT-PCR, les ADNc ont été générés avec l'amorce antisens du transcrit ( $R_c$ ), et amplifiés avec le couple  $F_c/R_c$  pour détecter le transcrit, et le couple  $F_{is}/R_c$  pour détecter le prolongement de transcrit issu du  $P_{out}$ . Flèche Rouge : Annotation des IS481. Flèche bleue : Annotation des gènes. Flèche noire : Promoteurs. Flèche orange : Transcrits. T- : Témoin négatif de PCR. T+ et IST+ : Témoin positif de PCR de la région génomique ( $F_c/R_c$ ) et de la région de l'IS ( $F_{is}/R_c$ ). L : Marqueur de taille. AS : PCR avec  $F_c/R_c$ . IS : PCR avec  $F_{is}/R_c$ . RT T- : Témoin négatif de RT.

18323, en comparaison avec les autres souches. Ces divergences génomiques illustrent l'activité des IS481 et des remaniements génomiques possibles. Elles pourraient de ce fait expliquer les divergences de localisation génomique spécifique des IS481 entre les différentes souches de *Bordetella pertussis*.

## ii. La diversité transcriptomique des IS481 inter-souche

La localisation génomique spécifique de certaines IS481 chez différentes souches de *Bordetella pertussis* pourrait générer des divergences transcriptomiques des prolongements de transcription issus de ces IS. Pour valider les diversités transcriptomiques entre différentes souches du pathogène, les ARN totaux de la souche D420 cultivée en condition classique de culture (en phase Bvg<sup>+</sup>) ont été séquencés (Matériel et Méthode 4). Les prédictions de prolongements de transcription issues des IS481 ont été comparées aux localisations génomiques spécifiques des IS481 entre les souches D420 et BpSM, souche dérivée de Tohama I résistante à la streptomycine.



**Figure 34 :** Validation des variations de prolongements de transcription issus de  $P_{out}$  orientés en sens des gènes en aval entre les souches BpSM et D420.

Les ADNc ont été générés avec une amorce antisens de l'ARNm de *smpB* ( $R_s$ ), et amplifiés avec le couple d'amorces  $F_s/R_s$  pour vérifier la transcription du gène, et le couple d'amorce  $F_{is}/R_s$  pour tester la transcription de l'ARNm par le prolongement de transcription issue du  $P_{out}$ . La légende de la figure est identique à celle décrite en figure 33.

Parmi les 9 localisations génomiques spécifiques de la souche Tohama I, seules 3 IS481 montrent des prolongements de transcrite issus de P<sub>out</sub>, 2 sont orientés en sens de gènes pour transcrire un ARNm et un transcrite est orienté en antisens du gène en aval. De la même manière, parmi les 15 localisations génomiques spécifiques de la souche D420, 9 IS481 montrent des prolongements de transcription issus de P<sub>out</sub>, 5 d'entre eux se prolongent en sens des gènes en aval, et 4 en antisens des gènes.

Un antisens du gène *tgt*, codant pour une ARNt ribosyltransférase, a été validé par RT-PCR (Figure 33). Cet ARN est transcrit par le P<sub>out</sub> de l'IS481 localisée en aval du gène *tgt* et est spécifique de la souche BpSM, aucune IS481 n'étant présente en aval de ce gène chez la souche D420. De la même façon, le transcrite du gène *smpB*, codant pour une protéine liant l'ARN messenger-transfert, est transcrit chez les deux souches et a été validé par RT-PCR (Figure 34). Cependant, l'ARNm de *smpB* est transcrit à la fois par son propre promoteur, et par le P<sub>out</sub> d'une IS481 localisée en amont du gène chez la souche D420, absent chez la souche BpSM.

### **3. Discussion : Une régulation transcriptomique par les IS481 ?**

L'analyse du profil d'expression des séquences d'insertion de *Bordetella pertussis* a permis de mettre en évidence des prolongements de transcription, s'étendant en dehors des extrémités 5' et 3' de ces éléments génétiques mobiles que sont les IS. Ces prolongements proviennent de deux promoteurs portés par les IS, le P<sub>in</sub>, permettant la transcription de l'ARN messenger de la transposase, et le P<sub>out</sub>, permettant la transcription d'un ARN régulateur orienté en antisens de cet ARNm (Case et al., 1988, 1990; Kittle et al., 1989; Ma and Simons, 1990). Dans certains cas, les IS peuvent former de nouveaux promoteurs forts, appelé P<sub>hyb</sub>, pouvant également impacter sur la transcription des gènes environnants (Prentki et al., 1986).

Les transcrits peuvent s'étendre au-delà des IS et impacter sur la transcription des gènes proches, il a déjà été montré que ces promoteurs pouvaient fournir de nouveaux ARN messagers, notamment pour les gènes *katA* et *bteA* (Ciampi et al., 1982; DeShazer et al., 1994; Han et al., 2011; Safi et al., 2004). Notre analyse du transcriptome primaire montre que la transcription de gènes à partir des promoteurs fonctionnels des IS est un phénomène commun, dû à l'important nombre d'IS présent dans le génome de *Bordetella pertussis*. Un total de 46 gènes serait ainsi co-transcrit avec l'ARN messenger de la transposase, dont le gène *trpS*, codant pour l'ARNt ligase du tryptophane. De la même façon, 70 gènes seraient transcrits par un P<sub>out</sub> d'une IS, en plus du gène *katA*. C'est la cas, par exemple, des gènes *bfr*, impliqués dans le métabolisme des métaux lourds, et le gène *tyrS*, codant pour une ARNt ligase de la tyrosine (Grundy and Henkin, 1993; Honarmand Ebrahimi et al., 2015).

L'un des résultats majeurs de cette étude concerne de nouveaux prolongements de transcription encore non documentés à ce jour. Un total de 40 transcrits se prolongeant en région intergénique a pu être dénombré, ces transcrits ont été détectés en tant que nouveaux candidats ARN régulateurs non codant (23 transcrits d'un P<sub>out</sub>, 17 transcrits d'un P<sub>in</sub>). De plus, les transcrits les plus représentés de cette catégorie sont orientés en

antisens des gènes environnant, comptabilisant un total de 196 prolongements (150 transcrits d'un  $P_{out}$ , 38 transcrits d'un  $P_{in}$  et 6 transcrits d'un  $P_{hyb}$ ). Ces derniers pourraient avoir une fonction régulatrice sur les gènes environnant les IS481, similaire à celle des ARN régulateurs indépendants orientés en 3' des ARN messagers cibles ou à celle des 3'UTR chevauchant d'ARN messagers (Brantl, 2007; Georg and Hess, 2011; Lasa et al., 2012; Thomason and Storz, 2010; Toledo-Arana et al., 2009). Ces derniers pourraient induire la dégradation de l'ARN messenger cible par formation d'ARN double brin.

L'impact de ces éléments sur le génome de la bactérie est de plus en plus documenté, le nombre d'IS481 chez *Bordetella pertussis* implique une diversité génomique importante entre les souches de la même espèce (Brinig et al., 2006; Stibitz and Yang, 1999; Weigand et al., 2017). Les remaniements génomiques sont impliqués dans l'adaptation des pathogènes à leur environnement, mais seraient également impliqués dans la spécificité d'hôte (Safi et al., 2004; Siguier et al., 2015, 2014). Chez *Bordetella pertussis*, les remaniements génomiques seraient liés à la pression vaccinale pour contrer les défenses induites par le vaccin acellulaire, notamment par inactivation de gènes codant pour des protéines antigènes (Martin et al., 2015; Otsuka et al., 2012; Weigand et al., 2018; Williams et al., 2016; Xu et al., 2015). La localisation et le nombre d'IS481 sur le génome permet la transcription d'ARN messagers et de candidats ARN régulateurs et ces transcrits sont spécifiques d'une souche à l'autre. Les changements d'expression de certains gènes, notamment de gènes de virulence, pourraient s'expliquer par les prolongements de transcription des éléments génétiques mobiles dans une configuration donnée. Par exemple, les variations d'expression de certains gènes de virulence, comme *fim3* et *fhaB* observées chez la souche Bpe280 après seulement 12 passages en culture solide pourraient s'expliquer par les remaniements génomiques ayant entraîné des gains ou pertes de gènes ou encore par des changements au niveau des prolongements de transcription issus des IS qui provoquent une régulation différente (Brinig et al., 2006).

L'activation des IS en général n'est que très peu étudiée, cependant certaines conditions peuvent engendrer une activité plus importante du gène de la transposase. Il a été montré que l'activation des IS pouvait être liée à des facteurs de l'hôte, notamment à des protéines modifiant la topologie de l'ADN comme H-NS, des protéines liées à la réparation de l'ADN comme le système RecA et LexA, des protéines liées à la réplication de l'ADN comme DnaA ou encore les protéines du système de méthylation Dam (Nagy and Chandler, 2004). L'activation des IS par les facteurs intrinsèques de l'hôte impliquent des expositions provoquant des dommages à l'ADN, comme les stress UV ou les variations de températures extrêmes. L'activation des promoteurs  $P_{in}$  et  $P_{out}$  des IS481 pourrait être dépendante d'un stress provoqué chez *Bordetella pertussis*, notamment lors de la pression vaccinale. Les prolongements de transcription seraient régulés par les différentes configurations topologiques de l'ADN, et non pas par un facteur de transcription. Cette hypothèse est renforcée par les différentes forces des promoteurs  $P_{out}$  des éléments mobiles ISAbal, ISEcp1 et ISAbal25, ce qui a été également montré dans notre étude du transcriptome primaire (Figure supplémentaire

S6) (Amman et al., 2018; Kamruzzaman et al., 2015). L'activation des promoteurs et des prolongements de transcription serait ubiquitaire, et régulée uniquement en condition de stress.

L'important nombre de prolongements de transcription des IS orientés en antisens de gènes pourrait avoir un impact sur l'architecture globale du transcriptome de par leur expression et leur localisation. Les remaniements génomiques sont connus pour changer l'expression globale des gènes, mais également pour former ou interrompre les séquences d'ARN régulateurs (Brinig et al., 2006; Kopf et al., 2015; Raghavan et al., 2015). La formation de candidats ARN régulateurs orientés en antisens de gènes serait une nouvelle preuve de variation du transcriptome régulateur entre différentes souches d'une même espèce. Cette étude peut s'ouvrir également à d'autres espèces procaryotes, plusieurs bactéries possèdent des IS dont des prolongements de transcription ont déjà été observés. Par exemple, la souche pathogène *Shigella flexneri* 2457T possède dans son génome plus de 30 copies d'IS2 qui peuvent activer l'expression de gènes (Jellen-Ritter and Kern, 2001; Zaghloul et al., 2007).

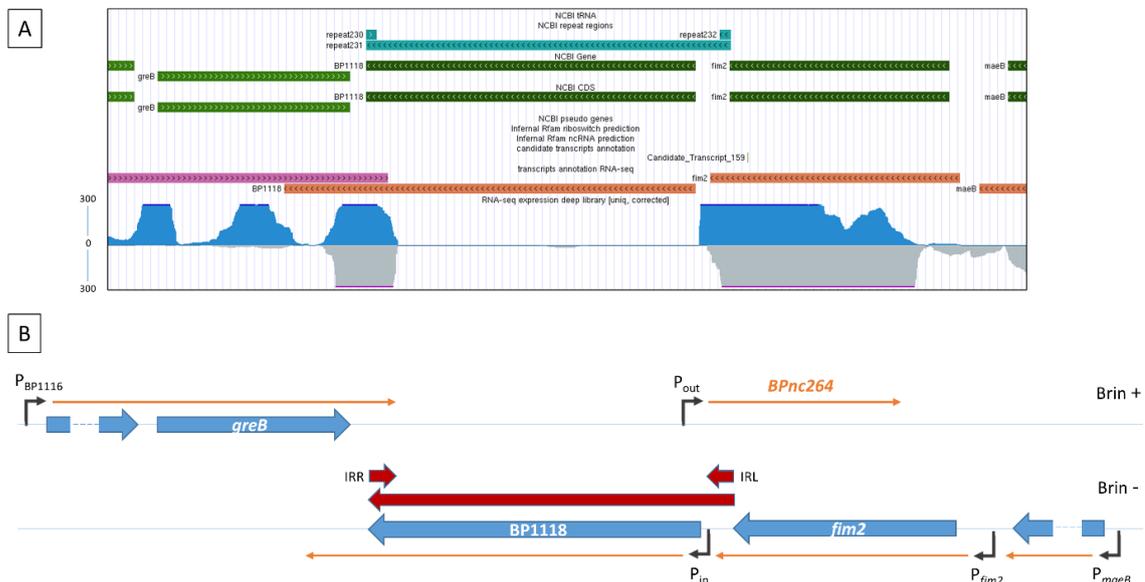
Les prolongements en antisens des IS pourraient également avoir un double rôle régulateur. Ces ARN sont orientés en antisens de l'ARN messenger de la transposase, pour lequel un rôle a déjà été montré, et du gène flanquant (Case et al., 1990). La double régulation potentielle de ces ARN a donc été étudiée à partir d'un ARN candidat, transcrit à partir du P<sub>out</sub> de l'IS481 BP1118, retrouvée en aval du gène *fim2*, dont la protéine est utilisée dans les composants du vaccin acellulaire.

## IV. Caractérisation de BPnc264, ARN régulateur orienté en antisens de BP1118 et de *fim2*

Les prolongements de transcription issus de P<sub>out</sub> des IS481 sont orientés en antisens de l'ARN messager de la transposase, et peuvent également être en antisens de l'ARN messager du gène localisé en amont de l'IS. Chez la souche Tohama I, il existe 150 candidats ARN régulateurs dans une telle configuration, parmi lesquels on retrouve par exemple un antisens de *dinP*, codant pour l'ADN polymérase IV, un antisens de *recR*, impliqué dans la réparation des dommages à l'ADN (Tableau Supplémentaire S2). L'existence et l'expression de tels candidats ARN est susceptible de varier d'une souche à une autre en fonction de la localisation des IS481, comme montré et discuté dans le chapitre 3.

### 1. Identification de BPnc264

Un des transcrits prédits dans notre liste de candidats ARN régulateur, appelé BPnc264, qui est issu du P<sub>out</sub> de l'IS BP1118 et qui est orienté en antisens d'un gène majeur de virulence, *fim2* (Tableaux supplémentaires S1 et S2), a été étudié afin d'établir son potentiel rôle régulateur sur *fim2*. Par ailleurs la fonction de BPnc264 sur la régulation de l'ARN messager de la transposase BP1118 peut être supposée par analogie avec des études menées chez une autre espèce procaryote et des caractéristiques prédites du transcrit.



**Figure 35:** Contexte génomique de BPnc264.

(A) : Résultats de séquençage centrés sur l'IS481 BP1118. (B) : Représentation du contexte génomique de l'IS481 et des transcrits observés. Les annotations sont représentées sur les brins positifs (+) et négatifs (-) de l'ADN (Lignes bleues). Flèche bleue : Annotation des séquences codantes de gènes. Flèches rouges : Annotation des IS481 et des régions répétées gauches (IRR) et droite (IRL). Flèches noires : Promoteurs. Flèches oranges : Transcrits observés par séquençage à ARN.

### i. Localisation génomique et répartition de BPnc264

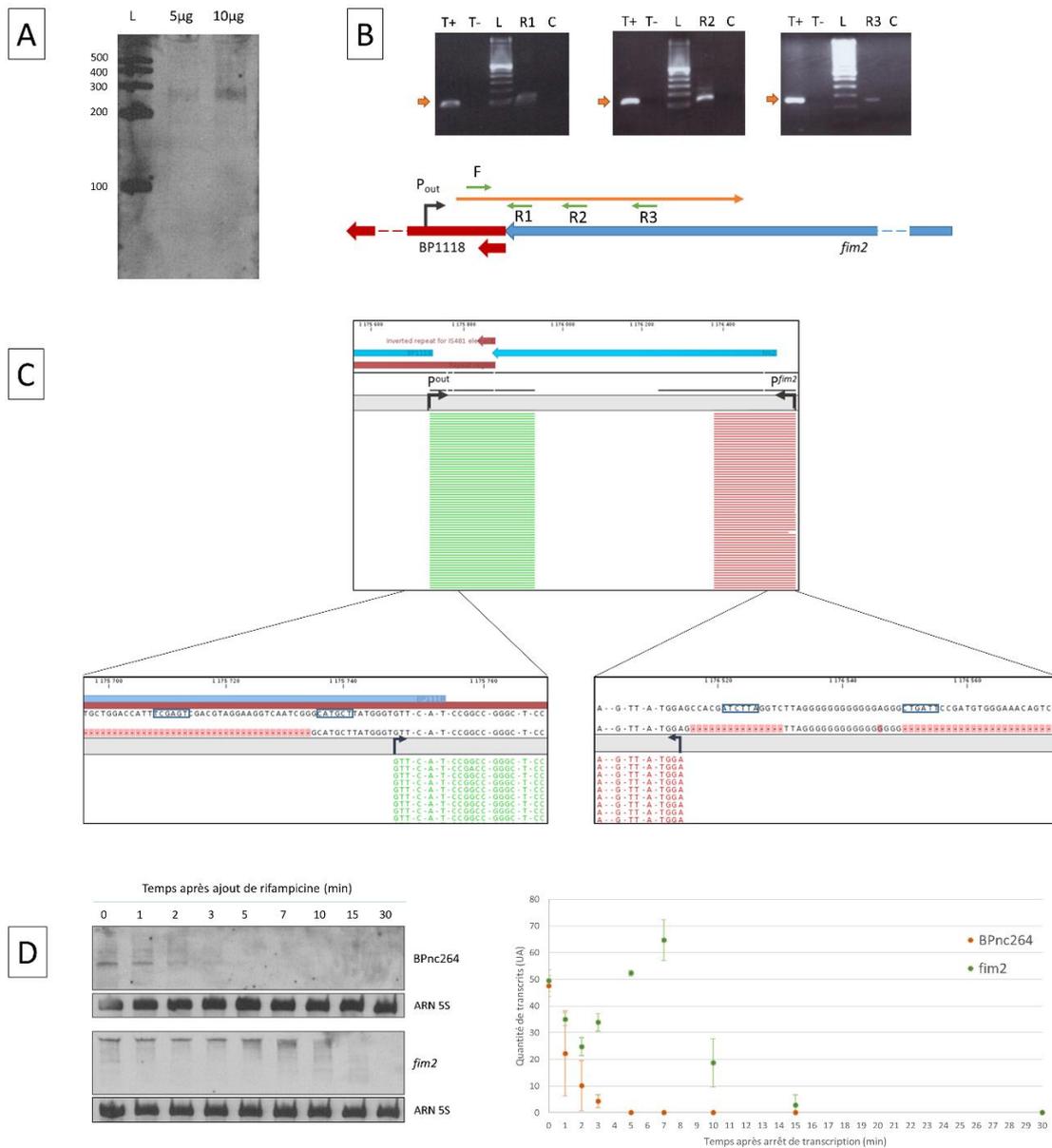
Le prolongement de transcription issu du P<sub>out</sub> de l'IS481 BP1118 s'étend en antisens du gène *fim2*, codant pour la fimbriae de sérotype 2, et dont la protéine est utilisée dans certains vaccins acellulaires (Figure 35) (Sato and Sato, 1999). BPnc264 est prédit à une taille de plus de 600 nucléotides, dont le départ de transcription démarre à 108 nucléotides de l'extrémité gauche de l'IS481 BP1118. Le chevauchement de BPnc264 avec l'ARN messager de la transposase est prédit à 19 nucléotides, et le chevauchement de BPnc264 avec l'ARN messager de *fim2* est prédit à plus de 500 nucléotides.

Le contexte génomique du gène *fim2* parmi les souches du genre *Bordetellae* a été déterminé par Blast-n sur les génomes des souches entièrement séquencées et annotées (Altschul et al., 1990). Celui-ci a été déterminé par alignement de la séquence génomique de BPnc264 prédite par nos résultats de séquençage à ARN (Figure Supplémentaire S7). La présence de l'IS481 en aval du gène *fim2* n'est retrouvée ni chez *Bordetella bronchiseptica*, ni chez *Bordetella parapertussis*. Cependant, la position de l'IS481 juste à côté de *fim2* est retrouvée de manière exacte chez toutes les souches séquencées à ce jour de *Bordetella pertussis*. L'expression d'un ARN antisens au transcrit de *fim2* (BPnc264) serait donc spécifique de la lignée *pertussis*.

### ii. Validations transcriptomiques et caractéristiques de BPnc264

La nature codante ou non-codante de BPnc264 a été prédite à l'aide du logiciel ORFfinder identifiant de potentielles séquences codantes (Matériel et Méthode 17) (Figure Supplémentaire 8). La séquence de BPnc264 contient 16 séquences codantes potentielles, dont 15 coderaient pour des peptides, et une pour une protéine de 177 acides aminés. Parmi ces prédictions, le codon d'initiation ATG a été retrouvé chez 7 d'entre-elles. Cependant, un alignement Blast-p parmi la banque de donnée du NCBI des séquences protéiques prédites ne montre aucun domaine connu à ce jour. La présence de site de fixation du ribosome a été prédite à partir du logiciel RBS calculator prenant en compte les 9 derniers nucléotides de la séquence en 3' de l'ARN 16S (Matériel et Méthode 17) (Salis, 2011). La séquence de BPnc264, bien qu'elle possède de potentiels cadres ouverts de lecture, n'aurait aucune séquence de fixation du ribosome et serait donc non codante.

BPnc264 a été validé par Northern Blot avec une sonde s'hybridant dans la région spécifique se prolongeant en antisens du gène *fim2* (Figure 36A) (Matériel et Méthode 7). Un transcrit stable a été détecté à une taille approximative de 280 nucléotides, inférieure à la taille prédite par séquençage. Le prolongement de transcription de BPnc264 a été validé également par RT-PCR (Figure 36B) (Matériel et Méthode 9). L'ADN complémentaire a été amplifié avec une amorce s'hybridant sur la séquence correspondant à l'extrémité de l'IS en aval du P<sub>out</sub> et une amorce s'hybridant sur la séquence correspondant à la région du gène *fim2*. Le signal est positif pour les trois amorces testées et indique donc que le transcrit BPnc264 provient bien d'un promoteur interne à l'IS481 BP1118.



**Figure 36:** Caractérisation transcriptomique de BPnc264.

(A) : Validation de BPnc264 par Northern Blot. L'ARN a été détecté avec 5 et 10µg d'ARN totaux. L : Marqueur de taille. (B) : Validation de la taille de BPnc264 par RT-PCT. Le plan de RT-PCR est détaillé dans le schéma. Les ADNc ont été générés avec les amorces R1, R2 ou R3, et amplifiés avec les couples F/R1, F/R2 ou F/R3. T+ : Témoin positif de PCR sur l'ADN génomique. T- : Témoin négatif de PCR. C : Contrôle négatif de transcription inverse. (C) : Validation du départ de transcription de BPnc264 et de *fim2* par 5'RACE. Les séquences vertes sont alignées sur le brin positif, et les séquences rouges sont alignées sur le brin négatif. Les flèches noires représentent les départs de transcription détectés par 5'RACE, et les rectangles bleus représentent les motifs -10 et -35 identifiés pour chaque départ. Figure générée par CLC Genomics Workbench v10. (D) : Test de stabilité de BPnc264 et de *fim2* en phase Bvg+. Les cultures ont été arrêtées à différents temps après ajout de rifampicine et 5µg d'ARN totaux de chaque culture a été migré sur gel. Le contrôle des quantités d'ARN a été vérifié par révélation de l'ARN 5S. L'intensité de chaque bande a été quantifiée à l'aide du logiciel Scion Image et représentés sous forme de graphique. La stabilité de chaque ARN a été testée en duplicatas biologiques.

Pour valider ces observations, les départs de transcription de BPnc264, de BP1118 et de *fim2* ont été déterminés par 5'RACE en séquençage à Haut-Débit (Figure 36C) (Matériel et Méthode 6). Le départ de transcription de l'IS481 BP1118 a déjà été décrit dans le Chapitre 3 des résultats.

Le TSS de BPnc264 se situe à 6 pb en aval du codon d'initiation de la traduction du gène de la transposase, position retrouvée par les P<sub>out</sub> des IS481 (Résultats Chapitre 3 Partie 1). Les motifs -10 et -35 déduits de cette position correspondent à ceux décrits précédemment pour le gène *katA* (DeShazer et al., 1994). La validité de la 5'RACE en séquençage à Haut Débit a été testée en déterminant le TSS du gène *fim2*, retrouvé à la base près par notre approche (Voir Résultats Chapitre 1 Partie 1) (Willems et al., 1990).

Le transcrit BPnc264 est orienté en antisens de deux gènes, et est probablement non codant. Pour déterminer l'intégrité de cet ARN régulateur potentiel, la demi-vie de BPnc264 et de l'une de ses cible, *fim2*, a été testée (Figure 36D) (Matériel et Méthode 16). Le nombre de copies d'IS481 étant supérieur à 240 chez la souche BpSM, la demi-vie de BP1118 n'a pas pu être testée pour des raisons de spécificité de sonde. En phase Bvg<sup>+</sup>, la demi-vie de l'ARN antisens est inférieure à une minute, contrairement à celle du transcrit de *fim2*, qui possède une demi-vie comprise entre 8 et 9 minutes. La faible demi-vie de BPnc264 comparée à celle de *fim2* semble indiquer que sa fonction potentielle n'aurait pas lieu dans cette phase. Des manipulations de stabilités en phase Bvg<sup>-</sup> et en phase Bvg<sup>i</sup> sont en cours de réalisation.

## **2. Double interaction et régulation de BP1118**

### **i. Régulation de BP1118 : Analogie au RNA-OUT de l'IS10**

Un rôle régulateur potentiel de BPnc264 a été exploré. La régulation du gène de la transposase a déjà été étudiée chez l'IS10 retrouvée chez *Escherichia coli* (Case et al., 1988, 1990; Ma and Simons, 1990; Ross et al., 2013; Simons and Kleckner, 1983). L'hybridation de l'ARN régulateur, appelé RNA-OUT, sur la séquence de fixation du ribosome bloque la traduction de l'ARN messenger, appelé RNA-IN, et entraîne sa dégradation prématurée.

La configuration des promoteurs impliqués dans cette régulation est retrouvée chez l'IS481 (Voir résultats chapitre 3). Le chevauchement entre les transcrits de BPnc264 et de BP1118 est de 19 nucléotides, contre 36 nucléotides entre le RNA-IN et le RNA-OUT de l'IS10 (Ma and Simons, 1990). La structure de BPnc264 a été prédite à l'aide du logiciel RNA Fold à la fois pour une structure exigeant le minimum d'énergie libre en Kcal/mol, et une structure exigeant un équilibre thermodynamique stable (Figure 37) (Matériel et Méthode 17). Cette dernière représente une structure centroïde, c'est-à-dire une structure avec le minimum de base libre accordée à toutes les structures dans l'ensemble thermodynamique. Une structure ARN stable est retrouvée à l'identique pour les deux types de prédiction.

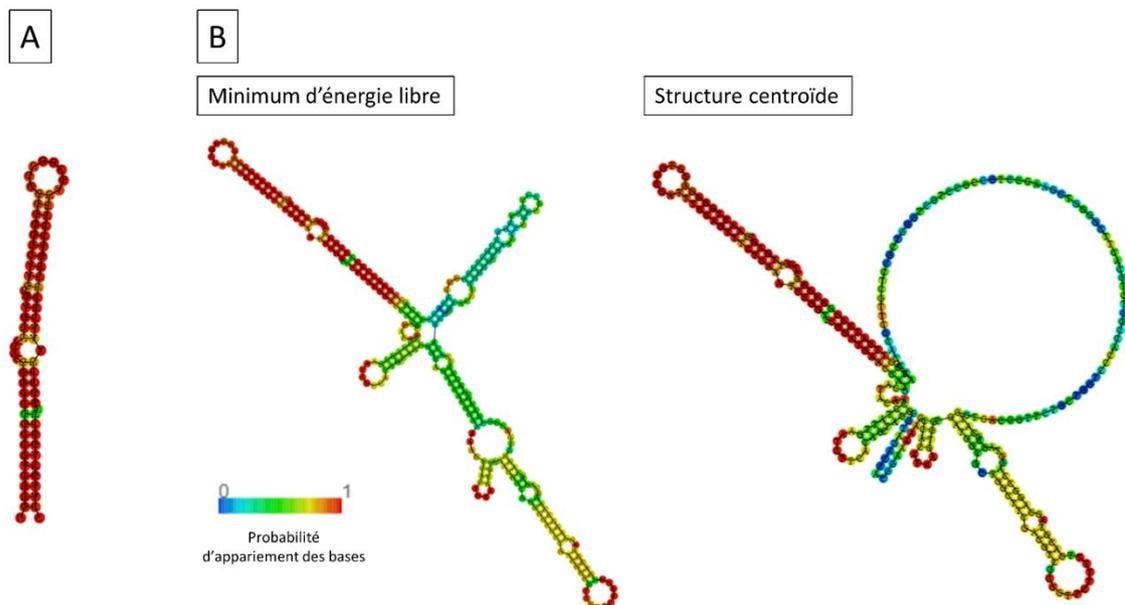
La structure transcrite du P<sub>out</sub> jusqu'à la région répétée gauche de BP1118 est retrouvée très stable, une structure identique a été observée pour les deux prédictions.

La forme tige-boucle est retrouvée également pour le RNA-OUT. La prédiction de structure, cumulée à la configuration génomique, renforce l'analogie entre le RNA-OUT de l'IS10 et BPnc264 (Kittle et al., 1989). Cependant, la structure en aval de l'IS diffère fortement entre les deux prédictions. La séquence de BPnc264 orientée en antisens de *fim2* est donc peu stable ce qui est en concordance avec la mesure de demi-vie de BPnc264 en phase Bvg<sup>+</sup>.

## ii. Interaction de BPnc264 avec BP1118 et *fim2*

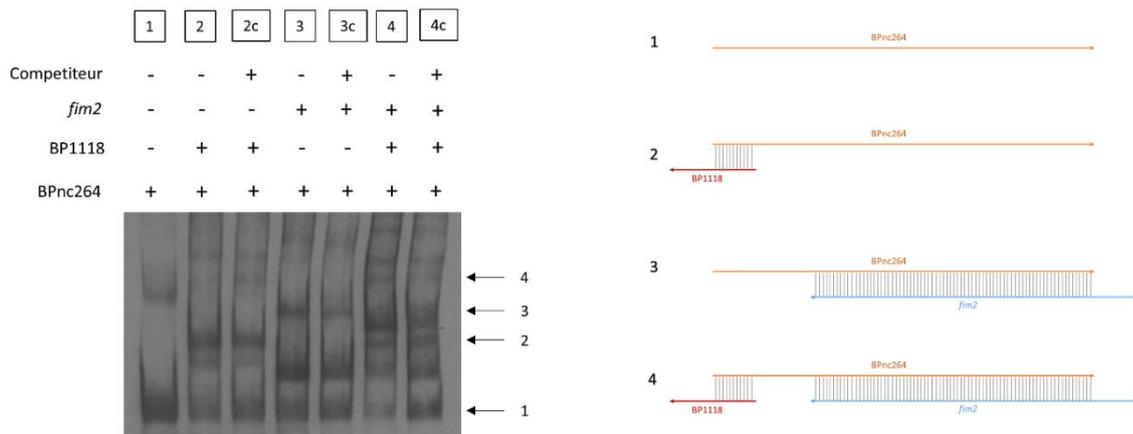
L'interaction de BPnc264 avec l'ARN messager de BP1118 et de *fim2* a été testée par des techniques de retard sur gel (Figure 38) (Matériel et Méthode 15). La séquence de BPnc264 a été clonée en aval d'un promoteur T7 et transcrite *in vitro* en incorporant un groupement biotine pour révéler les réactions d'hybridation et de retard sur gel (Matériel et Méthode 14). L'ARN de BPnc264 marqué à la biotine a été incubé avec soit l'extrémité 5' de l'ARN de BP1118, soit avec l'extrémité 3' de l'ARN de *fim2*, soit avec ces deux fragments d'ARN en même temps. Pour déterminer les bandes d'intérêt, un compétiteur BPnc264 non marqué a été utilisé pour chaque réaction.

