THÈSE

UNIVERSITE DE PAU ET DES PAYS DE L'ADOUR
École doctorale des Sciences Exactes et de leurs Applications
(ED211)

Présentée et soutenue
Le 09 décembre 2015

par **Justine ABELLA**

pour obtenir le grade de docteur
de l'Université de Pau et des Pays de l'Adour
Spécialité : Physiologie et biologie des organismes-populations-interactions
Microbiologie

IMPLICATION DES INTEGRONS DANS L'ADAPTATION DES COMMUNAUTES BACTERIENNES

MEMBRES DU JURY

RAPPORTEURS

• Pierre PEYRET	Pr / Université d'Auvergne, EA 4678 CIDAM
• Marie-Cécile PLOY	Pr / CHU de Limoges, Service de Bactériologie-Virologie-Hygiène ; INSERM U1092 ; Université de Limoges, UMR-S1092

EXAMINATEURS

• Philippe CUNY	Pr / Aix Marseille Université, CNRS/INSU, Université de Toulon, IRD, MIO, UM 110
• Pascal SIMONET	DR / Ecole Centrale de Lyon, CNRS UMR 5005, Laboratoire Ampère

DIRECTEURS

• Christine CAGNON	Dr / Université de Pau et des Pays de l'Adour, IPREM UMR CNRS 5254
• Robert DURAN	Pr / Université de Pau et des Pays de l'Adour, IPREM UMR CNRS 5254

IVERSITÉ
DE PAU ET DES
PAYS DE L'ADOUR

environnement
intégrase
intégrons
diversité
cassettes
contamination
séquences
gène
la
que
en
conditions
études
plus
leur
bactéries
milieux
bactériennes
antibiotiques
cliniques
montré
ou
pour
été
sont
milieu
contexte
possèdent
cette
adaptatifs
cela
avec
particulièrement
étudier
transfert
pétrole
important
entre
Huang
augmente
métabolisme
importante
fonctions
évaluer
horizontal
partir
types
qui
même
était
objectif
utilisation
abondance
bien
sélection
impact
ubiquistes
structures
effet
sur
nous
niveau
influencée
afin
est
résistances
microcosmes
pression
centaine
compétitives
métagénome
potentiel
évolution
réalisées
anciennes
possible
mutation
rôle
Lionel
présents
face
apprehender
soit
autres
intervenir
permettant
Ainsi
distinction
douce
encore
organismes
nouvelles
dehors
inconnues
dont
banques
Enfin
bactériens
grâce
temps
comparables
phylogénétique
nombreuses
éléments
voire
polluants
cherche
processus
colonizer
Pool
augmente

REMERCIEMENTS

Ces travaux ont été réalisés au sein de l'Equipe Environnement et Microbiologie de l'Institut des Sciences Analytiques et de Physico-chimie pour l'Environnement et les Matériaux de l'Université de Pau et des Pays de l'Adour. Mes remerciements vont en premier lieu au Pr. Robert Duran pour m'avoir accueillie dans cette équipe depuis mon Master Recherche.

Je tiens à remercier le Pr. Marie-Cécile Ploy et le Pr. Pierre Peyret d'avoir accepté d'être rapporteurs de cette thèse ainsi que le Pr. Philippe Cuny et le D.R. Pascal Simonet d'avoir accepté d'examiner ce travail. Je remercie également la Direction Générale de l'Armement et la Région Aquitaine de m'avoir accordé un financement pour cette thèse.

Je remercie également le Pr. Marie-Cécile Ploy d'avoir participé à mon comité de thèse qui s'est réuni à la fin de ma deuxième année de doctorat.

Mes remerciements vont également à mes directeurs de thèse, le Pr. Robert Duran et le Dr. Christine Cagnon pour m'avoir encadrée tout au long de ces 3 ans et même 4 pour Christine qui aura été ma maître de stage de M2. Merci pour votre confiance et pour la liberté que vous m'avez octroyée et qui m'ont permis d'assouvir ma curiosité. Merci à Robert pour son humour et ses taquineries mais aussi et surtout pour ses conseils judicieux surtout lors de la rédaction. Et désolée, toujours pas de Nature ni de Science… Merci à Christine d'avoir vécu les grands hauts et les quelques bas de cette thèse avec moi.

Merci à Philippe Goulas de m'avoir confié divers enseignements au cours de ma thèse. Ce fût une excellente expérience, enrichissante dont je garde un agréable souvenir.

Je remercie le Pr. Céline Brochier du Laboratoire de Biométrie et Biologie Evolutive (Université Lyon 1, CNRS, UMR 5558) de m'avoir accordé un peu de son temps lors de son passage à Pau pour discuter et me conseiller quant aux analyses phylogénétiques.

Je remercie Dušica et Ana de l'institut Ruđer Bošković de Zagreb pour notre collaboration qui aura mené à la rédaction d'un article.

Je remercie la plateforme GeT de Toulouse et en particulier Olivier Bouchez pour son aide précieuse.

Un grand merci au Pr. Pierre Peyret et à son équipe pour leur accueil chaleureux lors de mon séjour clermontois pour la réalisation de la capture de gène. Merci pour votre implication dans ces expérimentations qui auront permis d'obtenir des résultats importants. J'ai pris énormément de plaisir à travailler avec vous, avec vue sur le Puy de Dôme depuis la paillasse. Merci également pour votre aide et votre disponibilité lors de l'analyse des données et de

l'écriture des résultats. Un clin d'oeil à Auriane, Clémence et Rejane pour leur bonne humeur qui aura conduit à l'écriture d'un tube...

Un grand merci à tous les membres du projet Decapage, bretons, marseillais, toulousains et palois, avec qui j'ai adoré participer aux expérimentations et aux réunions de projet. Je garde un tas d'excellents souvenirs (mémorables !) de ces rencontres, souvent fatigantes, parfois enivrantes, qui auront débuté dès le premier jour de ma thèse par un « shoot » aux sulfures. Merci pour toutes ces discussions enrichissantes et tout ce que vous avez pu m'apporter qui m'ont permis d'avancer dans ma réflexion et mon travail (sans l'ombre d'une sardine).

Je remercie bien sûr tous les membres du laboratoire, enseignants-chercheurs, ingénieurs, techniciens, post-doc et doctorants qui ont eu la gentillesse de m'apporter leur aide et leurs conseils tout au long de ces trois ans. Je remercie également Karima et Maud pour leur aide lors des séances de TP que j'ai eu à encadrer.

Je remercie Béatrice pour ses nombreux conseils avisés au cours des réunions Melody, entre autres.

Je remercie Vanessa et son mari Maxime ainsi que Elisabeth et son mari Rudy pour m'avoir invitée à l'un des plus beaux jours de leur vie.

Merci à Anne Fahy qui m'aura légué le « flambeau » des intégrons dans les milieux d'eau douce. Merci à Viviane pour son aide technique dans les manipes.

Je tiens également à remercier Cristiana de m'avoir formée à l'art de la microbiologie. Un grand merci pour ta patience mais aussi pour ton soutien, pour tes encouragements et pour t'être réjouie des avancées de ma thèse.

Merci à Fanny, coloc' de chambre d'hôtel durant ces 2 premières années de thèse, d'avoir supporté « mes blonderies » lors des missions et des expérimentations.

Je remercie ma petite troupe de stagiaires : Cora, Audrey (la parfaite-petite), Laetitia et Noémie qui m'ont fait découvrir les joies de l'encadrement et de la correction de rapports de stage. Un grand merci pour toutes ces expériences enrichissantes à vos côtés et pour votre aide lors de vos stages. Il est certain que je n'aurai pas réussi à en faire autant sans vous.

Merci à Hélène pour nos petites discussions et tous les bons moments passés. Je te souhaite plein de bonnes choses pour l'avenir.

Une pensée particulière pour Fabinou et Cyrilou, toujours à la rescousse dans mes soucis de bio-info. Merci Fabien pour tes blagues déplacées, les fous-rires et la bonne humeur dans le

bureau durant ces 2 premières années. Merci Cyrilou pour le soutien et l'écoute durant cette 3e année. Je te souhaite une bonne continuation et n'oublie pas de lever le pied des fois, le chemin est long !

Je remercie mes proches et mes amis. Ces 3 années m'ont fait comprendre l'importance de s'entourer de quelques personnes spéciales. Ils ne sont pas beaucoup mais c'est amplement suffisant. Ils sont un tout qui m'a permis de changer, de grandir et de m'épanouir.

Je tiens à remercier en particulier deux filles extra qui me sont chères.

Je remercie Julie, ma biche, ma plus vieille amie. Il est bon de nous voir évoluer depuis 10 ans et de voir ce que l'on accomplit. Je suis heureuse d'être toujours à tes côtés pour te voir devenir mère, et quelle maman tu seras ! Hâte de pouvoir jouer la tata-gaga ! ☺

Un grand merci à ma folle, ma poule, mon âme sœur Laure (que je partage avec Fab !). Merci pour tout, tu auras été l'une des plus belles rencontres de ces 3 années. Franche, généreuse et pleine de joie de vivre, d'un soutien sans failles, toujours là pour une petite sortie nanas pour décompresser et chanter avec moi « Libéréééééeee, délivréééée *» !*

Je pense également à Yannick (le bizu), à Eric, à Fafa et à la petite Laure.

Merci à ma famille.

Merci à mon père, pour son soutien indéfectible même à des milliers de kilomètres. Merci de croire en moi et de m'avoir poussée à me dépasser et permise d'aller toujours plus loin en sécurité.

Merci à mon petit frère devenu tellement grand. Je suis immensément fière de l'homme généreux et accompli que tu deviens. Peu de grandes sœurs ont la chance que j'ai d'avoir un frero tel que toi.

Un grand merci à ma mère, mon pilier.

TABLE DES MATIERES

INTRODUCTION GENERALE **1**

I- SYNTHESE BIBLIOGRAPHIQUE **7**

1) ADAPTATION BACTERIENNE ET ELEMENTS GENETIQUES MOBILES 9

2) INTEGRONS : STRUCTURE ET FONCTIONNEMENT 12

2.1. La plateforme intégrative *14*

2.1.a. Le gène *intI* et son expression 14

i. Le gène *intI* 14

ii. La régulation de l'expression du gène *intI* 15

2.1.b. Le site de recombinaison *attI* : spécifique de son intégrase associée ? 17

2.1.c. Le promoteur des cassettes Pc 18

2.2. La région variable des intégrons *20*

2.2.a. Les cassettes de gène 20

i. Fonctions codées par les gènes contenus dans les cassettes 21

ii. Expression des gènes contenus dans les cassettes 21

2.2.b. Le site de recombinaison *attC* 22

2.3. L'intégrase d'intégron : une tyrosine recombinase particulière *25*

2.3.a. Structures primaire et secondaire 25

2.3.b. Structures tertiaire et quaternaire 30

2.3.c. Le mécanisme de la recombinaison 31

3) DIVERSITE DES INTEGRONS 32

3.1. Les intégrons dits « cliniques » *32*

3.1.a. Les intégrons de classe 1 32

3.1.b. Les intégrons de classe 2 33

3.1.c. Les intégrons de classe 3 35

3.2. Diversité des intégrons dans les environnements autres que cliniques *37*

3.2.a. Diversité des gènes d'intégrase 37

3.2.b. Diversité des cassettes dans les environnements autres que cliniques 39

II- MATERIEL ET METHODES **41**

1) ORIGINE DES ECHANTILLONS 43

1.1. Sédiments d'eau douce prélevés dans le bassin du Gave de Pau *43*

1.2. Sédiments côtiers prélevés dans la région de Rovinj (Croatie) *44*

1.3. Microcosmes de sédiments d'eau douce *45*

1.4. Mésocosmes de sédiments côtiers contaminés par du pétrole *46*

1.5. Bioréacteurs de sédiments côtiers *48*

1.6. Consortia *50*

2) METHODES MOLECULAIRES 52

2.1. Extraction des acides nucléiques *52*
2.1.a. Extraction de l'ADN métagénomique de sédiments 52
2.1.b. Extraction de l'ADN métagénomique à partir des consortia 52
2.2. Construction des banques de séquences de IntI et de cassettes de gène *52*
3.2.a. Amplification des gènes *intI* et des cassettes de gène à partir du métagénome 52
3.2.b. Purification et sélection par la taille des amplicons à cloner 53
i. Purification des produits de PCR sur le gène *intI* 53
ii. Sélection selon la taille des cassettes de gène amplifiées sur gel d'électrophorèse 54
3.2.c. Ligation et transformation 54
3.2.d. Criblage des clones et amplification des inserts 54
3.2.e. Séquençage selon la méthode Sanger et analyses des séquences 55
3.2.f. Analyses phylogénétiques des séquences de IntI 55
2.3. Quantification de gènes par PCR quantitative (qPCR) *56*
2.3.a. Protocole de PCR en temps réel 56
2.3.b. Réalisation d'un standard externe pour la quantification 56
2.3.c. Test de moyennes 2 à 2 58
2.4. Caractérisation de la structure des communautés par T-RFLP *58*
2.4.a. Obtention de fragments du gène de l'ARNr 16S marqués par fluorochrome 58
2.4.b. Digestion des fragments du gène de l'ARNr 16S marqués 59
2.4.c. Migration des fragments digérés du gène de l'ARNr 16S par électrophorèse capillaire 59
2.4.d. Traitement des données brutes 59
2.5. Etude du pool de cassette de gènes selon la taille *60*
2.5.a. Amplification des cassettes de gène à partir du métagénome 60
2.5.b. Migration des cassettes de gène amplifiées et marquées par électrophorèse capillaire et traitements des données brutes 60
2.6. Caractérisation de l'environnement génétique du gène intIOPS par capture de gène et séquençage haut débit *60*
2.6.a. Préparation des banques de fragments métagénomiques 62
2.6.b. Préparation des sondes de capture ARN biotinylées 63
i. Dessin des sondes 63
ii. Préparation des oligonucléotides utilisés pour la synthèse des sondes ARN 63
iii. Transcription *in vitro* (TIV) des sondes de capture 65
iv. Purification des sondes ARN biotinylées 66
2.6.c. Capture en solution 66
i. Hybridation entre les banques d'ADN et les sondes ARN biotinylées 66
ii. Elution des cibles capturées 67
2.6.d. Traitement des cibles capturées par billes AMPure 68
2.6.e. Ré-amplification 68
2.6.f. Séquençage MiSeq (Illumina) 69
2.6.g. Analyses des données MiSeq pour la reconstitution de la séquence de l'intégron IntIOPS 69
3) ANALYSES STATISTIQUES 70

3.1. Analyses des banques et indices de diversité *70*

3.2. Analyse multidimensionnelle non-métrique (nMDS) *71*

4) ANALYSES DES SEQUENCES DE CASSETTES ET DE INTIOPS 71

4.1. Caractérisation des sites attC et attI *71*

4.2. Caractérisation des cadres de lecture *72*

4.3. Caractérisation des régions promotrices *72*

III- ETUDE DE LA DIVERSITE DES INTEGRONS DANS DES ENVIRONNEMENTS NATURELS 73

1) INTEGRONS DANS DES SEDIMENTS D'EAU DOUCE 75

1.1. Diversité des intégrons dans des communautés bactériennes de sédiments d'eau douce soumis à différents niveaux de contamination *77*

1.2. Prédiction de la structure secondaire des sites de recombinaison attC *89*

1.3. Analyse phylogénétique des séquences d'intégrase *91*

2) DIVERSITE DES INTEGRONS DANS DES SEDIMENTS COTIERS NATURELS 94

3) CONCLUSIONS 105

IV- ETUDE DE LA DIVERSITE DES INTEGRONS DANS DES ENVIRONNEMENTS CONTROLES 107

1) DYNAMIQUE DE LA DIVERSITE DES INTEGRONS DANS DES SEDIMENTS D'EAU DOUCE 109

1.1. Introduction *109*

1.2. Diversité des intégrases et des cassettes de gène comparées à la structure des communautés *109*

1.2.a. La structure des communautés 109

1.2.b. La diversité des intégrases 110

i. Effet d'une contamination par du benzo[a]pyrène 111

ii. Effet d'une contamination par de l'arsenic 113

iii. Effet d'une contamination par triméthoprime 114

iv. Suivi du gène *intI1* 114

1.2.c. La diversité des cassettes de gène 115

1.2.d. Diversité des intégrons dans les microcosmes d'eau douce 116

1.3. Analyse phylogénétique des séquences d'intégrases d'intégron *118*

1.4. Conclusion *120*

2) DYNAMIQUE DE LA DIVERSITE DES INTEGRONS DANS DES SEDIMENTS COTIERS 121

2.1. Diversité des intégrases *121*

2.2. Suivi de l'abondance du gène intIOPS *124*

2.3. Etude de la dynamique des cassettes de gène dans les bioréacteurs *125*

2.3.a. Structure de la communauté 125

2.3.b. La diversité des cassettes de gène 126

2.4. Conclusion *128*

V- CARACTERISATION D'UN INTEGRON POTENTIELLEMENT ADAPTATIF 131

1) RECONSTITUTION DE LA SEQUENCE DE L'INTEGRON INTIOPS 133

2) CARACTERISATION DES DIFFERENTS ELEMENTS DE L'INTEGRON INTIOPS....135

2.1. Caractérisation du gène de l'intégrase d'intégron IntIOPS....135

2.2. Le site attI de l'intégron IntIOPS : attIOPS....135

2.3. La région promotrice de l'intégron IntIOPS....136

2.4. Les premières cassettes associées à l'intégron IntOPS....137

2.5. L'environnement génétique de IntIOPS....139

3) CONCLUSION....139

CONCLUSIONS GENERALES....141

REFERENCES BIBLIOGRAPHIQUES....149

ANNEXES....167

LISTE DES FIGURES

Figure I-1: Structure et fonctionnement d'un intégron **14**
Figure I-2 : Représentation schématique de la levée de la répression par LexA lors de la réponse SOS impliquant LexA. **16**
Figure I-3 : Alignement des sites attI1, attI2, attI3 des intégrons de classe1, 2 et 3 respectivement et du site attI de V. cholerae (attI4). **18**
Figure I-4 : Représentation schématique de 2 cassettes de gène insérées dans un intégron. **21**
Figure I-5 : Sites de recombinaison attC. **23**
Figure I-6: Région du site additionnel spécifique des intégrases d'intégron **27**
Figure I-7 : Alignements des séquences de IntI1, IntI2, IntI2 IntI3 et IntI4.* **29**
Figure I-8 : Structure tridimensionnelle du complexe formé par les 2 dimères de VchIntIA liés au site VCR db (site attC de V. cholerae) **30**
Figure I-9 : Représentation schématique des différentes étapes de la recombinaison réalisée par l'intégrase d'intégron entre le site attI et le site attC. **31**

Figure II-1 : Localisation des sites d'échantillonnage du bassin du Gave de Pau (d'après Abella et al., en révision). **43**
Figure II-2 : Localisation des sites d'échantillonnage dans la région de Rovinj **44**
Figure II-3 : Localisation géographique du site de prélèvement dans l'aber Benoît **47**
Figure II-4 : Dispositif expérimental recréant le cycle des marées mis en place pour l'incubation de sédiments côtiers non-déstructurés sous différents traitements de remaniement du sédiment (Hangar expérimental du CEDRE, Brest). **48**
Figure II-5 : A : Schéma descriptif du plan expérimental des incubations en bioréacteurs de chaque sédiment pollué issu des mésocosmes, sous des conditions oscillantes anoxie/oxie, d'anoxie permanente ou d'oxie permanente. B : Photographie d'un bioréacteur. C : Représentation schématique d'un bioréacteur. **49**
Figure II-6 : Schéma récapitulatif de l'incubation en bioréacteurs à l'obtention de cultures simplifiées de microorganismes se développant sur hydrocarbures **51**
Figure II-7 : Représentation schématique du principe utilisé pour caractériser l'environnement du gène cible à partir du métagénome. **61**

Figure III-1 : Schematic sampling map of the Gave de Pau basin.. ..... ***Erreur ! Signet non défini.***
Figure III-2: Metal concentrations in sediment (mg.kg-1) from each sample. ..... ***Erreur ! Signet non défini.***
Figure III-3 : Non metric multidimensional scaling plots of (A) bacterial community structure assessed by 16S rRNA gene T-RFLP; (B) IntI sequences libraries; and (C) gene cassette pool assessed by size profiles; based on a distance matrix calculated from Bray-Curtis similarity. ..... ***Erreur ! Signet non défini.***
Figure III-4 : Class 1 integrase (intI1) copy number, 16S rRNA gene copy number and 16S rRNA gene/intI1 copy number ratio for the various sites as assessed by quantitative PCR.. ..... ***Erreur ! Signet non défini.***
Figure III-5 : Characterization of ORFs in gene cassettes sequenced from Mar and Effluent metagenomes. ***.Erreur ! Signet non défini.***
Figure III-6 : Non metric multidimensional scaling plots of IntI sequence libraries ..... ***Erreur ! Signet non défini.***

Figure III-7 : Structure secondaire prédictive des sites attC présents à l'intérieur des cassettes séquencées.. ..... **90**
Figure III-8 : Arbre phylogénétique des OTUs les plus abondants et des 5 premières séquences obtenues après un blastP de chaque OTUs principaux **92**

Figure IV-1 : Analyse MDS de la structure des communautés bactériennes au cours du temps selon les conditions.. ..... **110**
Figure IV-2 : Indice de Shannon calculé à partir des données de séquences IntI, de T-RFs du gène de l'ARNr 16S et sur les données de taille de cassettes de gène à T0 et pour les 4 conditions à T358 jours. **112**
Figure IV-3 : Analyse MDS des banques de séquences d'intégrases au cours du temps selon les conditions. **112**
Figure IV-4 : Comparaison du pool des cassettes de gène au cours du temps selon les conditions en se basant sur l'analyse de la taille des cassettes. **115**
Figure IV-5 : Arbre phylogénétique des OTUs d'intégrase d'intégron obtenues représentant plus de 0.6% du total des séquences et de séquences connue. **119**
Figure IV-6 : Arbres phylogénétiques réalisés à partir de l'alignement des séquences IntI obtenues à partir des mésocosmes contaminés par du pétrole après 3 mois d'incubation et 9 mois. **121**
Figure IV-7 : Abondance relative de intIOPS dans les mésocosmes au cours du temps. **122**
Figure IV-8 : Arbres phylogénétiques réalisés à partir de l'alignement des séquences IntI obtenues à partir des suspensions des bioréacteurs de la communauté Pétrole au début de l'expérimentation (A) et à la fin (B) en condition anoxie permanente en bioréacteurs **123**
Figure IV-9 : Abondance relative du gène intIOPS dans les communautés bactériennes hydrocarbonoclastes au moment du prélèvement du sédiment pour la mise en place des bioréacteurs et dans le mésocosme contrôle sans contamination à la fin de l'expérimentation. **124**
Figure IV-10 : Abondance relative du gène intIOPS dans les bioréacteurs **125**
Figure IV-11 : Comparaison de la structure des 3 communautés au début et à la fin de l'expérimentation en bioréacteurs selon les conditions d'oxygénation **126**
Figure IV-12 : Comparaison de la diversité des cassettes de gène dans les 3 communautés au temps final selon les conditions d'oxygénations en se basant sur l'analyse de la taille des cassettes. **127**

Figure V-1 : Représentation schématique des différents éléments de l'intégron IntIOPS caractérisés. **134**
Figure V-2 : Alignement des séquences des sites attIOPS V1, V2 et V3 de IntIOPS et des sites attI des intégrons de classe 1, 2 et 3 et de Vibrio cholerae. **135**
Figure V-3 : Région située entre le début du gène intIOPS (en bleu) et le codon d'initiation de l'orf1 (en vert) de la région variable V1 caractérisée.. ..... **136**
Figure V-4 : Alignement de la région variable V1 de IntIOPS et des séquences génomiques de la souche δ-proteobacterium NaphS2. ..... **138**
Figure V-5 : Alignement de la séquence de l'orf5 de IntIOPS **139**

LISTE DES TABLEAUX

Tableau I-1 : Exemples de fonctions trouvées dans des éléments génétiques mobiles..... 10

*Tableau I-2 : Efficacité de l'excision des intégrases IntI1, IntI2*179E, IntI3 et SonIntIA au niveau des différents mutants du site attC_{dfrA1}. 24*

Tableau I-3 : Résidus catalytiques des tyrosine-recombinases (d'après Van Duyne 2002). 25

Tableau II-1 : Caractéristiques physico-chimiques des sédiments prélevés dans la région côtière de Rovinj. 45

Tableau II-2 : Concentrations finales des polluants dans les microcosmes, basées sur les données environnementales de la littérature 46

Tableau II-3 : Récapitulatif des amorces utilisées en PCR et qPCR. 53

Tableau II-4 : Séquences des oligonucléotides utilisés pour la réalisation des sondes utilisées pour la capture de gène de intIOPS. 64

Tableau II-5 : Couples d'amorces utilisés lors de la préparation de sondes et pour la ré-amplification des banques après capture 65

Tableau III-1 : Primers used in conventional and real time PCRs, expected amplicon size and annealing temperature ..... Erreur ! Signet non défini.

Tableau III-2 : Metal concentrations (mg.kg-1, dry weight) in sewage sludge from this study and from the literature ..... Erreur ! Signet non défini.

Tableau III-3 : Shannon diversity based on T-RFLP of 16S rRNA gene and gene cassette size profiles .... Erreur ! Signet non défini.*

Tableau III-4 : IntI library coverage and Shannon diversity (considering a 98% amino-acid identity threshold)..... Erreur ! Signet non défini.

Tableau IV-1 : Tableau descriptif des banques de séquences IntI. 111

Tableau V-1 : Récapitulatif du nombre et de la qualité des séquences pour chaque banque avant et après prétraitement. 134

Introduction générale

Les microorganismes, et plus particulièrement les bactéries, sont des organismes ubiquistes. La grande variété de métabolisme qu'ils possèdent leur permet de coloniser un grand nombre d'environnements, même les plus extrêmes y compris les milieux pollués. Une contamination du milieu, par des hydrocarbures, des métaux ou des antibiotiques, impose une pression de sélection importante et aura pour effet de modifier les communautés bactériennes du milieu. En effet, les bactéries qui possèdent un métabolisme leur permettant de résister aux nouvelles conditions environnementales voire même d'en tirer profit seront plus compétitives dans la communauté. Ces bactéries portent des gènes dits adaptatifs qu'elles peuvent posséder dans leur génome ou bien acquérir par mutation ou par transfert horizontal par exemple. Le transfert horizontal de gène est un processus important dans l'évolution des génomes bactériens faisant intervenir des structures appelées éléments génétiques mobiles tels que les plasmides, les phages, les transposons ou encore les intégrons.

Les intégrons sont connus pour être de formidables acteurs dans l'adaptation des bactéries face à l'utilisation des antibiotiques dans les milieux cliniques. Découverts dans ce contexte, il a pu être montré que leur diversité était homogène. En effet, on ne retrouve seulement que 3 séquences de l'intégrase et une centaine de cassettes de gène codant essentiellement des résistances à des antibiotiques. Les intégrons sont présents dans le génome de nombreuses bactéries et retrouvés dans le métagénome bactérien de très nombreux environnements, autres que cliniques, quel que soit leur niveau d'anthropisation. Structures génétiques bactériennes anciennes, datant d'avant l'utilisation des antibiotiques en thérapie, leur potentiel rôle adaptatif en dehors du contexte clinique n'est aujourd'hui pas élucidé. Ainsi, le but de ma thèse a été d'évaluer l'implication des intégrons environnementaux (*i.e*. en dehors du contexte clinique) dans l'adaptation bactérienne.