Une interaction entre BPnc264 et l'ARN messager de BP1118 est observée (Figure 38, puit n° 2). L'ajout d'un compétiteur (puits 2c) entraîne une diminution d'intensité de la bande, indiquée par la flèche noire 2. De plus, l'intensité du signal de BPnc264 non hybridé (flèche noire 1) s'intensifie en puits 2c comparé au puits 2. De la même manière, une interaction entre BPnc264 et l'ARN messager de *fim2* (flèche 3) est observée par un retard de migration plus important (puits 3 et 3c). L'ajout du compétiteur montre un effet similaire aux indices 2 et 2c. Une double hybridation a été testée entre les ARN de



**Figure 37 :** Prédiction de la structure de BPnc264.

(A): Prédiction de structure de l'extrémité 5' du transcrit jusqu'à la région répétée gauche de BP1118 avec le logiciel RNAfold. (B): Prédiction de la structure complète de BPnc264. Les structures ont été prédites avec une configuration exigeant le minimum d'énergie libre, et une exigeant un équilibre thermodynamique stable.



**Figure 38** : Résultats préliminaires d'interaction entre les transcrits de BPnc264, BP1118 et *fim2*. Le résultats de retard sur gel est retrouvé dans le pannel de gauche avec, pour chaque puit, les ARN engagés dans chaque réaction d'hybridation. Le compétiteur est l'ARN BPnc264 non biotinylé. Les chiffres de 1 à 4 correspondent aux réactions testées avec ou sans compétiteurs (c). Chaque réaction est schématisée par l'hybridation attendue dans le pannel gauche avec BPnc264 représenté en orange, BP1118 en rouge et *fim2* en bleu.

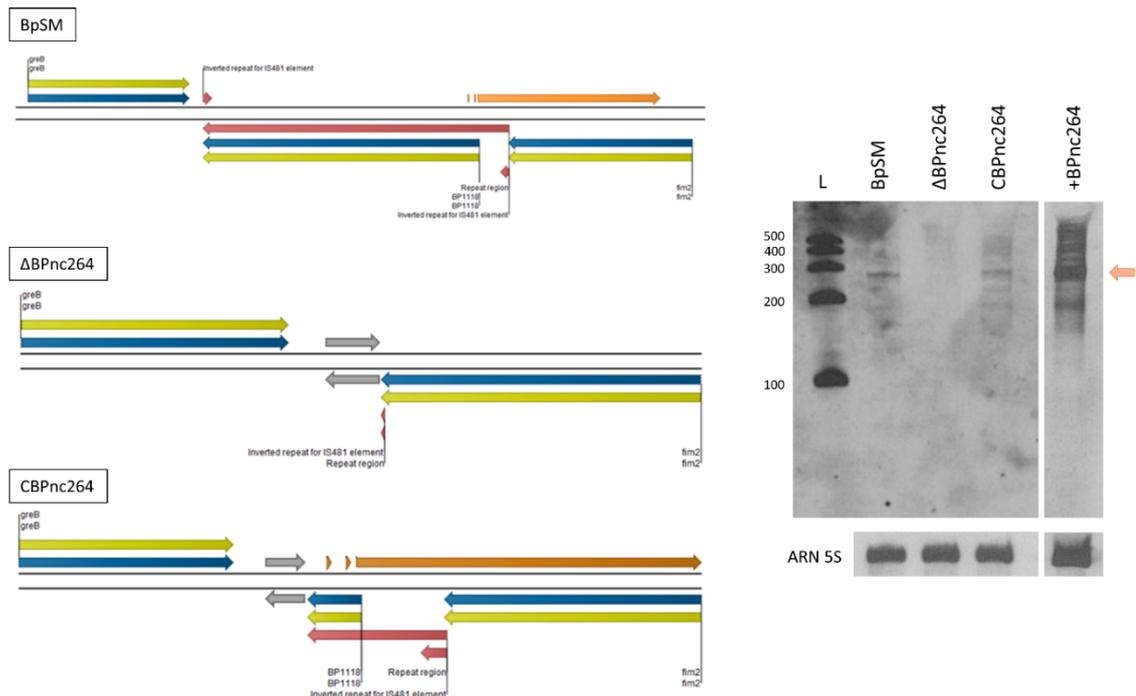
BPnc264, BP1118 et *fim2* en puits 4 et 4c. Les hybridations BPnc264/BP1118 et BPnc264/*fim2* ont été retrouvées, et la présence d'une bande supérieure indiquée par la flèche noire 4 a été observée. BPnc264 peut donc s'hybrider dans un même temps à la fois à l'ARN messager de BP1118 et à l'ARN messager de *fim2*. Ces résultats de retards sur gel restent préliminaires et doivent être renouvelés avec différentes quantité du compétiteur pour vérifier les bandes d'intérêts. Une dénaturation plus importante sera testée pour masquer le bruit de fond produit par les ultrastructures.

### 3. Régulation de l'ARN messager de *fim2* dans la cellule

BPnc264 est un ARN transcrit à partir du  $P_{out}$  de l'IS481 BP1118 et qui peut interagir avec les ARN messagers de BP1118 et de *fim2*. De par son analogie au RNA-OUT de l'IS10, on peut supposer qu'il déstabilise l'ARN messager de la transposase pour empêcher sa traduction et limiter les transpositions et les remaniements génétiques. (Case et al., 1990; Ma and Simons, 1990; Ross et al., 2013). Afin d'étudier l'interaction de BPnc264 avec le messager de *fim2*, l'expression du gène et la traduction de la protéine Fim2 ont été étudiées en absence de du transcrit antisens.

#### i. Construction des mutants de BPnc264

Pour étudier la régulation de BPnc264 sur *fim2*, des mutants de délétion et de complémentation ont été générés par double recombinaison homologue à partir de la souche BpSM (Matériel et Méthode 2). Le processus de recombinaison dans le génome nécessite deux séquences de recombinaison en amont et en aval de la région à muter. Les IS481 étant présentent en 246 copies sur le génome de la souche BpSM, une recombinaison homologue avec une région de BP1118 affecterait d'autres séquences d'insertion. Pour abolir l'expression de BPnc264, une délétion complète de BP1118 a



**Figure 39 :** Construction et validation des mutants de BPnc264.

La région génomique de BPnc264 chez les différents mutants est représentée dans le panel de gauche. Flèches rouges : Séquences d'Insertion. Flèches bleues : Gènes annotés. Flèches jaunes : Séquences codantes annotées. Flèches oranges : BPnc264 et les motifs -10 et -35 du  $P_{out}$  de BP1118. Flèches grises : Termineur Term2039 de *Bordetella bronchiseptica* RB50. Figure tirée de CLC Genomics Workbench v10. L'expression de BPnc264 chez les différents mutants a été validée par Northern Blot dans le panel de droite. L'ARN 5S a été révélé pour contrôler la quantité d'ARN engagée. L : Marqueur de taille. Flèche orange : Transcrit majoritaire de BPnc264.

donc été générée pour obtenir le mutant  $\Delta BP1118::Term2039RB50$ , appelé ici  $\Delta BPnc264$  (Figure 39). Aucun termineur de transcription n'a été prédit chez la souche BpSM pour les gènes *greB* et *fim2* par le logiciel TransTermHP (Matériel et Méthode 17) (Kingsford et al., 2007). Afin d'éviter un chevauchement de transcription entre ces deux gènes, le termineur de *greB* originel de *Bordetella bronchiseptica* RB50, appelé Term2039, a été inséré à la place du gène BP1118. Ce termineur Rho-indépendant est prédit comme une séquence forçant l'arrêt de la transcription sur les brins positif et négatif, le termineur Term2040 étant la séquence du brin opposée de Term2039.

Un mutant de complémentation a également été généré par double recombinaison homologue. La séquence de BP1118 étant complémentaire à 245 autres régions génomiques, uniquement le promoteur  $P_{out}$  de BP1118 et la séquence en aval ont été insérés à l'extrémité 3' de *fim2* pour obtenir le mutant  $\Delta BP1118::Term2039RB50-P_{out}BP1118$ , appelé pour l'étude CBPnc264 (Figure 39).

Enfin, un mutant de surexpression ectopique de BPnc264 a été généré par conjugaison, nommé  $\Delta BP1118::Term2039RB50/pBBR1MCS::P_{out}BP1118$ , appelé ici +BPnc264. Le plasmide inséré, le pBBR1MCS, contient la séquence du  $P_{out}$  de l'IS481 ainsi que la séquence codante du gène *fim2* en aval du promoteur orientée en antisens.

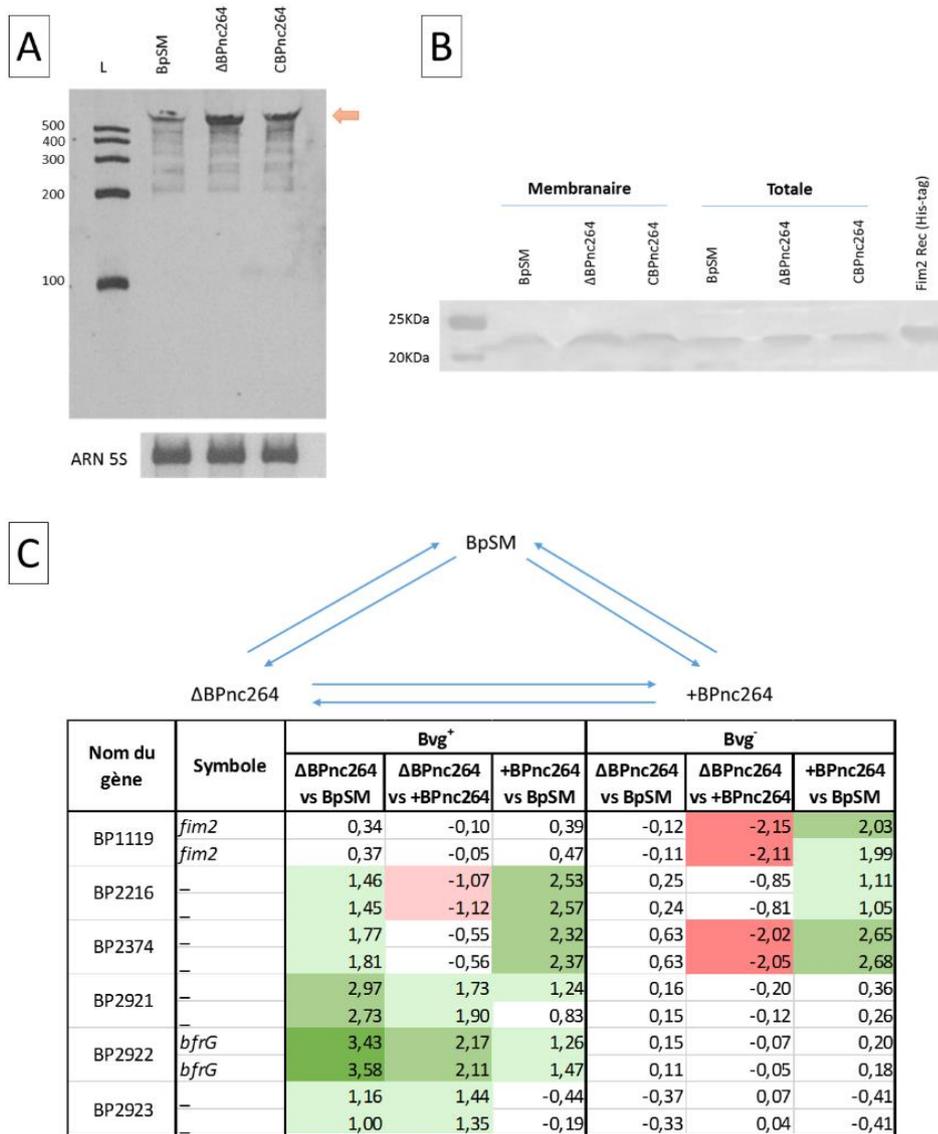
Les séquences génomiques des différents mutants ont été vérifiées par séquençage Sanger et l'expression de BPnc264 a été validée par Northern Blot (Figure 39) (Matériel et Méthode 2 et 7). Le mutant de surexpression montre également un transcrite majoritaire d'une taille de 280 nucléotides, mais également plusieurs tailles inférieures et supérieures. Ceci est probablement dû à des terminaisons de transcription précoces et tardives.

## ii. Etudes de la régulation de *fim2* en phases Bvg<sup>+</sup> et Bvg<sup>-</sup>

Les constructions des mutants ayant été validées au niveau génomique et au niveau transcriptomique, des études de l'expression de *fim2* en condition standard de culture, c'est-à-dire en phase Bvg<sup>+</sup>, ont été entreprises. Le transcrite de *fim2* semble plus exprimé par Northern Blot en phase Bvg<sup>+</sup> chez le mutant de délétion  $\Delta$ BPnc264 comparé à son expression chez la souche sauvage BpSM et le mutant complétement CBPnc264 (Figure 40A) (Matériel et Méthode 7). Cependant, la quantité d'ARN messager de *fim2* en phase Bvg<sup>+</sup> entre ces trois souches ne varie pas significativement en qRT-PCR (Matériel et Méthode 10). Un gène de référence, le gène *purH* dont l'expression est invariable, a été sélectionné pour calculer les  $\Delta\Delta$ Ct. Le  $\log_2$  entre les souches  $\Delta$ BPnc264 et BpSM calculé est de 0,43.

La production totale de protéine Fim2 entre les 3 souches a également été révélée par Western Blot (Figure 40B) (Matériel et Méthode 12). Un anticorps polyclonal a été hybridé pour détecter la forme monomérique de Fim2 à partir des extraits protéiques totaux, et à partir d'une fraction insoluble correspondant aux protéines ancrées à la membrane, isolées après ultracentrifugation. La quantité de protéine ne varie pas entre les 3 souches parmi les extraits membranaires et les extraits totaux. BPnc264 ne régule ni la stabilité de l'ARN messager de *fim2*, ni sa traduction en phase de virulence Bvg<sup>+</sup>.

L'expression de l'ensemble des gènes annotés des souches de référence BpSM, du mutant de délétion  $\Delta$ BPnc264 et du mutant de surexpression +BPnc264 a été analysée par puces à ADN en suivant un plan d'analyse croisée des lames entre les trois souches (Figure 40C) (Matériel et Méthode 11). De ce fait, un gène potentiellement affecté par l'absence de BPnc264 adopterait un profil d'expression opposé entre les mutants  $\Delta$ BPnc264 et +BPnc264. Les transcriptomes ont été analysés à partir de cellules cultivées en absence et en présence de sulfate de magnésium à 50mM pour obtenir des bactéries en phases Bvg<sup>+</sup> et en phase Bvg<sup>-</sup> respectivement. Cependant, le transcrite du gène *fim2* ne varie pas entre les trois souches dans ces deux phases. La production 4 fois plus importante de transcrits de *fim2* chez le mutant +BPnc264 par rapport aux souches BpSM et  $\Delta$ BPnc264 pourrait provenir de promoteurs endogènes du plasmide pBBR1MCS, donnant un faible taux de transcription basal de *fim2* (Figure supplémentaire S9). Les profils des autres gènes modulés en phase Bvg<sup>+</sup> ou Bvg<sup>-</sup> ne suivent pas un schéma de régulation dépendant de la mutation de BPnc264, mais pourraient être affecté par la culture. Par exemple, le gène BP2374 est plus exprimé chez  $\Delta$ BPnc264 que chez BpSM, mais n'est pas retrouvé moins exprimé chez le mutant +BPnc264.



**Figure 40** : Etude de l'expression transcriptomique et protéique de *fim2* chez les différents mutants de BPnc264.

(A) : Northern Blot de détection de l'ARN messager de *fim2* chez les souches BpSM, ΔBPnc264 et CBPnc264. L'ARN 5S a été révélé pour contrôler la quantité d'ARN engagée. L : Marqueur de taille. (B) : Western Blot de détection des monomères de Fim2 chez les souches BpSM, ΔBPnc264 et CBPnc264. Les protéines ont été extraites et séparées en une fraction soluble et insoluble, correspondant aux protéines ancrées à la membrane. Les extraits totaux ont également été révélés. La révélation d'une protéine recombinante de Fim2 (Fim2-Rec), marquée avec une histidine, a permis de détecter la bande d'intérêt. (C) : Analyses croisées des expressions transcriptomiques déterminées par puces à ADN des souches BpSM, ΔBPnc264 et +BPnc264 en condition Bvg<sup>+</sup> et Bvg<sup>-</sup>. Les valeurs sont exprimées en Log<sub>2</sub> et ont été déterminées à partir de deux sondes par lame. Le plan d'analyse croisée est schématisé au dessus du tableau, chaque flèche bleue représentant une analyse de lame. Les cases vertes montrent une expression plus importante de la souche testée par rapport à la souche de référence (Log<sub>2</sub>>1), et à une expression plus faible pour les cases rouges (Log<sub>2</sub><1).

### iii. Etudes de la régulation de *fim2* en phase Bvg<sup>i</sup>

BPnc264 n'agit donc pas sur la régulation de Fim2 ni au niveau transcriptomique ni au niveau protéique en phase Bvg<sup>+</sup> et en phase Bvg<sup>-</sup>. Ce gène de virulence est un gène de classe II, il code pour une adhésine et son taux de transcription augmente progressivement jusqu'à une expression maximale durant la phase Bvg<sup>i</sup> (Voir le chapitre 2 partie 2 de l'introduction) (Melvin et al., 2014). L'expression du gène *fim2* a donc été également déterminée lors de cette phase. Le sulfate de magnésium permet de réguler le degré d'activité du système à deux composants BvgAS, régulant également le gène *fim2* (Decker et al., 2012; Dupré et al., 2015; Scarlato and Rappuoli, 1991). Différentes phases Bvg<sup>i</sup> ont donc été obtenues en modulant la concentration finale de sulfate de magnésium dans le milieu de culture pour obtenir un profil de transition de phase de virulence (50mM pour la phase Bvg<sup>-</sup> ; 40, 30, 25, 20, 15 et 10mM pour la phase Bvg<sup>i</sup> ; 0mM pour la phase Bvg<sup>+</sup>).

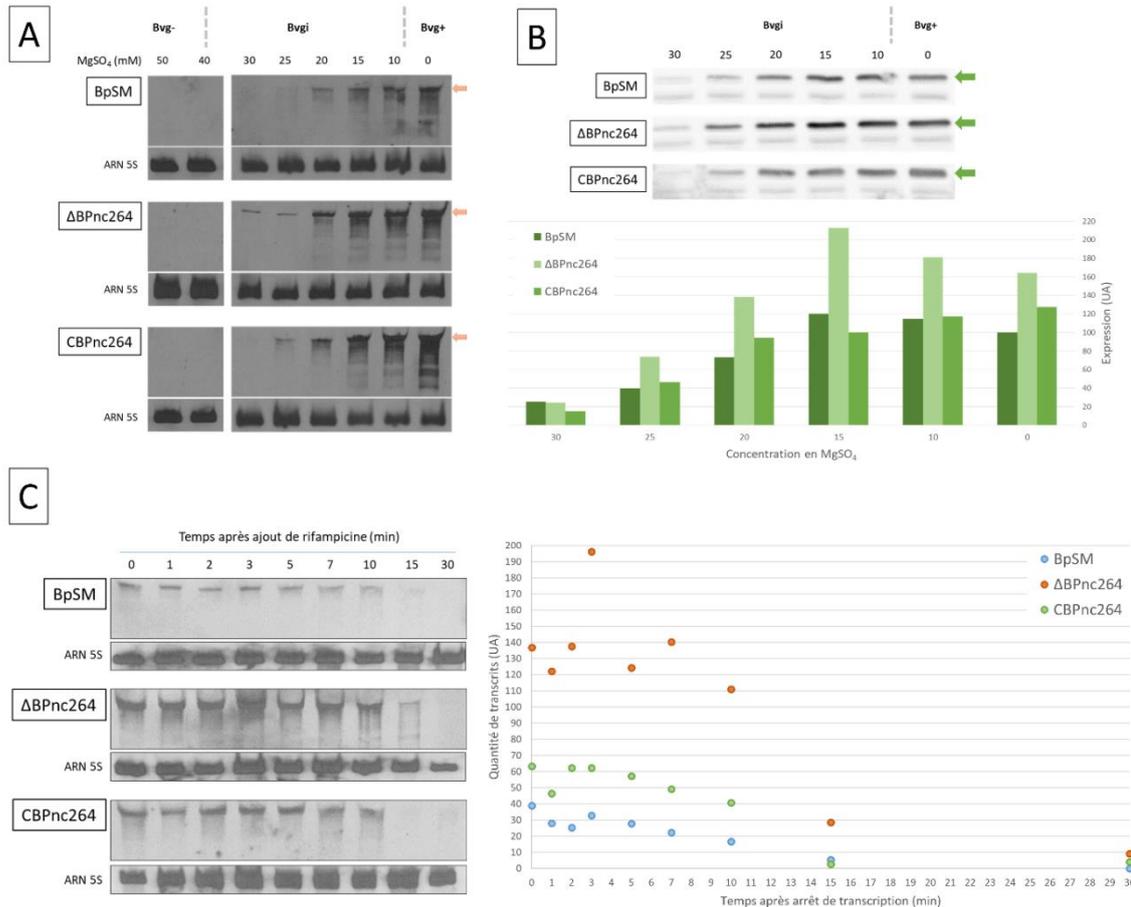
L'expression du gène *fim2* a été détectée par Northern Blot chez les souches BpSM, ΔBPnc264 et CBPnc264 cultivées en présence des différentes concentrations de sulfate de magnésium (Figure 41A) (Matériel et Méthode 7). L'ARN messager de *fim2* est détecté à une concentration de 20mM jusqu'à 0mM de sulfate de magnésium pour la souche BpSM, alors qu'il est déjà détecté à partir de 30mM pour le mutant ΔBPnc264. La détection redescend à une concentration de 25mM pour le mutant de complémentation CBPnc264. L'expression du gène *fim2* est donc altérée dans les premières phases de virulence Bvg<sup>i</sup>.

Afin de mesurer l'effet de cette différence de transcription sur la traduction de l'ARN messager de *fim2*, les trois souches ont été cultivées en présence de 30, 25, 20, 15, 10 et 0mM de sulfate de magnésium pour une évaluation en Western blot (Figure 41B) (Matériel et Méthode 12). La quantité de protéine Fim2 semble plus importante chez le mutant ΔBPnc264 à 25, 20 et 15mM de sulfate de magnésium. Pour confirmer ces observations, l'intensité des bandes a été quantifiée avec le logiciel Scion Image, utilisé pour quantifier différentes nuances de couleurs et de noir et blanc, et notamment déjà utilisé dans plusieurs domaines en biologie (Goodwin and Hsiang, 2010; Sotanaphun et al., 2009; Tolviva et al., 2006). La quantité de monomères de Fim2 est doublée chez le mutant ΔBPnc264 à 25, 20 et 15mM par rapport à la souche BpSM. L'effet de la délétion est complétement à ces concentrations chez le mutant CBPnc264.

L'augmentation de la transcription de *fim2* en phase Bvg<sup>i</sup> chez le mutant ΔBPnc264 pourrait indiquer que BPnc264 a un rôle déstabilisant sur l'ARN messager de *fim2* et provoque ainsi une dégradation du transcrite et ceci de manière importante en phase intermédiaire (Bvg<sup>i</sup>). La stabilité de ce transcrite a donc été testée sur les 3 souches cultivées en présence de 20mM de sulfate de magnésium et par ajout de rifampicine (Figure 41C) (Matériel et Méthode 16). Celui-ci a une demi-vie proche de 8min pour la souche BpSM, et elle augmente à environ 13min chez le mutant ΔBPnc264. Ces résultats sont partiellement complétés chez le mutant CBPnc264 pour lequel la demi-vie du transcrite est d'environ 11min. L'ajout uniquement de la séquence correspondant au P<sub>out</sub> et à la région non codante de BP1118 chez le mutant CBPnc264 pourrait expliquer la

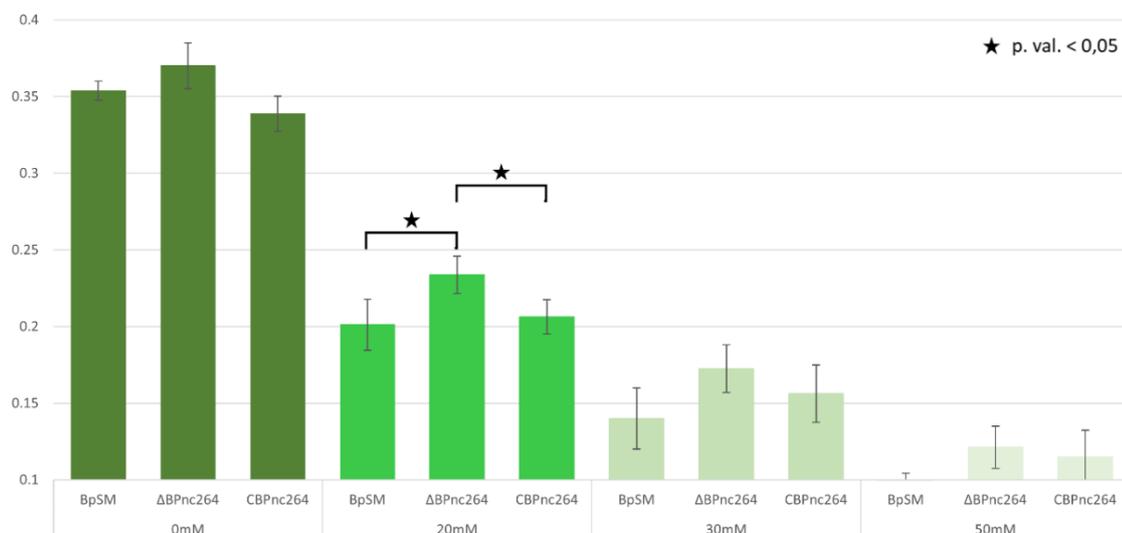
complémentation partielle, la séquence complète de BP1118 ne peut en effet pas être réinsérée pour des raisons d'homologie de séquence.

Pour déterminer l'impact de BPnc264 sur la régulation de *fim2* au sein de la cellule, des études préliminaires de formation de biofilms ont également été entreprises (Figure 42) (Matériel et Méthode 13). Les trois souches forment très peu de biofilm en présence de 50mM de sulfate de magnésium, correspondant à la phase Bvg<sup>-</sup>. Aucune différence



**Figure 41 :** Impact de BPnc264 sur la transcription et la traduction du gène *fim2* en phase Bvg<sup>I</sup>.

(A) : Northern Blot de détection du transcrite de *fim2* sur des ARN extraits de cellules cultivées avec différentes concentrations en sulfate de magnésium. La flèche orange indique l'ARN messager de *fim2*. L'ARN 5S a été révélé pour contrôler la quantité d'ARN engagé. Les phases de virulence sont représentées avec les différentes concentrations de MgSO<sub>4</sub> en mM. (B) : Western Blot de détection des monomères de Fim2 à partir d'extraits protéiques totaux sur des cultures supplémentées de différentes concentrations de sulfate de magnésium. Fim2 est indiqué par les flèches vertes. Une protéine contrôle a également été révélée pour contrôler la quantité engagée. Les différences d'intensité des bandes ont été quantifiées avec le logiciel Scion Image et représentées sous forme d'histogramme. (C) : Test de stabilité et calcul de la demi-vie de l'ARN messager de *fim2* chez les souches BpSM, ΔBPnc264 et CBPnc264 cultivées en présence de 20mM de sulfate de magnésium et de rifampicine. L'intensité des bandes a également été calculée avec le logiciel Scion Image et représenté sous forme de graphique.



**Figure 42 :** Tests de formation de biofilm des souches BpSM, ΔBPnc264 et CBPnc264 en absence et en présence de 50, 30, 20mM de sulfate de magnésium. Les écarts-type ont été déterminés à partir de 8 réplicats techniques pour chaque souche et chaque condition. Les seuils de significativités ont été calculés par un test de Mann-Whitney disponible sur le site : <http://www.socscistatistics.com/tests/mannwhitney/> Barres noires : Populations ayant des valeurs significativement différentes. Etoile : Seuil de confiance de  $p < 0,05$ .

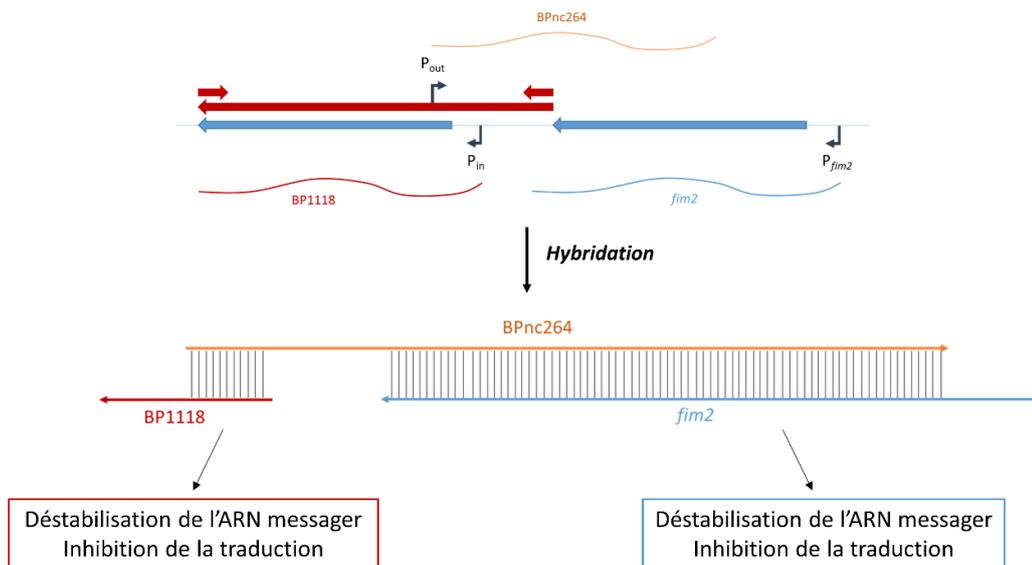
n'a pu être observée également en phase Bvg<sup>+</sup> à 0mM. Cependant, une augmentation significative de la formation de biofilms a pu être observée en présence de 20mM de sulfate de magnésium pour la souche mutante ΔBPnc264 par rapport aux souches BpSM et au mutant complémenté CBPnc264.

#### **4. Discussion : Les IS481, un réservoir de double ARN régulateurs potentiels**

Les ARN non codant orientés en antisens de gènes font partis d'une catégorie d'ARN régulateurs dont l'étude est en pleine essor (Chang et al., 2015; Georg and Hess, 2011; Lasa et al., 2012; Pinel-Marie et al., 2014; Wen et al., 2013). Ces transcrits agissent régulièrement sur une unique cible pour réguler l'activité codée par le gène à une ou plusieurs conditions favorables ou non à son expression. BPnc264 est un ARN régulateur indépendant orienté en antisens de deux gènes, BP1118 et *fim2*. L'existence d'ARN régulateurs bactériens orientés en antisens de deux gènes et capables d'agir sur ces 2 gènes n'a pas encore été répertoriée. Nous avons montré que BPnc264 est capable de s'hybrider à la fois à l'ARN messager de BP1118 et à l'ARN messager de *fim2*, ce qui en ferait donc, à notre connaissance, le premier cas du genre. De par son homologie avec le RNA-OUT produit par l'IS10 chez *Escherichia coli*, il déstabiliserait l'ARN messager du gène de la transposase porté par l'IS481 BP1118 (Case et al., 1990). Cette régulation, commune aux ARN transcrits par le P<sub>out</sub> des IS481, inhibe la traduction de la transposase et empêche un remodelage important du génome (Ellis et al., 2015; Ma and Simons, 1990). Par ailleurs nous venons de montrer que l'interaction de BPnc264 avec l'ARN messager de *fim2*, la seconde cible, entraîne une dégradation prématurée du transcrit

messenger qui affecte la quantité de protéine Fim2 traduite. La régulation de *fim2* ayant lieu en phase de transition de virulence Bvg<sup>i</sup>, la régulation par BPnc264 pourrait intervenir de manière spécifique dans les phases précoces de la colonisation du tractus respiratoire de l'hôte (Figure 43).

La fimbriae permet au pathogène de supprimer la réponse inflammatoire et d'adhérer aux cellules épithéliales, notamment pour coloniser le tractus respiratoire inférieur comme montré chez la souris (Melvin et al., 2014). La régulation de *fim2* par BPnc264 pourrait limiter la production de *pili* pour éviter l'adhésion lors de la phase de transmission Bvg<sup>i</sup> lorsque la bactérie n'est pas encore dans le tractus respiratoire inférieur. Fim2 constituant le corps du *pili*, BPnc264 pourrait entraîner un retard dans la formation complète de la fimbriae (Voir Introduction Chapitre 2 Partie 2). Cette hypothèse est renforcée par l'observation de la formation précoce de biofilms chez un mutant déficient dans la production de l'ARN régulateur (Figure 42). La faible stabilité de BPnc264 permettrait de réguler finement cette production d'antigène lors des premières phases de l'infection. BPnc264 n'ayant aucun impact sur *fim2* en phase Bvg<sup>+</sup>, la transcription abondante de *fim2* passerait au-dessus de la régulation de BPnc264, permettant la formation de biofilms et donc l'adhésion du pathogène aux cellules épithéliales pour prévenir des dégagements de mucus (Cattelan et al., 2016; Scheller et



**Régulation des réarrangements génomiques**

**Régulation des phases précoces de la colonisation ?**

**Figure 43 :** Schéma d'interaction et de régulation de BPnc264.

L'ARN régulateur BPnc264 (Orange) peut s'hybrider sur son extrémité 5' à l'ARN messenger de BP1118 (Rouge) pour réguler le remodelage du génome et sur son extrémité 3' à l'ARN messenger de *fim2* (Bleu) pour le déstabiliser et inhiber sa traduction. La régulation de *fim2* pourrait être impliquée dans les phases précoces de la colonisation du tractus respiratoire. Flèches rouges : Annotation des IS481. Flèches Bleues : Annotation des séquences codantes. Flèches noires : Promoteurs P<sub>out</sub>, P<sub>in</sub> et P<sub>fim2</sub>.

al., 2015). Cette régulation est spécifique de *Bordetella pertussis* car il n'y a pas d'homologue de BPnc264 chez *Bordetella bronchiseptica* ou *parapertussis*, laissant supposer que BPnc264 pourrait jouer un rôle, avec d'autres facteurs de virulence, dans la spécificité d'hôte du pathogène et la colonisation du tractus respiratoire de l'être humain. L'impact global de BPnc264 sur la colonisation du tractus respiratoire sera à l'avenir déterminé par des tests de formation de biofilm, des tests d'adhésion *in vitro* sur des cellules épithéliales, et également des tests de production de cellules immunitaires, de colonisation et de localisation *in vivo* du pathogène et des mutants de l'ARN régulateur lors d'infection chez le modèle murin.

# Discussion et perspectives

Les travaux détaillés dans cette thèse ont permis d'élucider certains aspects transcriptomiques de *Bordetella pertussis* par détermination du transcriptome primaire, première étude de ce cas chez ce pathogène. L'analyse des données a permis de mettre en évidence une architecture complexe propre à chaque gène et de déterminer un ensemble de gènes exprimés à l'état de virulence, déjà mis en évidence auparavant (Hot et al. 2003; Cummings et al. 2006; Moon et al. 2017; Deora et al. 2001). Les techniques de séquençage des ARN totaux combinées à des analyses de prédictions bioinformatiques ont permis de mettre en évidence 477 structures polycistroniques exprimées en phase Bvg<sup>+</sup> contre 838 structures monocistroniques ayant pour la majorité d'entre eux des caractéristiques transcriptomiques particulières. Une proportion de 0,90% d'ARN messager dépourvu de séquence non traduite a pu être identifiée, ainsi que 120 longs 5'UTR correspondant pour certains d'entre eux à des structures régulatrices comme des riboswitches ou des 5'UTR chevauchant. Plusieurs transcrits présentent également des prolongements en 3'UTR, dont certains chevauchent les transcrits en aval, reflétant une régulation potentielle de gènes ayant des fonctions opposées (Brantl 2007; Lasa, Toledo-Arana, and Gingeras 2012; Thomason and Storz 2010). De plus, plus de 900 ARN régulateurs indépendants, appelés « small RNA » en anglais, ont été prédits en phase Bvg<sup>+</sup> chez le pathogène.

La proportion de candidats ARN régulateurs et de structures régulatrices indépendantes identifiés est comparable à celles d'autres pathogènes (Sharma et al. 2010; Toledo-Arana et al. 2009). L'étude de ces candidats, de leurs fonctions et modes d'action devrait permettre d'élargir nos connaissances sur la régulation de la virulence de *Bordetella pertussis* et potentiellement d'identifier de nouveaux éléments liés à son adaptation à la pression vaccinale. Deux exemples soulevés par notre étude du transcriptome primaire concernent le chevauchement des transcrits des gènes *bvgAS* et *bvgR* (Résultats Chapitre 1 Partie 2), et la validation des prédictions d'ARN transcrits en antisens des gènes de virulence (Résultats Chapitre 2 Partie 2). L'extension de notre annotation du transcriptome primaire à différentes phases de virulence permettrait d'étoffer la liste de transcrits régulateurs potentiellement impliqués dans la régulation de la virulence de *Bordetella pertussis*. Des thermosenseurs, non identifiés dans notre étude, pourraient être prédits en comparant notre annotation actuelle avec l'annotation du transcriptome de la souche BpSM en phase Bvg<sup>-</sup> cultivée à 20°C (Dupré et al. 2015; Dupré et al. 2013). De la même manière, les ARN régulateurs impliqués dans la virulence du pathogène seraient identifiés et validés par analyse du transcriptome de la bactérie cultivée en présence de 25mM (Bvg<sup>i</sup>) ou de 50mM (Bvg<sup>-</sup>) de sulfate de magnésium et chez un mutant  $\Delta bvgA$  (Bvg<sup>-</sup>). Enfin, l'annotation du transcriptome du mutant  $\Delta risA$  nous permettrait de prédire les ARN régulateurs ayant un profil *vrg* ou *vag* dépendant de ce facteur de transcription.

L'analyse du transcriptome primaire suivi des validations biologiques ont également permis de montrer l'existence d'ARN régulateurs transcrits par les IS481 chez *Bordetella pertussis*. Ces ARN sont majoritairement transcrits à partir du P<sub>out</sub>, orienté en antisens de l'ARN messager de la transposase, et dont la transcription s'étend en antisens des gènes en aval. Leur fonction potentielle de double ARN régulateur a pu être explorée à

partir du transcrit BPnc264, régulant la production de transposase et de fimbriae lors des étapes intermédiaires de virulence. Ce système de régulation, encore non documenté, montre que les séquences d'insertions impactent sur la régulation des gènes environnant leur séquence. Le nombre important de ces ARN issus des IS et orientés en antisens de deux gènes (150 pour la souche Tohama I) sous-entend une régulation globale complexe du pathogène, certains candidats ARN régulateurs étant prédits en antisens de gènes impliqués dans le métabolisme (exemples : *metX*, *aroE*), dans la réplication et la réparation de l'ADN (exemples : *recR*, *dinP*) ou encore dans la virulence (exemple : *fim2*) (Tableau Supplémentaire S2).

L'étude de BPnc264, orienté en antisens des transcrit du gène de la transposase BP1118 et du gène de virulence *fim2*, a permis de caractériser son rôle régulateur. Cependant, l'impact biologique de ce transcrit est à compléter par une étude approfondie de l'adhésion des cellules impactées dans son expression *in vitro* et *in vivo*. L'avantage de la régulation de l'expression du gène *fim2* chez *Bordetella pertussis* pourrait également être déterminé par insertion de l'IS481 BP1118 chez *Bordetella bronchiseptica* et *Bordetella parapertussis* en aval du gène *fim2* et des tests de formation de biofilms et d'adhésion *in vitro*. Les prolongements de transcription des IS481 pourraient avoir un rôle dans l'adaptation de *Bordetella pertussis* à la pression vaccinale par régulation des gènes environnant ces îlots génétiques. Les remaniements génomiques étant fréquents chez certaines souches, la détermination de la régulation des promoteurs  $P_{in}$  et  $P_{out}$  des IS481 permettrait d'élucider l'impact de ces transcrits sur l'adaptation du pathogène (Brinig et al. 2006). Les éléments génétiques mobiles au sein des génomes bactériens sont majoritairement activés en cas de stress, les IS481 pourraient de ce fait être régulées en fonction de la topologie de l'ADN génomique (Nagy and Chandler 2004). De cette manière, il est envisageable de caractériser l'importance transcriptomique des IS481 en induisant des remaniements génétiques en condition de stress extrême (Cultures sous UV, en présence de faibles concentrations d'antibiotiques, induction d'un choc osmotique), et de caractériser d'un point de vue génomique et transcriptomique les cellules survivantes.

L'importance des ARN transcrits par les IS est encore plus significative si l'on considère leurs localisations et leurs spécificités génomiques au sein de différentes souches d'une même espèce. *Bordetella pertussis* ayant un génome contenant plus de 230 IS481, la production de ces doubles ARN régulateurs varie d'une souche à une autre de par leurs positions génomiques (exemple : le candidat ARN orienté en antisens du gène *tgt* transcrit spécifiquement chez la souche Tohama I par rapport à la souche D420 – Voir Résultats Chapitre 3 Partie 2) (Parkhill et al., 2003; Stibitz and Yang, 1999). Les variations génomiques seraient donc également en lien avec une variation d'un réseau de régulation complexe donné par les séquences d'insertion. Cette régulation assurée par les transcrits produits par les éléments génétiques mobiles pourrait également se retrouver chez d'autres agents pathogènes dont le génome comprend de nombreuses IS, affectant à la fois la plasticité de leur génome et le réseau de régulation de leur transcriptome (Siguier et al., 2014, 2015; Vandecraen et al., 2017).

# Matériel et Méthodes

## 1. Cultures des souches et suivie de croissance

### i. Souches et Plasmides

Les souches et les plasmides utilisés dans cette étude sont détaillés dans le tableau ci-dessous, avec une description des plasmides utilisés, le génotype des souches étudiées, et la source ou la référence décrivant le plasmide ou la souche :

Souches et plasmides	Description	Sources et Références
<i>B. pertussis</i>		
BpSM	Souche Tohama I dérivée résistante à la streptomycine ; Bvg <sup>+</sup> ; Sm <sup>R</sup>	Antoine and Locht 1990
D420	Souche clinique circulante; Bvg <sup>+</sup>	Boinett et al. 2015
$\Delta Hfq$	Souche BpSM délétée du gène <i>Hfq</i> ; Bvg <sup>-</sup> ; Sm <sup>R</sup>	Bibova et al., 2013
$\Delta bvgA$	Souche BpSM délétée du gène <i>bvgA</i> ; Bvg <sup>-</sup> ; Sm <sup>R</sup>	Coutte et al. 2016
$\Delta risA$	Souche BpSM délétée du gène <i>risA</i> ; Bvg <sup>+</sup> ; Sm <sup>R</sup>	Coutte et al. 2016
$\Delta BPnc264$	$\Delta BP1118::Term2039RB50$ ; Bvg <sup>+</sup> ; Sm <sup>R</sup>	Cette étude
CBPnc264	$\Delta BP1118::Term2039RB50-P_{out}BP1118$ ; Bvg <sup>+</sup> ; Sm <sup>R</sup>	Cette étude
+BPnc264	$\Delta BP1118::Term2039RB50/pBBR1MCS::P_{out}BP1118$ ; Bvg <sup>+</sup> ; Sm <sup>R</sup> Cm <sup>R</sup>	Cette étude
<i>E. coli</i>		
DH5 $\alpha$	Souche favorisant les transformations	Invitrogen
TOP10	Souche favorisant les transformations	Invitrogen
SM10	Souche de conjugaison ; RP4-2-Tc::Mu $\lambda$ pir ; Km <sup>R</sup>	Simon, Priefer, and Pühler 1983
Plasmides		
pJQ200mp18-rpL	Vecteur de conjugaison ; Gm <sup>R</sup>	Quandt and Hynes 1993
pJQ200- $\Delta BPnc264$	pJQ200mp18-rpL avec l'insert de délétion de BP1118	Cette étude
pSS1129	Vecteur de conjugaison ; Gm <sup>R</sup> Am <sup>R</sup>	Stibitz 1994
pSS1129-CBPnc264	pSS1129 comportant l'insert de complémentation génomique de BPnc264	Cette étude
pBBR1MCS	Vecteur d'expression ; Cm <sup>R</sup>	Kovach et al. 1995
pBBR-BPnc264	pBBR1MCS avec l'insert d'expression de BPnc264	Cette étude
pUCIDT-KAN- $\Delta BPnc264$	Vecteur de réception portant l'insert $\Delta BPnc264$	IDT
pUCIDT-KAN-CBPnc264	Vecteur de réception portant l'insert CBPnc264	IDT
pUCIDT-KAN-+BPnc264	Vecteur de réception portant l'insert +BPnc264	IDT
pCR2.1-TOPO	Vecteur de clonage TA ; Km <sup>R</sup> Am <sup>R</sup>	Invitrogen
pCR-BPnc264	pCR2.1-TOPO avec la séquence de BPnc264	Cette étude
pCR-BP1118	pCR2.1-TOPO avec la séquence en 5' transcrite de BP1118	Cette étude
pCR- <i>fim2</i>	pCR2.1-TOPO avec la séquence en 3' transcrite de <i>fim2</i>	Cette étude

## ii. Cultures liquides et suivi de croissance

Les souches de *Bordetella pertussis* ont été cultivées sur milieu solide Bordet Gengou (BG), supplémenté à 10% de sang de mouton défibriné, pendant 48 heures à 37°C (Bordet and Gengou, 1906). Les bactéries ont été adaptées dans 10 ml de milieu liquide Stainer-Scholte modifié, pendant 24 heures à 37°C sous agitation, avant ensemencement des cultures à une  $DO_{600nm}$  finale de 0,1 dans les mêmes conditions (Stainer and Scholte, 1970). Pour les validations par Northern Blot, les cultures ont été arrêtées en phase précoce de croissance ( $DO_{600nm} = 1,5$  à  $1,8$ ), en phase exponentielle de croissance ( $DO_{600nm} = 2$  à  $2,5$ ) ou en phase tardive de croissance ( $DO_{600nm} = 3$  à  $3,5$ ). Pour les expériences sur les ARN (Northern blot, RNA-seq, RT-PCR, microarrays), 8ml de cultures homogénéisées avec 2ml d'un mélange de phénol / éthanol (5:95, vol:vol) ont été centrifugées 8 minutes à 4°C 4000rpm et conservées à -80°C. Pour les expériences protéiques, les cultures ont directement été centrifugées et stockées dans les mêmes conditions.

Pour les modulations phénotypiques, le milieu de culture liquide Stainer Scholte a été supplémenté à 10, 15, 20, 25, 30, 40 ou 50mM en sulfate de magnésium (concentration finale) pour obtenir le pathogène en phase  $Bvg^i$  et  $Bvg^-$ . Les différents mutants étudiés ont été cultivés dans les mêmes conditions modulées ou non. La croissance de la souche BpSM et du mutant  $\Delta Hfq$  ont été suivies par mesure continue de la densité optique en utilisant le système Elocheck (Biotronix) à partir d'une culture ensemencée à une  $DO_{600nm} = 0,1$  en milieu Stainer-Scholte modifié et supplémenté avec 100  $\mu\text{g/ml}$  de streptomycine.

## 2. Clonage de plasmide et génération des mutants

Les mutants ont été générés par double recombinaison homologue. Les inserts ont été synthétisés et clonés dans le plasmide pUCIDT-KAN par IDT. Les plasmides ont été transformés par choc thermique chez des *E. coli* DH5 $\alpha$ , cultivées dans 2mL de milieu Luria-Bertani (LB) (Conda #1231) supplémenté de 50  $\mu\text{g/mL}$  de kanamycine (Sezonov et al., 2007). Les plasmides ont été extraits par MiniPrep (Thermoscientific #K0502) en suivant les recommandations du fournisseur. Les vecteurs de conjugaison et les vecteurs portant l'insert digérés avec les enzymes de restriction *HindIII* pendant 1h à 37°C suivi d'une extraction organique au phénol-chloroforme-isoamyl alcool (PCI) (25:24:1 ; vol:vol) (Fluka #77617) et d'une digestion avec l'enzyme *BamHI* pendant 1h à 30°C (Invitrogen #ER0051 #ER0501). Les inserts et vecteurs digérés ont été extraits après électrophorèse avec le kit GeneJet Gel Extraction Kit (Thermo Fisher Scientific #K0691) et ligaturés avec le kit Fast-Link™ DNA Ligation Kit (Epicentre #LK0750H). Les vecteurs ont d'abord été transformés chez des *E. coli* DH5 $\alpha$  et/ou TOP10, extraits par MiniPrep, puis transformés une nouvelle fois chez des *E. coli* SM10 par choc thermique. La présence de l'insert a été vérifiée par PCR avec les amorces du tableau 7.

Les souches  $\Delta\text{BPnc264}$  et  $\text{CBPnc264}$  ont été obtenues par double recombinaison homologue, les vecteurs de conjugaison ont été transférés par conjugaison avec les

souches SM10 possédant le plasmide sur milieu BG supplémenté de 100µM de MgCl<sub>2</sub>. Les souches de *Bordetella pertussis* mutées ont été sélectionnées par pression antibiotique. Les plasmides pSS1129 et pJQ200mp18-rpIS possédant un gène de résistance à la gentamycine et un gène de sensibilité à la streptomycine, le premier événement de recombinaison intégrant le plasmide fournit une sensibilité partielle à la streptomycine et une résistance à la gentamycine, alors que le second événement induit une résistance à la streptomycine et une sensibilité à la gentamycine. Chaque clone obtenu a été stocké dans un mélange de milieu liquide SS et de glycérol (1:1).

La délétion / insertion génomique des clones obtenus du mutant a été amplifiée par PCR sur l'ADN génomique (Matériel et Méthode 9), et vérifiée en séquençage Sanger avec les amorces ci-dessous (tableau 7). L'absence de mutations secondaires a été vérifiée par séquençage de l'ADN génomique entier des mutants ΔBPnc264 et CBPnc264 sur le séquenceur Illumina Nextseq 500. Les bibliothèques de séquençage ont été construites avec le kit Nextera XT DNA Library Preparation Kit (Illumina #FC-131-1024) et les séquences ont été alignées avec le logiciel CLC Genomics Workbench v10.

Nom de l'amorce	Sens	Séquence 5' → 3'	Taille (nt)	Tm (°C)	Qualité de l'amorce (%)
pSS1129	>	ccccgaaaagtgccacctgac	21	60.9	97
	<	cgtagatgcacatcgttgcgctc	23	60.5	93
ΔCBPnc264-vec	>	ccaccgtgctctacatggacc	21	59.8	87
	<	gcttgacttgccggcatacaa	21	58.2	95
ΔCBPnc264-rec	>	gcgattacctgtacggcaagaagc	24	60.2	96
	<	ccgaccactgactacagcacc	21	59.6	87

**Tableau 7:** Amorces de vérification de présence de l'insert dans les vecteurs de conjugaison

### 3. Extraction d'ARN et contrôle qualité

Les culots bactériens arrêtés en présence de phénol ont été lysés avec 400µg de lysozyme (Sigma Aldrich #L6876). Les ARN totaux ont été extraits avec le kit TRI Reagent (Ambion #AM9738) en suivant les recommandations du fournisseur. L'ADN génomique a été digéré en traitant 30µg d'ARN totaux à la DNase I (Sigma Aldrich #AMPD1-1KT) à température ambiante pendant 10 min. La quantité et la pureté des échantillons ont été déterminées par mesure de la densité optique au spectrophotomètre Nanodrop 2000 (Thermo Fisher), et l'intégrité des ARN a été contrôlée par électrophorèse capillaire au Bioanalyzer 2100 (Agilent Technologies). Les ARN utilisés pour les expériences suivantes ont été sélectionnés avec une valeur d'intégrité (RIN, RNA Integrity Number) supérieure à 9.

### 4. Séquençage profond des ARN totaux

Les ARN totaux des souches BpSM et D420 ont été traités avec le kit Ribo-Zero™ RNA removal Kit (Epicentre #MRZH11124) en suivant les recommandations du fournisseur pour appauvrir la préparation en ARN ribosomiques. La quantité et la qualité des ARN ont été déterminées respectivement par Nanodrop 2000 et Bioanalyzer 2100. Les bibliothèques de séquençage ont été construites à partir du kit TruSeq® Stranded Total

RNA Library Prep (Illumina #20020598) et séquencés sur le séquenceur HiSeq® 2500. Les séquences obtenues ont été alignées à l'aide du logiciel CLC Genomics Workbench v10 sur le génome de *Bordetella pertussis* Tohama I (NC002929.2).

## 5. Séquençage différentiel des ARN totaux

### i. Traitements enzymatiques et construction des librairies

Cinq microgrammes d'ARN totaux ont été traités avec 1U de 5' monophosphate exonucléase (TEX) (Epicentre #TER51020) pendant 1h à 30°C. La réaction a été stoppée avec 1µl d'EDTA 100mM pH8 et purifiée avec une extraction organique PCI. Les ARN ont été dosés au Nanodrop 2000 et une quantité égale d'ARN traités (TEX+) et non traités (TEX-) ont été incubés avec 1U d'ARN 5' polyphosphatase (Epicentre #RP8092H) pendant 1h à 37°C. La réaction a été arrêtée par une nouvelle extraction organique PCI.

Les librairies ont été construites à partir de 500ng d'ARN traités TEX+ et TEX- avec le kit Ion Total RNA-seq Kit v2 (Thermo Fisher #4475936) en suivant les recommandations du fournisseur. Les librairies ont été purifiées par deux fois avec les billes Agencourt AMPure XP (Beckman Coulter #A63880) (1:1,8 ; vol:vol). La qualité des ARN après fragmentation et la qualité des ADN complémentaires obtenus ont été déterminés par Bioanalyzer 2100. La PCR en émulsion, l'enrichissement des billes, et la réaction de séquençage ont été faites à partir des kits Ion PGM™ Hi-Q™ View OT2 Kit et Ion PGM™ Hi-Q™ View Sequencing Kit (Thermo Fisher #A29900 #A30044) au départ de 9pmol de chaque librairie. Les billes enrichies ont été séquencées sur une puce 318v2 (Thermo Fisher #4488150) sur le séquenceur Ion Torrent PGM.

### ii. Alignement des séquences et prédictions des TSS

Des triplicats biologiques ont été analysés pour la souche BpSM cultivée en présence de 50mM de sulfate de magnésium (Phase Bvg<sup>-</sup>), et quatre répliques pour la souche BpSM cultivée en condition classique (Phase Bvg<sup>+</sup>). La qualité des librairies a été contrôlée par fastqc, et les adaptateurs A et P1 ont été enlevés des séquences en utilisant le 'fastq quality trimmer' de l'outil 'fastx' avec une qualité de score Phred de 22 minimum. Les séquences de taille inférieures à 16 bases ont été supprimées. Les librairies ont été alignées sur le génome de *Bordetella pertussis* Tohama I (NC\_002929.2) avec l'outil Segemehl avec les paramétrages par défaut (Otto et al., 2014).

Chaque réplique de séquençage différentiel d'ARN a été analysé avec le logiciel TSSAR avec une taille de pic minimale de 2. L'enrichissement relatif du signal des 4 librairies TEX+ de la phase Bvg<sup>+</sup> et des 3 librairies TEX+ de la phase Bvg<sup>-</sup> ont été fusionnés en accord avec la loi de fusion des tests de probabilité de Fisher et corrigés par la méthode de Benjamini et Hochberg (Benjamini and Hochberg, 1995; Fisher, 1925). Les positions présentant un enrichissement de départ de séquence dans les librairies TEX+ comparés aux librairies TEX- avec une significativité de 0,1 ont été considérées comme TSS. Les TSS séparés de moins de 10pb ont été fusionnés et représentés par le pic le plus proéminent. Chaque TSS ainsi déterminé a été numéroté.

## 6. 5'RACE en séquençage Haut-Débit

Les départs de transcription ont été validés par 5'RACE en séquençage à Haut Débit comme décrit par Beauregard et al. 2013. Brièvement, 30µg d'ARN totaux ont été traités avec 1U de 5' monophosphate exonucléase (TEX) (Epicentre #TER51020) pendant 1h à 30°C, purifiés par une extraction organique PCI puis traités avec 1U d'ARN 5' polyphosphatase (Epicentre #RP8092H) pendant 1h à 37°C suivie d'une nouvelle extraction organique PCI. Un adaptateur ARN synthétique a été ligaturé à l'extrémité 5' des ARN par utilisation de 40U d'ARN T4 ligase 1 (BioLabs #M0437M) et en présence de 10 µM d'adaptateur pendant 1h à 37°C avant une nouvelle purification PCI. La séquence de l'adaptateur ARN est répertoriée dans le tableau ci-dessous (tableau 8). Des ADN complémentaires des l'ensemble des transcrits ont été obtenus par transcription inverse en utilisant des amorces de séquences aléatoires de 8 nucléotides (Thermo Fisher #8080127) et le kit SuperScript III (Thermo Fisher #18080044) en suivant les recommandations du fournisseur. Les ADN complémentaires ont ensuite été amplifiés par PCR avec la polymérase Herculase Fusion II (Agilent #600675) en suivant les recommandations du fournisseur. La première amorce est spécifique de la séquence de l'adaptateur ARN et la seconde amorce est spécifique de la séquence de la cible (tableau 8). Les amorces ont été sélectionnées à l'aide du logiciel Fast PCR (Kalendar et al., 2014). Les séquences A et P1 tronqué reconnues par le séquenceur PGM ont été ligaturées aux produits PCR avec le kit Ion Plus Fragment Library (Thermo Fisher #4471252) en suivant

Nom du primer / adaptateur	Sequence (5' -> 3')	Taille (nt)	Tm(°C)	Qualité du primer (%)
A (Ion Torrent)	CCATCTCATCCCTGCGTGTCTCCGACTCAG	33nt	-	-
P1T (Ion Torrent)	CCTCTCTATGGGCAGTCGGTGAT	23nt	-	-
Adaptateur ARN	ATGCGCGAATTCCTGTAGAACGAACACTAGAAG	33nt	62,4	82
<i>ptxA</i>	TGGCGGGAGGATCGTCGGC	19nt	64,1	91
<i>fim2</i>	ATGACGCAGGTGGTGTCTGGTG	21nt	61,7	93
<i>bipA</i>	GGACCAATGACCAGACAACACG	22nt	58,3	90
<i>bfr</i>	CAGACAGCGCGTAAAGCTGG	21nt	60,8	89
<i>katA</i>	GGGTCTTGTTGGTCATTGCGT	21nt	58	93
BP0840_1	GGCAGCGAGCAGAGTCTTTT	20nt	57,6	82
BP0840_2	TTGCATCAAAAATCGGCAGGAC	22nt	56,6	91
BP3322	TGCGGGGAGCGAATGCAATC	20nt	60,6	84
SAH	TCCCGCTTCATGCGACAAAGGT	22nt	61	94
SAM-α	CGCAGGAATACAGGAGCAACG	21nt	58,3	87
Glycine	GCGGTATGGGTACCTGAGCGA	22nt	60,7	93
AS-talB	CCAGCGAGAACTGTCCGAGG	21nt	59,8	94
P <sup>in</sup> IS481	CCGCTTCAGGCACACAAACT	20nt	58,3	91
BPnc180	CGAGCAGTTCGAGTTTCGCGT	21nt	59,6	96
BPnc264	CCTACGTCAAAAAGAACGGCGA	22nt	57,5	90
BPnc266	GCTGGCGCTGCTTTGAATGC	20nt	60,6	86
BPnc385	CCATTGCCTGGGCCTTGATGG	21nt	61,3	95
BPnc418	GACGACAGCGCCTGAGGGCA	20nt	64,1	86
BPnc468	TGTACAGGCTGGTCTTCAGCCA	22nt	59,8	94
BPnc576	GGGGCGAACTATAGCACAAC	21nt	56,8	83
BPnc917	CGATCCACGAAGTCAGGCACCT	22nt	60,8	97
BPnc962	TGCTAGGTGCCCGCAAGCATC	22nt	64,1	92

**Tableau 8:** Séquence des adaptateurs et des amorces utilisées en 5'RACE

les recommandations du fournisseur. Les produits PCR obtenus ont été purifiés par deux fois avec les billes Agencourt AMPure XP (Beckman Coulter #A63880) (1:1,8 ; vol:vol) et séquencés sur une puce de séquençage 314 v2 avec les kits Hi-Q du séquenceur Ion Torrent PGM (Thermo Fisher #4488144 #A29900 #A30044).

Après séquençage, les séquences des adaptateurs ont été enlevées des séquences d'intérêt avec le logiciel 'cutadapt' (Martin, 2011). Les séquences ont été alignées sur le génome de *Bordetella pertussis* Tohama I (NC\_002929.2) avec le logiciel CLC Genomics Workbench v10. Chaque cible a été validée pour un minimum de 15 séquences alignées sur la même position. Les TSS multiples séparés d'un ou deux nucléotides ont été fusionnés et représentés à la position du TSS le plus proéminent.

## **7. Northern Blot**

### **i. Sélection des sondes de détection : Fast PCR**

Les sondes biotinylées en 5' utilisées pour la détection de transcrits par Northern Blot ont été sélectionnées avec le logiciel FastPCR et sont répertoriées dans le tableau 9 (Kalendar et al., 2014).

### **ii. Protocol et « striping »**

Les ARN totaux ont été dénaturés pendant 4min à 70°C, et 5µg d'ARN totaux par puit ont été migrés sur un gel dénaturant formé de 10% acrylamide:bis-acrylamide (37,5:1 ; vol:vol) et 8M d'urée dans un tampon de TBE 0,5X. Les ARN ont été transférés sur une membrane BrightStar Plus (Thermo Fisher #AM10100) pendant 100min à 4°C à ampérage constant de 90mA, puis ancrés à cette membrane par action des UV dans un UV-Stratalinker 2400. Le tampon Northern Max (Thermo Fisher #AM8677) a été utilisé pour l'hybridation les oligonucléotides biotinylés sur l'ARN cible. Les membranes ont été révélées par chimioluminescence avec le kit BrightStar Biodetect (Thermo Fisher #AM1930). Pour la majorité des Northern Blot, la quantité d'ARN déposé par puit a été contrôlée par révélation de l'ARN 5S (sonde disponible dans la publication Hot et al. 2011). Après révélation des cibles, les sondes ont été désappariées de leur cible par incubation de la membrane dans une solution de Striping (5mM Tris pH8, 0,2mM EDTA, 0,1X Denhardt's (Thermo Fisher #750018)). Après trois lavages au TE 1X pH7,3, l'hybridation et la révélation de la sonde de l'ARN 5S ont été réitérés de la même façon.

## **8. Extraction d'ADN génomique**

L'ADN génomique des souches de *Bordetella pertussis* a été extrait à partir du kit QIAamp DNA Mini Kit (Qiagen #51304). Les culots bactériens ont été resuspendus dans 160µl de lysozyme 10mg/ml et 40µl de protéinase K avant de suivre les recommandations du fournisseur. Les membranes des colonnes d'extraction ont été éluées avec de l'eau pure certifiée sans RNase ni DNase (Thermo Fisher #10977035). La quantité et la pureté de l'ADN extrait ont été déterminées au Nanodrop 2000.