A l'origine des travaux présentés dans cette thèse, une importante diversité des intégrons dans des environnements autres que les milieux cliniques était suspectée que cela soit au niveau de l'intégrase ou des cassettes de gènes. Cette diversité pourrait être influencée par les conditions environnementales. Au niveau de l'intégrase, environ 400 classes -ou

phylotypes- d'intégrase ont déjà été identifiées à partir de banques de séquences issues d'ADN métagénomique ou grâce à l'analyse des génomes contenus dans les bases de données. Une première analyse phylogénétique des séquences d'intégrase de divers environnements a pu mettre en évidence une distinction entre les intégrases de milieux marins et les intégrases de milieux terrestres (sol et eau douce). De plus, la comparaison de banques de séquences entre 2 ports avec une intense activité humaine a montré des diversités divergentes des intégrases, même si les 2 sites pouvaient être comparables en termes de sources de contamination et de conditions environnementales. Ceci suggèrent une possible influence géographique sur la diversité des intégrases. Enfin, Lionel Huang a montré une baisse de la diversité des intégrases face à une contamination par du pétrole, associée à l'augmentation de l'abondance d'une séquence particulière, nommée IntIOPS. L'ensemble de ces observations suggère donc un impact des conditions environnementales sur la diversité des intégrases.

Concernant les cassettes de gène, un effort plus important de séquençage a été réalisé et a permis de découvrir au total 5000 séquences à partir d'environnements variés. Or, ces études ont montré que bien que le nombre de cassettes séquencées était important, cela n'était pas suffisant pour appréhender la diversité des cassettes dans sa globalité. Ainsi, une estimation indiquerait qu'un échantillon donné peut contenir entre 2000 et 3000 cassettes de gène différentes dont la plupart coderait des fonctions encore inconnues. Ceci montre l'importante diversité de fonctions que représente le pool de cassettes de gène environnementales qui est à ce jour majoritairement inexploré et inconnu. Par ailleurs, les études sur la diversité des cassettes dans l'environnement ont montré que le pool de cassettes pouvait être influencé par le milieu et les conditions environnementales et l'hypothèse de l'existence d'écotypes de cassettes a été émise. Enfin, des fonctions codées par les cassettes de gène ont pu être identifiées et certaines sont en lien avec les contaminants présents dans le milieu. Ces études suggèrent donc que les intégrons, grâce à leurs capacités de capture et d'expression des fonctions portées par les cassettes de gène, pourraient avoir un rôle dans la réponse adaptative des communautés bactériennes aux conditions environnementales et en particulier à la présence de polluants.

Les objectifs de cette thèse étaient de mieux comprendre et appréhender l'immense diversité des intégrons environnementaux et les facteurs qui peuvent influencer leur diversité. Pour cela, dans un premier temps, nous avons cherché à évaluer la diversité des séquences d'intégrase à partir de différents environnements avec différents types et niveaux de contamination afin d'enrichir nos connaissances et réaliser une base de données de séquences d'intégrase plus conséquente. Puis, nous avons cherché à évaluer l'impact d'une contamination chimique du milieu sur la diversité des intégrons. Pour cela, des études *in situ* et en microcosmes ont été menées. Ces études ont été réalisées dans deux types d'environnements différents, eau douce et côtiers, et avec différents types de contaminants afin de voir si les conclusions obtenues peuvent être généralisées. Les études en microcosmes avaient également pour objectifs de nous permettre de découvrir des séquences dont l'abondance augmente au cours du temps face à un polluant donné, afin d'identifier d'éventuels nouveaux intégrons adaptatifs. Enfin, le dernier objectif de cette thèse était de caractériser l'intégron portant le gène *intIOPS*, découvert par Lionel Huang lors de sa thèse, afin de comprendre pourquoi cette séquence devient prépondérante face à une contamination par du pétrole.

I- Synthèse bibliographique

1) Adaptation bactérienne et éléments génétiques mobiles

Les procaryotes sont des organismes capables de coloniser l'ensemble des environnements terrestres, même les plus extrêmes, grâce une grande variété de métabolismes. De plus, les communautés bactériennes sont capables de se maintenir face à des modifications de l'environnement, même drastiques, grâce à différents mécanismes d'adaptation qu'elles ont développés. Outre la sélection des bactéries les mieux adaptées ou l'expression de fonctions permettant une meilleure réponse aux nouvelles conditions environnementales, un moyen pour les bactéries de s'adapter sont des modifications de leur génome. Ces modifications peuvent résulter de mutations spontanées. Mais une façon d'acquérir de nouvelles fonctions plus rapidement est le transfert horizontal (Springael et Top, 2004). Celui–ci peut être défini comme étant le transfert de matériel génétique entre deux organismes d'une même génération, qui peuvent être d'espèces ou de genres différents. Le transfert horizontal fait intervenir des éléments génétiques mobiles divers qui vont être transférés/échangés selon plusieurs mécanismes. Chez les bactéries, l'entrée d'ADN exogène peut se faire selon 3 mécanismes : la conjugaison, par transfert de plasmides conjugatifs ; la transduction, par infection phagique ; ou la transformation permettant l'entrée d'ADN nu (Frost *et al.,* 2005). Une fois l'ADN exogène rentré dans la cellule receveuse, il doit pouvoir se maintenir. Si le matériel génétique ainsi transféré est situé sur un plasmide, il sera répliqué de manière autonome et pourra être maintenu dans le patrimoine génétique de la bactérie. Dans le cas contraire, il ne sera maintenu que s'il s'intègre sur le chromosome bactérien (Thomas et Nielsen, 2005). Cette intégration peut se faire par recombinaison (Lovett *et al.,* 2002), mais plus généralement par le biais d'éléments génétiques mobiles tels que les phages lysogéniques, les transposons ou encore les intégrons (Bordenstein et Reznikoff, 2005). Toutefois, il est important pour la survie d'un organisme et de sa descendance que l'insertion de matériel génétique n'altère pas l'intégrité du génome. Ceci nécessite donc des structures génétiques élaborées tels que les intégrons au niveau desquels l'ADN exogène va pouvoir être intégré sans perturber le reste du génome.

Ainsi, les éléments génétiques mobiles prennent une part non négligeable dans l'adaptation des bactéries. Ceci est illustré par de nombreux exemples de gènes potentiellement

adaptatifs contenus dans ces structures génétiques. Le Tableau I-1 présente une liste non exhaustive de tels gènes répondant à des contaminations chimiques et organiques abordées dans cette thèse, comme les métaux, les antibiotiques, les molécules de la famille des BTEX (benzène, toluène, éthylène, xylène), ou encore les hydrocarbures aromatiques polycycliques (HAP).

Tableau I-1 : Exemples de fonctions trouvées dans des éléments génétiques mobiles

Gène(s)	Eléments génétiques mobiles		Fonctions codées	Références
	Nom	Type		
Résistance à des antibiotiques				
accC1	Classe 1	intégron	Résistance à la gentamicine	(Stokes et Hall, 1991)
aadA2			Résistance à la streptomycine et à la spectinomycine	
cmlA			Résistance au chloramphénicol	
dfrA1	Classe 2	intégron	Résistance au triméthoprime	((Hansson *et al.,* 2002))
sat			Résistance à la streptothricine	
aadA1			Résistance à la spectinomycine	
mef	Tn2009	transposon	Résistances aux macrolides	(Grosso *et al.,* 2007)
cat	Tn5252	transposon	Résistance au catéchol	(Korona-Glowniak *et al.*, 2015)
Gènes *bla*	pM91	plasmide conjugatif	Résistance aux β-lactamases	(Giakkoupi *et al.,* 2015)
Dégradation de polluants organiques				
Gènes *xyl*	pWW53	plasmide	Catabolisme du toluène	(Yano *et al.,* 2007)
bphBC	pNL1	plasmide	Catabolisme du biphényl	(Romine *et al.,* 1999)
bphABCDHIJK	pSK4	plasmide	Catabolisme de PCB	(Ilori *et al.,* 2015)
smr	pSM52	plasmide	Pompe à efflux des composés ammonium quaternaires	(Costa *et al.,* 2013)
cbaABC	Tn5271	transposon	Catabolisme de chlorobenzoate	(Nakatsu *et al.,* 1991)
tcbAB	Tn5280	transposon	Catabolisme du chlorobenzène	(Van der Meer *et al*., 1991)
bedDC1C2BA	Tn5542	transposon	Catabolisme du benzène	(Fong *et al.*, 2000)
		cassettes de gène	Catabolisme de BTEX, HAP	(Koenig *et al.,* 2009)
Résistance aux métaux				
opéron *mer*	pRYC103T24.1	plasmide	Résistance au mercure	(Novais *et al.,* 2010)
opéron *ars*	pI258	plasmide	Résistance à l'arsenic	(Ji et Silver, 1992)
opéron *mer*	Tn21	transposon	Résistance au mercure	(Liebert *et al.,* 1999)
opéron *ars*	Tn21	transposon	Résistance à l'arsenic	(Tuffin *et al.,* 2005)

L'acquisition de résistances à des antibiotiques est le mécanisme d'adaptation impliquant des éléments génétiques mobiles le mieux étudié chez les bactéries. Les gènes de résistance à des antibiotiques sont nombreux et très largement disséminés au sein des communautés bactériennes dans le contexte clinique mais aussi dans l'environnement. Les communautés bactériennes environnementales peuvent alors jouer le rôle de réservoirs de gènes de

résistance et de lieu pour leur propagation (Biyela *et al*., 2004). L'acquisition et la dissémination de multi-résistances à des antibiotiques représentent un problème sanitaire majeur qui ne se limite donc pas seulement aux bactéries pathogènes dans les milieux hospitaliers. Les éléments génétiques mobiles portant ces gènes de résistance sont divers : on retrouve parmi eux des plasmides, des transposons et aujourd'hui surtout des intégrons en milieu clinique et dans des contextes environnementaux anthropisés (Tableau I-1 ; (Cambray *et al.,* 2010; Gillings, 2014; Stalder *et al.,* 2014) montrant que ce sont de formidables outils pour l'adaptation bactérienne.

Dans l'environnement, les bactéries peuvent être soumises à d'autres pressions de sélection exercées par des molécules organiques comme les BTEX, les HAP ou les pesticides. Ces composés peuvent être métabolisables par certaines bactéries et cette capacité peut représenter un avantage sélectif. Plusieurs gènes clefs peuvent être impliqués dans ce catabolisme et peuvent être retrouvés ensemble sur des éléments génétiques mobiles sous forme d'opérons ou non (Tableau I-1). Même si l'acquisition et la dissémination de gènes de dégradation est un sujet moins étudié que celui des résistances à des antibiotiques, il semble que les éléments génétiques mobiles impliqués dans l'acquisition et la dissémination de telles voies métaboliques soient de même nature que ceux impliqués dans l'émergence de multi-résistances à des antibiotiques (Tableau I-1).

Les métaux peuvent être considérés comme des contaminants même si certains métaux sont nécessaires au métabolisme des bactéries, comme cofacteurs enzymatiques, composants de macromolécules comme les cytochromes ou accepteurs d'électrons par exemple. En effet, au-delà d'une certaine concentration propre à chaque métal, ils sont toxiques pour les organismes. Les bactéries doivent alors mettre en place des mécanismes de résistance pour survivre. Ces mécanismes peuvent être des pompes à efflux ou bien des transformations de type méthylation. Certains gènes de résistance à des métaux sont retrouvés sur des éléments génétiques mobiles tels que des plasmides ou des transposons (Tableau I-1).

Le rôle d'éléments génétiques mobiles tels que les plasmides ou les transposons dans l'adaptation bactérienne a été largement étudié. Nous savons aujourd'hui qu'ils sont des

acteurs importants dans la réponse adaptative aux contaminations du milieu. Les intégrons sont quant à eux considérés aujourd'hui comme les principaux acteurs responsables de l'acquisition et de la dissémination de résistances à des antibiotiques en particulier chez les pathogènes Gram- (Gillings, 2014) mais aussi chez les Gram+ (Deng *et al.,* 2015). Toutefois, même si ils ont été rencontrés dans de nombreux environnements non cliniques, peu de chose est connu quant à leur implication dans l'adaptation des communautés bactériennes aux modifications environnementales.

2) Intégrons : structure et fonctionnement

Les intégrons ont été découverts à la fin des années 80 chez des bactéries pathogènes Gram- (Stokes et Hall, 1989). Toutefois, ce sont des structures anciennes, présentes chez les bactéries bien avant l'utilisation des antibiotiques à des fins thérapeutiques (Mazel, 2006). Les intégrons sont des éléments génétiques rencontrés chez certaines bactéries (environ 10 à 17% des bactéries dont le génome a été séquencé (Boucher *et al.,* 2007; Cambray *et al.,* 2010) et ont été identifiés dans de nombreux environnements différents, aux niveaux d'anthropisation variables (Elsaied *et al.,* 2007, 2011, 2014; Hardwick *et al.,* 2008; Wright *et al.,* 2008; Rosewarne *et al.,* 2010; Byrne-Bailey *et al.,* 2011; Gaze *et al.,* 2011). Ils sont capables d'acquérir et de réarranger des cadres de lecture présents dans des structures appelées cassettes de gène et de les convertir en des gènes fonctionnels en assurant leur expression (Cambray *et al.,* 2010).

Les intégrons sont composés d'une partie comprenant un gène *intI* codant une intégrase IntI, suivi d'un site de recombinaison *attI*, au niveau duquel sont intégrées les cassettes de gène (Figure I-1). Ces deux éléments représentent une structure de base simple et constante qui fait des intégrons une plateforme efficace pour l'intégration site-spécifique de matériel génétique nouveau et potentiellement adaptatif. Le reste de l'intégron est une région variable puisqu'il contient une succession de cassettes de gènes dont les séquences peuvent être très diverses en taille et en nombre. Ces cassettes sont composées de région codante, généralement une par cassette, associées à un site de recombinaison *attC* (Hall et Collis,

1995 ; Figure I-1). L'intégrase IntI codée par les intégrons réalise une recombinaison particulière capable d'interagir avec 2 sites de recombinaison différents, contrairement aux autres tyrosines recombinases. Ainsi, la réaction d'intégration d'une cassette de gène intervient généralement entre un site *attC* compris dans une cassette de gène circulaire et le site *attI* de l'intégron (Collis *et al.,* 1993 ; Figure I-1), alors que l'excision d'une cassette intervient préférentiellement par recombinaison entre deux sites *attC* (Collis *et al.,* 1993 ; Gestal *et al.,* 2011 ; Figure I-1).

De par la grande diversité des fonctions portées par les cassettes de gène, les intégrons représentent un atout non négligeable pour les bactéries porteuses, leur permettant une plasticité du génome importante par l'acquisition de gènes potentiellement adaptifs sans perturbation du génome receveur. De plus, les intégrons sont des structures qui peuvent être mobiles quand ils sont portés par des plasmides, par des transposons ou dans des îlots génomiques (Gillings, 2014). Enfin lorsque que l'intégron est chromosomique, un gène codant une transposase ou une recombinase a régulièrement pu être identifié à proximité (Gillings, 2014) suggérant que même si ils ne sont pas portés par des plasmides ou des transposons, ils pourraient se disséminer. Ainsi, ils peuvent donc s'avérer être de formidables acteurs dans l'adaptation des communautés bactériennes.

Les intégrons ont été largement étudiés dans les milieux cliniques où 3 classes d'intégrons (classe 1, 2 et 3) sont principalement retrouvées et définies selon la séquence de leur intégrase. Ces 3 classes d'intégrons ont la caractéristique commune de porter essentiellement des cassettes de gène codant des résistances à des antibiotiques. D'autres intégrons ont été également bien caractérisés : il s'agit entre autre des intégrons chromosomiques de *Vibrio cholerae*, VchIntIA et de *Shewanella oneidensis*, SonIntIA.

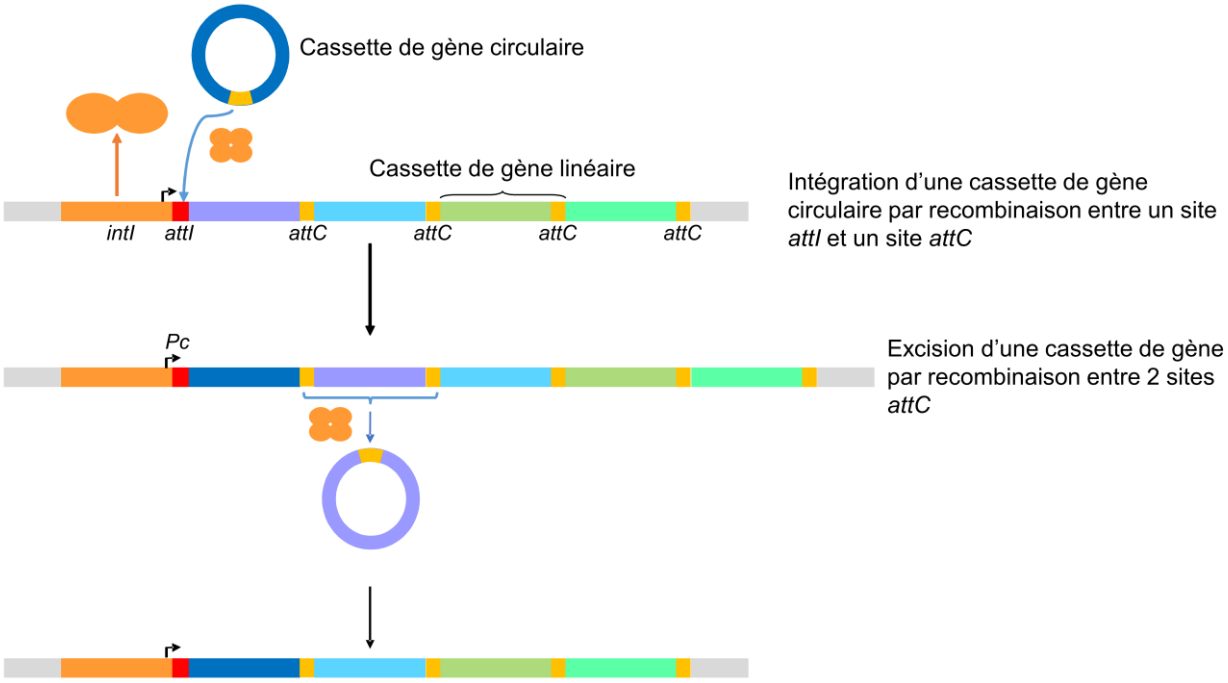


Figure I-1: Structure et fonctionnement d'un intégron (d'après Abella *et al.,* 2015).

2.1. La plateforme intégrative

2.1.a. Le gène *intI* et son expression

i. Le gène *intI*

Le premier élément des intégrons est le gène *intI* qui code l'intégrase IntI. Généralement, le gène *intI* a une orientation anti-sens par rapport aux cassettes, mais il peut être orienté en sens inverse chez des intégrons retrouvés chez certaines *Spirochaetes*, *Plantomycetes*, *Chlorobi* et *Cyanobacteria* (Boucher *et al.,* 2007).

Le gène *intI* présente des variations dans sa séquence (Mazel, 2006). De ce fait, les intégrases d'intégron peuvent ne partager que 35% d'identité protéique, ce qui indique une longue évolution de ces séquences (Larouche et Roy, 2011). C'est par ailleurs en se basant sur ces variations que différentes classes d'intégrons, notamment chez les bactéries pathogènes, ont commencé à être distinguées de manière arbitraire selon un seuil de 98% d'identité protéique (Hall, 2012). Toutefois, ce seuil n'étant pas toujours respecté pour la définition de nouvelles séquences (Elsaied *et al.,* (2011) se basent sur un seuil de 100% d'identité

protéique), il n'existe donc pas de véritable règle pour la classification des intégrons et de leur intégrase, principalement en dehors du contexte clinique.

Enfin, la comparaison des arbres basés sur le gène *intI* et ceux basés sur l'ARNr 16S ou le gène *rpoB* des bactéries porteuses d'intégrons, montre une phylogénie non congruente, ce qui suggère que ces éléments sont soumis au transfert horizontal (Boucher *et al.,* 2007; Gillings, 2014).

ii. La régulation de l'expression du gène *intI*

L'expression du gène *intI* est assurée par le promoteur *Pint*. Chez certains intégrons, ce promoteur est régulé par la réponse SOS (Guerin *et al.,* 2009; Cambray *et al.,* 2011; Krin *et al.,* 2014).

La réponse SOS est un système de régulation retrouvé dans un grand nombre de bactéries répondant aux dommages de l'ADN. Cette réponse est induite lorsqu'une quantité anormale d'ADN monocaténaire est présente dans la cellule (Baharoglu et Mazel, 2014). Lorsque la bactérie n'est pas soumise à un stress entraînant des dommages de l'ADN, la protéine LexA est fixée au niveau d'un site appelé « LexA box », situé à proximité du site de fixation de l'ARN polymérase des gènes régulés par la réponse SOS. La présence de LexA sur son site de liaison va empêcher la fixation de l'ARN polymérase et ainsi réprimer l'expression de ces gènes. Le site de fixation de LexA est un motif de l'ADN qui consiste en un palindrome de 16 pb, CTGTATATATATACAG chez *E. coli* (Walker, 1989). Lorsque la réponse SOS est enclenchée par une quantité anormale d'ADN simple brin présent dans la cellule, une seconde protéine impliquée dans la réponse SOS, la protéine RecA, est activée par sa fixation sur l'ADNsb. La protéine RecA activée va induire l'auto-clivage de LexA, qui va alors perdre sa capacité de répression (Baharoglu et Mazel, 2014).

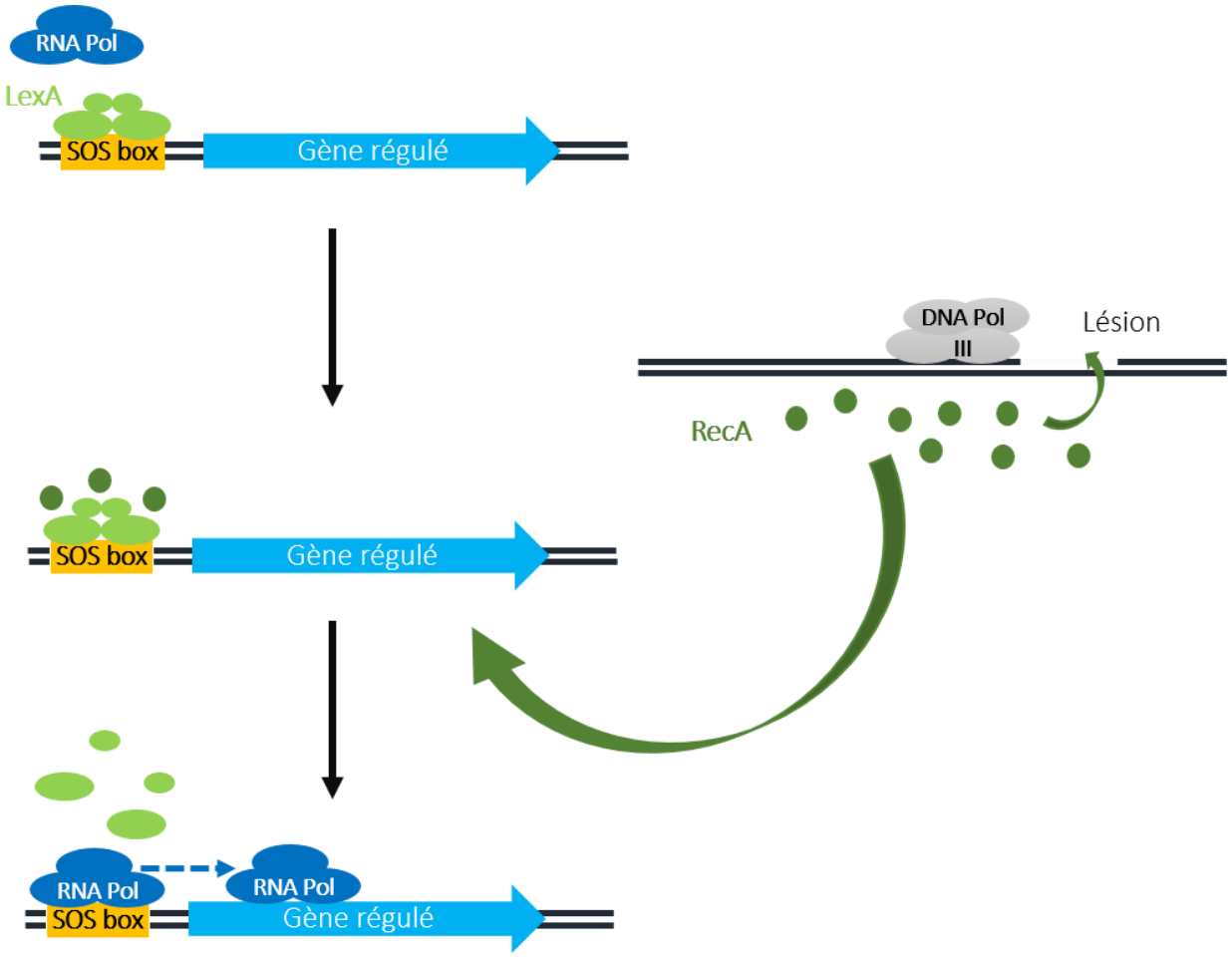


Figure I-2 : Représentation schématique de la levée de la répression par LexA lors de la réponse SOS impliquant LexA.

En 2009, Guerin et ses collaborateurs ont démontré la présence d'un motif conservé entre les positions -10 et +1 du promoteur PintI des intégrons de plusieurs *Vibrionaceae* et des intégrons de classes 1, 2, 3. Ce motif présente une forte homologie avec le site de liaison de LexA de *E. coli*, fortement conservé chez les β et γ-*proteobacteria.* Une étude sur les séquences de la région promotrice des intégrases d'intégron disponibles dans les bases de données en 2011 a permis de démontrer qu'un motif homologue au site de liaison de LexA était présent chez de nombreux intégrons (Cambray *et al.,* 2011). Ce motif est répandue parmi les intégrons des *Vibrionaceae* et aussi chez les intégrons des environnements marins (Cambray *et al.,* 2011). Chez les bactéries provenant d'environnement d'eau douce ou de sol dont les intégrases forment un groupe phylogénétique, il semblerait que la régulation par LexA était également présente chez l'ancêtre de ces intégrons (Cambray *et al.*, 2011). Néanmoins, le site de liaison de LexA est moins conservé et semble avoir été perdu au fur et

à mesure, à l'exception notable des intégrons de classes 1 et 3 (Cambray *et al*., 2011). La régulation du promoteur de *intI* par le système SOS a été démontrée *in vitro* pour l'intégrase *intI1* de *E. coli* (Guerin *et al.*, 2009 ; Cambray *et al*., 2011), ainsi que pour les intégrases d'intégrons chromosomiques de *V. cholerae* (Guerin *et al.,* 2009; Baharoglu *et al.,* 2010; Baharoglu *et al.,* 2012) et de *V. parahaemolyticus* (Cambray *et al.,* 2011). La levée de la répression par LexA de l'expression de l'intégrase a également été démontrée lors de l'entrée d'ADN simple brin dans la cellule par conjugaison ou transformation (Baharoglu *et al.*, 2010 ; 2012).