Nom de la sonde	Sens	Séquence 5' → 3'	Taille (nt)	Tm (°C)	Contenance en GC (%)
BPnc008	>	cgttcccttctgttggcgacac	23	62.4	60.9
BPnc082	>	cgcaacaagcggcggttgcg	24	69.6	70.8
BPnc180	>	cggattccactgacggcgga	21	60.9	61.9
BPnc181	>	ccagctaccgtcaaggtcgg	20	59.8	65
BPnc218	<	gccactgattgctccacca	20	59.9	60
BPnc233	>	ttgtgtcggcacggctttccgc	23	64.3	60.9
	<	cgccatgcaaaaaggcacc	20	60.7	60
BPnc264	<	gtggctcaccctagctgtg	20	56.6	60
	<	gtcaaaaagaacggcgagctcgaagccagcgccatcaccacttatgtgggttttccgtg	60	72	53.3
BPnc344	<	cggtgatacgcatacgccaag	20	58.6	60
BPnc363	<	gcgattgtactcgttgggcac	22	60.8	59.1
BPnc370	<	ggtgaacctgatagagggtccgc	24	61.2	58.3
BPnc375	<	gtcggccatgattctctct	20	56.3	55
BPnc384	>	acgcaaaccgtcgtccacact	22	63.1	59.1
	>	gcggtgaatctgtcaggtcgg	22	62	63.6
	<	cgagccttactccggcactggttc	24	63.4	62.5
BPnc392	>	tatccttgccgctcggcctgact	24	65.3	62.5
BPnc425	>	gcgaaccgaaaggcggtacagt	23	65.3	65.2
BPnc451	>	atccgcaatgggtccgtacc	20	59.4	60
	<	ttggggtacggaccattgcggtac	25	64.5	60
BPnc452	<	cgggcccccgaacgacaagggc	23	69.2	73.9
BPnc465	>	acaggcaccggcctcctggtt	23	68.5	69.6
BPnc468	>	cctgatgttggcgtgtacggg	21	60.4	61.9
BPnc469	>	cgaagtccaaggtcggcttctt	21	59.4	57.1
BPnc494	>	caagatccggctatcggcggtc	23	63.3	60.9
BPnc519	>	cgtaaggcggcgtgtacgg	20	63.2	70
BPnc560	<	atcgatgcagaaagtgggcag	21	57.6	52.4
BPnc627	<	agcgacaagcccggaggggcaaa	25	70.2	68
BPnc642	>	gttgaccgctgcgccaatc	20	61.9	65
	<	ggccgatagatcgggagagc	20	59.9	65
BPnc651	<	tcgcaaacggcggttttctt	23	67.1	65.2
BPnc656	>	cttgtctgggtcgggcaac	21	61.6	61.9
BPnc670	>	gcgctcatctctcgggtctg	23	62.4	60.9
BPnc748	>	ccatctcagccgctgtttg	20	58.8	60
BPnc807	>	ggtatccatgcttgcggccttggg	25	64.2	60
BPnc811	<	ccagccctgcttgcggaag	20	61.7	65
BPnc861	>	agaaaatcccgcgccttgcac	23	60.5	52.5
BPnc863	<	ggaagataatccaaattgcgacaggc	26	58.6	46.2
BPnc875	>	gcaagatcgactactgtcccgg	23	61.5	60.9
BPnc878	>	tttcaacgacggacaagggcg	23	63.7	60.9
BPnc923	>	acgttgtgatctggacggc	20	60.3	60
BPnc968	>	gagcaggtcgccttggat	20	60.1	60
BPnc969	>	atctggtaccggatggcctc	21	60.5	61.9

**Tableau 9:** Séquence des sondes biotinyllées utilisées pour la détection de transcrits par Northern Blot

## 9. PCR et RT-PCR

### i. Transcription inverse et amplification PCR

Les prolongements de transcription ont été validés par RT-PCR. Les ADN complémentaires ont été générés par transcription inverse à partir de 5µg d'ARN totaux traités à la DNase, une amorce orientée en antisens de la cible (voir tableau ci-dessous) et avec le kit SuperScript III (Thermo Fisher #18080044) en suivant les indications du fournisseur. Une réaction de transcription inverse a été faite sans enzyme pour vérifier l'absence d'ADN génomique pour chaque test.

N° Chapitre	Nom de l'amorce	Sens	Séquence 5' → 3'	Taille (nt)	Tm (°C)	Qualité de l'amorce (%)
CH 1	<i>ptlH</i>	>	catgcaacaccgaattgcca	21	58.9	97
		<	gctccaccaccacgacgata	20	58.7	86
	BPt51	>	gcgcccctctacttttccc	20	60.4	74
		<	gggctttctggaatctggct	20	57.3	89
	<i>bvgAS</i>	>	cctcaccgaaatcgacggc	20	60.6	95
		<	tcaggctgatcttgatcggg	21	59.2	93
	<i>bvgR</i>	>	tatctgctggcgaccatgtg	21	59.1	98
		<	ggataattggcgccatctcg	21	58.4	91
	SAM riboswitch	>	cgattgcctgatccgcttg	20	56.8	83
		<	tcagaccttttagctggat	20	52.1	80
	<i>metX</i>	>	cctgtcctgaatccggcttg	21	60.5	96
		<	cagttcgtagcattgagggt	21	56.7	95
	Glycine Riboswitch	>	ctgtcacgatgtcgcgagg	21	62.6	89
		<	gcggtatgggttacctgagcga	22	60.7	93
	<i>gcvT</i>	>	catgtccgaccgctcaaag	21	61.4	93
		<	gggtctctccagttcgagc	21	60.9	97
	TPP riboswitch	>	gatcacgcaggggtactcgt	21	60.8	97
		<	gcaactggaatacgtcccagc	22	60.7	96
	<i>thiC</i>	>	cgttgcccaaatcccgaagt	22	61.1	95
		<	gttcttctgccccgaacatg	22	60.8	94
	FMN riboswitch	>	caaagcaccctcgctcccaag	22	62.9	96
		<	gaattgcaccccctgaaga	22	62	95
	<i>ribB</i>	>	tccaacgtctcccaccct	21	62.4	92
		<	atcaggctggcttcttctcgcg	23	62.9	94
	SAH riboswitch	>	gtcgccaggctcaggaaatcct	22	61.2	83
		<	aggtgagcgccgttctcgcg	20	65.1	81
	<i>acyH</i>	>	gaccgagatcccggcctgatg	22	64.2	91
		<	ccggtctggatcgtcatgtca	22	61.1	93
	<i>yybP-ykoY</i> riboswitch	>	agtaacctctatccccgga	22	56	87
		<	gggtacgaaaccaccaagtgt	21	57.2	92
BP3410	>	cgagttctctgcaacggcgtc	22	60.2	97	
	<	ccatgaacgtgacggcgtttc	22	61.2	94	
Cobalamine riboswitch	>	accggtgctttcggcaagtaacg	22	61.9	94	
	<	cagggccttctgtttccgt	21	61.7	86	
BP3595	>	gccaaacctcccgaactcg	21	62.8	94	
	<	cagcgtatcgctcgtatcac	21	60.2	94	
ARNt-tyrosine riboswitch	>	cggaatcttccgactcagtcg	22	59.8	98	
	<	gacgcgataccaaccgacg	20	60.2	94	
<i>thrS</i>	>	taacttccggatggctcgag	22	63.4	97	
	<	ccgtcacgggaccgggatac	20	61.7	91	
CH 2	BPnc363	>	agcagtagacaatcgccccct	21	62.3	92
		<	ttggaaccgcccctgggggta	21	65.1	93
	<i>ppsA</i>	>	ttcgttcgagcagctccgatg	22	61.7	98
		<	ttgcctctaccgagtcacg	21	60.5	96
	BPnc513	>	tgtatcgcaacttcgccgt	21	60.8	94
		<	gtaagacagcgctggtcac	21	60.2	93
	BP2055	>	ccgacaagcccagtgcaag	21	63.1	93
		<	aattgacggcccgagacttc	22	61.7	95
	BPnc375	>	gcgcttttctcaaggaga	20	57.8	81
		<	gctcaccttgctgctgctc	20	61.5	85
	<i>ptsI</i>	>	tccaatgccttcgggtcaag	21	60.3	97
		<	tcgagcgtatcgggctgatg	21	60.3	96
BPnc451	>	ctggggcaagtgggggatc	20	62.9	88	
	<	gcggttttctcatttgggtacgga	26	61.2	95	
<i>ptpA</i>	>	gagggttcggcaaggtgctc	21	60.8	96	
	<	aggccgctgtattgctgg	21	64.4	89	
CH 3 et 4	<i>P<sub>out</sub></i> IS481	>	ccggttcgtagaacatgca	21	59.9	90
CH 3	BPnc247	>	cgatgcctcgcatattctc	21	60.8	90
		<	tgaacacctgcacaacctgc	21	58.5	97
	<i>smpB</i>	>	ctacgaagctggcatggtcctg	22	59.9	98
<		cgcacgatgacgtactttcc	21	58.3	84	
CH 4	BPnc264 R1	<	tgggttttccgtgctacc	21	56.9	93
	BPnc264 R2	<	ctcctactcaaaaagacggcga	24	59	92
	BPnc264 R3	<	ctgacaggtacgattccaa	21	57.4	86

**Tableau 10:** Séquence des amorces utilisées en PCR et RT-PCR

Les ARN ont été hydrolysés en fin de réaction par ajout de 1M NaOH pendant 15min à 70°C. Les ADN complémentaires, génomiques, ou plasmidiques ont été amplifiés par PCR avec le kit Herculese Fusion II (Aglient #600675) en suivant les recommandations du fournisseur. Les séquences d'intérêts ont été amplifiées en 35 cycles de PCR, et analysées par électrophorèse sur un gel d'agarose 1% en tampon TBE 0,5X.

## ii. Sélection des amorces : Fast PCR

Les amorces utilisées en PCR et en transcription inverses sont listées dans le tableau 10 avec les chapitres de résultats concernés. Chaque amorce a été sélectionnée avec le logiciel FastPCR (Kalendar et al., 2009). La spécificité de chaque amorce a été testée par Blast et validée par PCR sur l'ADN génomique des souches BpSM et/ou D420 (Altschul et al., 1990).

## 10. RT-PCR quantitative

### i. Protocol de qRT-PCR

La réaction de PCR quantitative a été faite en deux étapes. Cinq microgrammes d'ARN totaux ayant subi à une double digestion à la DNase I ont été traités à la SuperScript III (Thermo Fisher #18080044) avec une amorce spécifique des cibles (Voir tableau 11). Les ADN complémentaires ont été dilués au 1/1000 et amplifiés avec le kit Brilliant III Ultra-Fast SYBR® Green QPCR Master Mix (Agilent #600882). Les réactions de PCR ont suivi un programme de 35 cycles de dénaturation (95°C, 20 sec) et d'amplification + élongation (60°C, 30 sec) avec l'appareil Rotorgene 6000 Real-Time PCR (Corbett). Chaque ARN cible a été testé en triplicats techniques et biologiques.

Nom de l'amorce	Sens	Séquence 5' → 3'	Taille (nt)	Tm (°C)	Qualité de l'amorce (%)
BPnc264	>	gtgatggcgctggcttcgac	20	61.3	93
	<	gcttcgaccccgaagtgcag	20	60.8	93
<i>fim2</i>	>	gccaagacgcccgtgtacc	20	61.1	94
	<	atgacgcaggtggtgtcgggtg	21	61.7	93
<i>purH</i>	>	acatggatac gatcggccgctg	22	61.1	91
	<	gaaggtgcagtcgggcttgg	20	61	96

**Tableau 11:** Séquence des amorces utilisées en RT-PCR quantitative

### ii. Calculs des différences d'expression : $\Delta\Delta Ct$

Les cycles seuils (Ct) déterminés par qRT-PCR ont été utilisés pour calculer des différences d'expression du même gène entre deux conditions. La valeur du Ct de *purH*, gène de ménage ne variant dans aucune condition de culture, a été retirée de la valeur du Ct de BPnc264 ou du Ct de *fim2* pour obtenir la valeur du  $\Delta Ct$  de chaque souche. La valeur du  $\Delta Ct$  de la souche BpSM a été soustraite à la valeur du  $\Delta Ct$  des souches  $\Delta BPnc264$  et  $\Delta CBPnc264$  pour obtenir la valeur du  $\Delta\Delta Ct$ . Les données ont été transformées ensuite en  $\text{Log}_2$ .

## **11. Puces à ADN**

### **i. Marquages et hybridation**

Le phénotype des cellules cultivées en présence de sulfate de magnésium et l'expression globale des gènes des mutants ont été validés par puces à ADN à partir de lames produites par le laboratoire (Hot et al. 2003). Les ADN complémentaires ont été marqués à la cyanine 3 (Cy3) ou à la cyanine 5 (Cy5) lors de la réaction de transcription inverse. Cinq microgrammes d'ARN totaux ont été incubés 2h à 45°C avec 500 µM de dATP, dTTP et dGTP ; 300 µM de dCTP (Thermo Fisher #10297018) ; 200 µM de Cy3-dCTP ou de Cy5-dCTP Amersham (GE Healthcare #PA53022 #PA55022) ; 16U de RNasin (Promega #N2511), 300 mM d'amorces hexanucléotidiques aléatoires (Roche #11277081001) et 400U de SuperScript III (Thermo Fisher #18080044). Les ARN ont été hydrolysés avec 5µl de 1M NaOH pendant 15min à 65°C et les ADN complémentaires marqués ont été purifiés sur colonne avec le kit QIAquick PCR Purification (QIAGEN #28104).

Les puces à ADN ont été lavées dans une solution de 0.2% de SDS et bloquées pendant 15min dans une solution réductrice composée de 0.75X de PBS, 75% d'éthanol et de 65mM de borohydrure de sodium. Un microgramme d'ADN complémentaire Cy3 et d'ADN complémentaire Cy5 provenant de deux souches ou conditions de cultures différentes ont été hybridés aux sondes de la puce à ADN dans un tampon contenant 40% de formamide désionisé, du Denhardt's 5X, 50mM Tris pH 7,4, 0,1% SDS, 1mM pyrophosphate de sodium et 5X SSC pendant 16H à 52°C. Les lames ont été lavées 5min dans une solution de 2X SSC / 0,2% SDS, puis 10min dans une solution 0,5X SSC, 5min dans une solution 0,05X SSC et 2min dans une solution 0,01X SSC. Les lames ont été scannées avec l'InnoScan 700 et le logiciel Mapix v3.1 (Innopsys).

### **ii. Analyses différentielles : Logiciel R et progiciel LIMMA**

Chaque réplica biologique a été analysé à partir de deux lames ayant les marqueurs fluorescents inversés entre les deux conditions (1<sup>ère</sup> lame : Test-Cy3 / Référence-Cy5 ; 2<sup>ème</sup> lame : Test-Cy5 / Référence-Cy3). Les données ont été normalisées et déterminées à partir du progiciel LIMMA fonctionnant avec le langage statistique R v2.15.0 (Smyth et al., 2003). L'intensité des fluorochromes pouvant varier entre chaque lame, les données ont été normalisées entre les deux lames ayant les marqueurs fluorescents inversés. Les ratios obtenus en fonction des intensités moyennes ont été représentés en MA-plot afin d'appliquer une normalisation de type « loess » (Smyth and Speed, 2003). Les différences d'expression entre deux conditions ont été calculées à partir d'un t-test basée sur une estimation de Bayes des variances (Long et al., 2001). Les variations d'expression de chaque gène sont basées sur des triplicats biologiques.

## **12. Révélation des protéines par Western Blot**

### **i. Extraction des protéines totales et membranaires**

Les cultures de *Bordetella pertussis* ont été centrifugées pendant 8min à 4°C à 4000rpm immédiatement après avoir atteint une  $DO_{600nm} = 2$ . Les culots ont été resuspendus dans 3ml de 50mM Tris pH8 contenant un cocktail d'inhibiteurs de protéases « cComplete »(Sigma Aldrich #11697498001) et 10 µg/ml de DNase I. Les cellules ont été lysées par trois passages dans la pompe à pression French Press (Thermo Fisher) à 1000 Pa et centrifugées pendant 15min à 4°C à 4000rpm pour éliminer les débris cellulaires. Une partie du surnageant (Correspondant à la fraction des protéines totales) a été centrifugée pendant 1h à 4°C à 45000rpm et le culot (correspondant à la fraction des protéines membranaires) a été resuspendu dans 50 µl de 50mM Tris pH8.

## ii. Western Blot

Les protéines des fractions totales et membranaires ont été révélées par Western Blot par migration sur un gel d'électrophorèse composé d'un gel de concentration des échantillons (6% acrylamide : bis-acrylamide (29:1, vol:vol), 250mM Tris pH6,8, 0.1% SDS) et d'un gel de séparation des échantillons (12% acrylamide : bis-acrylamide (29:1, vol:vol), 250mM Tris pH8,8, 0.1% SDS). Les échantillons protéiques ont été dénaturés pendant 10min à 95°C dans un tampon de Laemmli (Sigma Aldrich #S3401) et migrés dans un tampon de SDS-Page (25mM Tris pH8,8, 200mM Glycine, 0.1% SDS). Les ARN ont été transférés à 4°C sur une membrane de nitrocellulose Hybond-C extra (GE Healthcare #10564755) dans un tampon de transfert composé de 25mM Tris pH7,6, 192mM Glycine et 10% éthanol). Les membranes ont été bloquées dans une solution de lait 10% pendant 1H puis lavées avec un mélange de 0.01% Tween20 et 1X PBS. L'anticorps primaire anti-Fim2 a été hybridé pendant 1h (Alpha Diagnostic International #FIM-25S) et l'anticorps secondaire anti-IgG de lapin (Thermo Fisher #65-6120) conjugué à une peroxydase de raifort a été hybridé pendant 1h également. Les protéines ont été révélées avec le kit ECL Prime Amersham (GE Healthcare #RPN2232) en suivant les indications du fournisseur et l'appareil Amersham Imager 680 (GE Healthcare).

## 13. Mesure de la formation de biofilms

### i. Cultures et lavages

Les souches BpSM, ΔBPnc264 et CBPnc264 ont été cultivées sur milieu BG solide pendant 48H et adaptées en 10ml de milieu liquide pendant 24H. Chaque culture a étéensemencée dans 800µl de milieu SS classique ou supplémenté de 20, 30 ou 50mM de sulfate de magnésium à  $DO_{600nm}=0,1$  en plaque Abgene 96 puits constitués de polypropylènes (Thermo Fisher #AB0564). Après 96h d'incubation à 37°C sans agitation et trois lavages à l'eau milliQ, les cellules ont été colorées avec 900µl de Crystal Violet 0,1% pendant 15min avant d'être lavées une nouvelle fois à l'eau. La plaque a été séchée pendant 3h et les cellules ont été resuspendues dans 1ml d'éthanol 95% avant de prendre les  $DO_{600nm}$ . Pour chaque condition, la densité optique d'un contrôle négatif sans ensemencement a été déterminée.

## ii. Tests de Mann-Whitney

Chaque souche et chaque condition a été testée avec 8 répliques techniques. Les différences significatives entre deux souches ont été testées par un U-test de Mann-Whitney disponible à l'adresse : <http://www.socscistatistics.com/tests/mannwhitney> (Vermeulen et al., 2015). Les différences significatives ont été calculées pour une valeur P inférieure à 0,01 et 0,05 avec une hypothèse bidirectionnelle.

## 14. Transcription *in vitro*

Les séquences en 3' de *fim2* (218 pb), en 5' de BP1118 (108 pb) et de BPnc264 (280 pb) ont été amplifiées avec les amorces listées dans le tableau 12 en aval d'un promoteur T7:

Chaque séquence a été clonée dans un plasmide pCR2.1-TOPO (Matériel et Méthode 1) avec le kit TOPO-TA cloning (Thermo Fisher #451641) en suivant les recommandations du fournisseur. Les plasmides ont été transformés par choc thermique chez des *E. coli* DH5 $\alpha$  cultivés dans 2ml de milieu LB et extraits par MiniPrep (Thermoscientific #K0502). Les séquences ont été réamplifiées par PCR à partir des plasmides extraits avec le kit Herculase Fusion II (Agilent #600675) et purifiées par extraction des bandes d'intérêt sur gel.

Les ARN ont été transcrits avec le kit Ampliscribe T7-Flash Biotin-RNA (Epicentre #ASB71110) en suivant les recommandations du fournisseur. Brièvement, 500 ng de produit PCR ont été incubés avec 10mM de DTT, 16U de RNasin, 2 $\mu$ l d'ARN polymérase T7 et 7,5mM de chaque NTP pendant 2H à 37°C. Pour la biotinylation de BPnc264, l'échantillon a été incubé avec 4,5mM d'UTP non marqué et 3mM d'UTP-biotinylé. Les fragments PCR ont été digérés par addition de 1U de DNase I pendant 15min à 37°C, et les ARN ont été migrés sur un gel dénaturant de 6% Acrylamide : Bis-acrylamide (37,5:1 ; vol:vol) contenant 8M d'urée. Les transcrits ont été découpés du gel, élués sur la nuit dans un mélange d'acétate de sodium 0,5mM pH5,2, SDS 0.1%, EDTA 1mM, phénol 0,1%, et purifiés par une extraction organique PCI.

Nom de l'amorce	Sens	Séquence 5' → 3'	Taille (nt)	Tm (°C)	Qualité de l'amorce (%)
<i>fim2</i> - start	>	taatacgactcactatagggcggtgcaggtacggatttc	39	65.4	81
<i>fim2</i> - stop	<	ctaggggtagaccacggaaa	20	55.4	89
BP1118 - start	>	taatacgactcactataggggagaaactggaaatcgccacc	41	65.1	86
BP1118 - stop	<	gctggaccatttcgagtcgacg	22	60	91
BPnc264 - start	>	taatacgactcactataggggttcatccggccggctcc	40	68.8	63
BPnc624 - stop	<	catgggtcggaacgagggcgac	21	63	88

**Tableau 11:** Séquence des amorces utilisées pour l'amplification des séquences des transcrits de *fim2*, BP1118 et BPnc264

## 15. Tests de retard sur gel

Pour chaque réaction, 5nM de sonde biotinylée de BPnc264 ont été hybridées avec 5nM de chaque cible BP1118 et/ou *fim2*. Chaque réaction a également été testée avec l'ajout de 5nM de compétiteur, correspondant au transcrit BPnc264 non biotinylé. Les

transcrits ont été dénaturés pendant 2min à 70°C dans un tampon d'hybridation contenant 10mM de Tris pH8, 50mM de chlorure de sodium, 50mM de chlorure de potassium et 10mM de chlorure de magnésium, puis hybridés progressivement par diminution de la température de 0,1°C par seconde jusqu'à 37°C, température à laquelle les réactions ont été incubées 30min. Les mélanges d'ARN ont ensuite été migrés sur un gel non dénaturant contenant 5% d'Acrylamide : Bis-Acrylamide (37,5:1 ; vol:vol) pendant 80min à 4°C à 180V dans un tampon TBE 1X. Les ARN ont ensuite été transférés sur une membrane de nylon et révélés par chimioluminescence (Matériel et Méthode 7).

## **16. Tests de stabilité des ARN**

### **i. Culture et Northern Blot**

*Bordetella pertussis* BpSM, ΔBPnc264 et CBPnc264 ont été cultivés dans 100ml de milieu SS en absence ou en présence de 20mM de sulfate de magnésium jusqu'à une  $DO_{600nm}=2$ . Les cultures ont été divisées en 10 fois 10ml et 100μg/ml de rifampicine a été ajouté au milieu. Les bactéries ont été arrêtées en présence de phénol / éthanol à 0min, 1min, 2min, 3min, 5min, 7min, 10min, 15min et 30min après ajout de l'antibiotique. Les ARN ont été extraits (Matériel et Méthode 3) et les transcrits de BPnc264 et *fim2* ont été révélés par Northern Blot (Matériel et Méthode 7).

### **ii. Calcul de la demi-vie**

L'expression des transcrits vus par Northern Blot ont été calculées en unités arbitraires à partir du logiciel Scion Image (Goodwin and Hsiang, 2010; Sotanaphun et al., 2009; Tolivia et al., 2006). Les résultats ont été représentés sous forme de graphique et la demi-vie a été calculée en prenant la valeur d'intensité de bande initiale ( $t=0$ ) divisée par deux et rapportée sur l'axe des abscisses par la courbe. Le test de stabilité en condition Bvg<sup>+</sup> représente des duplicatas de cultures biologiques.

## **17. Analyses bioinformatiques**

### **i. Prédiction des terminateurs de transcription : TransTermHP**

Les prédictions de terminateurs Rho-indépendants ont été téléchargées sur le site TransTermHP à l'adresse : <http://transterm.cbcb.umd.edu/index.php> (Kingsford et al., 2007). Les prédictions présentent un score de confiance de structure en tête d'épingle supérieur ou égal à 30 et ont été mises à jour le 7 septembre 2006.

### **ii. Correction du bruit de fond de séquençage et prédiction de l'architecture du transcriptome primaire**

Depuis l'apparition des séquençages orientés, il est estimé que 0,5 à 10% des séquences issues de séquençage de l'ARN ne seraient pas alignées sur le bon brin (Levin et al., 2010). Ces alignements erronés peuvent biaiser l'annotation *de novo* des transcrits orientés en antisens. Pour corriger cette fraction de séquences erronées, le

ratio de la couverture du génome entre les deux brins a été calculé sur des fenêtres de 250 bases. Le brin avec la couverture la plus faible est divisé par la somme des couvertures du brin positif et négatif. La probabilité de chaque ratio de couverture des brins peut être estimée si la distribution minimale et maximale suit une loi de Poisson. Le ratio calculé serait alors distribué de façon ubiquitaire sur le génome. Les séquences erronées seraient ainsi détectées par une augmentation significative des ratios. La correction à apporter a été fixée à la valeur maximale observée sur le génome. La correction à apporter a été quantifiée à 1.4% à partir des résultats de séquençage à ARN différentiel.

L'architecture du transcriptome primaire a été déduite à partir de la couverture du génome prenant en compte les alignements de séquences erronés. Les informations sur les positions de TSS et des annotations *in silico* des terminateurs Rho-indépendants ont été incluses. La couverture du génome dans une fenêtre de 3 bases a été transformée en  $\log_2$  pour y appliquer deux valeurs seuils. Une valeur supérieure ou égale à 4 permet de démarrer une transcription, s'étendant dans les deux directions. Pour déterminer la fin de la transcription, une deuxième valeur seuil de 2,25 est appliquée. Une valeur descendant en dessous de ce seuil permet de terminer la transcription de la prédiction. Au niveau des TSS et des terminateurs de transcription, la valeur seuil a été fixée à 3,125 pour favoriser un début ou une fin de transcription sans l'imposer. L'annotation *de novo* a été fusionnée pour les prédictions localisées au sein d'un même gène annoté. De cette manière, les ARN mono- et polycistroniques, des longs 5' et 3'UTR et des transcrits indépendants ont été déterminés à partir des TSS, des terminateurs de transcription Rho-indépendant, des résultats de séquençage et des annotations des séquences codantes.

### iii. Prédiction de structures opéroniques : DOOR

Les prédictions des structures opéroniques par DOOR 2.0 ont été téléchargées dans la banque de donnée en ligne à l'adresse : <http://csbl.bmb.uga.edu/DOOR/> (Mao et al., 2014).

### iv. Prédiction des motifs de promoteurs : MEME

Les motifs -10 et -35 des promoteurs ont été prédits à partir du logiciel MEME disponible à l'adresse : <http://meme-suite.org/tools/meme> (Bailey and Elkan, 1994). Les séquences en amont des TSS ont été utilisées pour déterminer les motifs des promoteurs. Chaque motif est associé à une valeur de probabilité E (« E-value ») et au nombre de fois où il est retrouvé en amont des départs de transcription.

### v. Prédiction de riboswitches et des ARN connus : Infernal

Les structures ARN déjà connues ont été retrouvées par le logiciel Infernal (El Korbi et al., 2014; Nawrocki, 2014; Nawrocki and Eddy, 2013). Le logiciel recherche des familles d'ARN et des structures déjà répertoriées par homologie de séquence et prédiction de structures. Les analyses du génome de la souche Tohama I ont été téléchargées et sont retrouvées dans la banque de données Rfam à l'adresse :

[http://rfam.xfam.org/search?q=UP000002676%20AND%20entry\\_type:%22Family%22](http://rfam.xfam.org/search?q=UP000002676%20AND%20entry_type:%22Family%22)  
2

**vi. Prédictions bioinformatiques de candidats ARN régulateurs : NAPP, SIPHT, RNAz**

Le programme RNAz permet de détecter des structures ARN stables par comparaison des génomes d'espèces proches (Gruber et al., 2010). Pour notre étude, 18 génomes de Burkholdériales ont été analysés (*Bordetella pertussis* Tohama I, *Bordetella petrii* DSM12804, *Bordetella pertussis* CS, *Bordetella parapertussis* Bpp5, *Bordetella pertussis* 18323, *Bordetella parapertussis* 12822, *Bordetella bronchiseptica* RB50, *Bordetella bronchiseptica* MO149, *Bordetella bronchiseptica* 253, *Taylorella equigenitalis* MCE9, *Taylorella equigenitalis* ATCC 35865, *Taylorella asinigenitalis* MCE3, *Pusillimonas* T77, *Advenella kashmirensis* WT001, *Achromobacter xylosoxidans* NBRC 15126, *Achromobacter xylosoxidans* A8, *Herbaspirillum seropedicae* SmR1, *Janthinobacterium* Marseille). Les génomes ont été alignés de façon multiple avec Clustal W v2 à partir de la séquence annotée de l'ARN 16S pour construire un arbre phylogénétique guide (Larkin et al., 2007). Cet arbre guide a été utilisé avec l'outil Multi Z pour construire l'alignement multiple des 18 génomes (Blanchette et al., 2004). Ces alignements ont été entrés dans le pipeline RNAz, découpés en blocs de 120 bases recouvrant entre 3 et 7 espèces, et analysés avec le modèle de composition dinucléotidique. Les séquences ayant une valeur P supérieur à 0,9 ont été gardées dans la liste de candidats ARN régulateurs.

Les programmes NAPP et SIPHT disposent d'une banque de données, comme TransTerm HP, dont les génomes déposés au format GeneBank ont déjà été analysés (Livny, 2012; Ott et al., 2012). Les prédictions issues du logiciel NAPP et du logiciel SIPHT ont été téléchargées et ajoutées à la liste de candidats ARN régulateurs. Les prédictions des trois logiciels bioinformatiques ont été assemblées en une annotation *in silico* cumulée de candidats ARN régulateurs.

**vii. Prédiction de la distribution des ARN chez les Burkholdériales : RNAlieN et Infernal**

Les familles de transcrits ont été prédits avec le logiciel RNAlieN, disponible à l'adresse : <http://rna.tbi.univie.ac.at/rnalieN/> (Eggenhofer et al., 2016). Un modèle de covariance pour chaque candidat ARN a été prédit par cet outil, reflétant la conservation de la séquence et de la structure. Des homologues ont été recherchés ensuite avec le logiciel Infernal parmi le genre Burkholdériales, et ont été comptés pour déterminer la disparité génétique du candidat chez ces espèces (Nawrocki and Eddy, 2013)

**viii. Prédiction des structures des ARN : RNAfold**

Les structures secondaires des ARN et leur stabilité ont été prédites par le logiciel RNAfold disponible à l'adresse : <http://rna.tbi.univie.ac.at/> (Hofacker, 2003). Les séquences prédites des candidats ARN régulateurs ont été analysées avec les paramètres par défaut du logiciel. Les structures exigeant un minimum d'énergie libre

et les structures exigeant un équilibre thermodynamique ont été prédites, incluant les appariements faibles GU (Rare mais existant chez les ARN procaryotes) et évitant l'apparition de bases isolées.

**ix. Prédiction de la fonction des gènes : String et Blast**

La fonction des gènes codant pour une protéine hypothétique a été prédite par Blast-p avec le logiciel Blast (Altschul et al., 1990). Les gènes homologues rapportés chez d'autres espèces procaryotes ont été vérifiés à l'aide du logiciel String, prédisant les réseaux d'interaction entre les protéines (Szklarczyk et al., 2015).

**x. Distribution des IS481**

La distribution des IS481 chez 21 souches séquencées et annotées de *Bordetella pertussis* a été déterminée à partir de la séquence en 5' de ces éléments génétiques mobiles. Les séquences démarrant du +1 de transcription du P<sub>out</sub> jusqu'à 200 nucléotides en aval de la région répétée gauche de l'IS481 (302 nucléotides) ont été extraites et alignées par Blast-n entre-elles (Altschul et al., 1990). Les séquences uniques de chaque souche ont ainsi pu être répertoriées.

**xi. Détermination des disparités génomiques entre les souches de *Bordetella pertussis* : Mauve**

Les disparités génomiques entre les souches de *Bordetella pertussis* a été déterminé avec le logiciel Mauve pour les souches Tohama I (NC\_002929.2), CS (NC\_017223), D420 (NZ\_LN849008) et 18323 (NC\_018518) (Darling et al., 2004). Les séquences ont été alignées avec les paramètres par défaut en alignement progressif entre la souche Tohama I et les autres.

**xii. Prédiction des séquences codantes : ORFfinder et RBS calculator**

Les prédictions de séquences codantes de BPnc264 ont été déterminées avec le logiciel ORFfinder disponible à l'adresse : <https://www.ncbi.nlm.nih.gov/orffinder/> (Rombel et al., 2002). L'analyse a été paramétrée pour les génomes bactériens, et dont le codon d'initiation peut-être le codon « ATG » et d'autres codons alternatifs (Ex : GTG, TTG, ...). Les séquences protéiques ont ensuite été analysées une par une avec le logiciel RBS calculator, prenant en compte la séquence en amont du codon d'initiation de la traduction et les 9 derniers nucléotides de la séquence de l'ARN 16S (Salis, 2011). Le logiciel est disponible à l'adresse : <https://www.denovodna.com/software/>.

## Bibliographie

- Abriata, L.A., Albanesi, D., Dal Peraro, M., De Mendoza, D., 2017. Signal Sensing and Transduction by Histidine Kinases as Unveiled through Studies on a Temperature Sensor. *Acc. Chem. Res.* 50, 1359–1366.
- Akerley, B.J., Miller, J.F., 1993. Flagellin gene transcription in *Bordetella bronchiseptica* is regulated by the BvgAS virulence control system. *J. Bacteriol.* 175, 3468–79.
- Akerley, B.J., Monack, D.M., Falkow, S., Miller, J.F., 1992. The bvgAS locus negatively controls motility and synthesis of flagella in *Bordetella bronchiseptica*. *J. Bacteriol.* 174, 980–90.
- Althouse, B.M., Scarpino, S. V., 2015. Asymptomatic transmission and the resurgence of *Bordetella pertussis*. *BMC Med.* 13, 146.
- Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic local alignment search tool. *J. Mol. Biol.* 215, 403–10.
- Amman, F., D'Halluin, A., Antoine, R., Huot, L., Bibova, I., Keidel, K., Slupek, S., Bouquet, P., Coutte, L., Caboche, S., Loch, C., Vecerek, B., Hot, D., 2018. Primary transcriptome analysis reveals importance of IS elements for the shaping of the transcriptional landscape of *Bordetella pertussis*. *RNA Biol.* 6286, 1–9.
- Amman, F., Wolfinger, M.T., Lorenz, R., Hofacker, I.L., Stadler, P.F., Findeiß, S., 2014. TSSAR: TSS annotation regime for dRNA-seq data. *BMC Bioinformatics* 15, 89.
- Antoine, R., Loch, C., 1990. Roles of the disulfide bond and the carboxy-terminal region of the S1 subunit in the assembly and biosynthesis of pertussis toxin. *Infect. Immun.* 58, 1518–26.
- Bailey, T.L., Elkan, C., 1994. Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proceedings. Int. Conf. Intell. Syst. Mol. Biol.* 2, 28–36.
- Beall, B., 1998. '1 o iron-regulated putative ferric siderophore receptor genes in *Bordetella bronchiseptica* and *Bordetella pertussis*. *Res. Microbiol* 149, 189–201.
- Beauregard, A., Smith, E., Petrone, B., Singh, N., Karch, C., McDonough, K., Wade, J., 2013. Identification and characterization of small RNAs in *Yersinia pestis*. *RNA Biol.* 10, 397–405.
- Benjamini, Y., Hochberg, Y., 1995. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J. R. Stat. Soc. Ser. B* 57, 289–300.
- Bibova, I., Hot, D., Keidel, K., Amman, F., Slupek, S., Cerny, O., Gross, R., Vecerek, B., 2015. Transcriptional profiling of *Bordetella pertussis* reveals requirement of RNA chaperone Hfq for Type III secretion system functionality. *RNA Biol.* 12, 175–85.
- Bibova, I., Skopova, K., Masin, J., Cerny, O., Hot, D., Sebo, P., Vecereka, B., 2013. The RNA chaperone Hfq is required for virulence of *bordetella pertussis*. *Infect. Immun.* 81, 4081–4090.
- Bisgard, K.M., Pascual, F.B., Ehresmann, K.R., Miller, C.A., Cianfrini, C., Jennings, C.E., Rebmann, C.A., Gabel, J., Schauer, S.L., Lett, S.M., 2004. Infant pertussis: Who was the source? *Pediatr. Infect. Dis. J.* 23, 985–989.
- Blanchette, M., Kent, W.J., Riemer, C., Elnitski, L., Smith, A.F.A., Roskin, K.M., Baertsch, R., Rosenbloom, K., Clawson, H., Green, E.D., Haussler, D., Miller, W., 2004. Aligning multiple

- genomic sequences with the threaded blockset aligner. *Genome Res.* 14, 708–715.
- Boinett, C.J., Harris, S.R., Langridge, G.C., Trainor, E.A., Merkel, T.J., Parkhill, J., 2015. Complete Genome Sequence of *Bordetella pertussis* D420. *Genome Announc.* 3, e00657-e15.
- Bordet, J., Gengou, O., 1906. Le Microbe de la Coqueluche. *Les Ann. l'Institut Pasteur* 20, 731–741.
- Boucher, P.E., Stibitz, S., 1995. Synergistic binding of RNA polymerase and BvgA phosphate to the pertussis toxin promoter of *Bordetella pertussis*. *J. Bacteriol.* 177, 6486–91.
- Boulanger, A., Moon, K., Decker, K.B., Chen, Q., Knipling, L., Stibitz, S., Hinton, D.M., 2015. *Bordetella pertussis* fim3 gene regulation by BvgA: phosphorylation controls the formation of inactive vs. active transcription complexes. *Proc. Natl. Acad. Sci. U. S. A.* 112, E526-35.
- Bouthier de la Tour, C., Armengaud, J., Dulermo, R., Blanchard, L., Devigne, A., de Groot, A., Ludanyi, M., Sommer, S., 2015. The abundant and essential HU proteins in *Deinococcus deserti* and *Deinococcus radiodurans* are translated from leaderless mRNA. *Microbiology* 161, 2410–2422.
- Brantl, S., 2007. Regulatory mechanisms employed by cis-encoded antisense RNAs. *Curr. Opin. Microbiol.* 10, 102–9.
- Brantl, S., Jahn, N., 2015. SRNAs in bacterial type I and type III toxin-antitoxin systems. *FEMS Microbiol. Rev.*
- Brennan, R.G., Link, T.M., 2007. Hfq structure, function and ligand binding. *Curr. Opin. Microbiol.* 10, 125–133.
- Brinig, M.M., Cummings, C.A., Sanden, G.N., Stefanelli, P., Lawrence, A., Relman, D.A., 2006. Significant gene order and expression differences in *Bordetella pertussis* despite limited gene content variation. *J. Bacteriol.* 188, 2375–82.
- Brock, J.E., Pourshahian, S., Giliberti, J., Limbach, P.A., Janssen, G.R., 2008. Ribosomes bind leaderless mRNA in *Escherichia coli* through recognition of their 5'-terminal AUG. *RNA* 14, 2159–2169.
- Burns, E.H., Norman, J.M., Hatcher, M.D., Bemis, D.A., 1993. Fimbriae and determination of host species specificity of *Bordetella bronchiseptica*. *J. Clin. Microbiol.* 31, 1838–44.
- Byrd, M.S., Mason, E., Henderson, M.W., Scheller, E. V, Cotter, P.A., 2013. An improved recombination-based in vivo expression technology-like reporter system reveals differential *cyaA* gene activation in *Bordetella* species. *Infect. Immun.* 81, 1295–305.
- Case, C.C., Roels, S.M., González, J.E., Simons, E.L., Simons, R.W., 1988. Analysis of the promoters and transcripts involved in IS10 anti-sense RNA control. *Gene* 72, 219–36.
- Case, C.C., Simons, E.L., Simons, R.W., 1990. The IS10 transposase mRNA is destabilized during antisense RNA control. *EMBO J.* 9, 1259–66.
- Caselli, A., Paoli, P., Santi, A., Mugnaioni, C., Toti, A., Camici, G., Cirri, P., 2016. Low molecular weight protein tyrosine phosphatase: Multifaceted functions of an evolutionarily conserved enzyme. *Biochim. Biophys. Acta - Proteins Proteomics* 1864, 1339–1355.
- Cattelan, N., Dubey, P., Arnal, L., Yantorno, O.M., Deora, R., 2016. *Bordetella* biofilms: a lifestyle leading to persistent infections. *Pathog. Dis.* 74, ftv108.
- Chain, P.S.G., Hu, P., Malfatti, S.A., Radnedge, L., Larimer, F., Vergez, L.M., Worsham, P., Chu,

- M.C., Andersen, G.L., 2006. Complete genome sequence of *Yersinia pestis* strains Antiqua and Nepal516: evidence of gene reduction in an emerging pathogen. *J. Bacteriol.* 188, 4453–63.
- Chandler, M., Fayet, O., Rousseau, P., Ton Hoang, B., Duval-Valentin, G., 2015. Copy-out–Paste-in Transposition of IS911: A Major Transposition Pathway. *Microbiol. Spectr.* 3.
- Chang, H., Replogle, J.M., Vather, N., Tsao-Wu, M., Mistry, R., Liu, J.M., 2015. A cis-regulatory antisense RNA represses translation in *Vibrio cholerae* through extensive complementarity and proximity to the target locus. *RNA Biol.* 12, 136–48.
- Chao, Y., Papenfort, K., Reinhardt, R., Sharma, C.M., Vogel, J., 2012. An atlas of Hfq-bound transcripts reveals 3' UTRs as a genomic reservoir of regulatory small RNAs. *EMBO J.* 31, 4005–19.
- Chen, Q., Decker, K.B., Boucher, P.E., Hinton, D., Stibitz, S., 2010. Novel architectural features of *Bordetella pertussis* fimbrial subunit promoters and their activation by the global virulence regulator BvgA. *Mol. Microbiol.* 77, 1326–1340.
- Chen, Q., Ng, V., Warfel, J.M., Merkel, T.J., Stibitz, S., 2017. Activation of Bvg-Repressed Genes in *Bordetella pertussis* by RisA Requires Cross Talk from Noncooperonic Histidine Kinase RisK. *J. Bacteriol.* 199, e00475-17.
- Cherry, J.D., 1996. Historical review of pertussis and the classical vaccine. *J. Infect. Dis.* 174 Suppl 3, S259-63.
- Ciampi, M.S., Schmid, M.B., Roth, J.R., 1982. Transposon Tn10 provides a promoter for transcription of adjacent sequences. *Proc. Natl. Acad. Sci. U. S. A.* 79, 5016–20.
- Condon, C., Putzer, H., Grunberg-Manago, M., 1996. Processing of the leader mRNA plays a major role in the induction of thrS expression following threonine starvation in *Bacillus subtilis*. *Proc. Natl. Acad. Sci. U. S. A.* 93, 6992–7.
- Cookson, B.T., Tyler, A.N., Goldman, W.E., 1989. Primary structure of the peptidoglycan-derived tracheal cytotoxin of *Bordetella pertussis*. *Biochemistry* 28, 1744–9.
- Cornia, P.B., Hersh, A.L., Lipsky, B.A., Newman, T.B., Gonzales, R., 2010. Does this coughing adolescent or adult patient have pertussis? *JAMA - J. Am. Med. Assoc.* 304, 890–896.
- Coutte, L., Huot, L., Antoine, R., Slupek, S., Merkel, T.J., Chen, Q., Stibitz, S., Hot, D., Loch, C., 2016. The multifaceted RisA regulon of *Bordetella pertussis*. *Sci. Rep.* 6, 32774.
- Croinin, T.O., Grippe, V.K., Merkel, T.J., 2005. Activation of the vrg6 Promoter of *Bordetella pertussis* by RisA. *J. Bacteriol.* 187, 1648–1658.
- Crowcroft, N.S., Stein, C., Duclos, P., Birmingham, M., 2003. How best to estimate the global burden of pertussis? *Lancet. Infect. Dis.* 3, 413–8.
- Cummings, C.A., Brinig, M.M., Lepp, P.W., Van De Pas, S., Relman, D.A., 2004. *Bordetella* Species Are Distinguished by Patterns of Substantial Gene Loss and Host Adaptation. *J. Bacteriol.* 186, 1484–1492.
- Dambach, M., Sandoval, M., Updegrove, T.B., Anantharaman, V., Aravind, L., Waters, L.S., Storz, G., 2015. The Ubiquitous yybP-ykoY Riboswitch Is a Manganese-Responsive Regulatory Element. *Mol. Cell* 57, 1099–1109.
- Darling, A.C.E., Mau, B., Blattner, F.R., Perna, N.T., 2004. Mauve: Multiple alignment of conserved genomic sequence with rearrangements. *Genome Res.* 14, 1394–1403.

- de Gouw, D., Diavatopoulos, D.A., Bootsma, H.J., Hermans, P.W.M., Mooi, F.R., 2011. Pertussis: a matter of immune modulation. *FEMS Microbiol. Rev.* 35, 441–74.
- De Lay, N.R., Garsin, D.A., 2016. The unmasking of “junk” RNA reveals novel sRNAs: from processed RNA fragments to marooned riboswitches. *Curr. Opin. Microbiol.* 30, 16–21.
- Decker, K.B., James, T.D., Stibitz, S., Hinton, D.M., 2012. The *Bordetella pertussis* model of exquisite gene control by the global transcription factor BvgA. *Microbiology* 158, 1665–76.
- Deora, R., 2002. Differential regulation of the *Bordetella bipA* gene: distinct roles for different BvgA binding sites. *J. Bacteriol.* 184, 6942–51.
- Deora, R., Bootsma, H.J., Miller, J.F., Cotter, P.A., 2001. Diversity in the *Bordetella* virulence regulon: transcriptional control of a Bvg-intermediate phase gene. *Mol. Microbiol.* 40, 669–83.
- Depardieu, F., Podglajen, I., Leclercq, R., Collatz, E., Courvalin, P., 2007. Modes and modulations of antibiotic resistance gene expression. *Clin. Microbiol. Rev.* 20, 79–114.
- DeShazer, D., Wood, G.E., Friedman, R.L., 1994. Molecular characterization of catalase from *Bordetella pertussis*: identification of the *katA* promoter in an upstream insertion sequence. *Mol. Microbiol.* 14, 123–30.
- Diavatopoulos, D.A., Cummings, C.A., Schouls, L.M., Brinig, M.M., Relman, D.A., Mooi, F.R., 2005. *Bordetella pertussis*, the causative agent of whooping cough, evolved from a distinct, human-associated lineage of *B. bronchiseptica*. *PLoS Pathog.* 1, e45.
- Diavatopoulos, D.A., Cummings, C.A., van der Heide, H.G.J., van Gent, M., Liew, S., Relman, D.A., Mooi, F.R., 2006. Characterization of a highly conserved island in the otherwise divergent *Bordetella holmesii* and *Bordetella pertussis* genomes. *J. Bacteriol.* 188, 8385–94.
- Dupré, E., Herrou, J., Lensink, M.F., Wintjens, R., Vagin, A., Lebedev, A., Crosson, S., Villeret, V., Locht, C., Antoine, R., Jacob-Dubuisson, F., 2015. Virulence regulation with Venus flytrap domains: structure and function of the periplasmic moiety of the sensor-kinase BvgS. *PLoS Pathog.* 11, e1004700.
- Dupré, E., Wohlkonig, A., Herrou, J., Locht, C., Jacob-Dubuisson, F., Antoine, R., 2013. Characterization of the PAS domain in the sensor-kinase BvgS: mechanical role in signal transmission. *BMC Microbiol.* 13, 172.
- Eggenhofer, F., Hofacker, I.L., Höner zu Siederdisen, C., 2016. RNAlie – Unsupervised RNA family model construction. *Nucleic Acids Res.* 44, 8433–8441.
- El Korbi, A., Ouellet, J., Naghdi, M.R., Perreault, J., 2014. Finding instances of riboswitches and ribozymes by homology search of structured RNA with Infernal. *Methods Mol. Biol.* 1103, 113–26.
- Ellis, M.J., Trussler, R.S., Haniford, D.B., 2015. Hfq binds directly to the ribosome-binding site of IS10 transposase mRNA to inhibit translation. *Mol. Microbiol.* 96, 633–50.
- Fernandez, R.C., 2012. Airborne Transmission of *Bordetella pertussis* Demonstrated in a Baboon Model of Whooping Cough. *J. Infect. Dis.* 206, 808–810.
- Fisher, R., 1925. Statistical methods for research workers, Biological monographs and manuals.
- Fry, N.K., Duncan, J., Malnick, H., Warner, M., Smith, A.J., Jackson, M.S., Ayoub, A., 2005.

- Bordetella petrii clinical isolate. *Emerg. Infect. Dis.* 11, 1131–3.
- Gangarosa, E.J., Galazka, A.M., Wolfe, C.R., Phillips, L.M., Gangarosa, R.E., Miller, E., Chen, R.T., 1998. Impact of anti-vaccine movements on pertussis control: the untold story. *Lancet* (London, England) 351, 356–61.
- Georg, J., Hess, W.R., 2011. cis-Antisense RNA, Another Level of Gene Regulation in Bacteria. *Microbiol. Mol. Biol. Rev.* 75, 286–300.
- Gerlach, G., von Wintzingerode, F., Middendorf, B., Gross, R., 2001. Evolutionary trends in the genus *Bordetella*. *Microbes Infect.* 3, 61–72.
- Giliberti, J., O'Donnell, S., Van Etten, W.J., Janssen, G.R., 2012. A 5'-terminal phosphate is required for stable ternary complex formation and translation of leaderless mRNA in *Escherichia coli*. *RNA* 18, 508–518.
- Goldman, W.E., Cookson, B.T., 1988. Structure and functions of the *Bordetella* tracheal cytotoxin. *Tokai J. Exp. Clin. Med.* 13 Suppl, 187–91.
- Goodwin, P.H., Hsiang, T., 2010. Quantification of fungal infection of leaves with digital images and Scion Image software. *Methods Mol. Biol.* 638, 125–135.
- Gorringe, A.R., Vaughan, T.E., 2014. *Bordetella pertussis* fimbriae (Fim): relevance for vaccines. *Expert Rev. Vaccines* 13, 1205–14.
- Grossman, A.D., Straus, D.B., Walter, W.A., Gross, C.A., 1987. Sigma 32 synthesis can regulate the synthesis of heat shock proteins in *Escherichia coli*. *Genes Dev.* 1, 179–84.
- Gruber, A.R., Findeiß, S., Washietl, S., Hofacker, I.L., Stadler, P.F., 2010. RNAz 2.0: improved noncoding RNA detection. *Pac. Symp. Biocomput.* 69–79.
- Grundy, F.J., Henkin, T.M., 1993. tRNA as a positive regulator of transcription antitermination in *B. subtilis*. *Cell* 74, 475–82.
- Guiso, N., 2006. Histoire de la coqueluche. *Med. Ther. Pediatr.* 9, 135–137.
- Guo, M.S., Updegrove, T.B., Gogol, E.B., Shabalina, S.A., Gross, C.A., Storz, G., 2014. MicL, a new E-dependent sRNA, combats envelope stress by repressing synthesis of Lpp, the major outer membrane lipoprotein. *Genes Dev.* 28, 1620–1634.
- Han, H.-J., Kuwae, A., Abe, A., Arakawa, Y., Kamachi, K., 2011. Differential expression of type III effector BteA protein due to IS481 insertion in *Bordetella pertussis*. *PLoS One* 6, e17797.
- Heikkinen, E., Kallonen, T., Saarinen, L., Sara, R., King, A.J., Mooi, F.R., Soini, J.T., Mertsola, J., He, Q., 2007. Comparative genomics of *Bordetella pertussis* reveals progressive gene loss in Finnish strains. *PLoS One* 2, e904.
- Higgs, R., Higgins, S.C., Ross, P.J., Mills, K.H.G., 2012. Immunity to the respiratory pathogen *Bordetella pertussis*. *Mucosal Immunol.* 5, 485–500.
- Hiramatsu, Y., Saito, M., Otsuka, N., Suzuki, E., Watanabe, M., Shibayama, K., Kamachi, K., 2016. BipA Is Associated with Preventing Autoagglutination and Promoting Biofilm Formation in *Bordetella holmesii*. *PLoS One* 11, e0159999.
- Hofacker, I.L., 2003. Vienna RNA secondary structure server. *Nucleic Acids Res.* 31, 3429–31.
- Hoffman, C., Eby, J., Gray, M., Heath Damron, F., Melvin, J., Cotter, P., Hewlett, E., 2017. *Bordetella* adenylate cyclase toxin interacts with filamentous haemagglutinin to inhibit biofilm formation *in vitro*. *Mol. Microbiol.* 103, 214–228.

- Honarmand Ebrahimi, K., Hagedoorn, P.L., Hagen, W.R., 2015. Unity in the biochemistry of the iron-storage proteins ferritin and bacterioferritin. *Chem. Rev.*
- Hoo, R., Lam, J.H., Huot, L., Pant, A., Li, R., Hot, D., Alonso, S., 2014. Evidence for a role of the polysaccharide capsule transport proteins in pertussis pathogenesis. *PLoS One* 9.
- Hot, D., Antoine, R., Renauld-Mongénie, G., Caro, V., Hennuy, B., Levillain, E., Huot, L., Wittmann, G., Poncet, D., Jacob-Dubuisson, F., Guyard, C., Rimlinger, F., Aujame, L., Godfroid, E., Guiso, N., Quentin-Millet, M.-J., Lemoine, Y., Loch, C., 2003. Differential modulation of *Bordetella pertussis* virulence genes as evidenced by DNA microarray analysis. *Mol. Genet. Genomics* 269, 475–86.
- Hot, D., Slupek, S., Wulbrecht, B., D'Hondt, A., Hubans, C., Antoine, R., Loch, C., Lemoine, Y., 2011. Detection of small RNAs in *Bordetella pertussis* and identification of a novel repeated genetic element. *BMC Genomics* 12, 207.
- Howson, C.P., Howe, C.J., Fineberg, H. V, 1991. Adverse Effects of Pertussis and Rubella Vaccines 382, 0–309.
- Humayun, M.Z., Zhang, Z., Butcher, A.M., Moshayedi, A., Saier, M.H., 2017. Hopping into a hot seat: Role of DNA structural features on IS5-mediated gene activation and inactivation under stress. *PLoS One* 12, e0180156.
- Inatsuka, C.S., Xu, Q., Vujkovic-Cvijin, I., Wong, S., Stibitz, S., Miller, J.F., Cotter, P.A., 2010. Pertactin is required for *Bordetella* species to resist neutrophil-mediated clearance. *Infect. Immun.* 78, 2901–9.
- Inouye, M., 1988. Antisense RNA: its functions and applications in gene regulation--a review. *Gene* 72, 25–34.
- Jellen-Ritter, A.S., Kern, W. V, 2001. Enhanced expression of the multidrug efflux pumps AcrAB and AcrEF associated with insertion element transposition in *Escherichia coli* mutants selected with a fluoroquinolone. *Antimicrob. Agents Chemother.* 45, 1467–1472.
- Jungnitz, H., West, N.P., Walker, M.J., Chhatwal, G.S., Guzmán, C.A., 1998. A second two-component regulatory system of *Bordetella bronchiseptica* required for bacterial resistance to oxidative stress, production of acid phosphatase, and in vivo persistence. *Infect. Immun.* 66, 4640–50.
- Kalendar, R., Lee, D., Schulman, A.H., 2009. FastPCR software for PCR primer and probe design and repeat search. *Genes Genomes and Genomics* 3, 1–14.
- Kalendar, R., Lee, D., Schulman, A.H., 2014. FastPCR software for PCR, in silico PCR, and oligonucleotide assembly and analysis. *Methods Mol. Biol.* 1116, 271–302.
- Kamruzzaman, M., Patterson, J.D., Shoma, S., Ginn, A.N., Partridge, S.R., Iredell, J.R., 2015. Relative Strengths of Promoters Provided by Common Mobile Genetic Elements Associated with Resistance Gene Expression in Gram-Negative Bacteria. *Antimicrob. Agents Chemother.* 59, 5088–91.
- Karimova, G., Ullmann, A., 1997. Characterization of DNA binding sites for the BvgA protein of *Bordetella pertussis*. *J. Bacteriol.* 179, 3790–2.
- Kendrick, P., Eldering, G., 1936. Progress Report on Pertussis Immunization. *Am. J. Public Health Nations. Health* 26, 8–12.
- Kendrick, P.L., 1942. Use of Alum-Treated Pertussis Vaccine, and of Alum-Precipitated Combined Pertussis Vaccine and Diphtheria Toxoid, for Active Immunization. *Am. J. Public*

- Health Nations. *Health* 32, 615–26.
- Khelef, N., Bachelet, C.M., Vargaftig, B.B., Guiso, N., 1994. Characterization of murine lung inflammation after infection with parental *Bordetella pertussis* and mutants deficient in adhesins or toxins. *Infect. Immun.* 62, 2893–900.
- Kim, S., Hwang, H., Kim, K.-P., Yoon, H., Kang, D.-H., Ryu, S., 2015. Hfq plays important roles in virulence and stress adaptation in *Cronobacter sakazakii* ATCC 29544. *Infect. Immun.* 83, 2089–98.
- Kimelman, A., Levy, A., Sberro, H., Kidron, S., Leavitt, A., Amitai, G., Yoder-Himes, D.R., Wurtzel, O., Zhu, Y., Rubin, E.M., Sorek, R., 2012. A vast collection of microbial genes that are toxic to bacteria. *Genome Res.* 22, 802–9.
- Kingsford, C.L., Ayanbule, K., Salzberg, S.L., 2007. Rapid, accurate, computational discovery of Rho-independent transcription terminators illuminates their relationship to DNA uptake. *Genome Biol.* 8, R22.
- Kinney, S.M., Boucher, P.E., Stibitz, S., Carbonetti, N.H., 1999. Analysis of BvgA activation of the pertactin gene promoter in *Bordetella pertussis*. *J. Bacteriol.* 181, 5234–41.
- Kittle, J.D., Simons, R.W., Lee, J., Kleckner, N., 1989. Insertion sequence IS10 anti-sense pairing initiates by an interaction between the 5' end of the target RNA and a loop in the anti-sense RNA. *J. Mol. Biol.* 210, 561–72.
- Klein, N.P., 2014. Licensed pertussis vaccines in the United States. History and current state. *Hum. Vaccin. Immunother.* 10, 2684–90.
- Kline, J.M., Lewis, W.D., Smith, E.A., Tracy, L.R., Moerschel, S.K., 2013. Pertussis: a reemerging infection. *Am. Fam. Physician* 88, 507–14.
- Kopf, M., Klähn, S., Scholz, I., Hess, W.R., Voß, B., 2015. Variations in the non-coding transcriptome as a driver of inter-strain divergence and physiological adaptation in bacteria. *Sci. Rep.* 5, 9560.
- Korkmaz, G., Holm, M., Wiens, T., Sanyal, S., 2014. Comprehensive Analysis of Stop Codon Usage in Bacteria and Its Correlation with Release Factor Abundance. *J. Biol. Chem.* 289, 30334–30342.
- Kovach, M.E., Elzer, P.H., Steven Hill, D., Robertson, G.T., Farris, M.A., Roop, R.M., Peterson, K.M., 1995. Four new derivatives of the broad-host-range cloning vector pBBR1MCS, carrying different antibiotic-resistance cassettes. *Gene* 166, 175–176.
- Kozak, M., 1991. A short leader sequence impairs the fidelity of initiation by eukaryotic ribosomes. *Gene Expr.* 1, 111–115.
- Kroger, C., Dillon, S.C., Cameron, A.D.S., Papenfort, K., Sivasankaran, S.K., Hokamp, K., Chao, Y., Sittka, A., Hebrard, M., Handler, 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., Hinton, J.C.D., 2012. The transcriptional landscape and small RNAs of *Salmonella enterica* serovar Typhimurium. *Proc. Natl. Acad. Sci.* 109, E1277–E1286.
- Kuchar, E., Karlikowska-Skwarnik, M., Han, S., Nitsch-Osuch, A., 2016. Pertussis: History of the Disease and Current Prevention Failure. In: *Advances in Experimental Medicine and Biology*. pp. 77–82.
- Kulenkampff, M., Schwartzman, J.S., Wilson, J., 1974. Neurological complications of pertussis inoculation. *Arch. Dis. Child.* 49.

- Lalaouna, D., Carrier, M.-C., Semsey, S., Brouard, J.-S., Wang, J., Wade, J.T., Massé, E., 2015. A 3' external transcribed spacer in a tRNA transcript acts as a sponge for small RNAs to prevent transcriptional noise. *Mol. Cell* 58, 393–405.
- Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., Mcgettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R., Thompson, J.D., Gibson, T.J., Higgins, D.G., 2007. Clustal W and Clustal X version 2.0. *Bioinformatics* 23, 2947–2948.
- Lasa, I., Toledo-Arana, A., Gingeras, T.R., 2012. An effort to make sense of antisense transcription in bacteria. *RNA Biol.*
- Latasa, P., García-Comas, L., Gil de Miguel, A., Barranco, M.D., Rodero, I., Sanz, J.C., Ordobás, M., Arce, A., Garrido-Esteba, M., 2018. Effectiveness of acellular pertussis vaccine and evolution of pertussis incidence in the community of Madrid from 1998 to 2015. *Vaccine*.
- Lesne, E., Dupré, E., Lensink, M.F., Locht, C., Antoine, R., Jacob-Dubuisson, F., 2018. Coiled-Coil Antagonism Regulates Activity of Venus Flytrap-Domain-Containing Sensor Kinases of the BvgS Family. *MBio* 9, e02052-17.
- Lesne, E., Krammer, E.-M., Dupre, E., Locht, C., Lensink, M.F., Antoine, R., Jacob-Dubuisson, F., 2016. Balance between Coiled-Coil Stability and Dynamics Regulates Activity of BvgS Sensor Kinase in *Bordetella*. *MBio* 7, e02089-15.
- Levin, J.Z., Yassour, M., Adiconis, X., Nusbaum, C., Thompson, D.A., Friedman, N., Gnirke, A., Regev, A., 2010. Comprehensive comparative analysis of strand-specific RNA sequencing methods. *Nat. Methods* 7, 709–715.
- Linz, B., Ivanov, Y. V., Preston, A., Brinkac, L., Parkhill, J., Kim, M., Harris, S.R., Goodfield, L.L., Fry, N.K., Gorrynge, A.R., Nicholson, T.L., Register, K.B., Losada, L., Harvill, E.T., 2016. Acquisition and loss of virulence-associated factors during genome evolution and speciation in three clades of *Bordetella* species. *BMC Genomics* 17.
- Livny, J., 2012. Bioinformatic Discovery of Bacterial Regulatory RNAs Using SIPHT. In: *Bacterial Regulatory RNA*. Humana Press, Totowa, NJ, pp. 3–14.
- Locht, C., 2016. Pertussis: Where did we go wrong and what can we do about it? *J. Infect.* 72 Suppl, S34-40.
- Locht, C., 2017. Will we have new pertussis vaccines? *Vaccine*.
- Locht, C., Mielcarek, N., 2014. Live attenuated vaccines against pertussis. *Expert Rev. Vaccines* 13, 1147–1158.
- Locht, C., Papin, J.F., Lecher, S., Debrie, A.-S., Thalen, M., Solovay, K., Rubin, K., Mielcarek, N., 2017. Live Attenuated Pertussis Vaccine BPZE1 Protects Baboons Against *Bordetella pertussis* Disease and Infection. *J. Infect. Dis.* 216, 117–124.
- Long, A.D., Mangalam, H.J., Chan, B.Y.P., Toller, L., Hatfield, G.W., Baldi, P., 2001. Improved statistical inference from DNA microarray data using analysis of variance and a Bayesian statistical framework. Analysis of global gene expression in *Escherichia coli* K12. *J. Biol. Chem.* 276, 19937–19944.
- Ma, C., Simons, R.W., 1990. The IS10 antisense RNA blocks ribosome binding at the transposase translation initiation site. *EMBO J.* 9, 1267–74.
- Mahon, B.P., Ryan, M.S., Griffin, F., Mills, K.H., 1996. Interleukin-12 is produced by macrophages in response to live or killed *Bordetella pertussis* and enhances the efficacy of an acellular pertussis vaccine by promoting induction of Th1 cells. *Infect. Immun.* 64,

5295–301.