Chez *E. coli* et *V. cholerae* des antibiotiques comme les fluoriquinolones, les β-lactames, le triméthoprime et le sulfamétoxazol induisent la réponse SOS (Beaber *et al.,* 2004; Blázquez *et al.,* 2012). Si une régulation induite par les antibiotiques intervient également dans l'expression des intégrases d'intégron, cela permettrait d'expliquer pourquoi ces structures sont devenues si efficaces pour l'acquisition de gènes de résistance à des antibiotiques.

Par ailleurs, il a été montré chez *V. cholerae* que l'expression de l'intégrase VchIntIA est également régulée par la répression catabolique qui met en jeu la fixation du complexe cAMP-CRP entre les positions -100 et -79 pb en amont du site d'initiation de la transcription de *VchintIA*. Ceci suggère que l'expression de cette intégrase pourrait être également influencée par les sources de carbone présentes dans l'environnement (Baharoglu *et al*., 2012).

Enfin, la présence de transcrits de *intI* a été recherchée dans des sédiments avec ou sans contamination par du pétrole. Ils n'ont pu être détectés que dans les microcosmes pollués, à certains temps. Ceci suggère que l'expression du gène *intI* pourrait être induite par une pollution pétrolière ou un stress dérivé de cette pollution (Huang, 2010).

2.1.b. Le site de recombinaison *attI* : spécifique de son intégrase associée ?

Le site de recombinaison *attI,* d'une longueur moyenne de 65 pb (Recchia, Stokes and Hall 1994; Partridge *et al.,* 2000), est peu conservé. Il est composé de 2 à 4 sites de fixation de l'intégrase : deux séquences répétées inverses L et R respectivement de 11 pb constituent le site cœur indispensable à la recombinaison (Collis *et al.,* 1998; Gravel, Fournier and Roy 1998;

Partridge *et al.,* 2000). Deux autres domaines DR1 et DR2 ont été identifiés dans un premier temps chez les intégrons de classe 1 mais des homologues ont pu être également caractérisés sur le site *attI3* des intégrons de classe 3 (Collis et Hall, 2004; Cambray *et al.,* 2010); Figure I-3). Les séquences L et R sont séparées par une région variable de 5 à 8 pb. Le site de liaison L est moins conservé que le site R et donc difficilement identifiable (Figure I-3). Le point de recombinaison est localisé dans la séquence R entre la guanine et les 2 thymines dans un triplé 5'-G↓TT-3' conservé (Hall *et al.,* 1991; Partridge *et al.,* 2000; Bouvier et al., 2005). Une délétion des nucléotides GT de ce motif rend la recombinaison impossible (Partridge *et al.,* 2000).

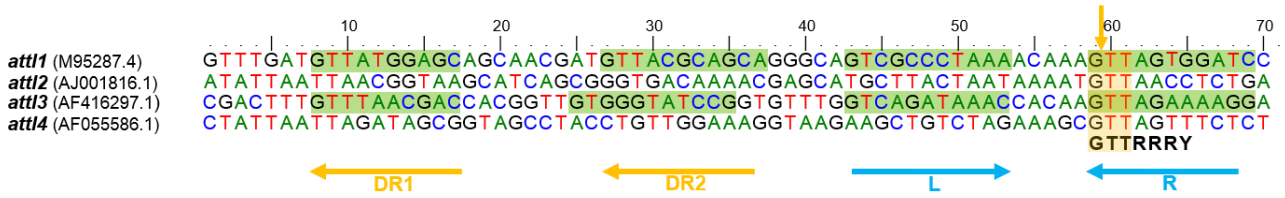


Figure I-3 : Alignement des sites *attI1*, *attI2*, *attI3* des intégrons de classe1, 2 et 3 respectivement et du site *attI* de *V. cholerae* (*attI4*). Les positions putatives des sites L et R, DR1 et DR2 sont représentés par des flèches horizontales. La séquence R est caractérisée par un motif GTTRRRY. Les sites de liaison des intégrases identifiés chez attI1 et attI3 sont surlignés en vert. Surligné en jaune, le triplé conservé GTT et le point de recombinaison indiqué d'une flèche verticale. *attI1*, *attI2*, *attI3* et *attI4* correspondent aux sites attI des intégrons de classes 1, 2, 3 et de VchIntIA respectivement.

Les sites *attI* diffèrent significativement d'un intégron à un autre, mais il semblerait que cette évolution se fasse de manière dépendante de l'intégrase à laquelle le site *attI* est lié (Cambray *et al.,* 2010). Une étude de recombinaisons croisées réalisées entre différentes intégrases et différents sites *attI* associés a montré que les intégrases reconnaissent préférentiellement le site *attI* qui leur est associé mais certaines peuvent également reconnaître le site *attI* d'autres intégrons. Ainsi, IntI1 de *E.coli* présente une grande versatilité puisqu'elle est capable de réaliser une réaction de recombinaison au niveau des sites *attI2* et *attI3* (Collis *et al.,* 2002) mais ne reconnait par le site *attI* associé à VchIntIA (Biskri *et al.,* 2005).

2.1.c. Le promoteur des cassettes Pc

Le troisième élément de la partie constante des intégrons est la région promotrice des cassettes, Pc, située généralement dans la région 5' du gène *intI* (Figure I-1). Un ou deux promoteurs, nommés Pc1 et Pc2, peuvent être retrouvés dans cette région (Cambray *et al.,*

2010 ; Vinué *et al.,* 2011). Les cassettes de gène étant généralement composées uniquement d'un cadre de lecture et d'un site *attC*, l'expression du gène est assurée par un de ces promoteurs externes. La majorité des études sur l'expression des gènes des cassettes ont été réalisées sur les intégrons de classe 1, où l'expression des cassettes est assurée par l'un des 2 promoteurs : Pc1, localisé dans le gène *intI1* ou Pc2, localisé dans le site *attI1* (Cambray *et al.,* 2010; Vinué *et al.,* 2011). Toutefois, des promoteurs assurant l'expression de cassettes de gène ont également été retrouvés dans les régions *attI* des intégrons de classe 2 (Lévesque *et al.,* 1994) et dans les gènes *intI* des intégrons de classe 3 et chez *Pseudomonas stutzeri* (Collis et Hall 1995; Coleman et Holmes 2005). Il semblerait donc que tous les intégrons portent un promoteur pour l'expression des cassettes au sein de la région *intI-attI*. Le promoteur Pc peut varier en séquence et ce polymorphisme va pouvoir affecter la séquence protéique de IntI1. Ainsi, il existe une corrélation inverse entre l'activité du promoteur et l'activité de l'intégrase (Cambray *et al.,* 2010; Jové *et al.,* 2010; Guérin *et al.,* 2011).

Chez les intégrons, le taux d'expression n'est pas le même pour tous les gènes portés par les cassettes : la cassette de gène la plus proche du site *attI* est la plus exprimée (Collis et Hall, 1995; Cambray *et al.,* 2010; Coyne *et al.,* 2010). Pour expliquer cela, plusieurs hypothèses ont été émises concernant des mécanismes modulant cette expression au niveau de la transcription et/ou de la traduction des cassettes de gène.

Chez les intégrons de classe 1, les transcrits à partir des promoteurs Pc peuvent couvrir plusieurs cassettes mais pas toujours le même nombre en fonction du réarrangement des cassettes (Collis et Hall, 1995; Cambray *et al.,* 2010). Ces résultats suggèrent que parmi les cassettes de gène, il pourrait y avoir des terminaisons prématurées de la transcription. Il a ainsi été proposé que le repliement de certains sites *attC* peut agir comme des terminateurs de transcription rho-indépendants (Collis et Hall, 1995).

Au niveau traductionnel, des sites de fixation de ribosomes (RBS) ont pu être identifiés pour certaines cassettes de gène mais pas pour toutes. La traduction des gènes est donc dépendante de la présence d'un motif RBS en amont des cassettes de gène en question. Chez les intégrons de classe 1, un court ORF précédé d'un RBS fonctionnel est localisé dans le site

attI1 et est retrouvé chez tous les transcrits à partir du promoteur Pc. Cet ORF nommé *orf1 1* chevauche le point de recombinaison du site *attI1*. Ainsi l'extrémité 3' de cet ORF dépend de la première cassette insérée. Lorsque la première cassette apporte un codon stop situé dans le même cadre de lecture que l'*orf1 1*, le complexe traductionnel responsable de l'expression de l'ORF1 1 va pouvoir progresser le long du gène porté par la cassette et le traduire en même temps que l'ORF1 1 (Hanau-Berçot *et al.,* 2002). La présence d'un court ORF dans le site *attC* pourrait également permettre l'augmentation de l'expression en modulant la processivité du complexe traductionnel (Jacquier *et al.,* 2009). Par ailleurs, une analyse mutationnelle des sites *attC* a révélé que la déstabilisation de la structure tige-boucle qu'ils pourraient prendre en se repliant, ne modifierait pas le taux de transcription des gènes situés en aval du site mais pourrait améliorer l'expression de ces gènes au niveau de la traduction. Le repliement de certains sites *attC* pourrait donc gêner la progression du ribosome le long des transcrits de plusieurs cassettes de gène, réduisant la traduction des peptides codés par les cassettes en aval (Jacquier *et al.,* 2009).

2.2. La région variable des intégrons

2.2.a. Les cassettes de gène

Les cassettes de gène sont de courts éléments mobiles généralement monogéniques composées d'un cadre de lecture et d'un site de recombinaison appelé *attC* qui est spécifiquement reconnu par IntI (Cambray *et al.,* 2010). Les cassettes de gène possèdent une organisation générale qui est la suivante : une section du site *attC* coupée durant l'insertion dans l'intégron comprenant le motif conservé GTTRRRY (avec R une purine, et Y une pyrimidine) ; une courte région non codante, généralement de moins de 10 nucléotides, qui peut contenir un site de fixation du ribosome ; un cadre de lecture encadré par un codon d'initiation de la traduction (ATG, GTG ou TTG) et un codon STOP, souvent situé dans le site *attC* ; et enfin le reste du site *attC* (Figure I-4). Des variations de cette structure générale peuvent être observées et impliquent principalement le nombre et l'orientation des cadres de lecture. La majorité des cassettes de gène contiennent un seul cadre de lecture, orienté de gauche vers la droite, mais des cassettes avec 2 cadres de lecture voir plus, sans cadre de

lecture ou avec un cadre de lecture orienté dans l'autre sens sont connues (Stokes *et al.,* 2001; Holmes *et al.,* 2003).

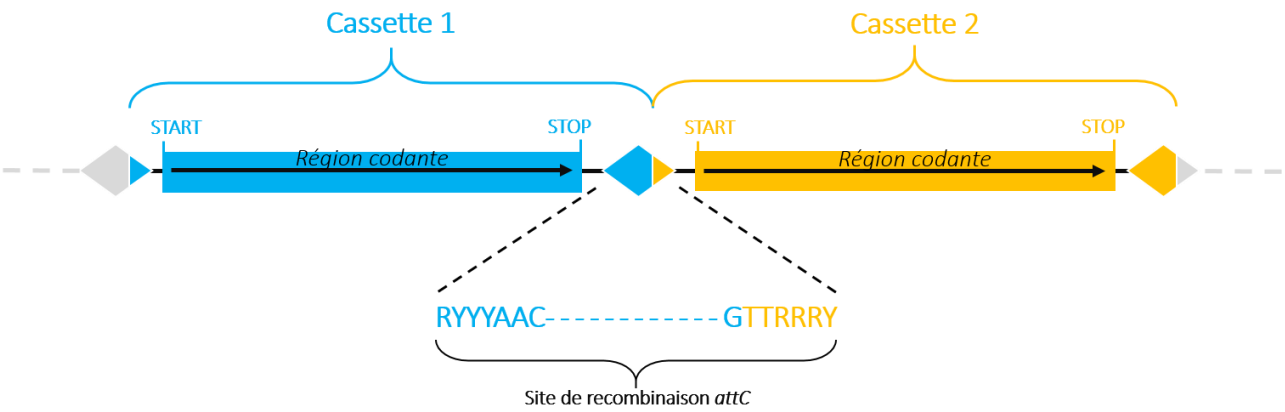


Figure I-4 : Représentation schématique de 2 cassettes de gène insérées dans un intégron (avec R une purine, et Y une pyrimidine)**.**

i. Fonctions codées par les gènes contenus dans les cassettes

Chez les intégrons cliniques, les cassettes présentent une importante homogénéité fonctionnelle puisqu'ils codent essentiellement des résistances à des antibiotiques (Cambray *et al.,* 2010). Un total de 132 cassettes de gène portant diverses résistances à des antibiotiques ont pu être identifiées en se basant sur un seuil de 98% d'identité nucléique (Partridge *et al.,* 2009). Ensemble, ces cassettes codent des résistances à la plupart des antibiotiques incluant les β-lactames, tous les aminoglycosides, le chloramphénicol, le triméthoprime, la streptothricine, la rifampicine, l'érythromycine, la fosfomycine, les quinolones et les antiseptiques de la famille des composés ammonium quaternaires (Fluit et Schmitz, 2004; Mazel, 2006; Partridge *et al.,* 2009).

L'analyse des cassettes portées par les intégrons en dehors d'un contexte clinique montre que les fonctions codées sont très diverses et pour la majorité encore inconnues (voir paragraphe I - 3.2.b).

ii. Expression des gènes contenus dans les cassettes

Comme vu précédemment, les cassettes de gène ne portent généralement pas de promoteur et leur expression est assurée par le promoteur *Pc* présent dans la région constante des

intégrons (Paragraphe I-2.3.c.). Toutefois, seules les cassettes les plus proches du promoteur *Pc* sont exprimées grâce à son activité.

Certains intégrons chromosomiques portent un grand nombre de cassettes de gènes (jusqu'à 150) comme c'est le cas chez le genre *Vibrio*. Il a été montré que la majorité des cassettes de gène étaient tout de même exprimées. Ceci est permis par l'existence de promoteurs internes à certaines cassettes de gène retrouvés tout au long de la région variable de l'intégron. Des cassettes, même les plus éloignées de la région constante de l'intégron peuvent ainsi être exprimées (Michael et Labbate, 2010). De plus, il semblerait que ces promoteurs ne soient pas régulés de la même façon. Par exemple, un stress oxydatif active l'expression de certains de ces promoteurs. Ainsi, des régions de cassettes de gène pourraient être exprimées de manière différente en fonction de stress environnementaux différents (Michael et Labbate, 2010).

2.2.b. Le site de recombinaison *attC*

A la différence des sites *attI*, les sites *attC* sont plus conservés. Ils partagent certaines caractéristiques communes qui leur permettent d'être reconnus par l'intégrase. Ils sont caractérisés par une région palindromique variable en taille et séquence qui va être à l'origine de la formation d'une structure cruciforme lorsque le site *attC* se retrouve sous forme simple brin (Stokes *et al.,* 1997; Johansson *et al.,* 2004). Le repliement sous cette forme secondaire permet de faire ressortir 2 bases extra-hélicales mésappariées : généralement une guanine et une thymine qui permettent l'orientation de la recombinaison réalisée par l'intégrase. On dénombre quatre domaines de liaison de l'intégrase nommés R'', L'', L' et R' (également nommés 1L-2L et 2R-1R (Stokes *et al.,* 1997 ; Figure I-5). Parmi ces domaines, seuls R'' et R' possèdent des séquences conservées, respectivement 5'-RYYYAAC et 5'-G↓TTRRRY (avec R une purine, et Y une pyrimidine, la flèche montre le point de recombinaison ; (Stokes *et al.,* 1997; Cambray *et al.,* 2010).

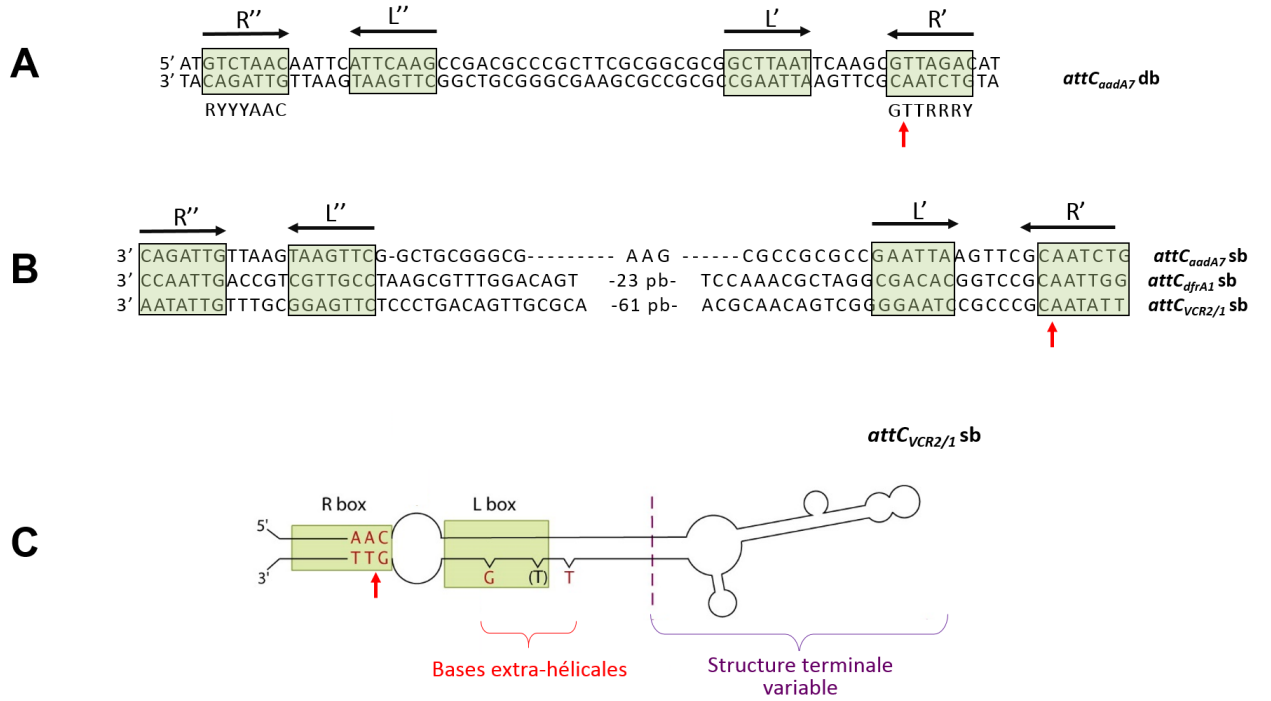


Figure I-5 : Sites de recombinaison *attC*. (**A**) Séquence du site *attCaadA7* double brin. (**B**) Alignement des séquences des brins anti-sens de 3 sites *attC*. (**C**) Représentation schématique de la structure secondaire que pourrait prendre le brin anti-sens du site $attC_{VCR2/1}$, avec en rouge le triplet GTT conservé et les bases extra-hélicales (modifié à partir de Cambray *et al.,* 2010). Les 4 domaines de liaison de l'intégrase nommés R'', L'', L' et R' sont indiqués par les flèches noires horizontales et surlignés en vert. Les flèches rouges représentent la position où a lieu la recombinaison.

La région centrale du site *attC* étant variable en taille (20 à 104 pb) et en séquences (Mazel 2006; Larouche et Roy, 2011), il semblerait que la conservation de la structure secondaire du site *attC* est plus essentielle pour l'activité de recombinaison que la séquence en elle-même, permettant ainsi une plus grande diversité des séquences reconnues et échangées (MacDonald *et al.,* 2006). Toutefois, une analyse phylogénétique des séquences du site *attC* des intégrons portés par différentes espèces du genre *Vibrio* a montré que les sites *attC* liés aux cassettes de gène n'étaient pas homogènes entre toutes les espèces mais tendent à être similaires entre souches ou espèces proches (Boucher *et al.,* 2007), suggérant une évolution non aléatoires de ces séquences.

Par ailleurs, il a été observé que les intégrases pouvaient aisément reconnaître certains sites *attC* et exciser les cassettes de gène à leur niveau alors que pour d'autres, les intégrases semblent montrer peu d'affinité. Pour comprendre cela, plusieurs études de mutagénèse dirigée sur des sites *attC* ont été réalisées afin de tester l'efficacité de différentes intégrases

à reconnaître et exciser les cassettes associées. La capacité de 4 intégrases (IntI1, IntI2*179E, IntI3, SonIntIA) à reconnaître le site $attC_{dfrA1}$ sur lequel des modifications de la séquences ont été réalisées, a été évaluée (Larouche et Roy, 2011 ; Tableau I-2). Deux paramètres semblent primordiaux dans l'efficacité d'excision des intégrases : l'identité des bases extra-hélicales et l'espace les séparant. Ainsi, le principal facteur favorisant l'activité de IntI1 est la présence d'une thymine en 1e base extra-hélicale et le premier facteur défavorisant semble être l'espace entre les 2 bases extra-hélicales, si il diffère de 6 bases. Une cytosine en 2e base extra-hélicale semble favoriser l'activité de IntI2 et de SonIntIA et un écart de base inférieur à 6 bases réduit presque totalement leur activité. La combinaison qui semble la plus favorable à IntI3 est T-N_6-C mais la combinaison A-N_6-G et un espace entre les base extra-hélicale de 5 bases réduisent fortement son activité (Larouche et Roy, 2011 ; Tableau I-2).

Par ailleurs, il semblerait que ces mutations n'aient pas le même impact en fonction de si elles sont réalisées sur le site *attC* en amont ou en aval de la cassette excisée (Larouche et Roy, 2011). En effet, une substitution d'une guanine par une cytosine sur la seconde base extra-hélicale quand la première base extra-hélicale est une thymine augmente l'efficacité d'excision de IntI1 (Bouvier *et al.,* 2009) alors que dans le cas d'un site *attC* en amont, aucune différence n'est observée (Larouche et Roy, 2011 ; Tableau I-2).

Tableau I-2 : Efficacité de l'excision des intégrases IntI1, IntI2*179E, IntI3 et SonIntIA au niveau des différents mutants du site $attC_{dfrA1}$.

Espace entre les bases extra-hélicale	Combinaisons testées	IntI1	IntI2*179E	IntI3	SonIntIA
7-8	A-N_{7-8}-C	--	-/+	--	-
	A-N_{7-8}-G	--	-	--	-
6	T-N_6-C	+/-	++	-	+/-
	T-N_6-G	+/-	+	-	+/-
	A-N_6-C	-	++	--	+/-
	A-N_6-G	--	+/-	--	-
5	A-N_5-C	--	--	--	---
	T-N_5-C	--	--	--	--

(--) efficacité comprise entre 0 et 20%, (-) efficacité comprise entre 20 et 40%, (+/-) efficacité comprise entre 40 et 60%, (+) efficacité comprise entre 60 et 80%, (++) efficacité comprise entre 80 et 100%. En rouge, les combinaisons étant les plus défavorables à l'efficacité de l'intégrase, en vert, les combinaisons étant les plus favorables à l'efficacité de l'intégrase (d'après Larouche et Roy, 2011).

2.3. L'intégrase d'intégron : une tyrosine recombinase particulière

2.3.a. Structures primaire et secondaire

L'intégrase codée par les intégrons appartient à la famille des tyrosine-recombinases à site spécifique dans laquelle on retrouve la recombinase Cre du bactériophage P1 (Abremski et Hoess, 1992), la recombinase Flp de *Saccharomyces cerevisiae* (Sadowski, 1995), l'intégrase du bactériophage λ (Landy 1989) et les recombinases bactériennes XerC et XerD (Arciszewska and Sherratt 1995). Les protéines de cette famille sont caractérisées par un domaine catalytique commun (Esposito et Scocca, 1997; Nunes-Düby *et al.,* 1998), contenant 4 acides aminés strictement conservés nécessaires à la catalyse, situés dans la région C-terminale : un triplé catalytique Arg-I,His-II,Arg-II et une tyrosine nucléophile (Argos *et al.,* 1986; Abremski et Hoess, 1992; Nunes-Düby *et al.,* 1998) ; Tableau I-3 ; Figure I-7). Deux autres résidus du site actif ont également été identifiés : un résidu tryptophane ou histidine (His/Trp-III) et un résidu lysine (Lys-β) (Nunes-Düby *et al.,* 1998). Cette dernière semble être en contact proche avec les substrats et pourrait influencer la réactivité des intermédiaires phospho-tyrosines formés pendant la réaction de recombinaison (Messier et Roy, 2001). La partie C-terminale de ces protéines montrent plus de similarités entre les protéines de la famille que la partie N-terminale (Nunes-Düby *et al.,* 1998).

Tableau I-3 : Résidus catalytiques des tyrosine-recombinases (d'après Van Duyne, 2002).

Protéines	Résidus catalytiques					
	Arg-I	Lys-β	His-II	Arg-II	His/Trp-III	Tyr
Cre	Arg173	Lys201	His289	Arg292	Trp315	Tyr324
Integrase λ	Arg212	Lys235	His308	Arg311	His333	Tyr342
XerC	Arg148	Lys172	His240	Arg243	His266	Tyr275
XerD	Arg148	Lys172	His244	Arg247	His270	Tyr279
Flp	Arg191	Lys223	His305	Arg308	Trp330	Tyr343
IntI1	Arg146	Lys171	His277	Arg280	His303	Tyr312
IntI2	Arg135	Lys160	His266	Arg269	His292	Tyr301
IntI3	Arg155	Lys180	His286	Arg289	His312	Tyr321
IntIVchA (IntI4)	Arg135	Lys160	His267	Arg270	His293	Tyr302

Quatre intégrases d'intégron partageant entre 42 et 58% d'identité protéique et similaires en taille (entre 317 et 340 acides aminés), ont particulièrement été bien étudiées et

caractérisées : il s'agit des intégrases IntI1, IntI2, IntI3 des intégrons de classes 1, 2 et 3 identifiés en milieu clinique et de l'intégrase IntIVchA aussi appelée IntI4, caractérisée dans l'intégron chromosomique de *Vibrio cholerae* (Stokes et Hall, 1989; Arakawa *et al.,* 1995; Clark *et al.,* 1997; Esposito et Scocca, 1997; Mazel *et al.,* 1998; Nunes-Düby *et al.,* 1998). Chez les intégrases d'intégron, le domaine C-terminal est constitué de 5 feuillets β et de 10 hélices α dont l'hélice α I_2, absente chez les autres protéines de la famille, et qui se situe dans une région appelé domaine additionnel (Figure I-6). Le reste du domaine C-terminal est structuralement similaire aux autres membres de la famille des tyrosine-recombinases.

Les intégrases d'intégron possèdent les 6 résidus conservés du site catalytique (Tableau I-3, Figure I-7). L'arginine Arg-I (R156 sur la Figure I-7) est située dans la boucle entre les hélices α F et G des intégrases d'intégron, les résidus His-II (H289) et Arg-II (R292) sont situées dans l'hélice α K et la tyrosine nucléophile (Y324) dans l'hélice α M. Le résidu His-III (H312) est situé dans la boucle entre les hélices α L et α M et la lysine Lys-β (K180) est située dans la boucle entre les feuillets β 2 et β 3. Des études de mutagénèses dirigées sur ces 6 résidus conservés ont montré qu'une mutation sur un seul de ces résidus chez IntI1 réduisait l'activité de recombinaison à un niveau très faible (Gravel *et al.,* 1998). Par ailleurs, l'acide aspartique 161 de IntI1 (en position 171 sur la Figure I-7) joue un rôle important dans l'interaction protéine/protéine des sous-unités de l'intégrase. Une mutation à cette position entraine une modification du nombre de recombinaisons *attI* x *attC* par rapport aux recombinaisons *attC* x *attC* en augmentant ou en diminuant la stabilité du complexe protéique (Demarre *et al.,* 2007).