- Malys, N., McCarthy, J.E.G., 2011. Translation initiation: Variations in the mechanism can be anticipated. *Cell. Mol. Life Sci.*
- Mann, P.B., Wolfe, D., Latz, E., Golenbock, D., Preston, A., Harvill, E.T., 2005. Comparative toll-like receptor 4-mediated innate host defense to *Bordetella* infection. *Infect. Immun.* 73, 8144–52.
- Mao, X., Ma, Q., Zhou, C., Chen, X., Zhang, H., Yang, J., Mao, F., Lai, W., Xu, Y., 2014. DOOR 2.0: presenting operons and their functions through dynamic and integrated views. *Nucleic Acids Res.* 42, D654-9.
- Marcus, A.-F. (1753-1816). A. du texte, 1821. *Traité de la coqueluche ou Bronchite épidémique, son diagnostic, sa nature et son traitement, par le docteur Adalbert Frédéric Marcus, à Bamberg et à Leipsig, en 1816 ; traduit de l'allemand par E.-L. Jacques,...*
- Martin, M., 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* 17, 10.
- Martin, S.W., Pawloski, L., Williams, M., Weening, K., DeBolt, C., Qin, X., Reynolds, L., Kenyon, C., Giambrone, G., Kudish, K., Miller, L., Selvage, D., Lee, A., Skoff, T.H., Kamiya, H., Cassidy, P.K., Tondella, M.L., Clark, T.A., 2015. Pertactin-negative *Bordetella pertussis* strains: evidence for a possible selective advantage. *Clin. Infect. Dis.* 60, 223–7.
- Melvin, J.A., Scheller, E. V., Miller, J.F., Cotter, P.A., 2014. *Bordetella pertussis* pathogenesis: current and future challenges. *Nat. Rev. Microbiol.* 12, 274–88.
- Menozi, F.D., Mutombo, R., Renauld, G., Gantiez, C., Hannah, J.H., Leininger, E., Brennan, M.J., Loch, C., 1994. Heparin-inhibitable lectin activity of the filamentous hemagglutinin adhesin of *Bordetella pertussis*. *Infect. Immun.* 62, 769–78.
- Merkel, T.J., Boucher, P.E., Stibitz, S., Grippe, V.K., 2003. Analysis of *bvgR* expression in *Bordetella pertussis*. *J. Bacteriol.* 185, 6902–12.
- Mielcarek, N., Debie, A.-S., Raze, D., Bertout, J., Rouanet, C., Younes, A. Ben, Creusy, C., Engle, J., Goldman, W.E., Loch, C., 2006. Live Attenuated *B. pertussis* as a Single-Dose Nasal Vaccine against Whooping Cough. *PLoS Pathog.* 2, e65.
- Miller, D., Madge, N., Diamond, J., Wadsworth, J., Ross, E., 1993. Pertussis immunisation and serious acute neurological illnesses in children. *BMJ* 307, 1171–6.
- Miralles, F., 2010. Compositional Properties and Thermal Adaptation of SRP-RNA in Bacteria and Archaea. *J. Mol. Evol.* 70, 181–189.
- Miyakoshi, M., Chao, Y., Vogel, J., 2015a. Regulatory small RNAs from the 3' regions of bacterial mRNAs. *Curr. Opin. Microbiol.* 24, 132–139.
- Miyakoshi, M., Chao, Y., Vogel, J., 2015b. Cross talk between ABC transporter mRNAs via a target mRNA-derived sponge of the GcvB small RNA. *EMBO J.* 34, 1478–92.
- Moll, I., Grill, S., Gualerzi, C.O., Bläsi, U., 2002. Leaderless mRNAs in bacteria: surprises in ribosomal recruitment and translational control. *Mol. Microbiol.* 43, 239–46.
- Moon, K., Bonocora, R.P., Kim, D.D., Chen, Q., Wade, J.T., Stibitz, S., Hinton, D.M., 2017. The BvgAS Regulon of *Bordetella pertussis*. *MBio* 8.
- Moreno López, M., 1990. [The genus *Bordetella*. By Manuel Moreno López, 1952]. *Enferm. Infecc. Microbiol. Clin.* 8, 480–5.

- Musumeci, L., Bongiorno, C., Tautz, L., Edwards, R.A., Osterman, A., Perego, M., Mustelin, T., Bottini, N., 2005. Low-molecular-weight protein tyrosine phosphatases of *Bacillus subtilis*. *J. Bacteriol.* 187, 4945–4956.
- Nagy, Z., Chandler, M., 2004. Regulation of transposition in bacteria. *Res. Microbiol.* 155, 387–398.
- Narlikar, L., 2014. Multiple novel promoter-architectures revealed by decoding the hidden heterogeneity within the genome. *Nucleic Acids Res.* 42, 12388–403.
- Nawrocki, E.P., 2014. Annotating functional RNAs in genomes using Infernal. *Methods Mol. Biol.* 1097, 163–97.
- Nawrocki, E.P., Eddy, S.R., 2013. Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics* 29, 2933–2935.
- Nicosia, A., Rappuoli, R., 1987. Promoter of the pertussis toxin operon and production of pertussis toxin. *J. Bacteriol.* 169, 2843–6.
- O'Donnell, S.M., Janssen, G.R., 2001. The initiation codon affects ribosome binding and translational efficiency in *Escherichia coli* of *cl* mRNA with or without the 5' untranslated leader. *J. Bacteriol.* 183, 1277–1283.
- Oliva, G., Sahr, T., Buchrieser, C., 2015. Small RNAs, 5' UTR elements and RNA-binding proteins in intracellular bacteria: impact on metabolism and virulence. *FEMS Microbiol. Rev.* 39, 331–49.
- Otsuka, N., Guiso, N., Bouchez, V., 2017. The length of poly(C) stretch in the *Bordetella pertussis* Pfm3 promoter determines the *vag* or *vrg* function of the *fim3* gene. *Microbiology*.
- Otsuka, N., Han, H.-J., Toyozumi-Ajisaka, H., Nakamura, Y., Arakawa, Y., Shibayama, K., Kamachi, K., 2012. Prevalence and Genetic Characterization of Pertactin-Deficient *Bordetella pertussis* in Japan. *PLoS One* 7, e31985.
- Ott, A., Idali, A., Marchais, A., Gautheret, D., 2012. NAPP: the Nucleic Acid Phylogenetic Profile Database. *Nucleic Acids Res.* 40, D205–D209.
- Otto, C., Stadler, P.F., Hoffmann, S., 2014. Lacking alignments? The next-generation sequencing mapper segemehl revisited. *Bioinformatics* 30, 1837–1843.
- Paddock, C.D., Sanden, G.N., Cherry, J.D., Gal, A.A., Langston, C., Tatti, K.M., Wu, K., Goldsmith, C.S., Greer, P.W., Montague, J.L., Eliason, M.T., Holman, R.C., Guarner, J., Shieh, W., Zaki, S.R., 2008. Pathology and Pathogenesis of Fatal *Bordetella pertussis* Infection in Infants. *Clin. Infect. Dis.* 47, 328–338.
- Papenfort, K., Förstner, K.U., Cong, J.-P., Sharma, C.M., Bassler, B.L., 2015. Differential RNA-seq of *Vibrio cholerae* identifies the *VqmR* small RNA as a regulator of biofilm formation. *Proc. Natl. Acad. Sci. U. S. A.* 112, E766–75.
- Park, J., Zhang, Y., Buboltz, A.M., Zhang, X., Schuster, S.C., Ahuja, U., Liu, M., Miller, J.F., Sebahia, M., Bentley, S.D., Parkhill, J., Harvill, E.T., 2012. Comparative genomics of the classical *Bordetella* subspecies: the evolution and exchange of virulence-associated diversity amongst closely related pathogens. *BMC Genomics* 13, 545.
- Park, J., Zhang, Y., Chen, C., Dudley, E.G., Harvill, E.T., 2015. Diversity of secretion systems associated with virulence characteristics of the classical *bordetellae*. *Microbiology* 161, 2328–40.

- Parkhill, J., Sebahia, M., Preston, A., Murphy, L.D., Thomson, N., Harris, D.E., Holden, M.T.G., Churcher, C.M., Bentley, S.D., Mungall, K.L., Cerdeño-Tárraga, A.M., Temple, L., James, K., Harris, B., Quail, M.A., Achtman, M., Atkin, R., Baker, S., Basham, D., Bason, N., Cherevach, I., Chillingworth, T., Collins, M., Cronin, A., Davis, P., Doggett, J., Feltwell, T., Goble, A., Hamlin, N., Hauser, H., Holroyd, S., Jagels, K., Leather, S., Moule, S., Norberczak, H., O'Neil, S., Ormond, D., Price, C., Rabbinowitsch, E., Rutter, S., Sanders, M., Saunders, D., Seeger, K., Sharp, S., Simmonds, M., Skelton, J., Squares, R., Squares, S., Stevens, K., Unwin, L., Whitehead, S., Barrell, B.G., Maskell, D.J., 2003. Comparative analysis of the genome sequences of *Bordetella pertussis*, *Bordetella parapertussis* and *Bordetella bronchiseptica*. *Nat. Genet.* 35, 32–40.
- Parkhill, J., Wren, B.W., Thomson, N.R., Titball, R.W., Holden, M.T.G., Prentice, M.B., Sebahia, M., James, K.D., Churcher, C., Mungall, K.L., Baker, S., Basham, D., Bentley, S.D., Brooks, K., Cerdeño-Tárraga, A.M., Chillingworth, T., Cronin, A., Davies, R.M., Davis, P., Dougan, G., Feltwell, T., Hamlin, N., Holroyd, S., Jagels, K., Karlyshev, A. V., Leather, S., Moule, S., Oyston, P.C.F., Quail, M., Rutherford, K., Simmonds, M., Skelton, J., Stevens, K., Whitehead, S., Barrell, B.G., 2001. Genome sequence of *Yersinia pestis*, the causative agent of plague. *Nature* 413, 523–527.
- Pedrolli, D., Langer, S., Hobl, B., Schwarz, J., Hashimoto, M., Mack, M., 2015. The *ribB* FMN riboswitch from *Escherichia coli* operates at the transcriptional and translational level and regulates riboflavin biosynthesis. *FEBS J.* 282, 3230–3242.
- Peterson, J.M., Phillips, G.J., 2008. Characterization of conserved bases in 4.5S RNA of *Escherichia coli* by construction of new F' factors. *J. Bacteriol.* 190, 7709–7718.
- Pinel-Marie, M.-L., Brielle, R., Felden, B., 2014. Dual toxic-peptide-coding *Staphylococcus aureus* RNA under antisense regulation targets host cells and bacterial rivals unequally. *Cell Rep.* 7, 424–35.
- Polaski, J.T., Holmstrom, E.D., Nesbitt, D.J., Batey, R.T., 2016. Mechanistic Insights into Cofactor-Dependent Coupling of RNA Folding and mRNA Transcription/Translation by a Cobalamin Riboswitch. *Cell Rep.* 15, 1100–1110.
- Poolman, J.T., 2014. Shortcomings of pertussis vaccines: why we need a third generation vaccine. *Expert Rev. Vaccines* 13, 1159–62.
- Prentki, P., Teter, B., Chandler, M., Galas, D.J., 1986. Functional promoters created by the insertion of transposable element IS1. *J. Mol. Biol.* 191, 383–93.
- Preston, A., Allen, A.G., Cadisch, J., Thomas, R., Stevens, K., Churcher, C.M., Badcock, K.L., Parkhill, J., Barrell, B., Maskell, D.J., 1999. Genetic basis for lipopolysaccharide O-antigen biosynthesis in *bordetellae*. *Infect. Immun.* 67, 3763–7.
- Quandt, J., Hynes, M.F., 1993. Versatile suicide vectors which allow direct selection for gene replacement in Gram-negative bacteria. *Gene* 127, 15–21.
- Queenan, A.M., Cassidy, P.K., Evangelista, A., 2013. Pertactin-Negative Variants of *Bordetella pertussis* in the United States. *N. Engl. J. Med.* 368, 583–584.
- Raghavan, R., Kacharia, F.R., Millar, J.A., Sislak, C.D., Ochman, H., 2015. Genome rearrangements can make and break small RNA genes. *Genome Biol. Evol.* 7, 557–66.
- Ran, W., Kristensen, D.M., Koonin, E. V., 2014. Coupling Between Protein Level Selection and Codon Usage Optimization in the Evolution of Bacteria and Archaea. *MBio* 5, e00956-14-e00956-14.

- Ren, G.-X., Guo, X.-P., Sun, Y.-C., 2017. Regulatory 3' Untranslated Regions of Bacterial mRNAs. *Front. Microbiol.* 8, 1276.
- Rombel, I.T., Sykes, K.F., Rayner, S., Johnston, S.A., 2002. ORF-FINDER: A vector for high-throughput gene identification. *Gene* 282, 33–41.
- Ross, J.A., Ellis, M.J., Hossain, S., Haniford, D.B., 2013. Hfq restructures RNA-IN and RNA-OUT and facilitates antisense pairing in the Tn10/IS10 system. *RNA* 19, 670–84.
- Safi, H., Barnes, P.F., Lakey, D.L., Shams, H., Samten, B., Vankayalapati, R., Howard, S.T., 2004. IS6110 functions as a mobile, monocyte-activated promoter in *Mycobacterium tuberculosis*. *Mol. Microbiol.* 52, 999–1012.
- Salis, H.M., 2011. The ribosome binding site calculator. *Methods Enzymol.* 498, 19–42.
- Sato, Y., Sato, H., 1999. Development of Acellular Pertussis Vaccines. *Biologicals* 27, 61–69.
- Savelkoul, P.H.M., De Kerf, D.P.G., Willems, R.J., Mooi, F.R., Van Der Zeijst, B.A.M., Gastra, W., 1996. Characterization of the fim2 and fim3 Fimbrial Subunit Genes of *Bordetella bronchiseptica*: Roles of Fim2 and Fim3 Fimbriae and Flagella in Adhesion. *Infect. Immun.* 64, 5098–5105.
- Scarlato, V., Rappuoli, R., 1991. Differential response of the bvg virulence regulon of *Bordetella pertussis* to MgSO<sub>4</sub> modulation. *J. Bacteriol.* 173, 7401–4.
- Scheller, E. V, Melvin, J.A., Sheets, A.J., Cotter, P.A., 2015. Cooperative roles for fimbria and filamentous hemagglutinin in *Bordetella* adherence and immune modulation. *MBio* 6, e00500-15.
- Schneider, D., Lenski, R.E., 2004. Dynamics of insertion sequence elements during experimental evolution of bacteria. *Res. Microbiol.* 155, 319–327.
- Sealey, K.L., Belcher, T., Preston, A., 2016. *Bordetella pertussis* epidemiology and evolution in the light of pertussis resurgence. *Infect. Genet. Evol.*
- Sealey, K.L., Harris, S.R., Fry, N.K., Hurst, L.D., Gorringer, A.R., Parkhill, J., Preston, A., 2015. Genomic analysis of isolates from the United Kingdom 2012 pertussis outbreak reveals that vaccine antigen genes are unusually fast evolving. *J. Infect. Dis.* 212.
- Seo, J.-H., Hong, J.S.-J., Kim, D., Cho, B.-K., Huang, T.-W., Tsai, S.-F., Palsson, B.O., Charusanti, P., 2012. Multiple-omic data analysis of *Klebsiella pneumoniae* MGH 78578 reveals its transcriptional architecture and regulatory features. *BMC Genomics* 13, 679.
- 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.
- Sezonov, G., Joseleau-Petit, D., D'Ari, R., 2007. *Escherichia coli* physiology in Luria-Bertani broth. *J. Bacteriol.* 189, 8746–8749.
- Sharma, C.M., Hoffmann, S., Darfeuille, F., Reignier, J., Findeiß, S., Sittka, A., Chabas, S., Reiche, K., Hackermüller, J., Reinhardt, R., Stadler, P.F., Vogel, J., 2010. The primary transcriptome of the major human pathogen *Helicobacter pylori*. *Nature* 464, 250–255.
- Sharma, C.M., Vogel, J., 2014. Differential RNA-seq: the approach behind and the biological insight gained. *Curr. Opin. Microbiol.* 19, 97–105.
- Sheridan, S.L., Frith, K., Snelling, T.L., Grimwood, K., McIntyre, P.B., Lambert, S.B., 2014. Waning vaccine immunity in teenagers primed with whole cell and acellular pertussis

- vaccine: recent epidemiology. *Expert Rev. Vaccines* 13, 1081–1106.
- Siguier, P., Gourgbeyre, E., Chandler, M., 2014. Bacterial insertion sequences: their genomic impact and diversity. *FEMS Microbiol. Rev.* 38, 865–891.
- Siguier, P., Gourgbeyre, E., Varani, A., Ton-Hoang, B., Chandler, M., 2015. Everyman's Guide to Bacterial Insertion Sequences. *Microbiol. Spectr.* 3, MDNA3-0030-2014.
- Simon, R., Priefer, U., Pühler, A., 1983. A broad host range mobilization system for in vivo genetic engineering: Transposon mutagenesis in gram negative bacteria. *Bio/Technology* 1, 784–791.
- Simons, R.W., Kleckner, N., 1983. Translational control of IS10 transposition. *Cell* 34, 683–691.
- Skoff, T.H., Kenyon, C., Cocoros, N., Liko, J., Miller, L., Kudish, K., Baumbach, J., Zansky, S., Faulkner, A., Martin, S.W., 2015. Sources of Infant Pertussis Infection in the United States. *Pediatrics* 136, 635–641.
- Smyth, G.K., Speed, T., 2003. Normalization of cDNA microarray data. *Methods* 31, 265–273.
- Smyth, G.K., Yang, Y.H., Speed, T., 2003. Statistical Issues in cDNA Microarray Data Analysis. In: *Functional Genomics*. Humana Press, New Jersey, pp. 111–136.
- Soane, M.C., Jackson, A., Maskell, D., Allen, A., Keig, P., Dewar, A., Dougan, G., Wilson, R., 2000. Interaction of *Bordetella pertussis* with human respiratory mucosa in vitro. *Respir. Med.* 94, 791–799.
- Sotanaphun, U., Phattanawasin, P., Sriphong, L., 2009. Application of scion image software to the simultaneous determination of curcuminoids in turmeric (*Curcuma longa*). *Phytochem. Anal.* 20, 19–23.
- Soumana, I.H., Linz, B., Harvill, E.T., 2017. Environmental origin of the genus *Bordetella*. *Front. Microbiol.* 8.
- 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.
- Stainer, D.W., Scholte, M.J., 1970. A simple chemically defined medium for the production of phase I *Bordetella pertussis*. *J. Gen. Microbiol.* 63, 211–20.
- Stazic, D., Voß, B., 2016. The complexity of bacterial transcriptomes. *J. Biotechnol.* 232, 69–78.
- Stenson, T.H., Allen, A.G., al-Meer, J.A., Maskell, D., Peppler, M.S., 2005. *Bordetella pertussis* *risA*, but Not *risS*, Is Required for Maximal Expression of Bvg-Repressed Genes. *Infect. Immun.* 73, 5995–6004.
- Stibitz, S., 1994. Use of conditionally counterselectable suicide vectors for allelic exchange. *Methods Enzymol.* 235, 458–465.
- Stibitz, S., 1998. IS481 and IS1002 of *Bordetella pertussis* create a 6-base-pair duplication upon insertion at a consensus target site. *J. Bacteriol.* 180, 4963–6.
- Stibitz, S., Yang, M.S., 1999. Genomic plasticity in natural populations of *Bordetella pertussis*. *J. Bacteriol.* 181, 5512–5515.
- Stockbauer, K.E., Fuchslocher, B., Miller, J.F., Cotter, P.A., 2001. Identification and characterization of BipA, a *Bordetella* Bvg-intermediate phase protein. *Mol. Microbiol.*

39, 65–78.

- Svensson, S.L., Sharma, C.M., 2016. Small RNAs in Bacterial Virulence and Communication. In: *Virulence Mechanisms of Bacterial Pathogens, Fifth Edition*. American Society of Microbiology, pp. 169–212.
- Szklarczyk, D., Franceschini, A., Wyder, S., Forslund, K., Heller, D., Huerta-Cepas, J., Simonovic, M., Roth, A., Santos, A., Tsafou, K.P., Kuhn, M., Bork, P., Jensen, L.J., Von Mering, C., 2015. STRING v10: Protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Res.* 43, D447–D452.
- Thomason, M.K., Storz, G., 2010. Bacterial Antisense RNAs: How Many Are There, and What Are They Doing? *Annu. Rev. Genet.* 44, 167–188.
- 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.
- Tolivia, J., Navarro, A., Del Valle, E., Perez, C., Ordoñez, C., Martínez, E., 2006. Application of photoshop and scion image analysis to quantification of signals in histochemistry, immunocytochemistry and hybridocytochemistry. *Anal. Quant. Cytol. Histol.* 28, 43–53.
- Tozzi, A.E., Celentano, L.P., Ciofi degli Atti, M.L., Salmaso, S., 2005. Diagnosis and management of pertussis. *CMAJ* 172, 509–15.
- Uhl, M.A., Miller, J.F., 1994. Autophosphorylation and phosphotransfer in the *Bordetella pertussis* BvgAS signal transduction cascade. *Proc. Natl. Acad. Sci. U. S. A.* 91, 1163–7.
- Uhl, M.A., Miller, J.F., 1996. Central role of the BvgS receiver as a phosphorylated intermediate in a complex two-component phosphorelay. *J. Biol. Chem.* 271, 33176–80.
- van den Berg, B.M., Beekhuizen, H., Willems, R.J., Mooi, F.R., van Furth, R., 1999. Role of *Bordetella pertussis* virulence factors in adherence to epithelial cell lines derived from the human respiratory tract. *Infect. Immun.* 67, 1056–1062.
- Vandecraen, J., Chandler, M., Aertsen, A., Van Houdt, R., 2017. The impact of insertion sequences on bacterial genome plasticity and adaptability. *Crit. Rev. Microbiol.* 43, 709–730.
- Vaughan, T.E., Pratt, C.B., Sealey, K., Preston, A., Fry, N.K., Gorringe, A.R., 2014. Plasticity of fimbrial genotype and serotype within populations of *Bordetella pertussis*: analysis by paired flow cytometry and genome sequencing. *Microbiology* 160, 2030–2044.
- Veal-Carr, W.L., Stibitz, S., 2005. Demonstration of differential virulence gene promoter activation in vivo in *Bordetella pertussis* using RIVET. *Mol. Microbiol.* 55, 788–98.
- Vermeulen, K., Thas, O., Vansteelandt, S., 2015. Increasing the power of the Mann-Whitney test in randomized experiments through flexible covariate adjustment. *Stat. Med.* 34, 1012–1030.
- Vesper, O., Amitai, S., Belitsky, M., Byrgazov, K., Kaberdina, A.C., Engelberg-Kulka, H., Moll, I., 2011. Selective Translation of Leaderless mRNAs by Specialized Ribosomes Generated by MazF in *Escherichia coli*. *Cell* 147, 147–157.
- Vogel, J., Luisi, B.F., 2011. Hfq and its constellation of RNA. *Nat. Rev. Microbiol.* 9, 578–89.
- Wadman, M., You, J., 2017. The vaccine wars. *Science* (80- ). 356, 364–365.

- Wagner, E.G.H., Romby, P., 2015. Small RNAs in Bacteria and Archaea. In: *Advances in Genetics*. pp. 133–208.
- Wang, A., Roth, J.R., 1988. Activation of silent genes by transposons Tn5 and Tn10. *Genetics* 120, 875–85.
- Warfel, J.M., Beren, J., Merkel, T.J., 2012. Airborne Transmission of *Bordetella pertussis*. *J. Infect. Dis.* 206, 902–906.
- Warfel, J.M., Edwards, K.M., 2015. Pertussis vaccines and the challenge of inducing durable immunity. *Curr. Opin. Immunol.*
- Warfel, J.M., Zimmerman, L.I., Merkel, T.J., 2014. Acellular pertussis vaccines protect against disease but fail to prevent infection and transmission in a nonhuman primate model. *Proc. Natl. Acad. Sci.* 111, 787–792.
- Weigand, M.R., Pawloski, L.C., Peng, Y., Ju, H., Burroughs, M., Cassidy, P.K., Davis, J.K., DuVall, M., Johnson, T., Juieng, P., Knipe, K., Loparev, V.N., Mathis, M.H., Rowe, L.A., Sheth, M., Williams, M.M., Tondella, M.L., 2018. Screening and genomic characterization of filamentous hemagglutinin-deficient *Bordetella pertussis*. *Infect. Immun.* IAI.00869-17.
- Weigand, M.R., Peng, Y., Loparev, V., Batra, D., Bowden, K.E., Burroughs, M., Cassidy, P.K., Davis, J.K., Johnson, T., Juieng, P., Knipe, K., Mathis, M.H., Pruitt, A.M., Rowe, L., Sheth, M., Tondella, M.L., Williams, M.M., 2017. The History of *Bordetella pertussis* Genome Evolution Includes Structural Rearrangement. *J. Bacteriol.* 199, e00806-16.
- Weinberg, Z., Barrick, J.E., Yao, Z., Roth, A., Kim, J.N., Gore, J., Wang, J.X., Lee, E.R., Block, K.F., Sudarsan, N., Neph, S., Tompa, M., Ruzzo, W.L., Breaker, R.R., 2007. Identification of 22 candidate structured RNAs in bacteria using the CMfinder comparative genomics pipeline. *Nucleic Acids Res.* 35, 4809–19.
- Wen, Y., Feng, J., Sachs, G., 2013. *Helicobacter pylori* 5' ureB-sRNA, a cis-Encoded Antisense Small RNA, Negatively Regulates ureAB Expression by Transcription Termination. *J. Bacteriol.* 195, 444–452.
- Wendelboe, A.M., Njamkepo, E., Bourillon, A., Floret, D.D., Gaudelus, J., Gerber, M., Grimprel, E., Greenberg, D., Halperin, S., Liese, J., Muñoz-Rivas, F., Teyssou, R., Guiso, N., Van Rie, A., 2007. Transmission of *Bordetella pertussis* to young infants. *Pediatr. Infect. Dis. J.* 26, 293–299.
- WHO, 2017. WHO Methods and data sources for country-level causes of death 2000-2015. *Glob. Heal. Estim. Tech. Pap.* 1–81.
- Willems, R., Paul, A., van der Heide, H.G., ter Avest, A.R., Mooi, F.R., 1990. Fimbrial phase variation in *Bordetella pertussis*: a novel mechanism for transcriptional regulation. *EMBO J.* 9, 2803–9.
- Williams, M.M., Sen, K., Weigand, M.R., Skoff, T.H., Cunningham, V.A., Halse, T.A., Tondella, M.L., CDC Pertussis Working Group, 2016. *Bordetella pertussis* Strain Lacking Pertactin and Pertussis Toxin. *Emerg. Infect. Dis.* 22, 319–322.
- Williamson, Y.M., Moura, H., Whitmon, J., Woolfitt, A.R., Schieltz, D.M., Rees, J.C., Guo, S., Kirkham, H., Bouck, D., Ades, E.W., Tondella, M.L., Carlone, G.M., Sampson, J.S., Barr, J.R., 2015. A Proteomic Characterization of *Bordetella pertussis* Clinical Isolates Associated with a California State Pertussis Outbreak. *Int. J. Proteomics* 2015, 536537.
- Wirsing Von König, C.H., 2014. Pertussis diagnostics: Overview and impact of immunization.

- Witt, M.A., Katz, P.H., Witt, D.J., 2012. Unexpectedly limited durability of immunity following acellular pertussis vaccination in preadolescents in a north American outbreak. *Clin. Infect. Dis.* 54, 1730–1735.
- Wurtzel, O., Sesto, N., Mellin, J.R., Karunker, I., Edelheit, S., Bécavin, C., Archambaud, C., Cossart, P., Sorek, R., 2012. Comparative transcriptomics of pathogenic and non-pathogenic *Listeria* species. *Mol. Syst. Biol.* 8, 583.
- Xu, Y., Liu, B., Gröndahl-Yli-Hannuksila, K., Tan, Y., Feng, L., Kallonen, T., Wang, L., Peng, D., He, Q., Wang, L., Zhang, S., 2015. Whole-genome sequencing reveals the effect of vaccination on the evolution of *Bordetella pertussis*. *Sci. Rep.* 5, 12888.
- Yim, H.H., Villarejo, M., 1992. *osmY*, a new hyperosmotically inducible gene, encodes a periplasmic protein in *Escherichia coli*. *J. Bacteriol.* 174, 3637–44.
- Zaghloul, L., Tang, C., Chin, H.Y., Bek, E.J., Lan, R., Tanaka, M.M., 2007. The distribution of insertion sequences in the genome of *Shigella flexneri* strain 2457T. *FEMS Microbiol. Lett.* 277, 197–204.
- Zaitsev, E.M., Poddubikov, A. V, Britsina, M. V, Mertsalova, N.U., Ozeretskovskaia, M.N., Bazhanova, I.G., Plekhanova, N.G., 2014. [Immunobiological properties of *Bordetella pertussis* lipopolysaccharide in the acellular pertussis vaccine]. *Zh Mikrobiol Epidemiol Immunobiol* 20–24.
- Zschiedrich, C.P., Keidel, V., Szurmant, H., 2016. Molecular Mechanisms of Two-Component Signal Transduction. *J. Mol. Biol.* 428, 3752–3775.

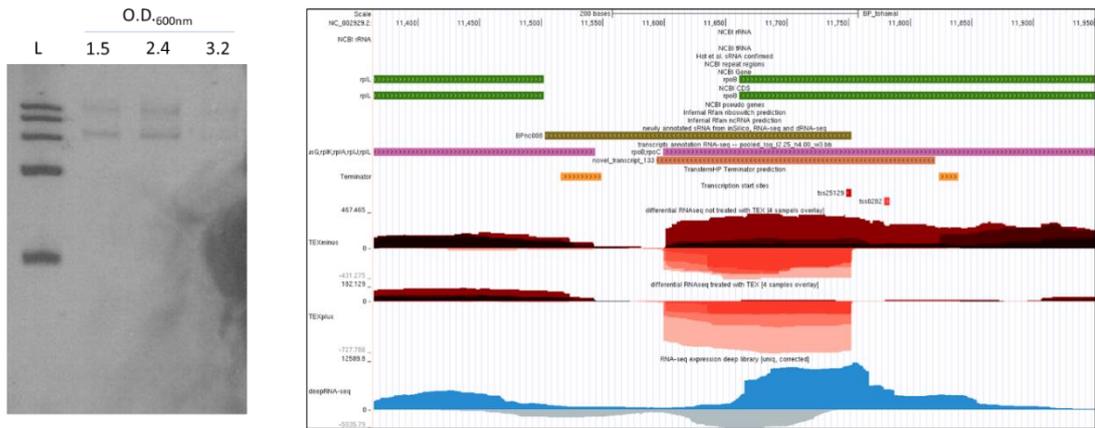
# Figures Supplémentaires

## Figure Supplémentaire S1 :

Validation de nouvelles prédictions d'ARN régulateurs par Northern Blot.