Le domaine additionnel est constitué d'une vingtaine d'acides aminés, avec un motif assez bien conservés de 15 résidus essentiellement non polaires (Messier et Roy, 2001, Figure I-6). Une délétion du motif ALER chez IntI1 n'altère pas la capacité de l'intégrase à se lier au site *attI1* mais la rend déficiente pour la recombinaison. Ce motif ne semble donc pas être impliqué dans la reconnaissance du site *attI* mais jouerait plutôt un rôle dans le positionnement et la stabilisation du site actif ou dans la reconnaissance et la liaison au site *attC* (Messier et Roy, 2001). Ce domaine particulier aurait un rôle dans l'assemblage

synaptique, et particulièrement dans la formation des poches hydrophobes qui permettent la stabilisation des 2 bases extra-hélicales impliquées dans la reconnaissance spécifique du brin anti-sens du site *attC* (MacDonald *et al.,* 2006). La reconnaissance de ces 2 bases extra-hélicales est assurée par 8 résidus. Cinq sont impliqués dans la reconnaissance de la guanine mésappariée : le tryptophane en position 240, le triplet Lys-Tyr-Pro en position 230 à 232 et l'acide aminé Arg/His/Trp en position 178 ; et 3 dans la reconnaissance de la thymine mésappariée : le doublet His-His en position 245-246 et la proline en position 253 (Figure I-7). Par ailleurs, les résidus Tyr240 et Lys230 semblent également impliqué dans la reconnaissance du site *attI* (Messier et Roy, 2001).

Figure I-6: Région du site additionnel spécifique des intégrases d'intégron, et absent chez les autres tyrosine-recombinases.

Le rôle d'autres résidus a également pu être caractérisé : le glutamate 131 jouerait un rôle dans l'affinité du site *attI* ce qui entrainerait une baisse de l'activité de recombinaison si il est muté (Messier et Roy, 2001) ; la phénylalanine 233 bien conservée chez toutes les intégrases d'intégron mais pas chez les autres tyrosines recombinases pourraient être également impliquée dans la reconnaissance du substrat mais aussi dans la stabilisation dans le repliement de la protéine. Par ailleurs, la cystéine 250 présente chez IntI2 et non chez Int1, IntI3 et VchIntIA est essentielle dans l'activité d'excision de l'intégrase de Shewanella mais n'est pas impliquée dans un pont disulfure notamment avec la cystéine 259 (Larouche et Roy, 2009). En outre, des différences d'efficacité d'excision entre différentes intégrases ont pu être montrées en fonction du site *attC* ciblé (Larouche et Roy, 2011, paragraphe 2.4.b.). Ces différences pourraient s'expliquer par des variations dans les séquences d'acides aminés de la région entre l'hélice α I2 et les feuillets β 4-5 des intégrases IntI1 et IntI3 par rapport aux

intégrases IntI4 et SonIntIA. Ainsi, certaines substitutions de résidus pourraient favoriser la versatilité de IntI4 et SonIntIA (Larouche et Roy, 2011).

Par ailleurs, des mutations sur les résidus voisins des acides aminés clés du site actif modifient l'efficacité de recombinaison de l'intégrase. Il est possible que ces mutations perturbent le placement correct de ces résidus. C'est le cas par exemple pour l'histidine His-III dont une mutation de la glycine voisine conservée pourrait avoir pour effet de modifier le placement de l'histidine si elle est remplacée par une arginine alors qu'une substitution par une alanine n'affecte pas l'activité de recombinaison (Messier et Roy, 2001).

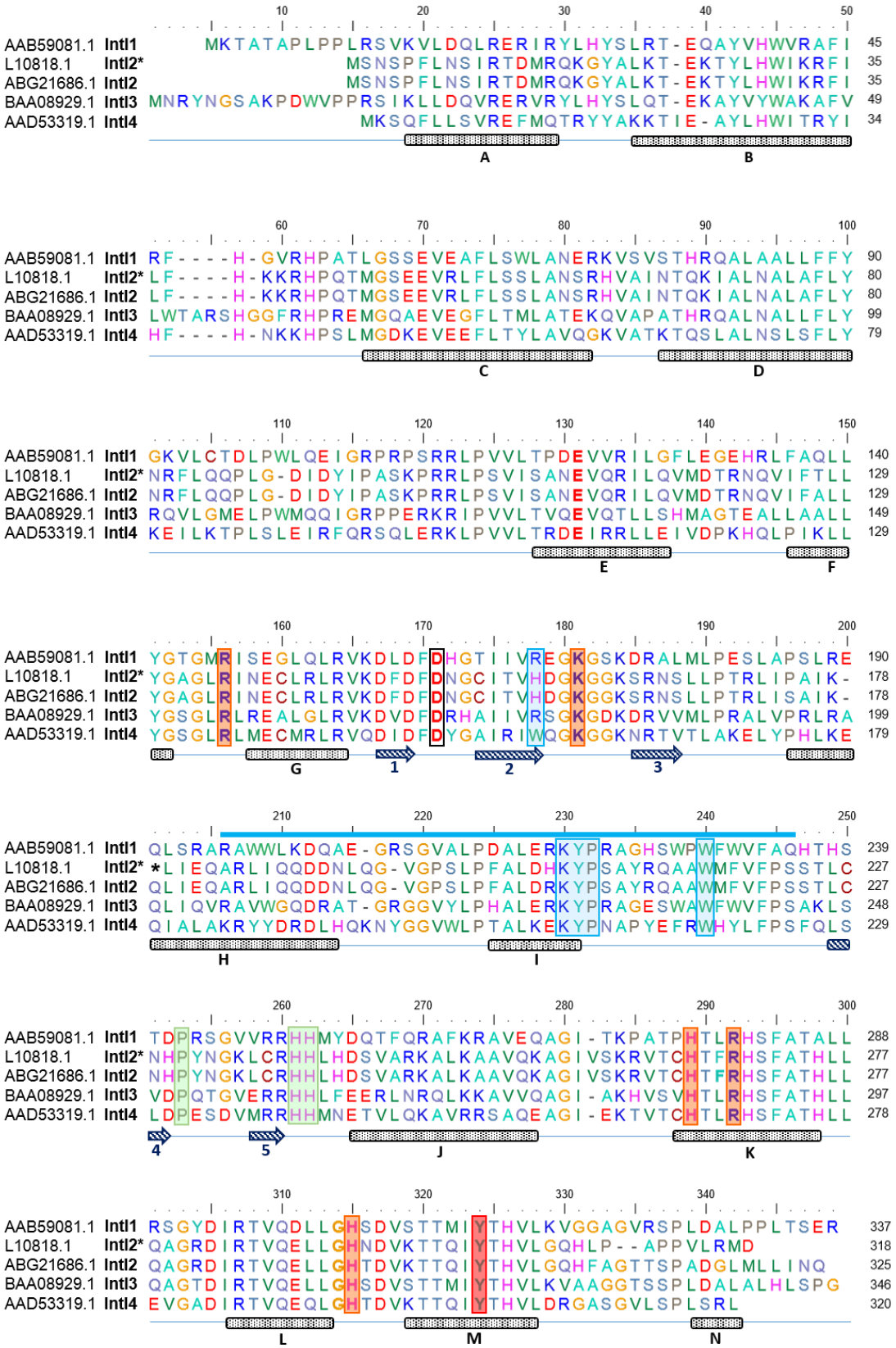


Figure I-7 : Alignements des séquences de IntI1, IntI2*, IntI2 IntI3 et IntI4. En dessous de l'alignement : les hélices α sont représentées par des rectangles grisés annotés par des lettres. Les feuillets β sont représentés par des flèches rayées bleues, annotées par des chiffres. Au-dessus de la séquence : le domaine additionnel est représenté par un trait bleu entre les résidus 206 et 246. La tyrosine nucléophile est surlignée en rouge ; les 5 autres résidus catalytiques sont surlignés en orange. Les résidus reconnaissant la guanine mésappariée sont surlignés en bleu, et ceux reconnaissant la thymine mésappariée en vert.

2.3.b. Structures tertiaire et quaternaire

La structure tridimensionnelle de l'intégrase de *Vibrio cholerae* VchIntIA associée à un site *attC* de *V. cholerae* a pu être caractérisée par cristallographie en 2006 par MacDonald et ses collaborateurs. L'intégrase est fonctionnelle sous sa forme tétramérique. Deux des quatre sous-unités catalysent la réaction de recombinaison, les deux autres interagissent de manière spécifique avec les sites de recombinaisons *attI* et *attC* (Figure I-8). Les deux sous-unités catalytiques clivent le brin entraînant la formation d'une jonction de Holliday.

Chaque monomère d'intégrase VchIntIA se replie en 2 domaines distincts. Le domaine N-terminal contient 4 hélices α organisées en motifs hélice-tour-hélice, les paires α A-α B et α C-α D étant perpendiculaires l'une de l'autre. Les hélices α B et α D sont en contact avec le grand sillon du site *attC* simple brin alors que l'hélice α A intervient dans le contact entre les sous unités protéiques du complexe synaptique. La structure tridimensionnelle a montré que les feuillets β 4 et 5 du domaine C-terminal de la sous-unité non attaquante interagissent avec la base extra-hélicale T qui s'insère entre les 2 histidines H250-H251 et la proline 245, alors que l'hélice α I2 de la sous-unité attaquante forme plusieurs contacts importants en *trans* avec la base extra-hélicale G (MacDonald *et al.,* 2006).

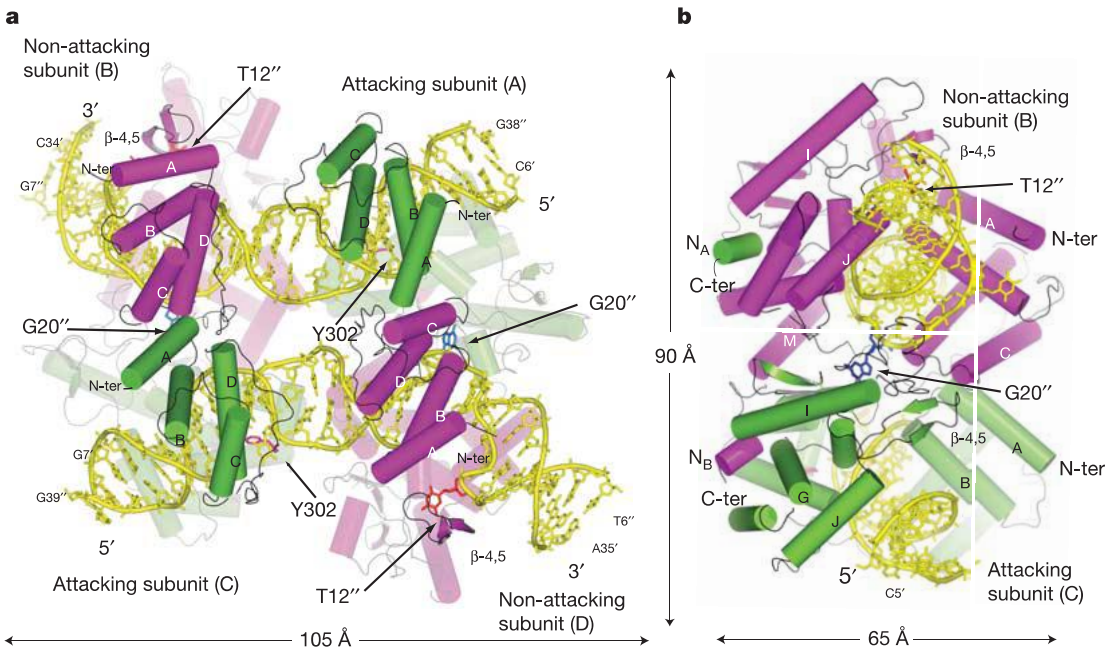


Figure I-8 : Structure tridimensionnelle du complexe formé par les 2 dimères de VchIntIA, en violet et vert, liés au site VCR db (site *attC* de *V. cholerae*) en jaune. (a) Vue des extrémités N-terminales du complexe. (b) Vue orthogonale du complexe. La thymine T12'' mésappariée extra-hélicale et la tyrosine Y302 sont en rouge. La guanine G20'' mésappariée est en bleu (d'après MacDonald *et al.*, 2006).

2.3.c. Le mécanisme de la recombinaison

Comme les autres membres de la famille, les intégrases d'intégron utilisent le mécanisme type des topoisomérases I pour le clivage. Quatre monomères sont impliqués dans la réaction de recombinaison site-spécifique dans laquelle l'échange d'un brin d'ADN conduit à la formation d'une jonction de Holliday. Pour la plupart des tyrosines recombinases, cet intermédiaire est résolu par l'échange du second brin. Dans le cas des intégrases d'intégron, le site *attC* étant simple brin, la résolution se fait grâce à la réplication de l'ADN (Mazel, 2006). Par ailleurs, contrairement aux autres membres de la famille des tyrosines recombinases, les intégrases d'intégron peuvent échanger l'ADN entre 2 sites avec des structures différentes : le site non-palindromique *attI* et le site palindromique *attC*.

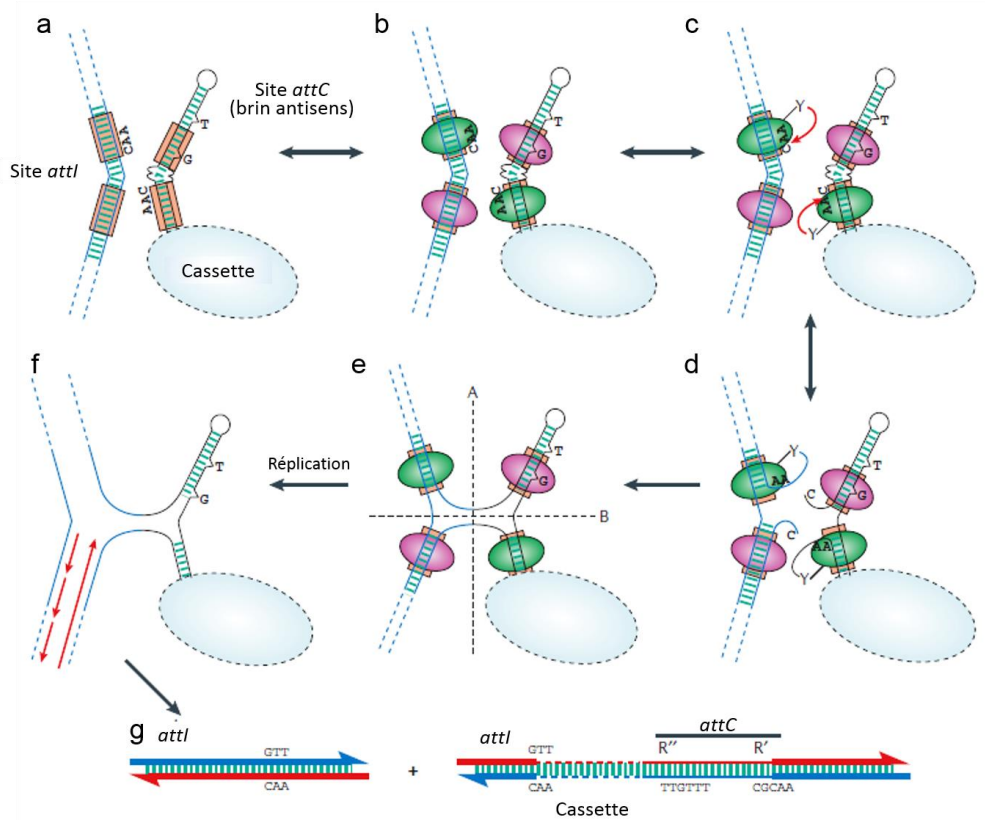


Figure I-9 : Représentation schématique des différentes étapes de la recombinaison réalisée par l'intégrase d'intégron entre le site *attI* et le site *attC*. Ce modèle utilise comme substrat un site *attC* simple brin replié grâce aux séquences palindromiques. Les étapes sont identiques à celles d'une recombinaison site-spécifique classique réalisée par les autres tyrosines recombinases jusqu'à la formation de la jonction de Holliday. (**a**) Les domaines de liaison de l'intégrase sur le site *attI* (R et L) et sur le site *attC* simple brin sont encadrés. Pour simplifier, seul le brin anti-sens du site *attC* impliqué dans la recombinaison est représenté. (**b**) Les quatre monomères de IntI se lient aux sites *attC* et *attI*. Les monomères verts correspondent aux sous-unités attaquantes, celles en rose correspondent aux sous-unités non attaquantes. (**c**) Les sous-unités attaquantes coupent les brins d'ADN grâce à la tyrosine du site catalytique pour former une liaison 3'-phosphotyrosine (**d**) et libéré le groupement 5'-hydroxyl. (**e**) La réaction de l'extrémité 5'hydroxyl avec l'autre phosphotyrosine permet l'échange d'un brin d'ADN entre les 2 sites et former l'intermédiaire de la jonction de Holliday. La résolution de cette jonction de Holliday se fait grâce à une étape de réplication (**f** et **g**) (modifié à partir de Mazel, 2006).

3) Diversité des intégrons

3.1. Les intégrons dits « cliniques »

Les intégrons de classe 1, 2 et 3 sont largement retrouvés dans les milieux cliniques mais on les trouve aussi dans d'autres environnements.

3.1.a. Les intégrons de classe 1

Les intégrons de classe 1 sont retrouvés chez 40 à 70% des bactéries gram-négatives pathogènes isolées dans un contexte clinique (Martinez-Freijo *et al.,* 1998; van Essen-Zandbergen *et al.,* 2007). La séquence nucléique des séquences *intI1* de ces intégrons est quasiment identique. Ceci suggère l'existence d'un ancêtre commun récent des intégrons de classe 1 cliniques (Stokes *et al.,* 2006 ; Gillings *et al.,* 2008). Ils portent essentiellement des gènes de résistance à des antibiotiques et on connaît à ce jour plus de 130 cassettes de gènes différentes codant une résistance (Partridge *et al.,* 2009). Ces intégrons dits cliniques sont associés à un dérivé du transposon Tn402 dont ils possèdent des éléments caractéristiques : des séquences inversées répétées *IRi* et *IRt* et un module de transposition *tni*. La majorité des intégrons de classe 1 cliniques ont perdu la fonction de transposition de Tn402 mais il semble très vraisemblable que l'ancêtre commun des intégrons de classe 1 cliniques soit un intégron provenant de l'environnement qui a été capturé par un transposon Tn*402* (Gillings, 2014).

En effet, les intégrons de classe 1 sont également retrouvés dans l'environnement. Ils ont été retrouvés dans de nombreuses β-*proteobacteria* non-pathogènes, incluant les genres *Hydrogenophaga*, *Aquabacterium*, *Acidovorax, Imtechium*, *Azoarcus* et *Thauera*. A la différence des intégrons de classe 1 cliniques, les intégrons de classe 1 environnementaux présentent des variations sur le gène *intI1* (avec un minimum de 95% d'identité nucléique ; Gillings *et al*., 2008). Par ailleurs, ces ils portent des cassettes de gène codant des fonctions inconnues (Stokes *et al.,* 2006; Gillings *et al.,* 2008, 2009). Ces intégrons de classe 1 d'origine environnementale sont pour la plupart situés sur le chromosome. Ils possèdent des séquences communes aux extrémités 5' et 3', respectivement de 107 pb après le codon stop du gène *intI1* et 43 pb derrière le dernier site de recombinaison *attC* (Gillings *et al*., 2009).

Au-delà de ces séquences, l'environnement génétique de ces intégrons environnementaux de classe 1 est variable (Gillings *et al.*, 2008).

Les intégrons de classe 1 sont retrouvés dans de très nombreux environnements impactés ou non par des activités humaines : eau douce (Wright *et al.,* 2008), sédiments aquatiques (Gillings *et al.,* 2008; Wright *et al.,* 2008; Rosewarne *et al.,* 2010; Nardelli *et al.,* 2012), sol (Nield *et al.,* 2001 ; Gillings *et al.,* 2008 ; Hardwick *et al.,* 2008), sortie de rejets hospitaliers et/ou urbains (Gaze *et al.,* 2011; Stalder *et al.,* 2012), estuaires et environnements côtiers (Gillings *et al.,* 2008 ; Wright *et al.,* 2008). Des études de l'abondance des intégrons de classe 1 (sans distinction entre les cliniques et les environnementaux) présents dans le métagénome des communautés bactériennes de différentes environnements ont été réalisées. Entre 1% et 10% des bactéries totales possèderaient un intégron de classe 1 d'après l'estimation à partir de divers environnements (Gillings *et al.,* 2008 ; Hardwick *et al.,* 2008 ; Wright *et al.,* 2008 ; Rosewarne *et al.,* 2010). C'est un taux basal que l'on retrouve naturellement dans les communautés bactériennes non soumises aux activités anthropiques. Il a également été observé que dans les environnements les plus impactés, les intégrons de classe 1 étaient plus abondants que dans des milieux moins impactés (Hardwick *et al.,* 2008 ; Wright *et al.,* 2008 ; Rosewarne *et al.,* 2010 ; Byrne-Bailey *et al.,* 2011 ; Gaze *et al.,* 2011; Gillings *et al.,* 2015). Une étude en microcosmes a montré que l'abondance des intégrons de classe 1 augmente face à une contamination par des antibiotiques mais également par des métaux dans des environnements d'eau douce après 7 jours de contamination (Wright *et al.,* 2008). Les intégrons de classe 1 peuvent être transportés par de larges éléments mobiles et associés à des transposons porteurs de gènes de résistance aux métaux (Liebert *et al.,* 1999 ; Wright *et al.,* 2008). Il pourrait donc y avoir une co-sélection des intégrons de classe 1 avec ces résistances aux métaux.

3.1.b. Les intégrons de classe 2

Si les intégrons de classe 1 sont aujourd'hui considérés comme les principaux acteurs dans l'acquisition et la dissémination de résistance à des antibiotiques ou des biocides par les bactéries, d'autres intégrons porteurs de ces résistances ont également été décrits. C'est le

cas des intégrons de classe 2 qui portent également des cassettes de résistance à des antibiotiques. Les intégrons de classe 2 se distinguent de ceux de classe 1 par la séquence de leur intégrase entre autre. Ainsi IntI1 et IntI2 partagent seulement 46% d'identité protéique (Hansson *et al.,* 2002).

Les intégrons cliniques de classe 2 sont associés au transposon Tn*7* et à ses dérivés (Ploy *et al.,* 2000; Hansson *et al.,* 2002). Ils possèdent une intégrase rendue non fonctionnelle à cause d'une mutation d'un codon guanine en un codon stop, résultant d'une substitution d'une thymine par une cytosine au nucléotide 444 (Hansson *et al.,* 2002). Du fait de leur intégrase inactive, ces intégrons ont perdu leur capacité à acquérir et à réorganiser leurs cassettes de gène. Ainsi, la succession de cassettes de gène portées par ces intégrons est fortement conservée (White *et al.*, 2001). Des variations dans les cassettes de gène ont cependant pu être observées (Ramírez *et al.,* 2010) et auraient pu être permises par l'activité d'une intégrase en *trans*. La présence simultanée d'un intégron de classe 1 et d'un intégron de classe 2 chez une même souche a été observée chez des bactéries multirésistantes à des antibiotiques : *Acinetobacter* (Ploy *et al.,* 2000; Wang, Dang and Ding 2008) ; *E.coli* (Biskri and Mazel 2003; Laroche *et al.,* 2009; Moura *et al.,* 2014) ; *Shigella flexneri* (Zhu *et al.,* 2011) ; *Aeromonas caviae* et *A. hydrophila* (Maravić *et al.,* 2013). Une étude par mutagénèse dirigée ayant pour but de tester l'activité de l'intégrase IntI2 clinique dont le codon stop a été remplacé par une glutamine a permis de montrer que la protéine IntI2 était capable de catalyser la recombinaison entre 2 sites *attC* permettant l'excision des cassettes de gène. La reconnaissance des sites *attI1* et *attI2* par IntI1 et IntI2 mutée a également été testée. Alors que IntI2 ne reconnaît seulement que son propre site *attI2*, IntI1 est capable de réaliser une recombinaison au niveau des 2 sites *attI1* ou *attI2*, (Hansson *et al.,* 2002). Ceci confirme donc la possibilité d'un réarrangement des cassettes chez les intégrons cliniques de classe 2 dont l'intégrase est non fonctionnelle, grâce à l'activité de l'intégrase d'un autre intégron, en particulier de classe 1.

Le fait que les intégrons de classe 2 cliniques possèdent une mutation dans le gène *intI2* suggère fortement qu'ils sont tous descendants d'un même et unique évènement de capture

par le transposon Tn7 (Gillings *et al.,* 2014). Toutefois, un nouveau variant de IntI2 a été récemment découvert chez une souche d'*E. coli* isolée chez un patient atteint d'une infection urinaire. Cet intégron a été recruté par un plasmide IncP. Le gène *intI2* porté par cet intégron présente 6 différences dans sa séquence avec le gène *intI2* retrouvé dans les intégrons associés à Tn*7*, dont le codon glutamine (CAA) qui supprime le codon stop caractéristiques des premiers intégrons de classe 2 cliniques étudiés. L'intégrase IntI2 codée par cet intégron serait donc capable de catalyser la réaction de recombinaison. Les cassettes de gène portées par cet intégron ont été caractérisées. Il possède un gène *drfA14*, codant une résistance au triméthoprime, ainsi qu'une cassette inconnue, codant une putative peptidase lipoprotéine-signal (Márquez *et al.,* 2008).

Une étude métagénomique réalisée sur des environnements non cliniques variés (sédiments marins, sols agricoles ou non, forêts tropicales et tempérées, sol enrichis en compost) présentant des niveaux d'anthropisation variables a permis de découvrir des intégrons de classe 2 potentiellement fonctionnels chez plusieurs *Firmicutes* (*Clostridium* et *Bacillus*) et *Bacteroidetes* (*Chitinophaga* et *Sphingobacterium*) (Rodríguez-Minguela *et al.,* 2009). Par ailleurs, cette étude a montré que le gène *intI2* est faiblement abondant dans les environnements peu ou pas anthropisés, suggérant donc que les activités humaines favorisent la présence des intégrons de classe 2 dans les métagénomes environnementaux. Un intégron de classe 2 potentiellement fonctionnel a également été décrit chez *Providencia stuartii* isolée à partir de déchets et de produits bovins. Les 9 cassettes de gène qui ont pu être caractérisées codent des fonctions inconnues, comme c'est souvent le cas pour les intégrons environnementaux (Barlow et Gobius, 2006).

3.1.c. Les intégrons de classe 3

Les intégrons de classe 3 ont également été découverts dans un contexte clinique. Ils ont été décrits pour la première fois au Japon chez une souche de *E. coli* (Arakawa *et al.,* 1995). L'intégrase IntI3 portée par cet intégron présente 61% d'identité protéique avec IntI1 et 47% avec IntI2. Les intégrons de classe 3 retrouvés chez des pathogènes sont associés à des résistances à des antibiotiques. L'intégrase IntI3 semble être moins active que l'intégrase de

classe 1 et que l'intégrase de classe 2 fonctionnelle (Gillings, 2014) ou capable de reconnaître d'autres sites *attC* plus spécifiquement que ceux des cassettes retrouvés chez les intégrons de classes 1 et 2 (Larouche et Roy, 2011). En effet, les sites *attC* caractérisés dans les intégrons de classe 3 diffèrent des premiers sites *attC* caractérisés chez IntI1 et IntI2. Même si la séquence spécifique de reconnaissance GTTRRRY est bien retrouvée, les sites *attC* portés par les intégrons de classe 3 sont généralement plus grands que ceux retrouvés chez les intégrons des classes 1 et 2 (67 pb contre 59 pb (Arakawa *et al.,* 1995)).

Les intégrons de classe 3 sont relativement communs au Japon où ils ont été disséminés chez des pathogènes humains (Correia *et al.,* 2003; Shibata *et al.,* 2003) mais sont peu communément retrouvés dans le reste du monde (van Essen-Zandbergen *et al.,* 2007; Laroche *et al.,* 2009 ; Kor *et al.,* 2013 ; Gillings, 2014). Certains des intégrons de classe 3 cliniques possèdent une histoire évolutive semblable à celle des intégrons de classe 1 : un intégron de classe 3 semble également avoir été capturé par un transposon Tn*402* mais dans une orientation inverse de celle des intégrons de classe 1. C'est le cas pour l'intégron de classe 3 retrouvé chez une souche de *Serratia marcescens* isolée au Japon (Collis *et al.,* 2002). Mais d'autres intégrons de classe 3 porteurs de résistances à des antibiotiques ont pu être caractérisés : ils sont portés par un plasmide de type IncQ, comme c'est le cas pour les intégrons retrouvés chez une souche de *Enterobacter cloacae* isolée en France (Barraud *et al.,* 2013) ou chez une souche de *E. coli* isolée en Suisse (Poirel *et al.,* 2010), ou bien ils sont portés par le plasmide p22K9 sur lequel est situé l'intégron de classe 3 isolé chez une souche de *Klebsiella pneumoniae* (Correia *et al.,* 2003).

Seulement 4 études font état de leur présence dans des milieux environnementaux : les 3 premières, grâce à l'amplification spécifique du gène *intI3* à partir de métagénomes, ont permis de le détecter dans des huîtres élevées en Virginie (USA) (Barkovskii *et al.,* 2010), dans des sédiments d'estuaire de Caroline du Sud (USA) (Uyaguari *et al.,* 2013) et dans des biosolides (Rodríguez-Minguela *et al.,* 2009). La 4e étude a caractérisé un intégron de classe 3 sur le chromosome de 2 souches de *Delftia* (Xu *et al.,* 2007). Ces intégrons sont associés à

des cassettes de gène codant des fonctions inconnues, comme c'est majoritairement le cas des intégrons dans l'environnement.

3.2. Diversité des intégrons dans les environnements autres que cliniques

L'étude des intégrons s'est longtemps limitée au contexte clinique et à l'analyse des intégrons de classe 1, 2 et 3. Or, la recherche d'intégrons dans des environnements divers a montré que les intégrons n'étaient pas seulement spécifiques aux pathogènes. De plus, cela a permis de les détecter dans de nombreux environnements tels que le sol, les sédiments, les biofilms ou l'eau, indépendamment de la présence de polluants, incluant les environnements non impactés par les antibiotiques (Nield *et al.,* 2001; Stokes *et al.,* 2001; Nemergut *et al.,* 2004; Elsaied *et al.,* 2007, 2011, 2014; Hardwick *et al.,* 2008; Koenig *et al.,* 2008, 2009; Wright *et al.,* 2008; Rodríguez-Minguela *et al.,* 2009; Huang, 2010; Michael et Andrew, 2010; Moura *et al.,* 2010; Gaze *et al.,* 2011). Il a pu être ainsi révélée une importante diversité des intégrons. Une partie de cette diversité a été mise en évidence par l'analyse des intégrons de bactéries isolées (Boucher *et al.,* 2007; Nemergut *et al.,* 2008), mais une plus grande variété de séquences a été révélée par des méthodes cultures indépendantes (Stokes *et al.*, 2001 ; Holmes *et al.*, 2003 ; Elsaied *et al.*, 2007, 2011 ; Koenig *et al.*, 2008, 2009 ; Huang, 2010 ; Michael et Andrew, 2010; Moura *et al.*, 2010).

3.2.a. Diversité des gènes d'intégrase

La diversité des intégrases d'intégron, seul élément permanent de l'intégron présentant un degré de conservation suffisant permettant sa détection dans les métagénomes, a été peu étudiée à ce jour. Seulement 6 études ont été réalisées sur le sujet dans les environnements marins et dans le sol, et aucune dans les environnements d'eau douce. Ces études ont permis la description de 3, 6, 11, 33, 37, 56 et 169 nouvelles séquences IntI, provenant respectivement du sol (Nield *et al*., 2001), de stériles (Nemergut *et al.,* 2004), de cheminées hydrothermales marines (Elsaied *et al*., 2007), de sédiments marins pollués (Elsaied *et al*., 2011), d'une carotte de sédiments extraite d'une fosse océanique (Elsaied *et al*., 2014), de sédiments côtiers (Huang, 2010) et de divers environnements terrestres, côtiers et marins

profonds (Rodríguez-Minguela *et al*., 2009). Au total, l'analyse de séquences d'intégrases obtenues à partir de ces études et de l'étude des génomes de bactéries séquencés (Boucher *et al.,* 2007) a révélé près de 400 phylotypes d'intégrase. Ceci suggère donc une très grande diversité des séquences d'intégrase, jusqu'alors sous-évaluée, du fait de la prédominance des études sur les intégrons dits cliniques (Rapa et Labbate, 2013). Or, les intégrons cliniques semblent loin d'être prépondérants dans l'environnement. En effet, lorsque des analyses métagénomiques basées sur la PCR ont été réalisées, les intégrons de classes 1, 2 et 3 n'ont pas pu être détectés dans les environnements marins (Elsaied *et al.,* 2007, 2011, 2014 ; Rodriguez-Minguela *et al.,* 2009 ; Huang, 2010), et une seule étude fait état de leur présence dans des environnements terrestres (Rodriguez-Minguela *et al.,* 2009). De plus, les analyses de raréfaction réalisées dans des ces études montrent qu'une part minime de la diversité des intégrases est estimée (Elsaied *et al.,* 2011). Ainsi, au cours de ces dernières années, certains auteurs ont pointé du doigt le manque de connaissances sur la diversité des intégrases et des intégrons en général (Rapa et Labbate, 2013 ; Gillings, 2014).

Il a toutefois pu être mis en évidence que la diversité des intégrases pourrait être influencée par l'environnement. Ainsi, la phylogénie de l'intégrase codée par des gènes *intI* portés par des bactéries dont le génome a été séquencé, a permis de mettre en évidence 3 grands groupes d'intégrases, basés sur l'environnement dans lequel ils sont retrouvés : (i) un groupe d'intégrases retrouvés chez les *Proteobacteria* du sol et d'eau douce, (ii) un groupe d'intégrases portés par des γ-*proteobacteria* de milieux marins et d'intégrases retrouvés dans des éléments intégratifs et conjugatifs (ICE) SXT et sur le plasmide pRSV1 des *Vibrio* et (iii) un groupe d'intégrase pour lesquelles le gène *intI* est orienté dans le même sens que le reste de l'intégron (Mazel, 2006 ; Boucher *et al.,* 2007 ; Cambray *et al.,* 2011 ; Gillings, 2014).

De plus, une analyse phylogénétique des intégrases obtenues entre 2 sites similaires situés dans deux importantes régions portuaires (Suez et Tokyo) a montré des embranchements distincts, suggérant également une influence géographique (Elsaied *et al.,* 2011).

A ce stade, les limites des conclusions que l'on peut tirer s'arrêtent là, montrant le manque de connaissances que nous avons sur les intégrons environnementaux, nous empêchant

d'appréhender les facteurs influençant la diversité des intégrons et par conséquent de comprendre leur rôle chez les bactéries de l'environnement.

3.2.b. Diversité des cassettes dans les environnements autres que cliniques

L'analyse des cassettes de gène obtenues à partir des génomes de *Vibrio* séquencés (Boucher *et al.,* 2007) et à partir de banques réalisées sur les métagénomes de différents environnements tels que les sols, sédiments côtiers contaminés, cheminées hydrothermales, sédiments d'une fosse marine, sédiments d'eau douce et d'estuaire (Stokes *et al.,* 2001 ; Koenig *et al.,* 2008, 2009 ; Elsaied *et al.,* 2007, 2011, 2014 ; Wright *et al.,* 2008) a permis d'identifier environ 5000 cadres de lecture contenus dans des cassettes de gène. Ces analyses montrent qu'environ 80% des cassettes de gènes codent de putatives protéines de fonctions encore inconnues. Par ailleurs, lorsqu'un domaine ou une fonction ont pu être identifiés, la diversité des fonctions codées s'est avérée importante. Elles correspondent à des fonctions cellulaires ou métaboliques (Stokes *et al*., 2001 ; Boucher *et al*., 2007 ; Elsaied et Maruyama, 2011) telles que des acétyl ou méthyl transférases, des méthylases, des phosphotransférases, thio-transférases ou des régulateurs transcriptionnels, pour n'en lister que quelques-unes. Peu de cassettes de gène habituellement retrouvées dans les intégrons dits cliniques ont pu être séquencées à partir de métagénomes environnementaux, quand les intégrons de classe 1 n'étaient pas spécifiquement ciblés (Rosser et Young, 1999 ; Stokes *et al.,* 2001 ; Koenig *et al.,* 2009 ; Elsaied *et al.,* 2011). Par ailleurs, des analyses de raréfaction réalisées sur les banques de clones ont montré que l'échantillonnage de séquences obtenues était loin d'être représentatif de la diversité totale en cassettes de gène (Stokes *et al*., 2001 ; Elsaied *et al*., 2007 ; Koenig *et al.*, 2008, 2009). Il a ainsi été estimé que dans un environnement donné, le nombre de séquences différentes pourraient être compris entre 2000 (pour un sol, Michael *et al.,* 2004) à 3000 (pour des échantillons côtiers, Koenig *et al.*, 2008). En comparaison, la totalité des cassettes de gènes portés par l'ensemble des intégrons cliniques connues à ce jour ne représentent qu'un peu plus d'une centaine de gènes (Partridge *et al*., 2009). Le réservoir de cassettes de gène environnementales est donc bien plus important que celui que représentent les intégrons cliniques et reste en grande majorité inexploré. Pourtant, il est

plus que probable que, parmi ces cassettes, on puisse retrouver des fonctions potentiellement adaptatives.

En effet, certaines fonctions portées par des cassettes de gène ont pu être reliées avec les conditions physico-chimiques (Nemergut *et al.,* 2004 ; Koenig *et al.,* 2008, 2009) Ceci suggère que certains intégrons pourraient être impliqués dans la réponse des bactéries face à des perturbations de leur milieu. En 2004, Michael *et al.,* ont observé des similarités dans le pool de cassettes de gène d'environnements proches et ont suggéré l'existence de groupes de cassettes de gène co-associés. De plus, Koenig *et al.,* (2008), ayant réalisé l'une des plus importantes études de séquençage de cassettes de gène faites à ce jour, a permis de montrer des cassettes de gène similaires dans 2 écosystèmes géographiquement éloignés mais présentant le même type de pression de contaminants. Ils ont ainsi émis l'hypothèse de l'existence d'un écotype de cassettes de gène. Cette hypothèse est soutenue par Wright et ses collaborateurs (2008) qui ont conclu que des pressions de sélection différentes mèneraient à des diversités en cassettes de gène différentes.

Ces corrélations entre fonctions et contaminants présents dans le milieu n'ont toutefois pu être mises en évidence que lorsqu'un large séquençage a été réalisé. Par ailleurs, ces approches de séquençages qui utilisent des amorces ciblant les sites *attC* ne permettent pas l'analyse de la première cassette de gène, celle qui devrait être le dernier évènement potentiellement adaptatif face à une pression de sélection (Huang *et al*., 2009 ; Elsaied *et al*., 2011). Malheureusement, l'importante dégénérescence du site *attI* et les variations de la séquence de *intI* rendent difficile une amplification de cette première cassette de gènes. Plusieurs approches ont été développées (Huang *et al*., 2009 ; Elsaied *et al.*, 2011, 2013) mais souvent difficiles à mettre en œuvre.

II- Matériel et méthodes

1) Origine des échantillons

1.1. Sédiments d'eau douce prélevés dans le bassin du Gave de Pau

Des sédiments ont été prélevés entre le 28 juin et le 5 juillet 2011 depuis la berge au niveau de plusieurs sites situés dans le bassin de l'Adour dans la région de Pau (43°32'N, 0°45'W). Ces lieux ont été choisis pour leur niveau d'anthropisation différents (Figure II-1) :

- boue active d'une station d'épuration (WWTP)
- sédiments prélevés à la sortie des eaux traitées de la station d'épuration (Effluent)
- sédiments d'un effluent industriel provenant d'un ancien site industriel dans lequel de forts taux de métaux ont été mesurés (Mar)
- sédiments d'une rivière ayant connu ponctuellement un taux élevé d'arsenic (Osm).

Des prélèvements ont également été effectués dans le Gave de Pau en aval de ces lieux (Gave1, 2 et 3, Figure II-1). Le site de référence pour cette étude est un ruisseau *a priori* non-anthropisé (Lou, 43°08'N, 0°07'W) qui est aussi le site choisi pour réaliser les microcosmes décrits ci-dessous (paragraphe III-1.3). Les sédiments ont été caractérisés en termes de concentrations en métaux (Abella *et al.*, 2015).

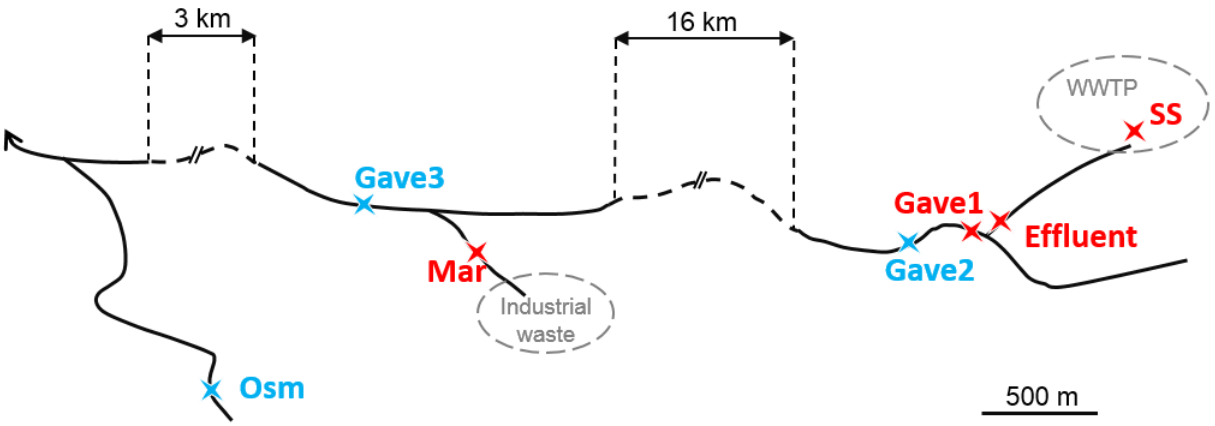


Figure II-1 : Localisation des sites d'échantillonnage du bassin du Gave de Pau (d'après Abella *et al.*, 2015).

1.2. Sédiments côtiers prélevés dans la région de Rovinj (Croatie)

Les sédiments ont été prélevés le 8 août 2011, au niveau de 4 sites situés dans la région côtière de Rovinj (Croatie) (Figure II-2). Les sites sont sous influence de différents types de pollution. Le site Luka est situé dans le port de la ville de Rovinj (environ 12 000 habitants, la population pouvant doubler durant la période touristique), utilisé pour les activités locales de pêche et de plaisance. Le site Marina est localisé proche de la ville de Rovinj entre le port local et un petit chantier naval. Le site Vani est situé au nord de Rovinj, à l'entrée du canal de Lim, aussi appelé baie de Lim. Le site Sredina est localisé au milieu de la baie de Lim et représente un environnement marin propre, non pollué et protégé, avec peu de tourisme et une faible activité de pisciculture (Figure II-2).

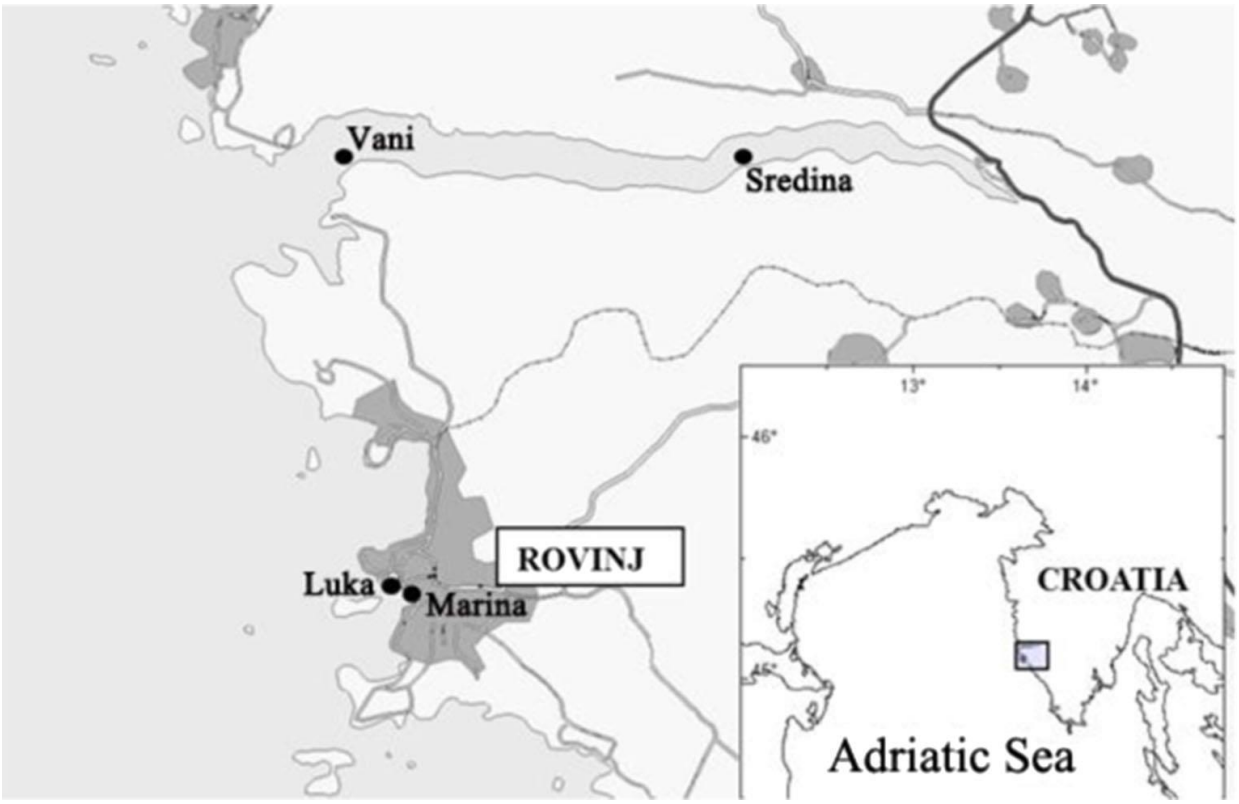


Figure II-2 : Localisation des sites d'échantillonnage dans la région de Rovinj : port de Luka (N45°04'46, E13°38'03), Marina (N45°04'37, E13°37'59), Sredina (N45°08'02, E13°37'31), et Vani (N45°07'49, E13°40'15) (d'après Duran *et al.,* 2015).

Les échantillons ont été prélevés depuis un bateau et les sédiments ont été récoltés en surface à une profondeur comprise entre 5 et 32 m. Une fois à bord, du sédiment des 5 premiers cm de la couche supérieure ont été répartis dans des cryotubes, rapidement

refroidis et stockés à -80°C. Les sédiments ont été caractérisés en terme de granulométrie et de concentrations en métaux et polluants (Duran *et al.,* 2015 ; Tableau I-2).

Tableau II-1 : Caractéristiques physico-chimiques des sédiments prélevés dans la région côtière de Rovinj. Dans un esprit de synthèse, tous les paramètres mesurés n'ont pas été reportés dans le tableau, et les données complètes sont publiées par Duran *et al.*, (2015).

Paramètres physico-chimiques	Luka	Marina	Vani	Sredina
Granulométrie				
Gravier (%)	13,04	0,74	0,227	0,41
Sable (%)	67,32	54,17	5,44	4,55
Composés argilo-limoneux (%)	19,64	45,09	94,29	95,04
Eau (%)	15,7	20,5	29,3	31,4
Métaux (mg/kg d.m.)				
As	9,22	23,44	8,12	13,85
Cd	0,26	0,09	0,08	0,09
Cu	69,95	30,59	13,77	18,45
Pb	24,38	6,95	1,83	1,35
Zn	115,60	50,66	71,75	88,00
Cr	22,49	26,31	76,11	98,14
Métaux totaux (mg/kg)	7593,2	11367,41	18820,18	24179,12
PAH (µg/kg)				
Benzo[a]pyrène	874,10	865,30	278,40	2,00
PAHs totaux (mg/kg)	10,609	9,867	4,555	0,028
PCB (mg/kg)				
Arochlore 1260	0,28	0,17	0,06	0,02

1.3. Microcosmes de sédiments d'eau douce

Des microcosmes ont été réalisés le 8 février 2011 à partir de sédiments prélevés dans un ruisseau forestier à priori non impacté par les activités anthropiques dans la région entre Lourdes, Pontacq et St Pé de Bigorre (43°08'N, 0°07'W) le 4 février 2011 et maintenus à 4°C jusqu'à la mise en place de l'expérimentation. 1750 mL de sédiment ont été tamisés en ajoutant 650 mL d'eau milliQ afin d'obtenir 2400 mL d'une suspension de sédiment homogène. Les microcosmes ont ensuite été mis en place avec 11 mL de suspension de sédiments, 5 mL d'eau du ruisseau pour les conditions contrôle ou 4 mL d'eau du ruisseau et 1 mL d'eau contenant un polluant. Le sédiment et l'eau ont été homogénéisés puis les microcosmes ont été laissés à stabiliser jusqu'à décantation. Chaque condition a été réalisée en 21 répétitions afin que des triplicatas soient sacrifiés à chaque temps d'échantillonnage. Les sédiments ont alors été conservés à -80°C jusqu'à extraction de l'ADN ou analyse

chimique. Les dates d'échantillonnage ont été définies à 0, 1, 7, 30, 90, 181 jours et 358 jours. Trois catégories de pollutions (métallique, organique et antibiotique) ont été réalisées : arsenic (50-50 arsénites et arséniate), benzo[a]pyrène et triméthoprime. Les concentrations ont été choisies en fonction des niveaux de pollution rapportés dans la littérature dans des sédiments environnementaux (Tableau II-2).

Tableau II-2 : Concentrations finales des polluants dans les microcosmes, basées sur les données environnementales de la littérature

Polluants	Concentration ($mg.L^{-1}$)	Données de la littérature
Arsenic	200	seuil de qualité = 22 $mg.kg^{-1}$ (US EPA, 2008) contaminations observées : sédiments côtiers 78 $mg.kg^{-1}$ (Rotkin-Ellman *et al.,* 2010)) drainage minier acide : 350 $mg.L^{-1}$ (Bertin *et al.,* 2011)
Benzo[a]pyrène	20	seuil de qualité = 1 $mg.kg^{-1}$ poids s (Johnson, Collier and Stein 2002) contaminations observées : sédiments côtiers 5,2 $mg.kg^{-1}$ (Oros, Collier and Simoneit 2007)
Triméthoprime	1	MIC = 1 $mg.L^{-1}$ (Vickers *et al.,* 2009) concentration de travail = 5 $mg.L^{-1}$

1.4. Mésocosmes de sédiments côtiers contaminés par du pétrole

Dans le cadre du projet ANR-CESA DECAPAGE (2011-2015 ; In depth caracterization of HC-degradation capacities of marine sediment microbial communities : adaptation, metabolic processes and influence of oxygenation regimes), des mésocosmes de sédiments ont été mis en place au CEDRE (Centre de Documentation, de Recherche et d'Expérimentations sur les pollutions accidentelles des eaux) à partir de janvier 2012. Les sédiments ont été prélevés à l'Aber-Benoît à Tréglonou (N 48° 33'12.40"; W 4° 32'8.69", Finistère ; Figure II-3). Ce site a été choisi car il est situé dans une réserve et considéré comme faiblement pollué par des hydrocarbures. En effet, même si l'Aber Benoit a été pollué par le pétrole de l'Amoco Cadiz en mars 1978, aucune trace de pollution par hydrocarbures n'a été détectée depuis 1990. Seuls des hydrocarbures biogéniques à raison de 4mg/g de sédiment s ont pu être détectés dans les 2 premiers cm du sédiment lors d'un précédent projet (projet ANR-SEST Dhyva 2007-2010 : Dégradation d'HYdrocarbures dans les VAsières). Par ailleurs, la salinité mesurée était d'environ 3,5‰ soit équivalente à celle de l'eau de mer.

Figure II-3 : Localisation géographique du site de prélèvement dans l'aber Benoît au niveau de la commune de Tréglonou (d'après Terrisse, 2014). : Limite de la zone sous influence océanique. : Site de prélèvement.

Douze mésocosmes ont été mis en place à partir de blocs de vases prélevés de manière à préserver leur structure et placés dans des bacs d'environ 30 L sur une membrane géotextile (Figure II-4). Chaque mésocosme est relié à un dispositif construit par le CEDRE alimenté en eau de mer filtrée sur sable et traitée aux UV permettant de simuler des cycles de marées de 12h et a été pollué ou non par du pétrole (OURAL étêté à 110°C) à raison de 24 ± 4 mg/g de sédiment s réparti sur les 2 premiers cm de sédiments. Les mésocosmes pollués ont été soumis, en triplicatas, à trois conditions différentes :

- 3 mésocosmes contaminés sans remaniement du sédiment (condition *Pétrole*) ;
- 3 mésocosmes contaminés par du pétrole et remaniés par hersage sur les 7 premiers centimètres de sédiments deux fois par semaine tout au long de l'incubation (condition *Pétrole + hersage*) afin de mimer le hersage, technique parfois utilisée comme stratégie de bio-remédiation de plages polluées par des produits pétroliers ;

- 3 mésocosmes contaminés par du pétrole et bio-augmentés par ajout d'organismes fouisseurs, *Hediste diversicolor* (1267 ±16 individus.m^{-2}) (condition *Pétrole + Hediste*) afin d'accroître la macrofaune bioturbatrice.