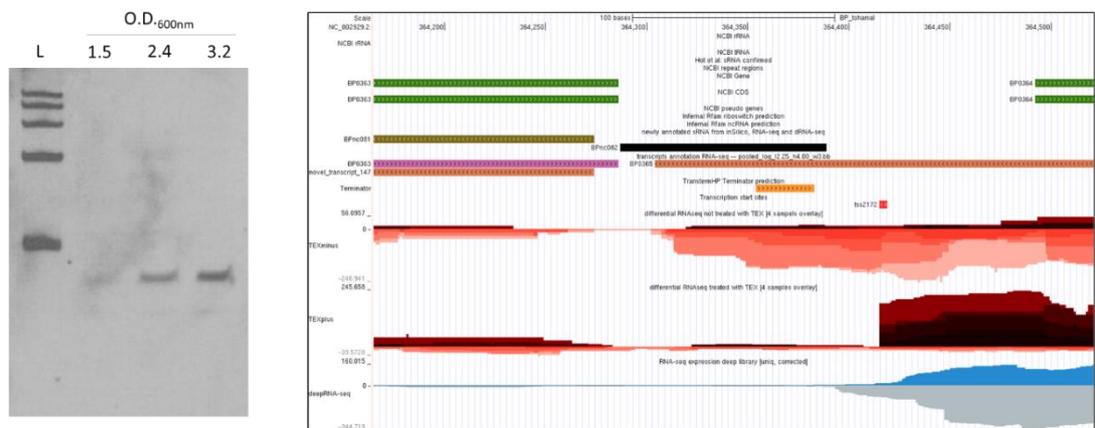
Chaque prédiction est indiquée par son nom « BPncXXX » en titre. Les figures contiennent à gauche le Northern Blot de détection du transcrit d'une culture arrêtée à différentes densités optiques (OD600nm) : en phase exponentielle précoce (1,5), en phase exponentielle tardive (2,4) et en phase tardive de croissance (3,2). Une capture d'écran positionnant la prédiction avec son nom « BPncXXX » (à droite), et les informations fournis par le tableau supplémentaire S1 (en bas).

### BPnc008



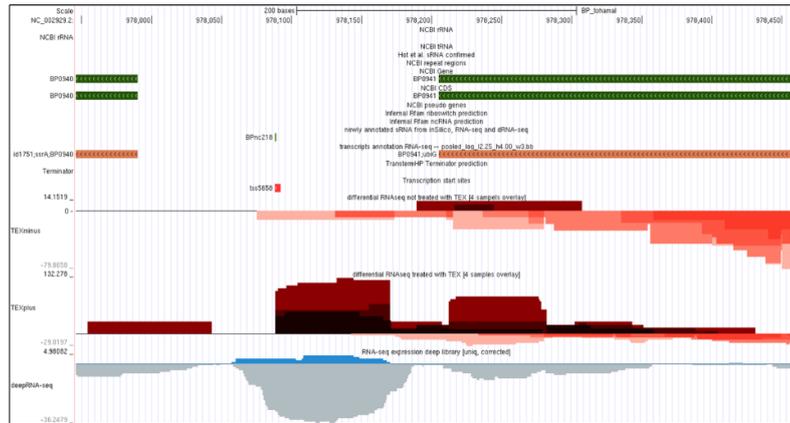
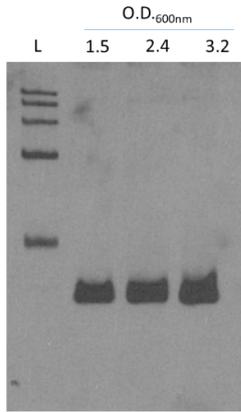
ID	chr	start	stop	strand	inSilico (methods details)	dRNAseq	RNAseq	upstream (distance [nt])	downstream (distance [nt])	antisense
BPnc008	NC_002929.2	11503	11752	-	found (NAPP, RNAz)	found	found	BP0017 (-8375)	---	rpoB

### BPnc082



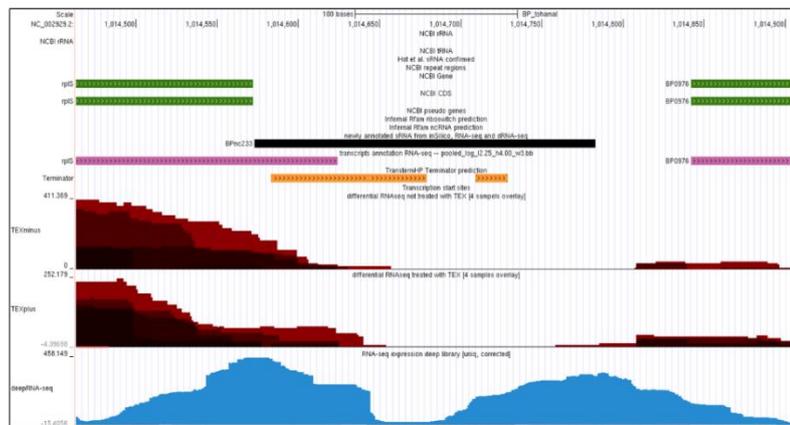
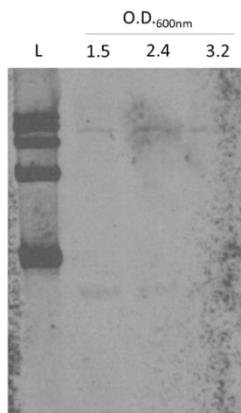
ID	chr	start	stop	strand	inSilico (methods details)	dRNAseq	RNAseq	upstream (distance [nt])	downstream (distance [nt])	antisense
BPnc082	NC_002929.2	364287	364389	0	found (SIPHT)	---	---	BP0363 (-2)	BP0364 (104)	---

## BPnc218



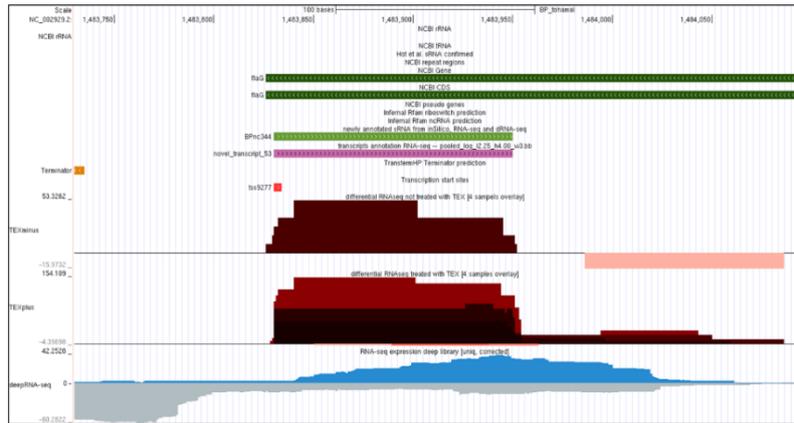
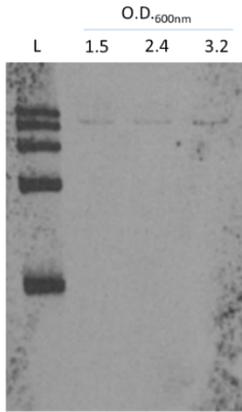
ID	chr	start	stop	strand	inSilico (methods details)	dRNAseq	RNAseq	upstream (distance [nt])	downstream (distance [nt])	antisense
BPnc218	NC_002929.2	978088	978089	+	---	found	---	BP0938 (-2265)	gyrA (2594)	---

## BPnc233



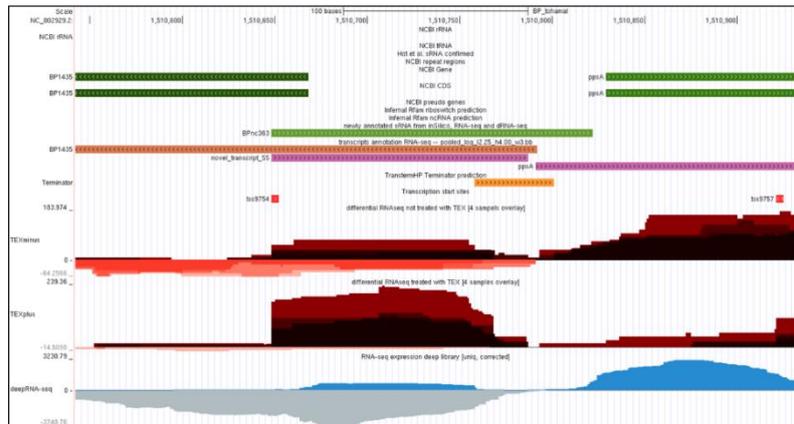
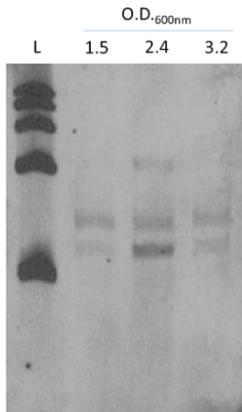
ID	chr	start	stop	strand	inSilico (methods details)	dRNAseq	RNAseq	upstream (distance [nt])	downstream (distance [nt])	antisense
BPnc233	NC_002929.2	1014573	1014783	0	found (NAPP, SIPHT)	---	---	rpsL (-2)	BP0976 (60)	---

## BPnc344



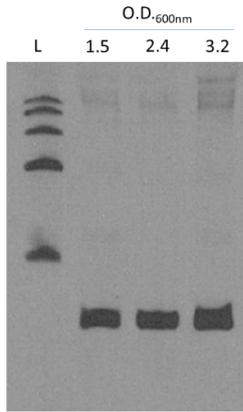
ID	chr	start	stop	strand	inSilico (methods details)	dRNAseq	RNAseq	upstream (distance [nt])	downstream (distance [nt])	antisense
BPnc344	NC_002929.2	1483830	1483950	+	---	found	found	fliE (-4366)	folC (517)	fliA

## BPnc363



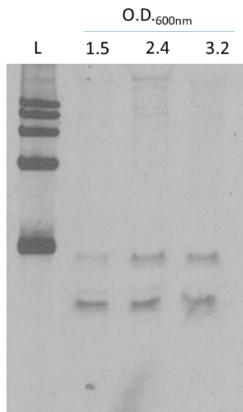
ID	chr	start	stop	strand	inSilico (methods details)	dRNAseq	RNAseq	upstream (distance [nt])	downstream (distance [nt])	antisense
BPnc363	NC_002929.2	1510648	1510822	+	found (RNAz)	found	found	rnhB (-820)	ppsA (8)	BP1435

### BPnc375



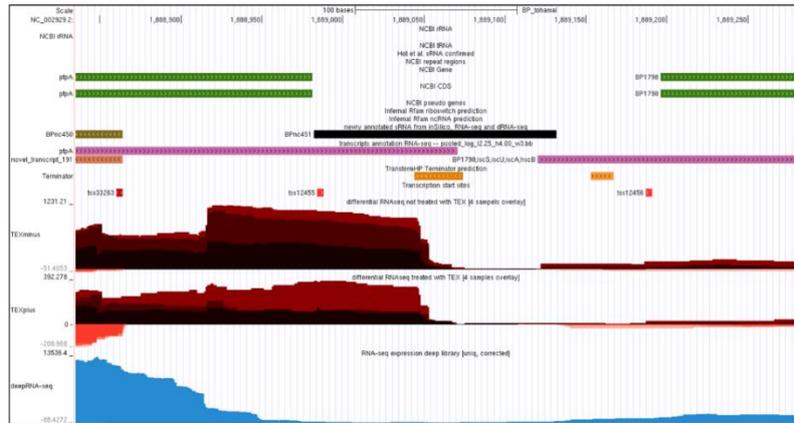
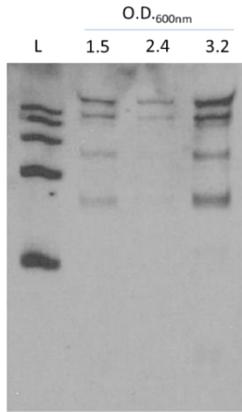
ID	chr	start	stop	strand	inSilico (methods details)	dRNAseq	RNAseq	upstream (distance [nt])	downstream (distance [nt])	antisense
BPnc375	NC_002929.2	1581071	1581316	+	found (SIPHT)	found	found	ptsI (-85)	BP1505 (659)	BP1503

### BPnc384



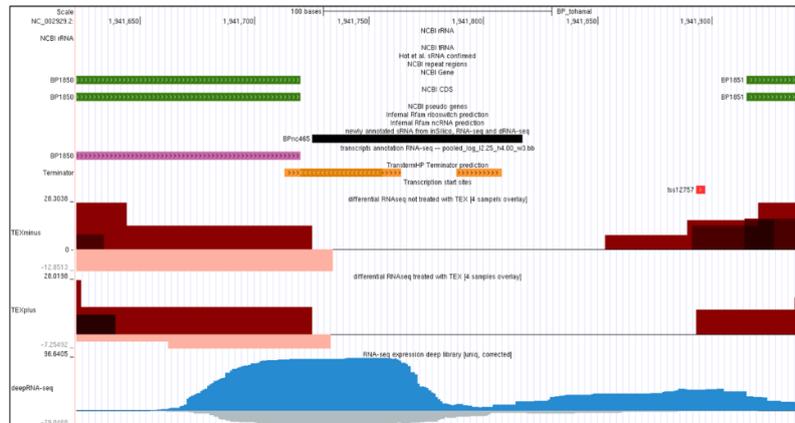
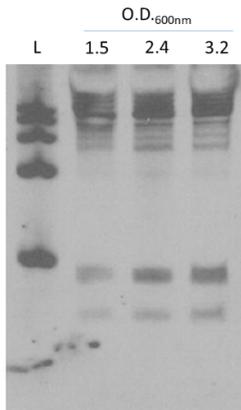
ID	chr	start	stop	strand	inSilico (methods details)	dRNAseq	RNAseq	upstream (distance [nt])	downstream (distance [nt])	antisense
BPnc384	NC_002929.2	1626732	1627269	+	found (NAPP, SIPHT)	found	found	BP1548 (-102)	dnaX (165)	---

## BPnc451



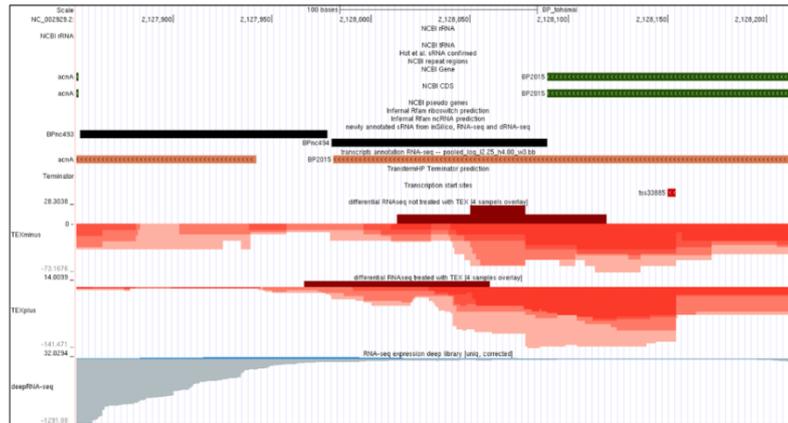
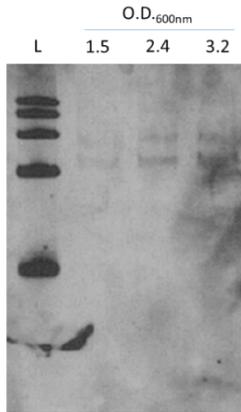
ID	chr	start	stop	strand	inSilico (methods details)	dRNAseq	RNAseq	upstream (distance [nt])	downstream (distance [nt])	antisense
BPnc451	NC_002929.2	1888982	1889132	0	found (NAPP, RNAz, SIPHT)	---	---	ptpA (-2)	BP1798 (65)	---

## BPnc465



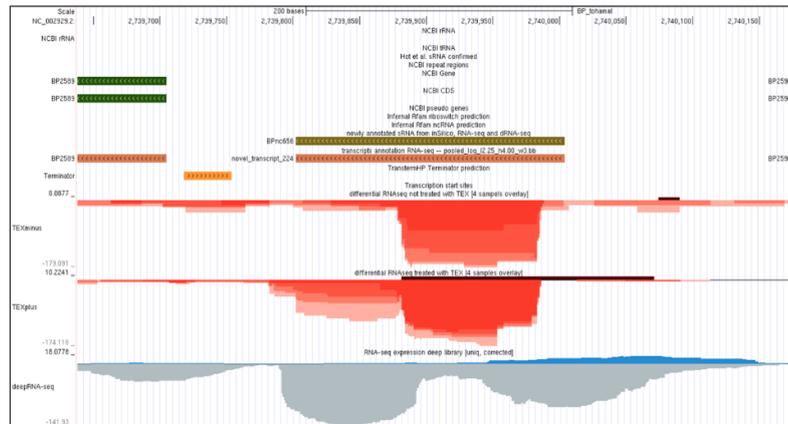
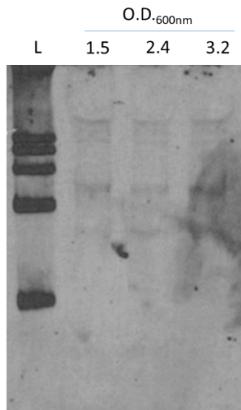
ID	chr	start	stop	strand	inSilico (methods details)	dRNAseq	RNAseq	upstream (distance [nt])	downstream (distance [nt])	antisense
BPnc465	NC_002929.2	1941725	1941817	0	found (SIPHT)	---	---	BP1850 (-6)	BP1851 (99)	---

## BPnc494



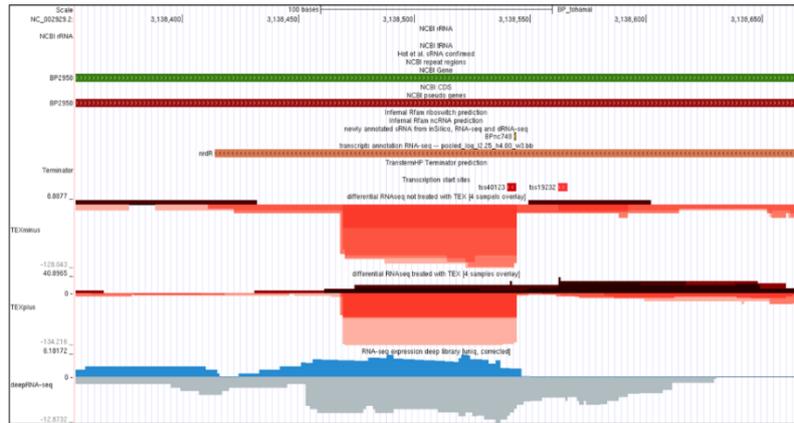
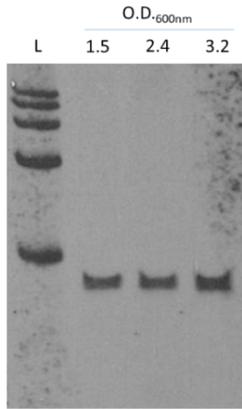
ID	chr	start	stop	strand	inSilico (methods details)	dRNAseq	RNAseq	upstream (distance [nt])	downstream (distance [nt])	antisense
BPnc494	NC_002929.2	2127980	2128089	0	found (SIPHT)	---	---	acnA (-129)	BP2015 (1)	---

## BPnc656



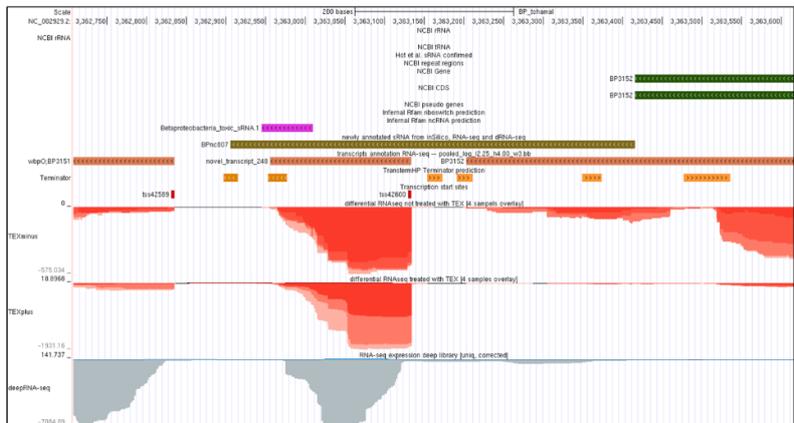
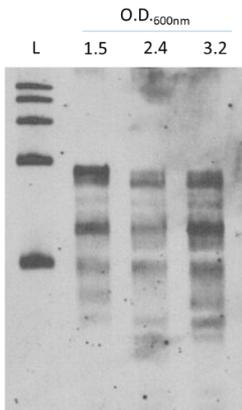
ID	chr	start	stop	strand	inSilico (methods details)	dRNAseq	RNAseq	upstream (distance [nt])	downstream (distance [nt])	antisense
BPnc656	NC_002929.2	2739802	2740004	-	---	---	found	BP2590 (-173)	BP2589 (98)	---

## BPnc748



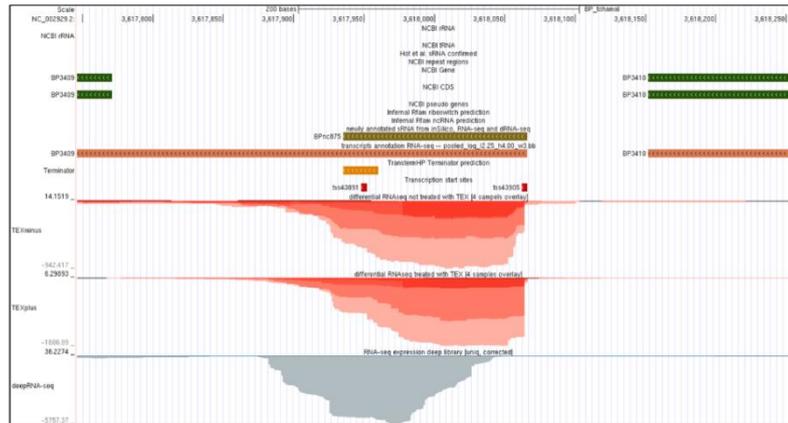
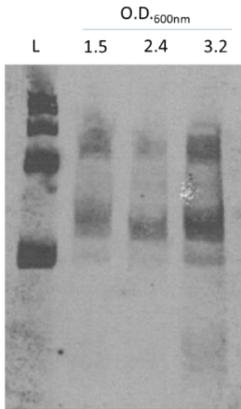
ID	chr	start	stop	strand	inSilico (methods details)	dRNAseq	RNAseq	upstream (distance [nt])	downstream (distance [nt])	antisense
BPnc748	NC_002929.2	3138543	3138544	-	---	found	---	nrdR (-155)	BP2948 (1573)	BP2950

## BPnc807



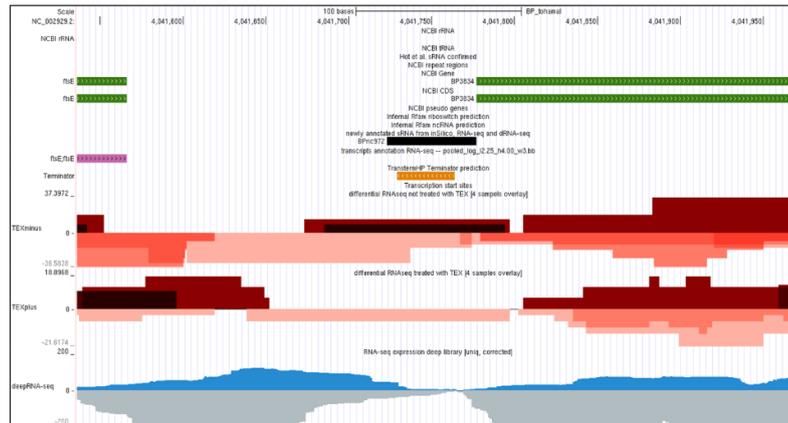
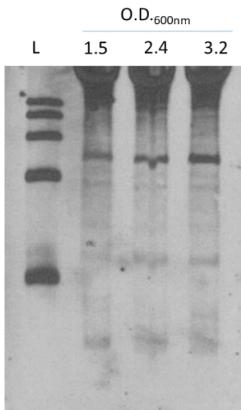
ID	chr	start	stop	strand	inSilico (methods details)	dRNAseq	RNAseq	upstream (distance [nt])	downstream (distance [nt])	antisense
BPnc807	NC_002929.2	3362906	3363416	-	found (NAPP, RNAz, SIPHT)	found	found	BP3152 (-1)	BP3151 (219)	---

### BPnc875



ID	chr	start	stop	strand	inSilico (methods details)	dRNAseq	RNAseq	upstream (distance [nt])	downstream (distance [nt])	antisense
BPnc875	NC_002929.2	3617935	3618066	-	found (NAPP, RNAz)	found	---	BP3410 (-87)	BP3409 (165)	---

### BPnc972



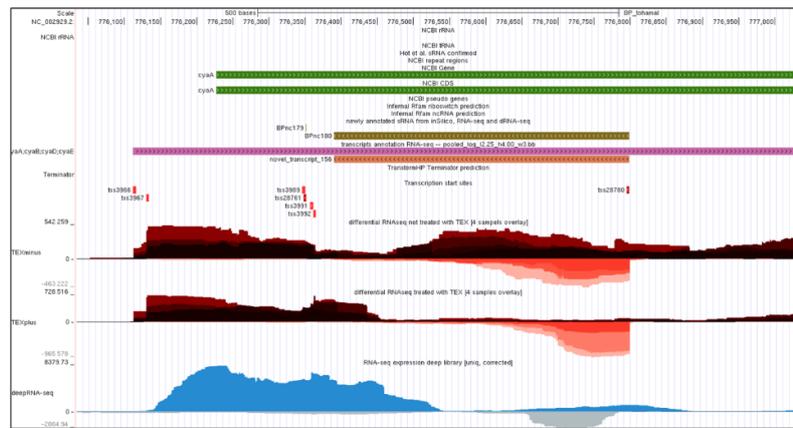
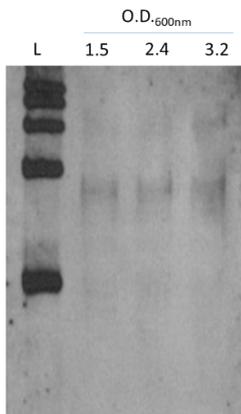
ID	chr	start	stop	strand	inSilico (methods details)	dRNAseq	RNAseq	upstream (distance [nt])	downstream (distance [nt])	antisense
BPnc972	NC_002929.2	4041723	4041777	0	found (SIPHT)	---	---	ftsE (-158)	BP3834 (1)	---

## Figure Supplémentaire S2 :

Validation des prédictions d'ARN régulateurs orientées en antisens de gènes majeurs de virulence par Northern Blot.

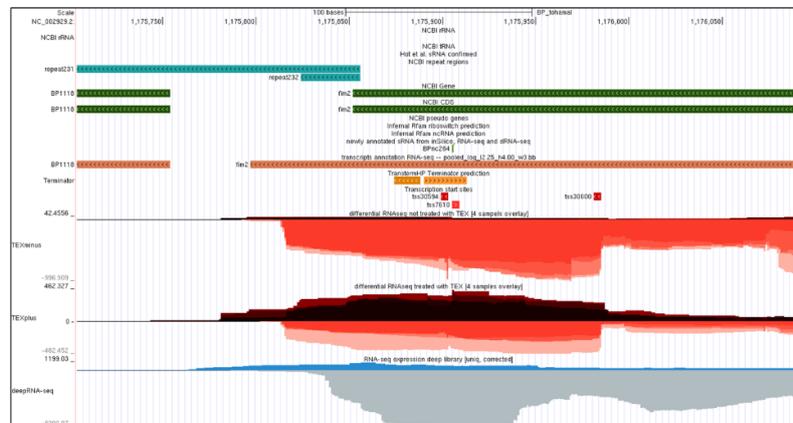
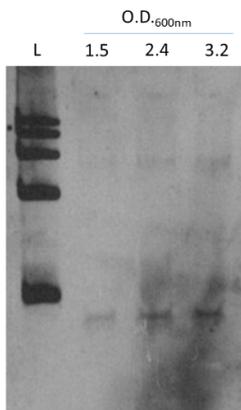
Chaque prédiction est indiquée par son nom « BPncXXX » en titre. Les figures contiennent à gauche le Northern Blot de détection du transcript d'une culture arrêtée à différentes densités optiques (OD600nm) : en phase exponentielle précoce (1,5), en phase exponentielle tardive (2,4) et en phase tardive de croissance (3,2). Une capture d'écran positionnant la prédiction avec son nom « BPncXXX » (à droite), et les informations fournis par le tableau supplémentaire S1 (en bas).