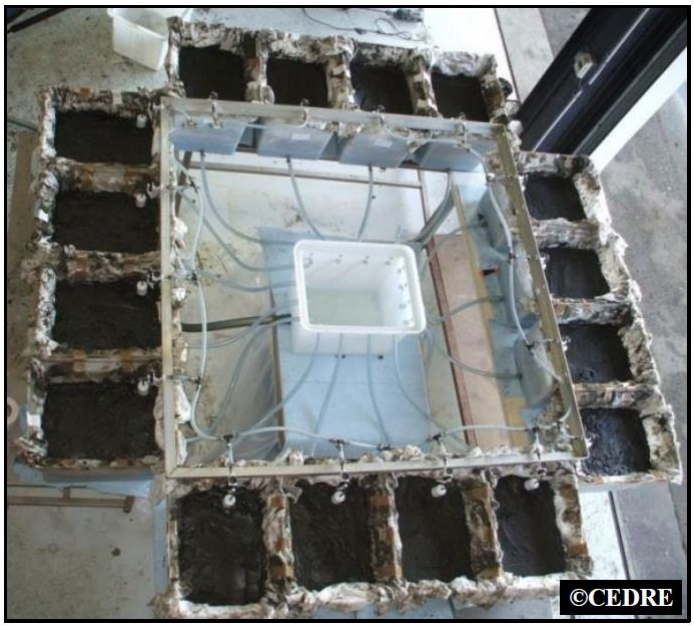


Figure II-4 : Dispositif expérimental recréant le cycle des marées mis en place pour l'incubation de sédiments côtiers non-déstructurés sous différents traitements de remaniement du sédiment (Hangar expérimental du CEDRE, Brest).

Des échantillons de sédiments ont été prélevés à 0, 1, 1,5, 2 mois d'incubation puis tous les mois jusqu'à la fin de l'incubation de chaque mésocosme. Un suivi des hydrocarbures (pristane et phytane) par GC-MS a été réalisé et les incubations ont été stoppées lorsqu'une biodégradation a pu être observée dans les mésocosmes. Ceci a permis l'obtention de 3 communautés bactériennes hydrocarbonoclastes et adaptées au pétrole. Les mésocosmes ont été stoppés lorsque la capacité de biodégradation a pu être observée : pour la condition *Pétrole + Hersage*, après 150 jours d'incubation, soit en juin 2012 ; pour la condition *Pétrole*, début octobre 2012 ; pour la condition *Pétrole + Hediste*, mi-novembre 2012. Au même moment, les mésocosmes contrôle sans contamination par du pétrole ont été arrêtés.

1.5. Bioréacteurs de sédiments côtiers

Une expérimentation en bioréacteurs a été réalisée sur les 3 communautés hydrocarbonoclastes obtenues à partir des incubations en mésocosmes décrits ci-dessus

(paragraphe III-1.4). Des suspensions de sédiments (20% m/v) provenant de la partie supérieure (6 premiers centimètres) des différents mésocosmes ont été réalisées par dilution dans du milieu d'eau de mer naturelle (eau de mer filtrée, 24 mM de bicarbonate, 0,4 $.10^{-3}$ M de phosphate de potassium, 0,2 $g.L^{-1}$ de chlorure d'ammonium et 1 $mL.L^{-1}$ de vitamine V7 (Pfennig *et al.,* 1981 ; pH = 7,2)). Les suspensions ont ensuite été incubées durant 15 jours en bioréacteurs, sous agitation permanente et soumises, en triplicatas, à des régimes d'oxygénation différents (Figure II-5) :

- anoxique (alimenté en N2)
- oxique (alimenté en O2)
- oscillation oxique/anoxique (O2 entre les 7ème et 8ème jours, et entre les 10ème et 11ème jours, N2 le reste du temps).

Un nouvel ajout de pétrole (5000 ppm par réacteur) a été réalisé au 5ème jour d'incubation (Figure II-5).

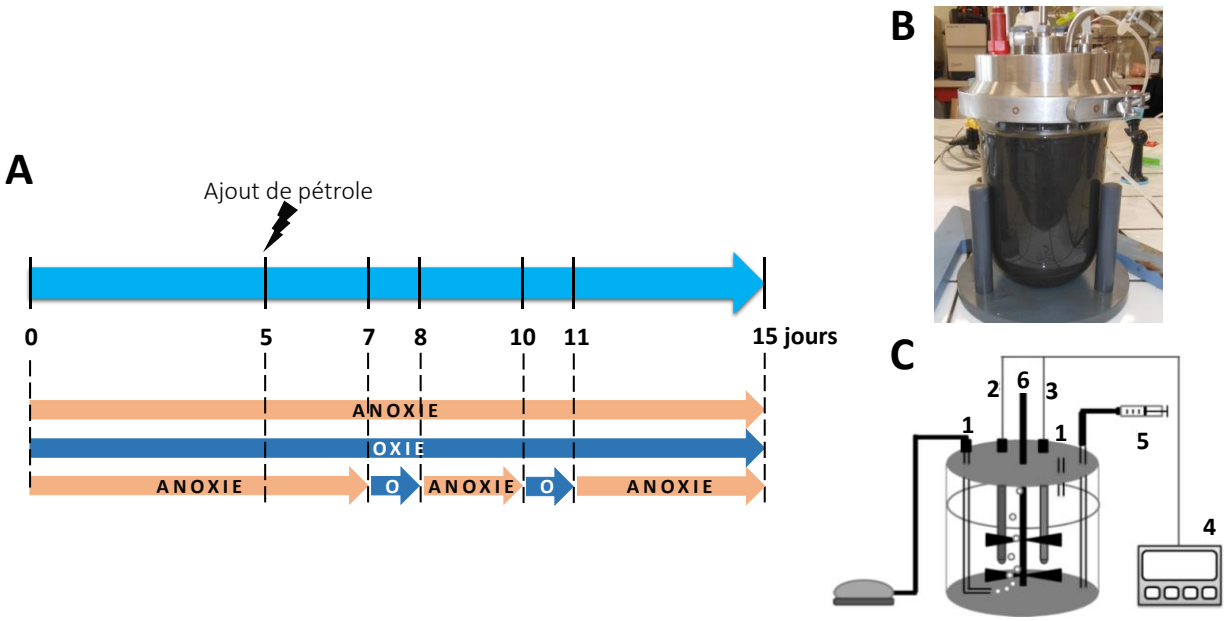


Figure II-5 : **A :** Schéma descriptif du plan expérimental des incubations en bioréacteurs de chaque sédiment pollué issu des mésocosmes, sous des conditions oscillantes anoxie/oxie, d'anoxie permanente ou d'oxie permanente. **B :** Photographie d'un bioréacteur. **C :** Représentation schématique d'un bioréacteur (modifié à partir de Terrisse *et al.,* 2014). 1, Entrée pour injection d'air ou d'azote gazeux filtré ; 2, Entrée pour la sonde de mesure de l'oxygène dissous et de la température ; 3, Entrée pour la sonde de mesure du pH et du potentiel d'oxydo-réduction ; 4, Emetteur relié aux sondes ; 5, Tube de prélèvement relié à une seringue ; 6, Système d'agitation.

Des échantillons des bioréacteurs ont été prélevés à différents temps au cours de l'incubation. Après ajout d'une solution stabilisatrice de l'ARN (tampon acétate à 1M saturé en phénol et éthanol 100%), ils ont été conservés à -80°C jusqu'à l'analyse.

1.6. Consortia

A partir des suspensions de sédiment des bioréacteurs en conditions oscillantes, des cultures d'enrichissement ont été réalisées afin d'obtenir des consortia en conditions de sulfato-réduction (Figure II-6). Pour cela, les suspensions de sédiment ont été diluées dans de l'eau de mer (décrit paragraphe I - 1.5) avec un facteur de 10^{-4} à 10^{-6}. Par la suite, les suspensions diluées de sédiments ont été réparties en flacons pénicilline fermés hermétiquement afin d'assurer des conditions anoxiques, avec comme source de carbone la fraction saturée ou aromatique du pétrole Oural (Vazquez et Mansoori, 2000). Les cultures ont été incubées à l'obscurité et à température ambiante. Deux repiquages ont ensuite été réalisés, avec 10% d'inoculum (v/v) (Figure II-4). A l'issu du deuxième repiquage, une extraction d'ADN et une PCR ciblant le gène *intOPS* ont été réalisées. Une amplification a été obtenue dans 31 des 32 consortia testés. Deux produits de PCR ont été séquencés afin de confirmer qu'il s'agissait bien du gène *intIOPS*. Par la suite, 4 consortia ont été sélectionnés en fonction des observations microscopiques des cultures : 3 consortia avec la fraction aromatique (nommés HC ARO E1, HC ARO A5 et HC ARO D4), et 1 consortium avec la fraction saturée (HC SAT H7). Ces consortia ont été repiqués dans 10 mL de milieu multi-substrat contenant 0,1 $mg.L^{-1}$ de fraction saturée ou aromatique, 4 $g.L^{-1}$ de Na_2SO_4, 0,2 $g.L^{-1}$ de NH_4Cl, 1 mM d'octanoate de sodium, 5 mM d'acétate de sodium, 5mM de lactate de sodium, 0,05 $g.L^{-1}$ d'extrait de levure, 1 $mL.L^{-1}$ de solution d'éléments traces SL12 (Overmann *et al.,* 1992), 1 $mL.L^{-1}$ de solution de sélénite tungstate (Widdel et Bak, 1992), 1 $mL.L^{-1}$ de résazurine, 0,4 mM de tampon phosphate de potassium pH 7.2, 1 $mL.L^{-1}$ de solution de vitamines V7 (Pfennig *et al.,* 1981), 150 µM de $Na_2S_2O_4$. Le consortium HC ARO D4 a été choisi pour être repiqué dans un volume de 300 mL milieu multi-substrats avec hydrocarbures avec une dilution au 100e dans le but d'extraire une grande quantité d'ADN pour caractériser les séquences de l'intégron porteur du gène *intIOPS*.

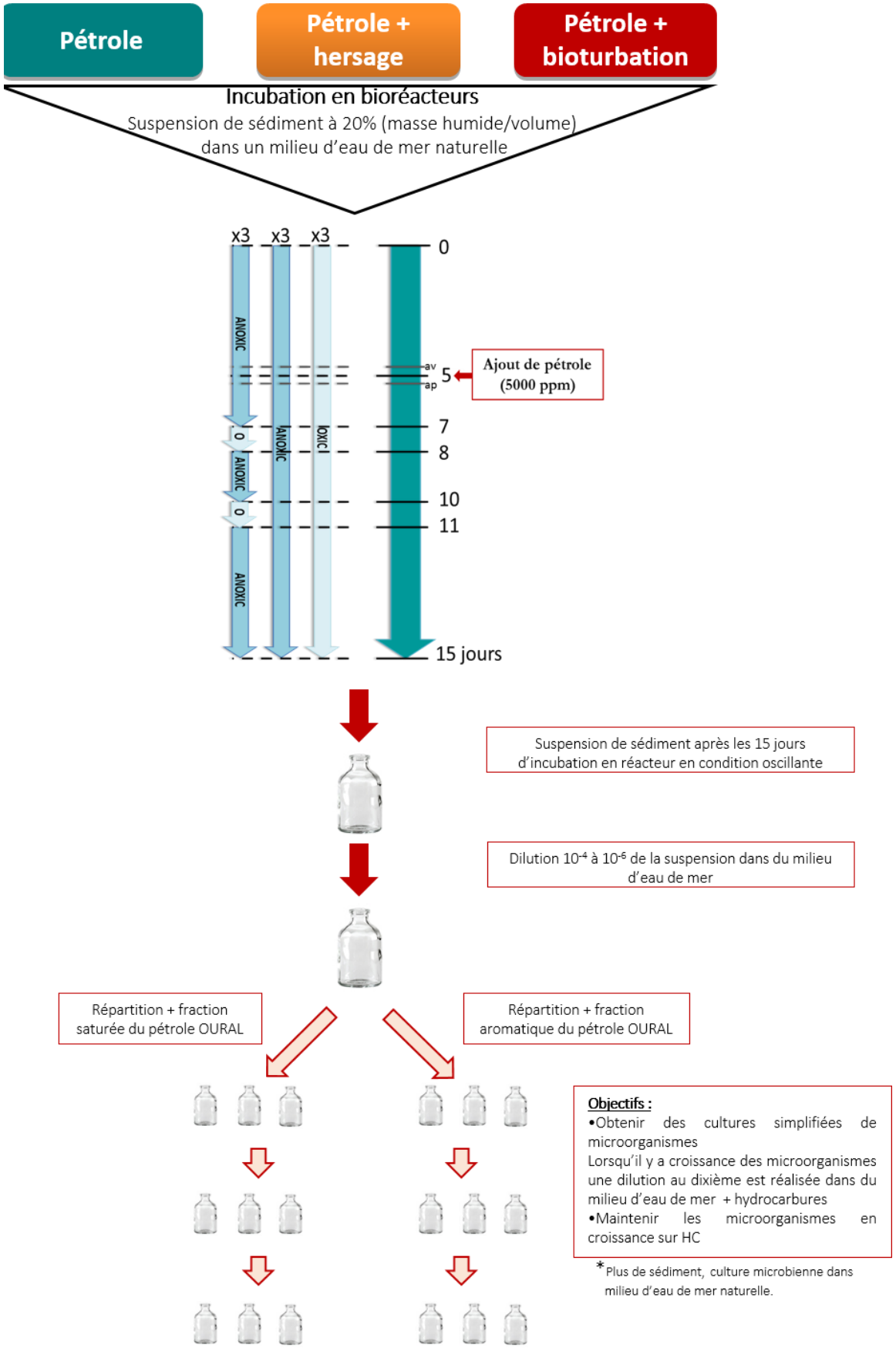


Figure II-6 : Schéma récapitulatif de l'incubation en bioréacteurs à l'obtention de cultures simplifiées de microorganismes se développant sur hydrocarbures (d'après Fanny Terrisse, données personnelles).

2) Méthodes moléculaires

2.1. Extraction des acides nucléiques

2.1.a. Extraction de l'ADN métagénomique de sédiments

L'ADN total a été extrait à partir de sédiments d'eau douce ou côtiers. Dans le cas d'un sédiment d'eau douce, l'extraction a été réalisée à partir de 0,70 g de sédiment. Pour les sédiments côtiers, 0,25 mg de sédiments ont été centrifugés à 5000 x g pendant 15 min et la phase aqueuse a ensuite été éliminée. L'extraction s'est ensuite déroulée de la même manière pour les 2 types de sédiment, à l'aide du kit commercial *PowerSoil® DNA Isolation Kit* (MoBio Laboratories) selon les recommandations du fournisseur. Les ADN métagénomiques ont été conservés à -20°C jusqu'à utilisation.

2.1.b. Extraction de l'ADN métagénomique à partir des consortia

L'extraction de l'ADN métagénomique à partir des consortia bactériens a été réalisée à l'aide du kit *PowerSoil™ DNA Isolation* (MoBio Laboratories). Trois tubes de 1,5 mL de culture cellulaire ont été centrifugés pendant 15 min à 10 000 x g puis le surnageant a été éliminé. Un des culots cellulaires a ensuite été resuspendu avec 100 µL de solution contenue dans un tube PowerBead puis le tout a été transféré dans le tube PowerBead. Les 2 autres culots ont été conservés à -20°C et utilisés si une quantité d'ADN plus importante était nécessaire. La suite du protocole a été réalisé selon les recommandations du fournisseur, sauf si le surnageant obtenu après la centrifugation après ajout de la solution C2, faisait plus de 600 µL, alors 2 tubes ont été utilisés pour les étapes avec les solutions C3 et C4. Une seule colonne a par la suite été utilisée pour les 2 tubes d'un même échantillon.

2.2. Construction des banques de séquences de IntI et de cassettes de gène

2.2.a. Amplification des gènes *intI* et des cassettes de gène à partir du métagénome

La réaction de PCR est préparée dans un volume réactionnel final de 50 µL contenant l'ADN matrice, 0,4 µM de l'amorce ICC32 et 4 µM de l'amorce ICC36 pour l'amplification du gène

intI, et 4 µM des amorces ICC1 et ICC4 pour l'amplification des cassettes de gène (Tableau II-3), 1X de *MasterMix Amplitaq Gold 360* (Applied Biosystems). Les amplifications PCR ont été réalisées dans un thermocycleur (PTC-100, MJ Research) selon le protocole suivant : une étape de dénaturation de l'ADN de 10 min à 95°C, puis 35 cycles comprenant 1 min de dénaturation à 95°C, 45 s d'hybridation à la température d'hybridation des amorces et 1 min d'élongation à 72°C, et enfin 10 min d'élongation finale à 72°C.

Tableau II-3 : Récapitulatif des amorces utilisées en PCR et qPCR.

Cible	Amorce	Séquence	Taille du fragment (pb)	$T_{hybridation}$ (°C)	Référence
intI	ICC32	GCC AAC GGN AAA CGK ATG C	~800	55	Inverse complémentaire de ILH72 (Gillings *et al.,* 2005)
	Int948R	NAR TAC RTG NGT RTA DAT CAT NGT			(Elsaied *et al.,* 2007)
Cassettes de gène	ICC1	GCS GCT KAN CTC VRR CGT TAG SG	variable	55	(Stokes *et al.,* 2001)
	ICC4	TCS GCT KGA RCG AMT TGT TAG VC			
intI1	qIntI1F	ACC AAC CGA ACA GGC TTA TG	285	57	(Wright *et al.,* 2008)
	qIntI1R	GAG GAT GCG AAC CAC TTC AT			
intIOPS	ICC62	GAC CCT CAA CCG AAC GGT TTC	424	50	Cette étude
	ICC81	CGA AAC TTC TGG ATC AAG TGA GGG			
ARNr 16S bactérien	338F	ACT CCT ACG GGA GGC AGC AG	181	55	(Lane, 1991)
	518R	ATT ACC GCG GCT GCT GC			
ARNr 16S bactérien	63F	CAG GCC TAA CAC ACT GCA AGT C	1300	58	(Marchesi *et al.,* 1998)
	1387R	GGG CGG WGT GTA ACA AGG C			

2.2.b. Purification et sélection par la taille des amplicons à cloner

i. Purification des produits de PCR sur le gène *intI*

Une amplification aspécifique d'un fragment d'environ 700 pb étant parfois observée, les amplicons sont purifiés sur gel d'électrophorèse à 1% d'agarose et la bande correspondant au fragment du gène *intI* d'une taille de 800 pb est découpée et purifiée avec le kit *GFXTM DNA and Gel Band Purification Kit* (GE Healthcare) selon les instructions du fournisseur.

ii. Sélection selon la taille des cassettes de gène amplifiées sur gel d'électrophorèse

La taille des cassettes de gène amplifiées variant de manière importante, il était important de s'affranchir du risque que les plasmides se referment seulement sur les amplicons de petites tailles ce qui empêcherait de pouvoir caractériser des cassettes de gène portant des potentiels cadres de lecture. Une sélection de taille a donc été réalisée par migration des produits de PCR sur un gel électrophorèse à 1%, puis découpe du gel à la taille choisie et enfin purification avec le kit *GFXTM DNA and Gel Band Purification Kit* (GE Healthcare) selon les instructions du fournisseur. Pour chaque produit de PCR, deux gammes de tailles ont été purifiés : de 400 à 600 pb et de 600 à 800 pb.

2.2.c. Ligation et transformation

Les clonages ont été réalisés à l'aide du kit commercial de clonage *pGEM-T easy* (ProMega) selon les recommandations du fournisseur. La ligation a été réalisée avec 2 µL de la solution de fragments obtenus après amplification et purification. La souche *E. coli* JM109 chimio-compétente a été utilisée comme cellule hôte. Pour la construction de banques de séquences de IntI et de cassettes de gène, la phase d'incubation en milieu Luria Bertani LB liquide (Tryptone, 10 $g.L^{-1}$; extrait de levure, 5 $g.L^{-1}$; NaCl, 5 $g.L^{-1}$) des clones après le choc thermique a été écourtée à 1 h au lieu de 1 h 30 comme conseillé dans le kit, afin d'éviter que certains clones commencent à se diviser et donc l'induction d'un biais dans l'abondance de certaines séquences. Les clones recombinants ont été sélectionnés sur milieu LB solide avec ajout d'ampicilline à 100 $g.mL^{-1}$, d'IPTG à 40 $g.mL^{-1}$et de X-gal à 40 $g.mL^{-1}$, après une nuit d'incubation à 37°C et 4 h à 4°C dans le but d'obtenir une coloration plus forte des colonies.

2.2.d. Criblage des clones et amplification des inserts

Afin de vérifier la présence du fragment d'intérêt dans les plasmides, une amplification de l'insert a été réalisée directement sur les cellules bactériennes, en parallèle du repiquage des clones sur milieu LB solide, avec ajout d'ampicilline (100 $g.mL^{-1}$), d'IPTG (40 $g.mL^{-1}$) et de X-gal (40 $g.mL^{-1}$). La réaction de PCR est préparée dans un volume réactionnel final de 50 µL contenant 0,4 µM des amorces M13F et M13R, 0,1 mM de dNTP, 1X de *ThermoPol Buffer* et

2,5 U de *Taq DNA Polymerase*. Les étapes de PCR sont les suivantes : 12 min à 95°C, 30 cycles de 1 min à 95°C, 45 s à 55°C et 1 min à 72°C, puis 10 min à 72°C. La taille et la quantité d'inserts amplifiés ont été contrôlées par électrophorèse sur gel d'agarose à 1%.

2.2.e. Séquençage selon la méthode Sanger et analyses des séquences

Le séquençage des produits PCR a été sous-traité par le prestataire *GATC Biotech* et a été réalisé sur chaque produit d'amplification à partir de la région du promoteur T7 du vecteur, après vérification sur gel de la qualité et de la quantité d'amplicons. Les séquences du gène *intI* et des cassettes de gène obtenues ont été éditées à l'aide du logiciel *Sequencer*. Les séquences de gènes *intI* ont ensuite été alignées grâce au logiciel *BioEdit*. Les séquences protéiques IntI ayant plus de 98% d'identité ont été regroupées en Unités Taxonomiques Opérationnelles (OTUs) à l'aide du logiciel DOTUR (Schloss et Handelsman, 2005). Les séquences de cassettes de gène ont quant à elles été analysées comme décrits au paragraphe I - 4).

2.2.f. Analyses phylogénétiques des séquences de IntI

Un blastP des séquences IntI obtenues au cours des différentes expérimentations a été réalisé contre les séquences connues dans la base de données nr. Un fichier contenant les 5 premières *hit*-séquences pour chaque séquence IntI analysées a été généré. A partir de ce fichier, les numéros d'accession ont été récupérés et collés dans un fichier .txt et un fichier fasta contenant toutes les séquences a pu être généré grâce à l'outil Batch Entrez disponible sur le site du NCBI.

Les séquences ont ensuite été alignées grâce au logiciel Muscle version v3.8.31 (Edgar, 2004). Les régions peu conservées ont ensuite été éliminées grâce au programme Gblocks (Castresana, 2000). Différents niveaux de stringence sont disponibles et ont été testés. Les paramètres par défaut ont été gardés. Les arbres ont ensuite été construits avec la méthode de l'estimation du maximum de vraisemblance (ou « *maximum-likelihood* ») et des valeurs de bootstrap ont été calculées pour 500 simulations à l'aide du logiciel MEGA4.1. Les arbres ont été enracinés avec une séquence d'une transposase Tnp, protéine qui n'appartient pas à la famille des tyrosines recombinases.

2.3. Quantification de gènes par PCR quantitative (qPCR)

2.3.a. Protocole de PCR en temps réel

Les amplifications en temps réel des gènes *intIOPS*, *intI1*, et du gène de l'ARNr 16S à partir des ADN métagénomiques ont été effectuées en duplicatas à l'aide du kit *Light Cycler 480 SYBR green Master I* (Roche) selon le protocole du fournisseur. La réaction a été réalisée dans un volume final de 10 µL contenant 2,5 µL d'ADN dilué au 1/10ème, au 1/50ème ou au 1/100ème après détermination du meilleur ratio de dilution, 0,4 µM de chacune des amorces (Tableau II-3) et 5 µL du mélange enzymatique. Les amplifications ont été réalisées sur l'appareil de PCR en temps réel *LightCycler 480* (Roche) selon le programme suivant : 5 min à 95°C, puis 40 cycles de 3 étapes : 15 s à 95°C, 15 s à la température d'hybridation des amorces et 20 s à 72°C à la fin desquelles la fluorescence est mesurée. Une courbe de dissociation des produits de PCR obtenus a été réalisée à la fin de chaque amplification. Pour cela, une rampe de température de 64°C à 95°C est effectuée durant laquelle la fluorescence est mesurée. Les résultats obtenus ont été analysés à l'aide du logiciel *LightCycler® Adapt Software* (Roche) qui permet la quantification des gènes ciblés à partir des gammes étalons d'ADN standard. La pureté des produits de PCR a été vérifiée par l'observation d'un unique pic de fusion après analyse des courbes de dissociation et par électrophorèse sur gel d'agarose.

2.3.b. Réalisation d'un standard externe pour la quantification

Un standard externe pour l'amplification de chaque gène d'intérêt a été réalisé à partir d'une solution d'ADN plasmidique recombinant contenant une séquence similaire à celle de la cible à quantifier et dont la concentration est connue. Cette solution est obtenue à partir d'une culture d'un clone obtenu après clonage d'un fragment de PCR du gène cible plus grand que le fragment amplifié en qPCR. Cette première PCR a été réalisée est préparée dans un volume réactionnel final de 50 µL contenant l'ADN matrice, c'est-à-dire de l'ADN métagénomique dans lequel le gène cible a été retrouvé (boue de station d'épuration pour *intI1* et pour le gène de l'ARNr 16S et ADN métagénomique extrait des mésocosmes de sédiments côtiers contaminés par du pétrole pour *intIOPS*), 0,4 µM des amorces (Tableau II-3), 1X de *MasterMix Amplitaq Gold 360* (Applied Biosystems). Les amplifications PCR ont été réalisées dans un thermocycleur (PTC-100, MJ Research) selon le protocole suivant : une étape de dénaturation

de l'ADN de 10 min à 95°C, puis 40 cycles afin d'obtenir une solution concentrée en fragments d'intérêts pour améliorer l'efficacité du clonage, comprenant 1min de dénaturation à 95°C, 45 s d'hybridation à la température d'hybridation des amorces et 1min d'élongation à 72°C, et enfin 10 min d'élongation finale à 72°C. Une purification sur gel et le clonage du fragment d'intérêt ont été réalisés de la même manière que décrit au paragraphe II - 2.2.

Après une nuit d'incubation à 37°C, une PCR de vérification de la présence d'un insert de la bonne taille a été réalisée sur 5 clones, en même temps qu'ils ont été repiqués sur milieu LB solide (comme décrit au paragraphe I - 2.2.d). 30 µL de produits d'amplification ont été ensuite séquencés en Sanger par le sous-traitant *GATC Biotech* et la séquence a été comparée aux séquences connues du gène cible, afin de confirmer l'identité du fragment amplifié et cloné. Les clones sélectionnés ont également été repiqués en milieu liquide, dans 30 mL de LB avec une concentration finale d'ampicilline de 100 $mg.mL^{-1}$ et incubés pendant 24 h à 37°C sous agitation (150 rpm).

Une extraction des plasmides a été réalisée à l'aide du kit *MiniPrep DNA* (QIAGEN) selon les instructions du fournisseur à partir de 3 mL de cultures liquides obtenues et centrifugées à 10 000g pendant 10 min. L'extraction a été vérifiée par électrophorèse sur gel d'agarose à 0,8%. Les plasmides ont ensuite été linéarisés par digestion réalisée en triplicatas dans un volume final de 20 µL pour 10 µL de solution plasmidique et 5 U d'enzymes *PstI* (NEB) pour *intI1* et le gène de l'ARNr 16S et *NdeI* (NEB) pour *intIOPS* qui possède un site de restriction *PstI* dans sa séquence. Les solutions de digestion ont été incubées pendant 3 h au bain marie à 37°C. Les plasmides linéarisés ont par la suite été précipités à -20°C durant la nuit dans un volume d'isopropanol 100% avec 1/10e d'acétate de sodium 3 M pH 5,2, puis centrifugés durant 15 min à 13 000 rpm le lendemain matin. Le surnageant a été retiré et les culots ont été lavés avec 300 µL d'éthanol 70% froid puis de nouveau centrifugés à 13 000 rpm durant 5 min. Le surnageant a été éliminé et les culots ont été incubés à température ambiante, bouchon ouvert, pour laisser évaporer les traces d'éthanol résiduelles. Enfin, les culots ont été resuspendus dans 50 µL d'eau stérile.

Les solutions plasmidiques linéarisées et purifiées ont été par la suite quantifiées au *Qubit® fluorometer* (Invitrogen) avec le kit *Quant-iT dsDNA BR Assay* (Invitrogen) et le nombre de copies du gène d'intérêt a été déterminé à l'aide la formule suivante :

$$Nombre\ de\ copies = \frac{6{,}023\ x10^{23} \times\ concentration\ de\ standard\ (ng.\mu L^{-1})}{MM\ (ng.mol^{-1})}$$

avec MM = masse molaire (en considérant que le poids moyen d'une paire de base est de 650 Daltons). Un logiciel permettant le calcul du nombre de copies à partir de la concentration mesurée est disponible sur le site de l'Université de Rhode Island à l'adresse suivante : http://cels.uri.edu/gsc/cndna.html.

2.3.c. Test de moyennes 2 à 2

Pour l'analyse statistique des résultats de qPCR, une analyse de la variance ou Anova a été réalisée afin de vérifier que les échantillons suivent une loi normale et sont indépendants. La normalité des résidus a été testée par un test de Shapiro. Une transformation logarithmique des données a parfois été nécessaire. Enfin, un test de Tukey HSD (*honest significant difference*) a été réalisé pour une comparaison 2 à 2 des moyennes. Ces tests ont été réalisés avec le logiciel R.

2.4. Caractérisation de la structure des communautés par T-RFLP

2.4.a. Obtention de fragments du gène de l'ARNr 16S marqués par fluorochrome

Une amplification du gène de l'ARNr 16S a été réalisée sur les sédiments étudiés. La réaction de PCR a été préparée dans un volume réactionnel final de 50 µL contenant 1 µL d'ADN métagénomique, 0,2 µM de l'amorce fluorescente 63F marquée en 5' avec le fluorochrome FAM et 0,2 µM de l'amorce 1387R (Tableau II-3), 1X de *MasterMix Amplitaq Gold 360* (Applied Biosystems). Les amplifications PCR ont été réalisées dans un thermocycleur (PTC-100, MJ Research) selon le protocole suivant : une étape de 10 min à 95°C, puis 30 cycles, comprenant 45 s à 95°C, 45 s à 55°C et 1 min à 72°C, et enfin 10 min à 72°C.

Les produits d'amplification ont par la suite été purifiés avec le kit *GFXTM DNA and Gel Band Purification* kit (GE Healthcare) selon les instructions du fournisseur.

2.4.b. Digestion des fragments du gène de l'ARNr 16S marqués

Une digestion des fragments PCR purifiés a été réalisée avec les enzymes de restriction *AluI* et *HhaI* (NEB) pendant 3 h à 37°C à l'obscurité. La réaction a été réalisée dans un volume de 10 µL contenant 100 ng de produits PCR purifiés, 5 U d'enzyme de restriction et 1X de tampon.

2.4.c. Migration des fragments digérés du gène de l'ARNr 16S par électrophorèse capillaire

Les produits de digestion ont été séparés par électrophorèse capillaire sur *ABI PRISMTM 3130 xl Genetic Analyzer* (Applied Biosystems). Un volume final de 10 µL a été préparé contenant 1 µL d'ADN digéré, 0,25 µL de marqueur de poids moléculaires (*GeneScanTM – 500 ROXTM Size Standard*, Applied Biosystems) et 8,75 µL de formamide. Après une injection électrocinétique de l'échantillon pendant 23 s dans le capillaire contenant le gel polymère Pop-7 (Applied Biosystems), la migration a été effectuée sous une tension de 15 kV pendant 20 min à 60°C.

2.4.d. Traitement des données brutes

Les profils obtenus ont été exportés sous forme d'une matrice de données intégrant pour chaque échantillon l'abondance (hauteur du pic, soit son intensité en unités de fluorescence) des pics détectés (T-RFs identifiés selon leur taille en pb). Les données ont été normalisées en divisant l'intensité de chacun des pics par l'intensité totale du signal (somme de l'intensité de tous les pics). Les pics dont l'intensité ne dépassait pas 1% de l'intensité totale ont été supprimés et une nouvelle normalisation a été réalisée. Les profils corrigés ont ensuite été alignés grâce au logiciel *T-align* (Smith *et al.,* 2005) avec un intervalle de confiance de 0,5 afin d'obtenir un tableau croisé dynamique. Celui-ci a par la suite été utilisé comme matrice pour l'analyse statistique de la structure des communautés.

2.5. Etude du pool de cassette de gènes selon la taille

2.5.a. Amplification des cassettes de gène à partir du métagénome

Des amplifications des cassettes de gène avec des amorces ciblant les sites *attC* ont été réalisées en triplicatas sur les sédiments étudiés. Les réactions de PCR ont été préparées dans un volume réactionnel final de 50 µL contenant 1 µL d'ADN métagénomique, 0,2 µM de l'amorce fluorescente ICC1 marquée en 5' avec le fluorochrome FAM et 0,2 µM de l'amorce ICC4 (Tableau II-3), 1X de *MasterMix Amplitaq Gold 360* (Applied Biosystems). Les amplifications PCR ont été réalisées dans un thermocycleur (PTC-100, MJ Research) selon le protocole suivant : une étape de 10 min à 95°C, puis 30 cycles, comprenant 45 s à 95°C, 45 s à 55°C et 1 min à 72°C, et enfin 10 min à 72°C.

2.5.b. Migration des cassettes de gène amplifiées et marquées par électrophorèse capillaire et traitements des données brutes

Les produits de digestion ont été séparés par électrophorèse capillaire comme décrit au paragraphe II-2.4.c. Les profils obtenus ont été exploités comme décrit au paragraphe II-2.4.d.

2.6. Caractérisation de l'environnement génétique du gène *intIOPS* par capture de gène et séquençage haut débit

Afin d'étudier l'environnement génétique du gène *intIOPS* à partir de la séquence de 800 pb connue, une approche par capture de gène a été utilisée. Pour cela, des sondes ARN biotinylées ont été synthétisées afin de cibler la séquence connue de *intIOPS* de manière spécifique. La capture de gène a été réalisée sur des fragments d'ADN métagénomique de 700 pb. Les fragments capturés ont par la suite été séquencés par un séquençage MiSeq 2x300 pb (Illumina), cette approche ne nécessitant pas de chevauchement des deux lectures. Un assemblage des séquences obtenues a ensuite été réalisé et a permis de reconstituer la séquence complète du fragment de *intIOPS* connu ainsi que l'environnement génétique du gène (Figure II-7).

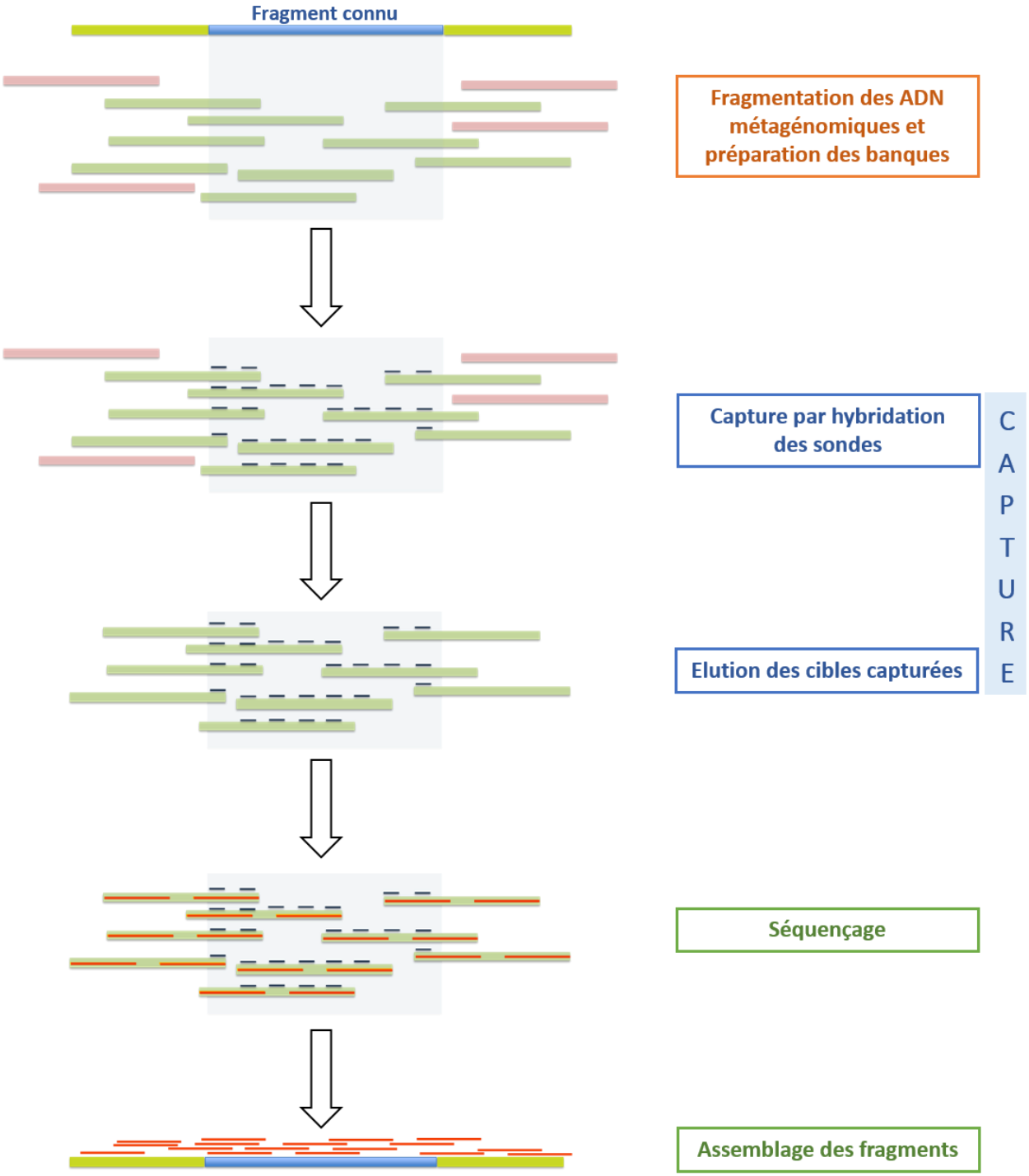


Figure II-7 : Représentation schématique du principe utilisé pour caractériser l'environnement du gène cible à partir du métagénome.

Cette expérience a été réalisée en collaboration avec le groupe animé par le Pr Pierre Peyret de l'EA CIDAM 4678 de l'Université d'Auvergne et avec la contribution de la plateforme Génomique du Génopole Toulouse Midi-Pyrénées (GeT) de Toulouse. Les sondes ont été dessinées par le Dr Nicolas Parisot et la synthèse des sondes de capture a été réalisée par Clémence Desfois et Auriane Bernard avant mon arrivée à l'EA-CIDAM. Les banques d'ADN

ont été préparées avec Olivier Bouchez de la plateforme GeT, les étapes de capture ont été réalisées lors de mon séjour à Clermont-Ferrand, et enfin le séquençage des banques a été réalisé par la plateforme GeT.

2.6.a. Préparation des banques de fragments métagénomiques

Dans un premier temps, l'ADN métagénomique de sédiments de mésocosmes, de suspension de bioréacteurs et de consortia ont été extraits comme décrit précédemment dans les paragraphes II-1.4, II-1.4 et II-1.6 respectivement et la qualité des ADN extraits a été vérifiée avec un lecteur Xpose™ 'Touch & Go' (Trinean). Une seconde étape de purification peut-être réalisée à l'aide du kit *PowerSoil* (MoBio Laboratories) selon le protocole suivant : après avoir ramené le volume d'ADN à 100 µL, 50 µL de *BeadSolution* (contenue dans un tube *BeadTube*) ont été ajoutés et le mélange a été agité par inversion. Puis 25 µL de solution C2 et 25 µL de solution C3 ont été ajoutés et le tube a été agité par inversion puis centrifugé pendant 2 min à 10 000 x g. Le surnageant a été transféré dans un nouveau tube et 2 volumes de solution C4 ont été ajoutés. Après avoir été mélangé au vortex, la solution a été déposée sur une colonne *SpinFilter* puis 500 µL de solution C5 avant d'être centrifugée 1 min à 10 000 x g. L'ADN a ensuite été élué avec 50 µL de solution C6 dans un nouveau tube.

Entre 200 et 300 ng d'ADN métagénomique a ensuite été fragmenté par la plateforme GeT grâce à la technologie Covaris (M220 Focused-ultrasonicator™) utilisant les ultrasons afin d'obtenir des fragments dont la distribution des tailles est comprise entre 100 et 1500 pb et centrée sur 500 pb.

Une sélection de taille a ensuite été réalisée en faisant migrer la solution d'ADN fragmenté sur un gel d'agarose 1% dans TBE à 100 V jusqu'à obtenir une bonne résolution. Le gel a ensuite été découpé de manière à prélever les ADN d'une taille centrée sur 700 pb puis l'ADN a été purifié avec le kit *GFX^{TM} DNA and Gel Band Purification* kit (GE Healthcare) selon les instructions du fournisseur.

Une banque Illumina (ajout des demi-adaptateurs) a ensuite été construite pour chaque échantillon par la plateforme GeT avec le kit *TruSeq Nano DNA Library Prep* (Illumina) puis amplifiée, purifiée et analysée sur puce Agilent.

2.6.b. Préparation des sondes de capture ARN biotinylées

i. Dessin des sondes

Les sondes dessinées font 50 nucléotides et ciblent la diversité de séquences de intIOPS obtenues lors des travaux de thèse du Dr Lionel Huang (Huang, 2010) à l'aide du logiciel KASpOD (Parisot *et al.,* 2012 ; http://g2im.u-clermont1.fr/kaspod/). Ce logiciel a utilisé 2 jeux de données : un fichier contenant toutes les séquences de *intIOPS* et un second fichier contenant un ensemble de séquences *intI* ne correspondant pas à *intIOPS*. Le logiciel sélectionne des sondes dans le premier jeu de données qui permettent de couvrir toute la diversité de ces séquences ciblées sans pour autant permettre de s'hybrider avec les séquences non-ciblées. Parmi l'ensemble de sondes sélectionnées par le logiciel une sonde toutes les 100 pb a été sélectionnée tout le long du gène.

ii. Préparation des oligonucléotides utilisés pour la synthèse des sondes ARN

La synthèse des oligonucléotides d'ADN utilisés pour la création des sondes ARN (avec à chaque extrémité l'adaptateur A en 5' et B en 3', Tableau II-4) a été réalisée par Eurogentec. Les oligonucléotides ont été reçus sous forme lyophilisées et reprises à 100 µM dans de l'eau « *nuclease free* ». Des aliquots d'oligonucléotides diluées à 10 µM ont été réalisées et congelés à -20°C.

Les oligonucléotides ont été amplifiés et le promoteur T7 a été ajouté. Les réactions de PCR ont été préparées dans un volume réactionnel final de 50 µL contenant 1 µL d'oligonucléotides, 0,2 µM des amorces T7-A et Rev-B, *Platinum® Taq DNA polymerase High Fidelity* 1X final (Invitrogen) (Tableau II-5), 2 mM de $MgSO_4$ et 0,2 mM de dNTPs. Les amplifications PCR ont été réalisées dans un thermocycleur (PTC-200, MJ Research) selon le protocole suivant : une étape de 2 min à 94°C, puis 35 cycles comprenant 30 s à 94°C, 30 s à 58°C et 20 s à 68°C, et enfin 5 min à 68°C. L'amplification des oligonucléotides a ensuite été

Tableau II-4 : Séquences des oligonucléotides utilisés pour la réalisation des sondes utilisées pour la capture de gène de *intIOPS*.
(Les valeurs de taille sont exprimées en nucléotides).

Dénomination	*Séquence complète des oligonucléotides*			*Taille de la sonde*	*Dégénérescence*	*Taille des oligonucléotides*	*Taille après PCR*	*Taille sonde finale*
	Adaptateur PCR A	*Séquence ADN de la sonde*	*Adaptateur PCR B*					
intIOPS-1_4	ATCGCACCAGCGTGT	AARGAGATGGGAAAGGACGAAGTGGAGGCTTTTTTAAGCGATTTGSCCGT	CACTGCGGCTCCTCA	50	4	80	112	89
intIOPS-100_64	ATCGCACCAGCGTGT	CTCTTTCTCTACRATCRRGTCCTGCAGAYGGACGTTYTRAARGAGATCAA	CACTGCGGCTCCTCA	50	128	80	112	89
intIOPS-200_4	ATCGCACCAGCGTGT	CAATGCGACGGYCCACCGTTCTCTCCTTTGATRAGGTGATGCGTTTGATC	CACTGCGGCTCCTCA	50	4	80	112	89
intIOPS-300_8	ATCGCACCAGCGTGT	GAGGGTCAATGACCTGGATTTTTCTTTAAATCAAATCYTGGYGAGAGACR	CACTGCGGCTCCTCA	50	8	80	112	89
intIOPS-400_8	ATCGCACCAGCGTGT	GGGATCRRAGAACACTTGAAATACGTGAAACACCTGCACGAARACGATCT	CACTGCGGCTCCTCA	50	8	80	112	89
intIOPS-500_2	ATCGCACCAGCGTGT	ATARAAAGGCGGAAAAGAGCTGGGGCTGGCAGTACGTGTTTCCCTCCAAA	CACTGCGGCTCCTCA	50	2	80	112	89
intIOPS-600_2	ATCGCACCAGCGTGT	CAGCGGTGTCAAGCGGCGTCACCATCTYCACCTGTCCTCGCTGAACAGCG	CACTGCGGCTCCTCA	50	2	80	112	89
ICC81_ext	ATCGCACCAGCGTGT	TGCAAAAACCGAAACTTCTGGATCAAGTGAGGGAAAAAATACGACTCAGG	CACTGCGGCTCCTCA	50	0	80	112	89

vérifiée par migration de 5 µL de produits sur gel d'agarose à 2% dans TBE à 100 V pendant 80 min. Les amplifications étant spécifiques, les 45 µL de produits de PCR restant ont été purifiés avec le kit MinElute PCR purification kit (Qiagen) sur des colonnes MinElute sorties à température ambiante 30 min avant utilisation. Si plusieurs bandes avaient été observées, le reste de produits de PCR auraient dû être purifiés sur gel avec le kit Qiaquick gel extraction (Qiagen) sur colonnes MinElute. La qualité des oligonucléotides amplifiés et purifiés a été vérifiée sur gel et la quantité au Nanodrop.

Tableau II-5 : Couples d'amorces utilisés lors de la préparation de sondes et pour la ré-amplification des banques après capture

Amorces	Séquences	T° hybridation (°C)	Etape
T7-A Rev-B	GGATTCTAATACGACTCACTATAGGGATCGCACCAGCGTGT CGTGGATGAGGAGCCGCAGTG	58	Préparation des sondes : ajout de promoteur T7
TS-PCR 1 TS-PCR 2	AAT GAT ACG GCG ACC ACC GAG A CAA GCA GAA GAC GGC ATA CGA G	58	Ré-amplification des banques

L'amplification des oligonucléotides a ensuite été vérifiée par migration de 5 µL de produits sur gel d'agarose à 2% dans TBE à 100 V pendant 80 min. Les amplifications étant spécifiques, les 45 µL de produits de PCR restant ont été purifiés avec le kit *MinElute* PCR purification kit (Qiagen) sur des colonnes *MinElute* sorties à température ambiante 30 min avant utilisation. Si plusieurs bandes avaient été observées, le reste de produits de PCR auraient dû être purifiés sur gel avec le kit *Qiaquick gel extraction* (Qiagen) sur colonnes *MinElute*.

La qualité des oligonucléotides amplifiés et purifiés a été vérifiée sur gel et la quantité au Nanodrop.

iii. Transcription *in vitro* (TIV) des sondes de capture

Un pool équimolaire des produits de PCR amplifiés-purifiés a été réalisé pour les oligonucléotides les moins dégénérés (intIOPS-1_4, intIOPS-200_4, intIOPS-300_8, intIOPS-400_8, intIOPS-500_2, intIOPS-600_2 et ICC81_ext) en fonction de leur dégénérescence. Pour calculer la dégénérescence de chaque sonde, le nombre de combinaisons possibles a

été calculé. Par exemple, une base Y correspond à 2 combinaisons possibles (C ou T). Les nombres de combinaisons pour chaque base dégénérée présente dans la séquence doivent être ensuite multipliés. Si une sonde possède dans sa séquences 3 bases dégénérées donnant chacune 2 combinaisons possibles, la dégénérescence totale de la sonde sera donc 2*2*2=8. Deux TIV ont ensuite été réalisées : une sur ce pool et une autre sur la sonde hautement dégénérée intIOPS-100_64.

La TIV a été réalisée sur 150 ng de pool ou de produit de PCR amplifié et purifié de la sonde intIOPS-100_64 avec le kit *MEGAScript* (Ambion) selon le protocole *synthesizing biotinylated aRNA*. Un mix a été réalisé à température ambiante contenant pour un volume réactionnel de 20 µL : tampon 1X final, 7,5 mM des ribonucléotides ATP, CTP et GTP, 5,625 mM de UTP et 1,875 mM de UTP biotinylé, 2 µL de rétrotranscriptase. Les réactions sont ensuite mise à incuber à 37°C dans un thermocycleur (PTC-200, MJ Research) pendant 6 h minimum voir sur la nuit. Le lendemain, 1 µL de TurboDNase sont ajoutés, puis les réactions sont remises à incuber pendant 30 min30min à 37°C. 82 µL d'eau ont été ensuite ajoutés (sans nucléases). Puis, 2 aliquots de 50 µL ont été préparés pour la purification.

iv. Purification des sondes ARN biotinylées

Les sondes biotinylées ont été purifiées avec le kit *RNeasy Plus Mini Kit* (Qiagen) en suivant les instructions du protocole « *Appendix D : Purification of Total RNA Containing Small RNAs from Cells* » en omettant l'étape D2 correspondant au passage sur la colonne *gDNA-Eliminator*. La concentration des sondes ARN biotinylées purifiées ont été ensuite dosée au nanodrop. Leur qualité a également été évaluée sur puce Agilent. Enfin, un pool équimolaire des sondes a été réalisé en prenant en compte la somme totale des dégénérescences dans le premier pool et dans la séquence de la dernière sonde.

2.6.c. Capture en solution

i. Hybridation entre les banques d'ADN et les sondes ARN biotinylées

Les réactions d'hybridation ont été réalisées dans un tube de 0,2 mL2mL. La première étape de dénaturation de l'ADN a été réalisée dans un volume de 16 µL avec 2,5 µg d'ADN de

sperme de saumon, 500 ng de la banque d'ADN fragmentée, amplifiée et purifiée, dans un thermocycleur (PTC-200, MJ Research), avec une étape de 5 min à 95°C puis maintenu à 65°C. Lorsque la température du thermocycleur était à 65°C depuis 5 min, celui-ci a été ouvert et les étapes suivantes ont été réalisées dans le thermocycleur sans sortir les tubes. 16 µL de tampon d'hybridation 2X (contenant 10X de solution SSPE, 10X de solution Denhardt's, 10 mM10mM de EDTA pH 8, et 0,20% de SDS) à 65°C ont été ajoutés ainsi que 500 ng du pool de sondes. La réaction d'hybridation a été laissée à 65°C pendant 24 h.

ii. Elution des cibles capturées

500 µg de billes M-280 Dynabeads (Life Technologies) ont d'abord été placées dans un tube de 1,5 mL placé sur un banc magnétique. Après sédimentation des billes, le tampon de stockage a été retiré. Les billes ont ensuite été rincées 3 fois avec 200 µL de la solution 1 (1M NaCl, 10 mM Tris-HCl pH 7,5, 1 mM EDTA), vortexées avant que le tube ne soit replacé sur le banc magnétique. Une fois les billes sédimentées, le surnageant a été retiré. Après, les 3 étapes de lavage, les billes ont été resuspendues dans 200 µL de solution 1.

Dans un second temps, les 32 µL de produits d'hybridation ont été ajoutés à la solution de billes de streptavidine lavées puis le tube a été incubé 30 min à 20°C en utilisant la roue d'un four à hybridation pour resuspendre les billes. Les billes ont ensuite été re-sédimentées sur banc magnétique et le surnageant a été éliminé. 500 µL de solution 2 (1X SSC, 0,1% de SDS) ont ensuite été ajoutés et le tube a été à nouveau mis à incuber 15 min à 20°C sur roue. Les billes sont ensuite re-sédimentées sur banc magnétique et les billes sont ensuite lavées 3 fois avec 500 µL de la solution 3 préchauffée à 65°C (0,1X SSC, 0,1% SDS). Les billes ont ensuite été resuspendues avec 500 µL de solution 3 et mises à incuber pendant 10 min à 65°C, en veillant à resuspendre les billes 1 à 2 fois au cours de l'incubation pour éviter qu'elles ne sédimentent dans le fond du tube. Le tube a ensuite été replacé sur le banc magnétique afin de faire sédimenter les billes et éliminer le surnageant. Les billes ont ensuite été resuspendues dans 50 µL de NaOH 0,1 M et la réaction a été incubée 10 min à 20°C. Enfin, le tube a été placé sur le banc magnétique pour sédimenter les billes et le surnageant contenant

les fragments cibles a été récupéré et ajouté à 70µL de Tris HCl 1M, pH 7,5 dans un nouveau tube.

2.6.d. Traitement des cibles capturées par billes AMPure

Les fragments capturés ont été ensuite traités avec des billes AMPure (Beckman Coulter). Pour cela, 120 µL de billes AMPure (1X final) ont été ajoutés à la solution contenant les fragments de capture. Après sédimentation des billes sur banc magnétique, le surnageant a été retiré et les billes ont été lavées 2 fois avec 500 µL d'éthanol 80%, en laissant incuber après le second lavage sur la paillasse, bouchon ouvert pour laisser s'évaporer les traces d'éthanol. Puis, 50 µL d'eau sans nucléases ont été ajoutés pour resuspendre les billes. Le tube a ensuite été replacé sur banc magnétique et une fois les billes sédimentées, le surnageant contenant les cibles capturées a été prélevé et transféré dans un nouveau tube.