### BPnc180



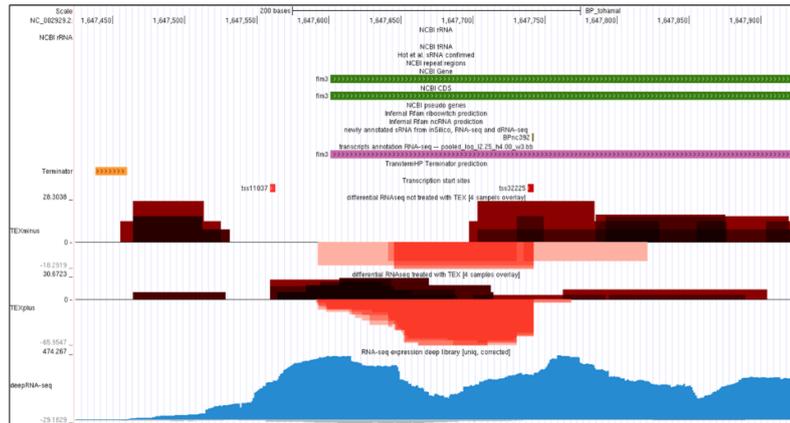
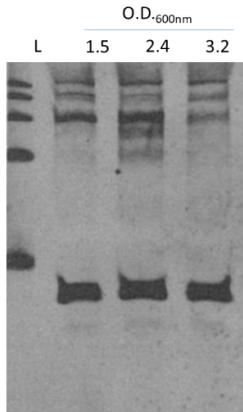
ID	chr	start	stop	strand	inSilico (methods details)	dRNAseq	RNAseq	upstream (distance [nt])	downstream (distance [nt])	antisense
BPnc180	NC_002929.2	776390	776799	-	---	found	found	BP0770 (-14667)	<i>cyaC</i> (642)	<i>cyaA</i>

### BPnc264



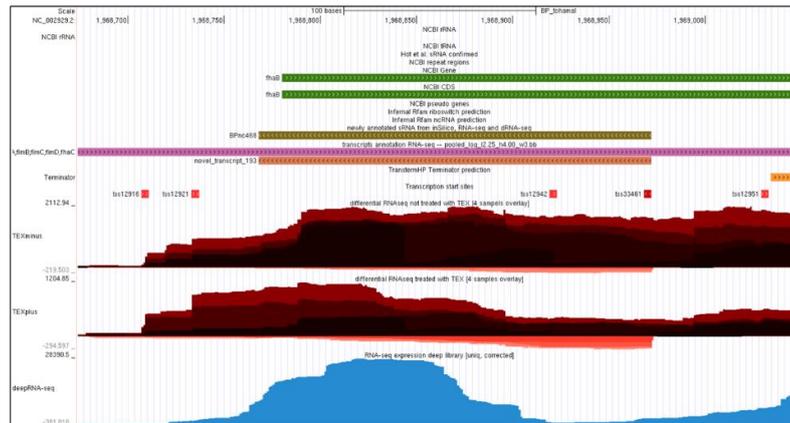
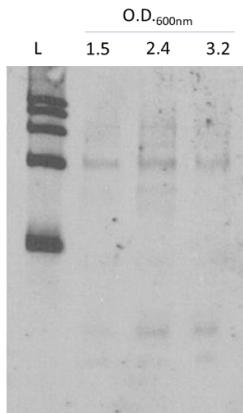
ID	chr	start	stop	strand	inSilico (methods details)	dRNAseq	RNAseq	upstream (distance [nt])	downstream (distance [nt])	antisense
BPnc264	NC_002929.2	1175905	1175906	+	---	found	---	greB (-1149)	BP1122 (6009)	<i>flh2</i>

## BPnc392



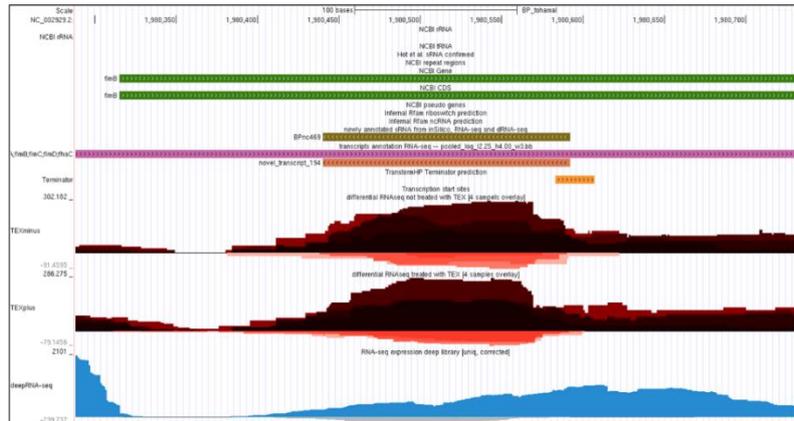
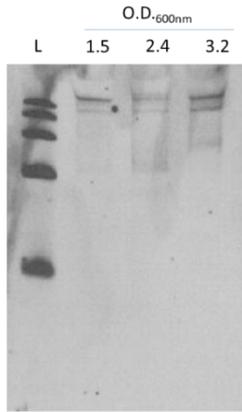
ID	chr	start	stop	strand	inSilico (methods details)	dRNAseq	RNAseq	upstream (distance [nt])	downstream (distance [nt])	antisense
BPnc392	NC_002929.2	1647741	1647742	-	---	found	---	BP1569 (-563)	BP1565 (4834)	fim3

## BPnc468



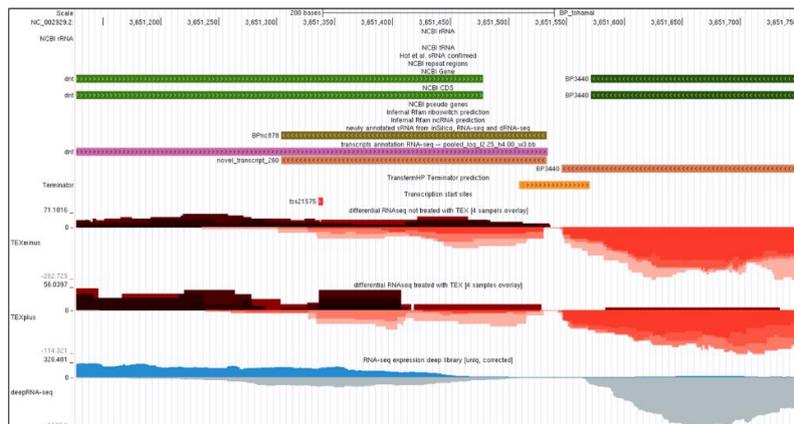
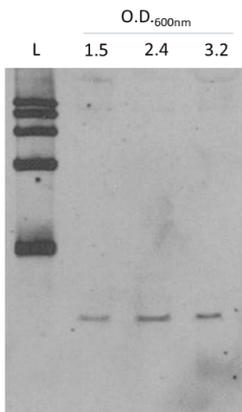
ID	chr	start	stop	strand	inSilico (methods details)	dRNAseq	RNAseq	upstream (distance [nt])	downstream (distance [nt])	antisense
BPnc468	NC_002929.2	1968768	1968972	-	---	found	found	BP1885 (-17646)	bvgA (415)	fhaB

## BPnc469



ID	chr	start	stop	strand	inSilico (methods details)	dRNAseq	RNAseq	upstream (distance [nt])	downstream (distance [nt])	antisense
BPnc469	NC_002929.2	1980440	1980592	-	---	---	found	BP1885 (-6026)	bvgA (12087)	fimB

## BPnc878

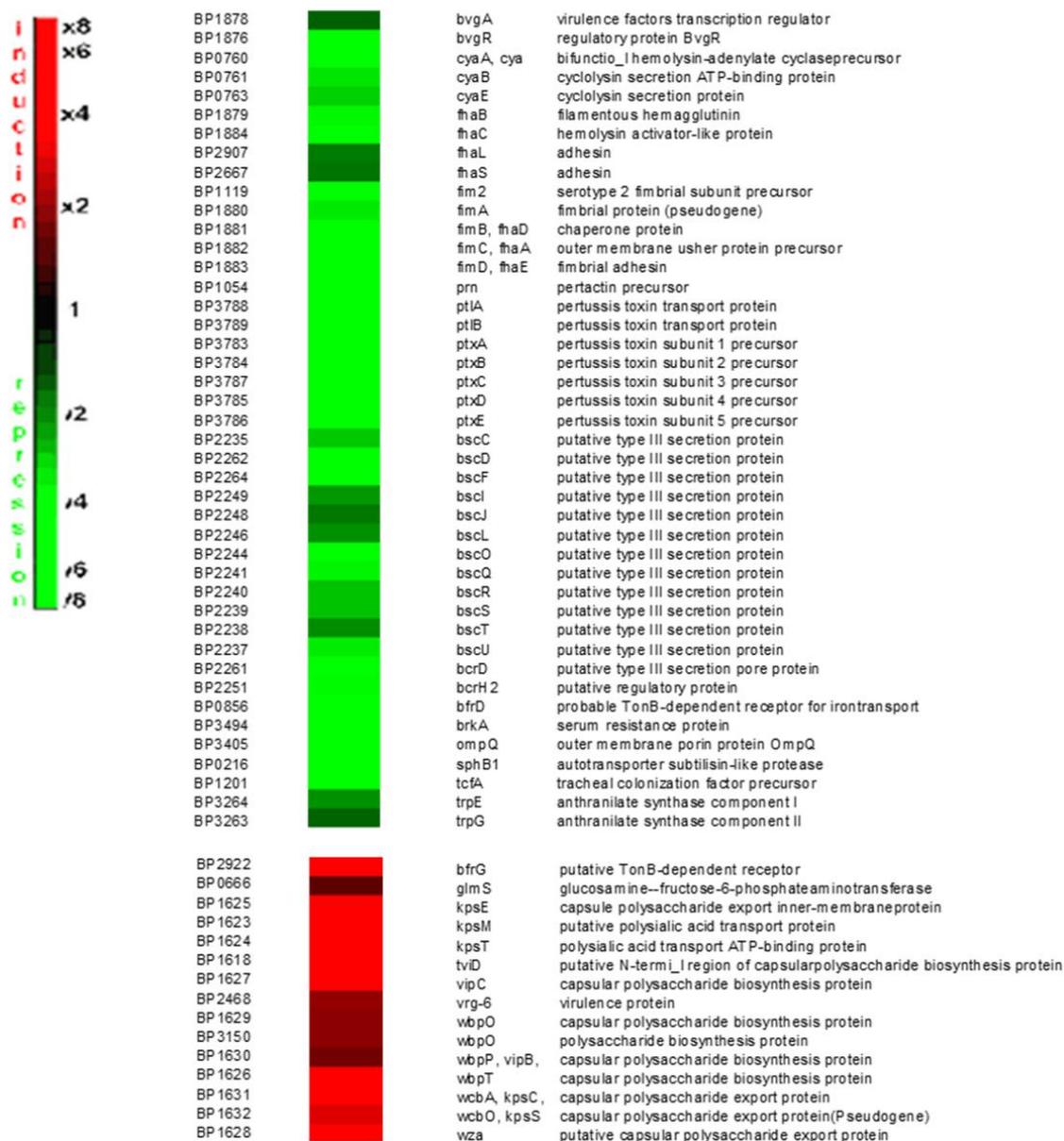


ID	chr	start	stop	strand	inSilico (methods details)	dRNAseq	RNAseq	upstream (distance [nt])	downstream (distance [nt])	antisense
BPnc878	NC_002929.2	3651304	3651533	-	---	---	found	BP3440 (-39)	BP3434 (8587)	dnt

### Figure Supplémentaire S3 :

Validation des triplicatas de culture en phase Bvg<sup>-</sup> par puces à ADN

Les ADNc générés à partir des ARN totaux des cultures en phase Bvg<sup>+</sup> et Bvg<sup>-</sup> ont été marqués à la cyanine 3 et/ou 5. Chaque triplicata a été analysé en « Dye-Swap » (inversion des fluorochromes) avec le progiciel Lima utilisé sous R. Les cartes d'expression ont été générées avec le logiciel MeV. Les rapports d'expression entre la condition Bvg<sup>-</sup> comparée à la condition Bvg<sup>+</sup> sont exprimés en Log<sub>2</sub>. Un facteur 8 (Rouge) correspond à 16 fois plus de transcrits en condition Bvg<sup>-</sup> par rapport à la condition Bvg<sup>+</sup>. A l'inverse, un facteur -8 (Vert) correspond à une expression 16 fois plus faible en phase Bvg<sup>-</sup> par rapport à l'expression en phase Bvg<sup>+</sup>.

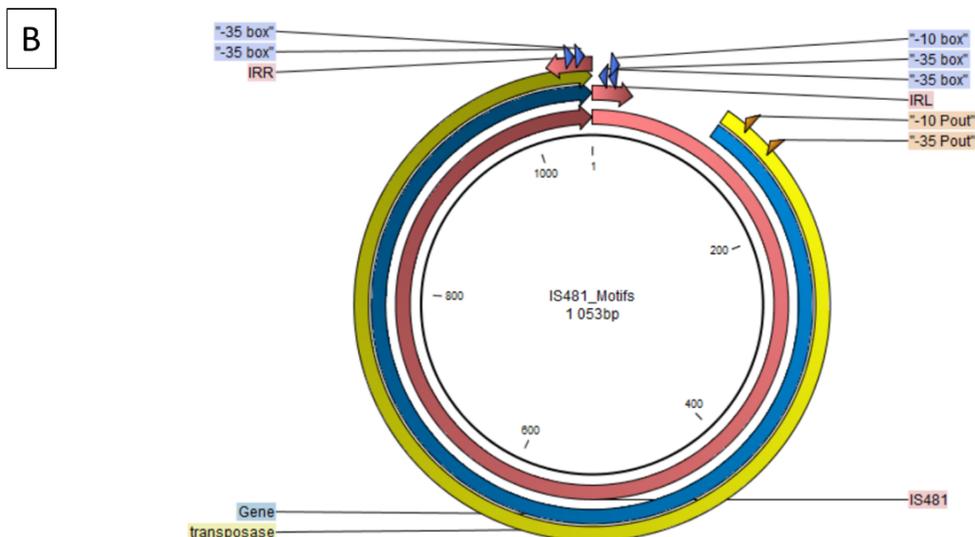
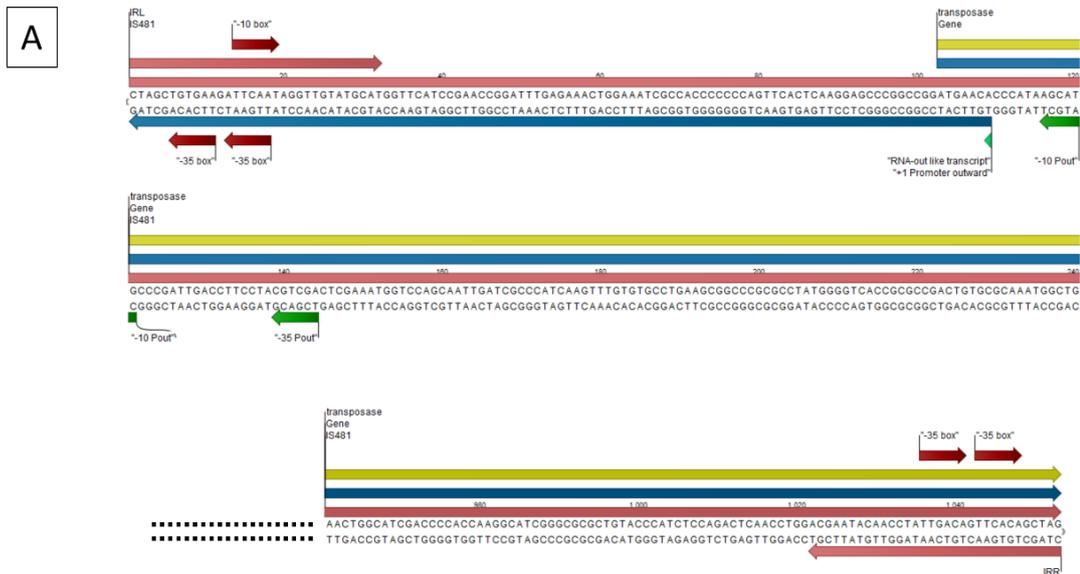


### Figure Supplémentaire S4 :

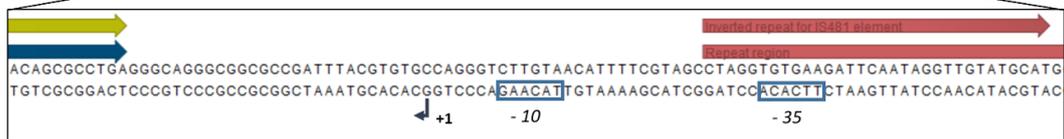
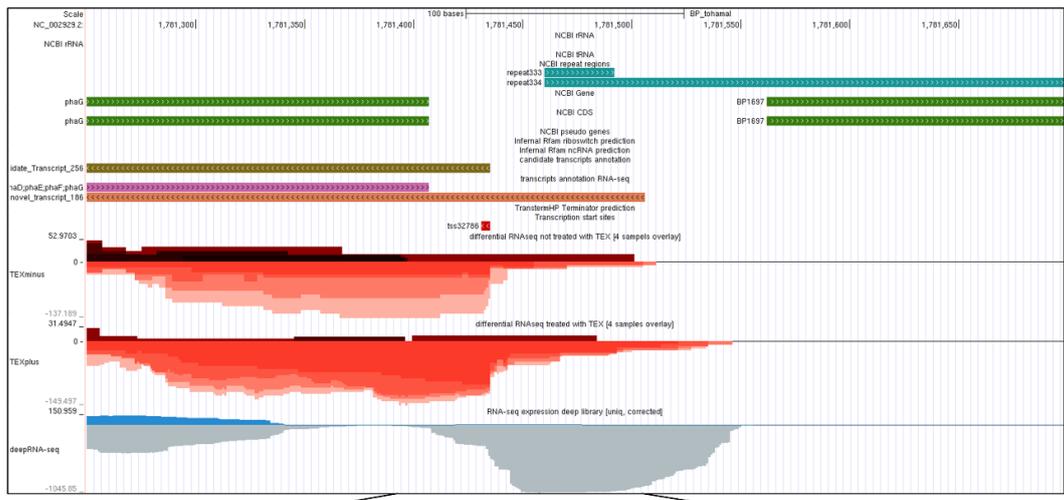
Motifs -35 et -10 des régions répétées gauches (IRL) et droites (IRR) des IS481 et la formation de  $P_{hyb}$ .

(A) : Séquences ADN tronquées des IS481. Flèche rouge : Annotation des IS481. Flèche jaune : Séquence codante. Flèche bleue : Annotation du gène. Flèche bleue marine : Transcrit antisens régulateur. Flèches vertes : Annotation du  $P_{out}$ . Flèches bordeaux : Motifs -10 et -35. (B) : Forme circulaire des IS481 et formation d'un promoteur fort permettant de transcrire abondamment le gène de la transposase.

(C) : Identification d'un promoteur  $P_{hyb}$  prédit par séquençage à ARN différentiel. La capture d'écran montre la région du gène *phaG* localisé en amont d'une IS481. Un TSS a été prédit dans la région intergénique orienté en antisens du gène *phaG* dont le promoteur a été formé par l'insertion de l'IS481. Le motif -10 se retrouve dans la région intergénique alors que le motif -35 est localisé dans la région répétée gauche de l'IS. Flèche noire : Départ de transcription. Rectangle bleu : Motifs -10 et -35 identifiés. Les figures ont été faites à partir du logiciel CLC Genomics Workbench v10.



C

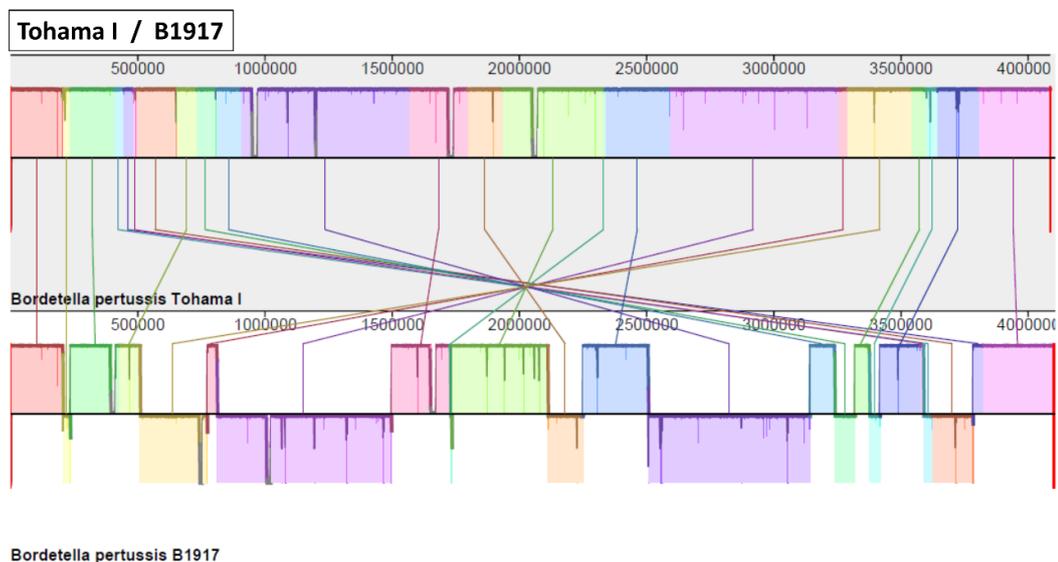
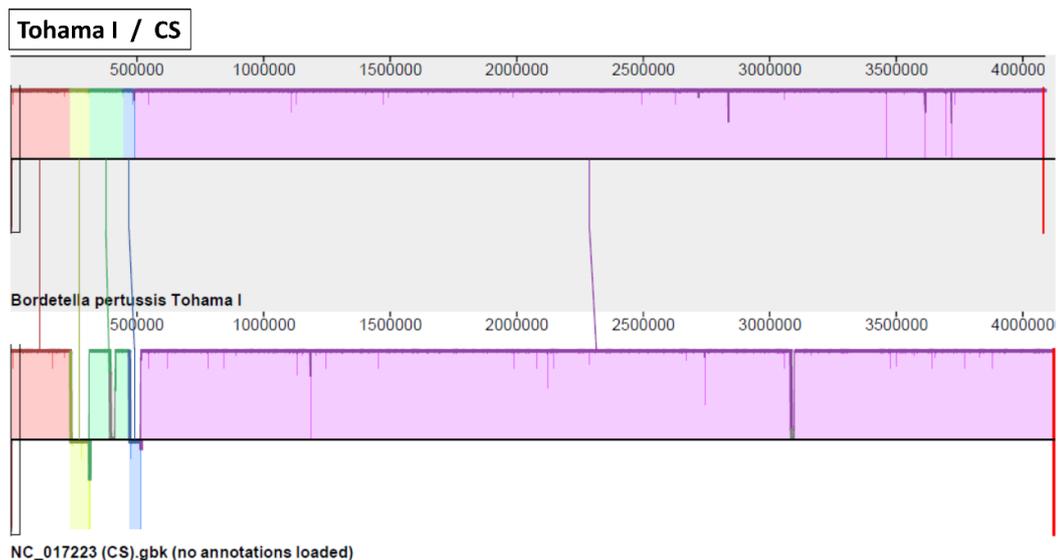


### **Figure Supplémentaire S5 :**

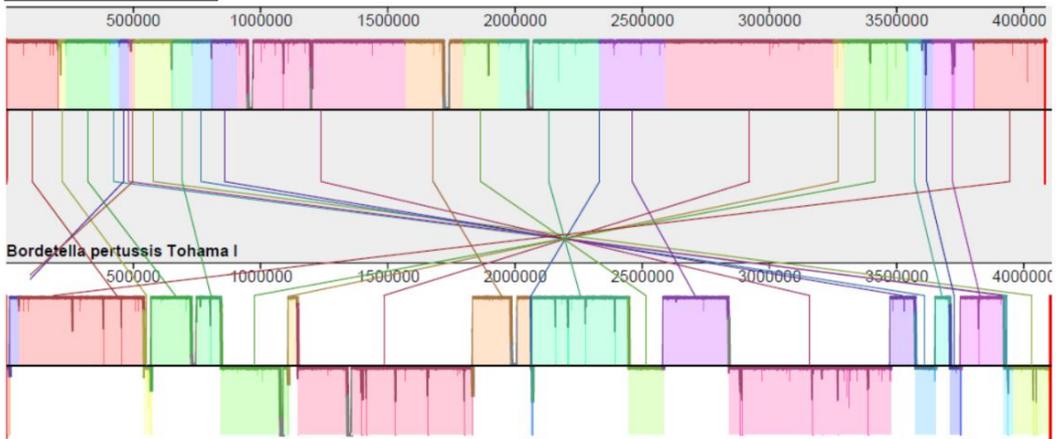
Alignement des séquences génomiques de la souche Tohama I contre les séquences de différentes souches de *Bordetella pertussis*.

Les génomes ont été alignés à l'aide du logiciel Mauve en utilisant les paramètres par défaut (Matériel et Méthode 19) (Darling et al., 2004). Les génomes sont représentés par une ligne noire horizontale. Les blocks de couleur représentent les portions du génome retrouvés sur le génome de la souche comparée. Les blocks au-dessus de la ligne horizontale représentent les séquences retrouvées dans le même sens, et les blocks en-dessous de cette ligne sont des séquences identiques inversées dans le génome. Chaque block est relié par une ligne indiquant sa position sur le génome en comparaison.

Chaque alignement est noté « Tohama I / XXX », indiquant une comparaison du génome de la souche Tohama I avec le génome de la souche XXX (CS, B1917, D420 ou 18323).

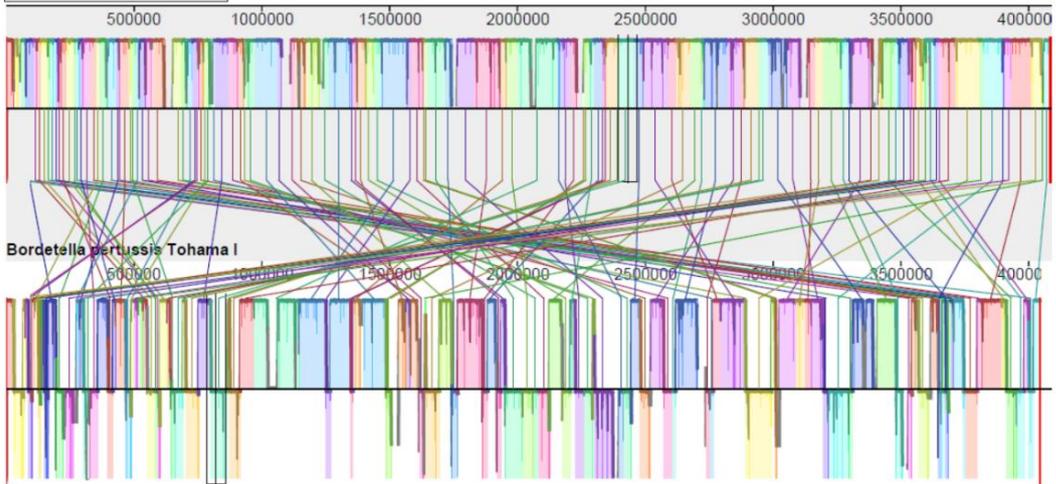


**Tohama I / D420**



*Bordetella pertussis* D420

**Tohama I / 18323**

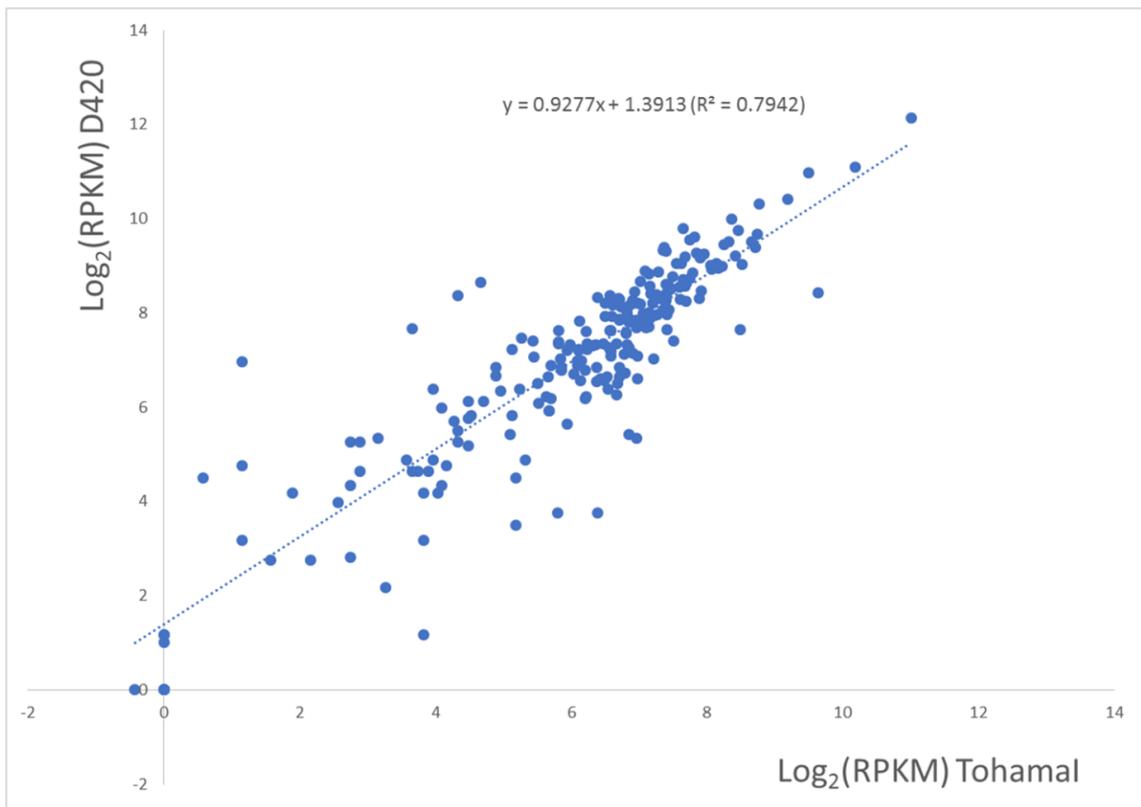


NC\_018518 (18323).gbk (no annotations loaded)

**Figure supplémentaire S6 :**

Corrélation des forces des promoteurs  $P_{out}$  des IS481 entre les souches Tohama I et D420.

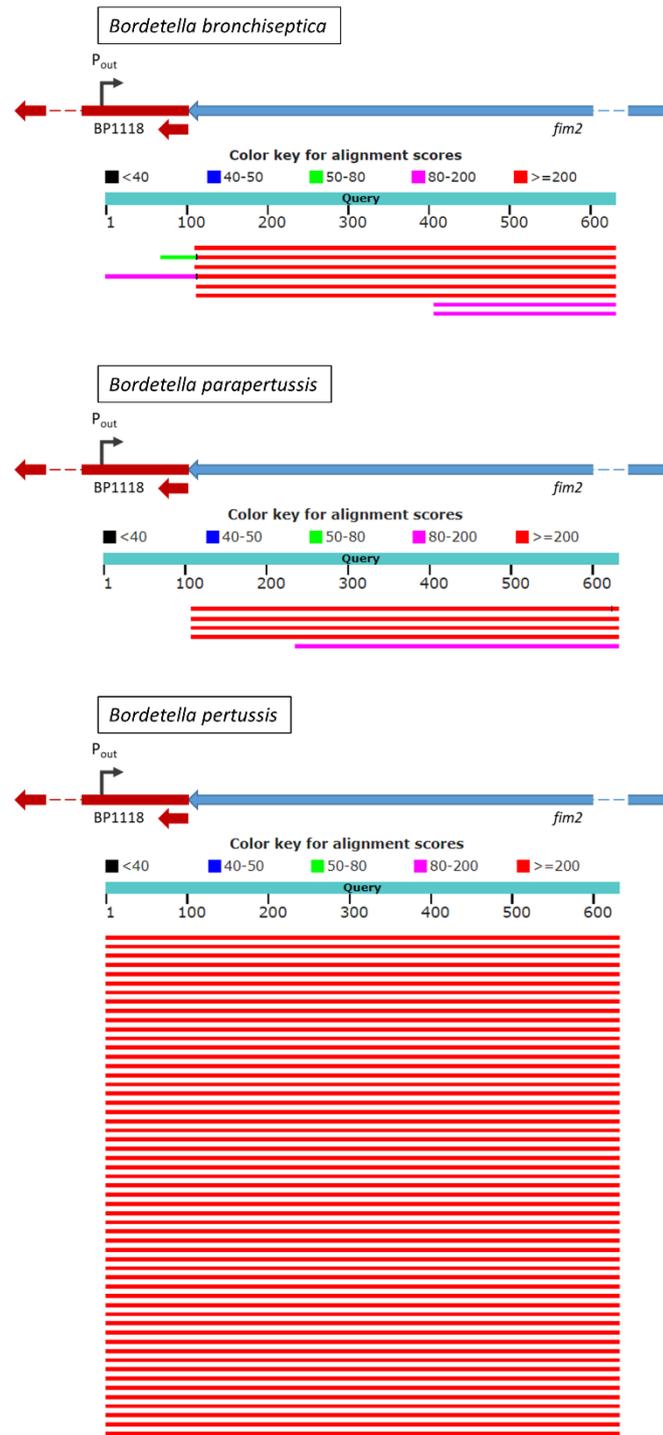
Les régions spécifiques des sondes informatiques (200pb en aval du  $P_{out}$ ) utilisées pour détecter les prolongements de transcription des IS481 ont été quantifiées par calcul de RPKM à partir des résultats de séquençage profond entre les souches Tohama I et D420. Les résultats sont exprimés en  $\text{Log}_2$ . Les résultats sont représentés sur le graphique ci-dessous. Une droite de régression a été tracée et le coefficient de corrélation a été calculé ( $R^2 = 0,79$ ).



### Figure Supplémentaire S7 :

Contexte génomique du gène *fim2* et de l'IS481 BP1118 chez *Bordetella pertussis*, *Bordetella bronchiseptica* et *Bordetella parapertussis*.

La séquence génomique de BPnc264, allant du départ prédit de transcription ( $P_{out}$  de l'IS481 BP1118) jusqu'à la fin de la prédiction (627 nucléotides), a été alignée sur les génomes entièrement séquencés de 5 souches de *Bordetella parapertussis*, 8 souches de *Bordetella bronchiseptica* et plus de 100 souches de *Bordetella pertussis*. Les résultats sont représentés par les alignements issus de Blast (Altschul et al., 1990). Un schéma de l'annotation génomique et de BPnc264 a été ajouté par rapport aux résultats du blast.



### **Figure Supplémentaire S8 :**

Prédiction de séquences codantes potentielles sur la séquence de BPnc264.

Les prédictions ont été faites avec le logiciel ORFfinder disponible à l'adresse : <https://www.ncbi.nlm.nih.gov/orffinder/>

Les cadres ouverts de lecture (ORF) ont été prédits sur le brin positif et négatif. Les prédictions alignées sur le transcrit de BPnc264 sont listées dans le tableau ci-dessous avec le numéro de cadre ouvert de lecture, les positions des codons d'initiation de la traduction et d'arrêt de la traduction, la taille de la séquence codante en nucléotides et la taille de la protéine en acides aminés (AA).

Numéro ORF	Brin ADN	Cadre ouvert de lecture	Position du codon d'initiation	Position du codon stop	Taille de l'ORF / Nombre d'AA
1	+	2	5	73	69   22
2	+	3	24	557	534   177
3	+	1	37	108	72   23
4	+	1	70	108	39   12
5	+	2	95	142	48   15
6	+	1	136	180	45   14
7	+	1	142	180	39   12
8	+	1	196	381	186   61
9	+	2	245	283	39   12
10	+	1	277	381	105   34
11	+	1	385	417	33   10
12	+	2	392	451	60   19
13	+	1	448	576	129   42
14	+	1	490	576	87   28
15	+	2	533	601	69   22
16	+	1	598	>627	57   18