2.6.e. Ré-amplification

Les fragments capturés ont ensuite été amplifiés en quintuplas. Les réactions de PCR ont été préparées dans un volume réactionnel final de 50 µL contenant 5 µL de solution contenant les cibles, 0,5 µM des amorces TS-PCR oligo 1 et 2 (Tableau II-5), *GC-rich PCR reaction buffer* 1X final (Roche),0,2 mM de *PCR Grade Nucleotide Mix*, 1 mM de $MgCl_2$ et 2U de GC-Rich enzyme Mix. Les amplifications ont été réalisées dans un thermocycleur (PTC-100, MJ Research) selon le protocole suivant : une étape de 4 min à 94°C, puis 25 cycles comprenant 30 s à 94°C, 1 min à 58°C et 1,5 min à 68°C, et enfin 3 min à 68°C. Les échantillons ont ensuite été conservés à 4°C. Un aliquot de 2 µL de chaque PCR a été prélevé pour être passé sur puce Agilent (Agilent Technologies) si besoin. Les 5 réactions d'amplification réalisées pour chaque banque ont ensuite été rassemblées et 5 µL de chaque pool ont été analysés sur puce Agilent. Les pools ont ensuite été séparés en 2 aliquots afin de réaliser un traitement AMPure.

Pour cela, des billes AMPure (1X final) ont été ajoutées aux aliquots et 2 lavages à l'éthanol ont été réalisés comme décrit précédemment. Les billes ont été resuspendues cette fois-ci dans 25 µL d'eau sans nucléases. Chaque aliquot traité a été analysé sur puce Agilent puis les aliquots ont été rassemblés dans un tube et le pool a été dosé au Nanodrop.

Un second cycle de capture a été ensuite réalisé de la même manière que décrit précédemment, à l'exception du traitement avec les billes AMPure qui a été réalisé avec un rapport de 0,8X de billes afin d'éliminer des fragments de tailles inférieures à 700 pb.

2.6.f. Séquençage MiSeq (Illumina)

Les banques capturées ont été séquencées en utilisant la technologie MiSeq 2x300pb nouvelle version V3 (Illumina). Les multiplexes ont été réalisés en utilisant des séquences index qui ont été ajoutées aux amorces anti-sens utilisées lors de la ré-amplification des banques avec 12 cycles, en ciblant les séquences des demi-adaptateurs. Une « *flowcell* » entière a été consacrée aux 5 banques, sans que d'autres échantillons n'aient été ajoutés, afin d'éviter le séquençage préférentiel de fragments plus petits, au détriment des fragments capturés d'une taille de 700 pb. Les produits de PCR ont été purifiés et chargés dans la cartouche Illumina MiSeq selon les instructions du fabricant. Après séquençage, la qualité a été vérifiée en utilisant le logiciel FastQC (Andrews).

2.6.g. Analyses des données MiSeq pour la reconstitution de la séquence de l'intégron IntIOPS

L'analyse des séquences a été réalisée grâce aux fonctions et programmes disponibles sur le cluster de calcul de la plateforme bioinformatique IBiSA de Toulouse. Dans un premier temps, les 2 fichiers correspondant aux 2 paires de lecture de chaque banque ont été concaténés (fonction *cat*). Le prétraitement des séquences a ensuite été réalisé à l'aide de l'outil prinseq. Les fins de séquences de mauvaise qualité (inférieure à Q15 sur une fenêtre de 2 nucléotides) ont été supprimées (- trim_qual_right 15 ; -trim_qual_window 2) ; les séquences de moins de 100 nucléotides ont été supprimées (- min_len 100) ; puis les séquences d'une qualité inférieure à une valeur de Q20 ont été supprimées (- trim_qual_right 20) et enfin, les séquences contenant plus de 1% de N ont été supprimées (- ns_max_p 1).

Un assemblage a ensuite été réalisé avec l'assembleur *de novo* IDBA-UD (Peng *et al.,* 2012). Une valeur de kmers est généralement définie (avec KmerGenie par exemple). Nous avons choisi une méthode multi-kmers avec des valeurs allant de 23 à 127. Les 53 fichiers de contigs obtenus ont ensuite été clustérisés, c'est-à-dire que les contigs regroupés en groupe de

séquences partageant 98% de similarité, et une séquence représentative du groupe a été sélectionnée grâce au programme CD-HIT (Huang *et al.,* 2010).

Le fichier de séquences représentatives a ensuite été transformé en base de données grâce à l'outil formatdb et un blastn des séquences connues de *intIOPS* a été réalisé contre cette base de données. Plusieurs blastn ont été faits au fur et à mesure que de nouvelles séquences de l'intégron été caractérisées. Enfin, l'assemblage des différents contigs d'intérêt a été réalisé manuellement grâce à l'identification des zones de chevauchement avec le logiciel BioEdit.

3) Analyses statistiques

3.1. Analyses des banques et indices de diversité

Le taux de recouvrement de la banque qui permet d'estimer la représentativité de l'échantillonnage a été mesuré à l'aide de l'estimateur de Good : $\hat{C} = 1 - \frac{\hat{S}_n^1}{n}$ avec $\hat{S}_n^1$ le nombre de clones représentés une seule fois dans une banque de n clones (Good, 1953). L'analyse de raréfaction (Simberloff, 1978) a été utilisée pour réaliser les courbes représentant le nombre d'OTUs attendu en fonction du nombre de séquences échantillonnées. La diversité des séquences obtenues a été analysée à l'aide de plusieurs méthodes fournies par le logiciel DOTUR.

L'estimateur de richesse Chao 1 : $\hat{S}_{Chao1} = \hat{S}^{\geq 1} + \frac{(\hat{S}^1)^2}{2\hat{S}^2}$ (Chao, 1984) avec $\hat{S}^{\geq 1}$ = le nombre de séquences différentes observées et $\hat{S}^z$ = le nombre de séquences observées z fois ; et l'indice de Shannon-Weaver : $H = -\sum_{i=1}^{S} p_i \ln p_i$ (Shannon 1948) avec p_i = la probabilité qu'une séquence, un t-RFs ou une taille donnée appartienne à l'espèce i ont été appliqués pour calculer respectivement le nombre de séquences non observées et la diversité des séquences de IntI, des cassettes de gène ou la diversité de la communauté bactérienne.

3.2. Analyse multidimensionnelle non-métrique (nMDS)

L'analyse des tables d'OTUs des banques de séquences IntI ainsi que des tables de T-RFs et de taille de cassettes de gène ont été traitées par analyse multidimensionnelle non-métrique ou nMDS, basée sur des distances de Bray-Curtis. Cette analyse présente l'avantage d'obtenir une représentation la plus fidèle possible des objets, dans un espace réduit à généralement 2 ou 3 dimensions. Pour cela, une méthode itérative est utilisée jusqu'à obtenir la meilleure représentation. Un coefficient de « stress » (Stress de Kruskal) indique le degré de contrainte imposé lors de la réduction dimensionnelle et permet de savoir si la représentation est fidèle ou non. Une valeur de stress inférieure à 0,1 décrit une représentation excellente ; un stress compris entre 0,1 et 0,25, une représentation satisfaisante ; entre 0,25 et 0,5, probablement aléatoire ; et supérieure à 0,5, une représentation de qualité médiocre.

Pour la réalisation de ces analyses, le logiciel Primer 6 version 6.1.16 (Clarke et Gorley, 2006) a été utilisé.

4) Analyses des séquences de cassettes et de IntIOPS

4.1. Caractérisation des sites *attC* et *attI*

Pour la caractérisation des sites *attC*, les motifs RYYYAAC et GTTRRRY ont été recherchés ainsi que des séquences répétées inversées. Une fois repérées, les séquences putatives des sites attC ont été analysées grâce à un outil de prédiction de structure secondaire. Les logiciels de prédiction pour l'ARN étant les plus courant, l'outil MFOLD 3.6 disponible sur le portail Mobyle @Pasteur, en réglant le paramètre « DNA » (section structure/2D_structure) a été utilisé. Le résultat de la prédiction n'étant pas toujours correct, chaque résultat a ensuite été vérifié voir repris manuellement afin d'obtenir la bonne structure secondaire laissant apparaître des bases mésappariées, généralement un G et un T, espacés de 4 à 7 nucléotides.

Afin de caractériser les sites *attI* des séquences *intIOPS*, le motif GTTRRRY a été recherché au voisinage du putatif point de recombinaison.

4.2. Caractérisation des cadres de lecture

Pour l'analyse des cassettes de gène, mais aussi de la séquence complète de *intIOPS*, les cadres de lecture (ou ORF pour *Open Reading Frame*) ont dû être recherchés. Pour cela, l'outil ORF finder disponible sur le site du NCBI a été utilisé pour déterminer tous les cadres d'ouverture. Un blastP peut-être réalisé à partir de cet outil sur l'ORF d'intérêt.

Les ORF prédits à partir des cassettes de gène ont par la suite été analysés grâce aux outils InterProScan, Psortb et SignalP disponibles sur le serveur ExPASy dans le but de caractériser de potentiels sites de protéines fonctionnelles, déterminer la localisation cellulaire putative de la protéine et de déterminer la présence d'un signal peptide.

4.3. Caractérisation des régions promotrices

Afin de faire une prédiction du promoteur Pc, pour l'expression des cassettes de gène, et du promoteur P*intIOPS*, pour l'expression du gène *intIOPS*, dans la région entre le début du gène *intIOPS* et le point de recombinaison du site *attIOPS*, des outils en ligne ont été utilisés : Promoter Prediction, Martin Reese, Lawrence Berkeley Laboratory, CA, U.S.A ; et BProm, *Softberry*.

III- Etude de la diversité des intégrons dans des environnements naturels

Introduction

Bien que plusieurs études ont été réalisées sur les intégrons environnementaux, c'est-à-dire en dehors du contexte clinique, il est encore difficile de conclure quant aux rôles des intégrons dans les communautés bactériennes environnementales. En effet, plusieurs études se sont focalisées sur les cassettes de gène en fonction des conditions environnementales et ont pu montrer des corrélations entre les fonctions portées par les cassettes et les contaminants présents dans l'environnement. Toutefois, la diversité des cassettes de gène est trop importante et la majorité porte des cadres de lecture codant des fonctions inconnues. L'intégrase a quant à elle été peu étudiée. Même si à l'heure actuelle environ 400 intégrases provenant d'environnements divers ont pu être caractérisées, les études montrent que la diversité de cette protéine est bien plus vaste.

Il est donc avéré aujourd'hui que la diversité des intégrons est bien plus importante dans l'environnement que dans les milieux cliniques et que cette diversité est loin d'être appréhendée dans sa totalité. Par ailleurs, il semble qu'elle soit soumise à l'influence des conditions environnementales et donc varie d'un environnement à l'autre. De par la richesse d'environnements existants, il est nécessaire de réaliser des études en échantillonnant un nombre le plus exhaustif possible d'environnements différents. Au cours de ces travaux de thèse, 2 études ont été réalisées sur des environnements aquatiques qui feront l'objet des deux sous-parties de ce chapitre : dans un premier temps, un cours d'eau douce et ses affluents, et dans un second des environnements côtiers.

1) Intégrons dans des sédiments d'eau douce

Pour cette première étude, nous avons choisi de travailler sur des sédiments prélevés dans le bassin du Gave de Pau. Cette rivière, dont la source la plus en amont se situe dans le cirque de Gavarnie dans les Hautes-Pyrénées, traverse des agglomérations de petite à moyenne taille telles que Argelès-Gazost, Lourdes ou Pau (7 377 habitants pour la communauté des communes de la Vallée d'Argelès-Gazost (2008) ; 22 142 habitants pour l'aire urbaine de Lourdes (2009) et 146 982 habitants pour la communauté d'agglomération de Pau-Pyrénées

(2011)). Les activités, autres qu'urbaines, exercées tout au long de son lit sont essentiellement d'origines agricoles et touristiques. On note toutefois, au niveau de sa source, d'anciennes mines de plomb et de zinc qui sont aujourd'hui fermées. On trouve également 3 complexes chimiques et industriels dont un situé en amont de la zone d'échantillonnage (Pierrefitte-Nestalas, 65) et un situé dans la zone d'échantillonnage (Marsillon, 64). Le Gave de Pau présente tout au long de son cours une contamination en zinc et organo-étains mais aussi des taux plus élevés en nickel et arsenic à certains niveaux, comme c'est le cas aux alentours de Pau (Bareille, communication personnelle). Une contamination en biphényles polychlorés (PCB) a également été révélée ces dernières années, entrainant des restrictions sur la pêche dans certaines zones du Gave de Pau.

Le débit du Gave de Pau, dont la moyenne est de 81,3 $m^3.s^{-1}$, présente d'importantes variations au cours de l'année, oscillant entre 120 $m^3.s^{-1}$ au printemps au moment de la fonte des neiges et 41 $m^3.s^{-1}$ en août (http://www.hydro.eaufrance.fr/selection.php). De plus, cette rivière possède l'une des plus importantes lames d'eau du bassin versant de France. Le Gave de Pau est donc sujet à une importante remobilisation du sédiment en période de crue et de fonte des neiges, et présente un faible débit en été alors que l'activité touristique sur son cours est importante. C'est donc une intéressante aire d'échantillonnage pour étudier l'impact d'activités humaines chroniques modérées sur l'adaptation des communautés bactériennes. Nous avons également choisi d'étudier comme témoin un sédiment prélevé au niveau d'une des nombreuses sources du Gave de Pau éloignée des activités humaines. Dans un premier temps, les résultats majeurs de cette étude seront présentés sous la forme d'une publication scientifique soumise au journal *FEMS Microbiology Ecology,* (actuellement en révision), suivie d'une analyse de la structure secondaire des sites de recombinaison *attC* qui ont pu être séquencés, et enfin une étude phylogénétique des séquences d'IntI.

1.1. Diversité des intégrons dans des communautés bactériennes de sédiments d'eau douce soumis à différents niveaux de contamination

Résumé

Le travail présenté ici vise à décrire la diversité des intégrons dans les communautés bactériennes des sédiments d'eau douce prélevés le long du Gave de Pau (France) avec différents niveaux et types de contaminants.

A partir du métagénome de chaque site d'échantillonnage, des banques de séquences IntI ont été construites. La diversité des cassettes de gène a également été caractérisée en se basant sur la taille des cassettes de gène. Les données de séquences IntI et de taille de cassettes de gènes ont été analysés par nMDS et comparées avec la structure de la communauté bactérienne évaluée par T- RFLP sur le gène de l'ARNr 16S.

322 nouvelles séquences IntI ont été identifiées révélant une importante diversité, plus grande que celle attendue précédemment dans des environnements non cliniques. La structure des communautés bactériennes n'a pas pu expliquer pleinement la diversité des intégrons. La diversité de l'intégrase semble liée au niveau de contamination tandis que la diversité des cassettes de gène répondrait au type de contaminants présents dans l'environnement. Par ailleurs, le gène *intI1* n'a pas pu être détecté dans les banques réalisées et a été suivi spécifiquement par qPCR. Son abondance est plus importante dans les milieux les plus impactés, notamment par des contaminations métalliques. Enfin, un séquençage des cassettes de gène à partir de 2 échantillons les plus impactés de l'étude a permis d'identifier 60 cassettes de gène différentes contenant un cadre de lecture. Dix-sept cassettes présentent une homologie significative avec des cassettes de gène séquencées dans des environnements côtiers pollués. Toutefois, la majorité des ORFs codent des fonctions encore inconnues.

Ces résultats en accord avec les précédentes études réalisées fournissent des éléments supplémentaires qui soutiennent l'hypothèse que les intégrons environnementaux seraient impliqués dans l'adaptation des communautés bactériennes en réponse à la présence de contaminants dans les milieux naturels.

RESEARCH ARTICLE

Integron diversity in bacterial communities of freshwater sediments at different contamination levels

Justine Abella, Anne Fahy, Robert Duran and Christine Cagnon*

Equipe Environnement et Microbiologie - MELODY group, IPREM UMR CNRS 5254, Université de Pau et des Pays de l'Adour, BP 1155, 64013 Pau cedex, France

*Corresponding author: Equipe Environnement et Microbiologie, IPREM UMR CNRS 5254, Université de Pau et des Pays de l'Adour, BP 1155, 64013 Pau cedex, France. Tel: +33-5-59-40-74 80; Fax: +33-5-59-40-74 94; E-mail: christine.cagnon@univ-pau.fr

One sentence summary: This work investigated integron diversity in bacterial communities from freshwater sediments and provided new further elements that argue the non-clinical integron would be involved in the adaptation of bacterial communities.

Editor: Patricia Sobecky

ABSTRACT

Integrons, genetic elements known to be involved in the adaptation of pathogenic bacteria, were first discovered in the clinical setting. However, they are ancient structures found in various environments. When clinical integrons have a low diversity, with three integrases and gene cassettes essentially encoding antibiotic resistance, in natural environments, integrons show a greater diversity, of both gene cassettes and integrases. Although a large number of gene cassettes from environmental samples have been identified, integrase diversity remains poorly documented, and has not yet been investigated in freshwater environments. The work presented here explores environmental integrons in sediments from a freshwater environment, with emphasis on integrases. Integron diversity in bacterial communities was analyzed at sampling stations with different contamination levels and contaminant types. A total of 684 integrase sequences were obtained and grouped into 322 previously undescribed integron classes, revealing a diversity wider than that previously expected in non-clinical environments. The bacterial community structures did not fully explain the integron diversity suggesting that integrase diversity could be influenced by contamination level, and that contaminant type could influence gene cassette diversity. These results provide further arguments for the involvement of integrons in the adaptation of bacterial communities in response to contaminants in natural environments.

Keywords: integron integrases; gene cassette pool; WWTP; metal; freshwater environment

INTRODUCTION

Integrons are bacterial genetic elements well known in clinical environments for their involvement in spreading antibiotic resistance among pathogenic bacteria (Hall 1997; Rowe-Magnus, Davies and Mazel 2002; Ploy *et al.* 2005). They were first discovered in this environment (Stokes and Hall 1989). However, they are ancient bacterial structures (Mazel *et al.* 1998; Rowe-Magnus and Mazel 2001) found in most bacterial communities (Nield *et al.* 2001; Nemergut, Martin and Schmidt 2004; Elsaied *et al.* 2007, 2011; Gillings *et al.* 2008; Koenig *et al.* 2009). Integrons contain a gene, *intI*, encoding an integrase, IntI, a recombination site, *attI* and an array of gene cassettes characterized by a coding sequence and a recombination site, *attC* (Hall and Collis 1995; Boucher *et al.* 2007; Cambray, Guerout and Mazel 2010). While *intI* and *attI* are the stable platforms of integrons, the gene cassette order and content varies: their sequences are diverse (Holmes

Received: 16 June 2015; **Accepted:** 4 November 2015

© FEMS 2015. All rights reserved. For permissions, please e-mail: journals.permissions@oup.com

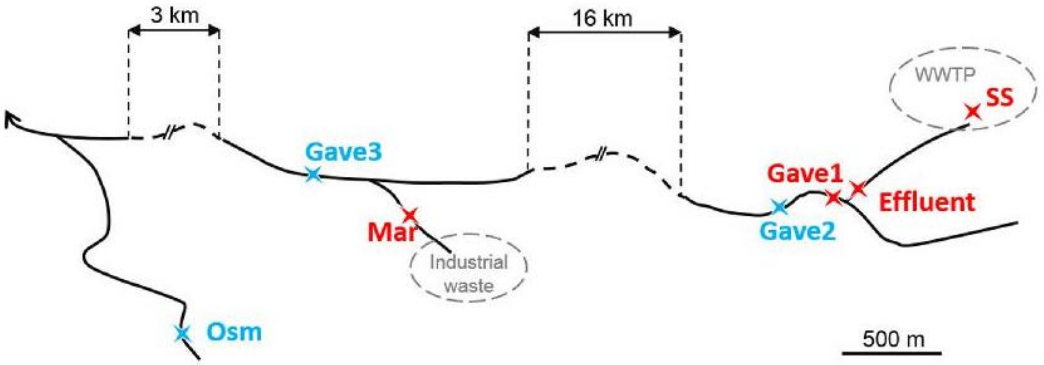


Figure 1. Schematic sampling map of the *Gave de Pau* basin. In blue, the weakly impacted sites and in red, the most impacted sites. Lou, the site external to the basin (~60 km away from the *Gave de Pau* River samples) is not represented here.

et al. 2003; Koenig *et al.* 2008) and their number can vary from none to over 100 (Mazel 2006; Gillings 2014). The integration of a new gene cassette is mediated by IntI, which catalyzes the recombination between *attC* of the gene cassette and *attI* of the integron (Collis *et al.* 2002; Demarre *et al.* 2007; Gestal, Liew and Coleman 2011). The expression of gene cassettes is driven by one or two promoters located within the *intI* sequence (Levesque *et al.* 1994; Collis and Hall 1995) or sometimes by cassette-borne promoters (Michael and Labbate 2010). Thus, integrons are both expression systems and gene cassette reservoirs (Holmes *et al.* 2003; Labbate, Case and Stokes 2009). Because of their capacity to acquire and spread genes conferring antibiotic resistance, they are involved in the adaptation and the evolution of bacteria, at least against antimicrobial pressure.

In clinical environments, integrons have a low diversity, with three *intI* sequences defining class 1, 2 and 3 integrons, and gene cassettes essentially encoding antibiotic resistance (Hall and Collis 1995; Ploy *et al.* 2000; Cambray, Guerout and Mazel 2010). Because of the importance of understanding the spreading process of antibiotic resistance in the environment, many surveys have focused on clinical integrons in other environments (for review, see Stalder *et al.* 2012). Class 1 integrons are widespread within and outside clinical environments, and are largely associated with antibiotic resistance (Henriques *et al.* 2006; Moura *et al.* 2007; Laroche *et al.* 2009; Byrne-Bailey *et al.* 2011; Gaze *et al.* 2011), but it has also been demonstrated that *intI1* abundance increases with ecological perturbations or metal contamination, suggesting that class 1 integrons may also be involved in adaptation to other perturbations (Hardwick *et al.* 2008; Wright *et al.* 2008; Rosewarne *et al.* 2010). However, integrons have a greater sequence diversity in natural environments than in clinical environments, both in gene cassettes and integrase genes. Part of this diversity was highlighted by the analysis of integrons from isolated bacteria (Boucher *et al.* 2007; Nemergut *et al.* 2008), but a much greater variety of sequences have been revealed by culture-independent methods (Stokes *et al.* 2001; Holmes *et al.* 2003; Elsaied *et al.* 2007, 2011; Koenig *et al.* 2008, 2009; Michael and Andrew 2010; Moura *et al.* 2010). Nevertheless, little is known about the role of these environmental integrons, especially with regard to bacterial adaptation.

Several studies focusing on gene cassette sequencing from environmental samples have shown that about 80% of sequences encoded unknown functions. When functions can be identified, they are diverse. Some functions were related to physicochemical conditions (Nemergut, Martin and Schmidt 2004; Koenig *et al.* 2008, 2009) suggesting that integrons can be involved in bacterial responses to various disturbances. In contrast, integrase diversity in natural environments has been poorly investigated to date; only six studies have reported on integrase diversity in soil and marine environments (Nield *et al.* 2001; Nemergut, Martin and Schmidt 2004; Elsaied *et al.* 2007, 2011, 2014, Rodríguez-Minguela *et al.* 2009), and none in freshwater environments. Together, integron sequences from isolated bacteria and these studies revealed more than 300 integron phylotypes, suggesting a larger diversity than that found in clinical environments (Rapa and Labbate 2013).

The present work aimed to explore environmental integrons in sediments from freshwater environments, with emphasis on *intI* diversity. Integron diversity was analyzed in bacterial communities along the *Gave de Pau* River (France) at sampling stations with different contamination levels and contaminant types, in order to investigate the impact of human activities on environmental integrons.

MATERIALS AND METHODS

Sampling

Sampling was performed in areas with varying levels of contamination. The selected sites (Fig. 1) were sediments of the *Gave de Pau* River (Gave1, Gave2 and Gave3) (downstream from the urban area of Pau, France; 43°32′ N, 0°45′W); activated sewage sludge (SS); sediment collected at the discharge of treated water of the wastewater treatment plant (WWTP) into the *Gave de Pau* River (Effluent); sediment from the effluent of a former industrial site (Mar); sediment from a tributary river going through a residential area (Osm); sediment collected at the source of a forest stream outside of the *Gave de Pau* basin, assumed to be little impacted by human activities (Lou; 43°08′N, 0°07′W). Samples were stored at −80°C until analysis. Metagenomic DNA was extracted from 0.70 g of sediment with PowerSoil DNA Isolation Kit (MoBio Laboratories) according to the manufacturer's instructions. Approximately 1.5 mL of sediment was sifted for chemical analysis.

Concentrations of metals

Concentrations of several metals (As, Cd, Cr, Cu, Sn, Pb, Zn) were determined in all samples. Approximately 0.4 g of the dry particle fraction between 0.45 and 62 μm was subjected to microwave digestion with 8 mL nitric acid and 2 mL hydrogen fluoride followed by inductively coupled plasma-mass spectrometry analysis as previously described (Point *et al.* 2007). Appropriate calibration standards and blanks were used in both digestion and analysis processes.

Table 1. Primers used in conventional and real time PCRs, expected amplicon size and annealing temperature.

Target	Primer	Sequence (5′→3′)	Amplicon size (bp)	Annealing temperature (°C)	References
16S rRNA	63F 1387R	CAGGCCTAACACATGCAAGTC GGGCGGWGTGTAACAAGGC	1325	56	Marchesi *et al.* (1998)
16S rRNA	338F 518R	ACTCCTACGGGAGGCAGCAG ATTACCGCGGCTGCTGC	181	55	Lane (1991)
intI	ICC32	GCCAACGGNAAACGKCATGC	800	55	Inverted complementary sequence of AJH72 (Gillings *et al.* 2005)
	intI948R	NARTACRTGNGTRTADATCATNGT			Elsaied *et al.* (2007)
attC	ICC1 ICC4	GCSGCTKANCTCVRRCGTTAGSC TCSGCTKGARCGAMTTGTTAGVC	Variable	55	Modified from HS286 and HS287 (Stokes *et al.* 2001)
intI1	qIntI1F qInt1R	ACCAACCGAACAGGCTTATG GAGGATGCGAACCACTTCCAT	~286	62	Wright *et al.* (2008)

Bacterial community structure

Bacterial community structure was determined by terminal restriction fragment length polymorphism (T-RFLP) of the 16S rRNA gene. PCR was performed in triplicate for each sample consisting of 10 ng of template DNA, 0.2 μM of 63F fluorescently labeled by FAM (carboxyfluorescein) and 1387R primers (Table 1) and 25 μL of Master Mix Amplitaq Gold 360 (Applied Biosystems) in a final volume of 50 μL. The amplification conditions were as follows: 10 min at 95°C, 35 cycles of 45 s at 95°C, 45 s at 55°C and 1 min at 72°C, and finally 10 min at 72°C. The PCR products were purified using GFX DNA and Gel Band Purification kit (GE Healthcare) according to the manufacturer's instructions. 100 ng of purified amplicons were separately digested by 5 U of *AluI* or *HhaI* restriction enzymes (New England Biolabs) at 37°C in a final volume of 10 μL for 3 h. 1 μL of digested products were mixed with 10 μL of deionized formamide (Applied Biosystems) and 0.25 μL of internal size standard (GeneScan-500 ROX Size Standard, 35–500bp, Applied Biosystems). They were denatured by heating at 95°C for 5 min before electrophoresis on an ABI Prism 3130xl Genetic Analyzer (Applied Biosystems) in GeneScan mode. Injection was performed electrokinetically at 15 kV for 30 s and the runs were completed within 30 min at 15 kV.

T-RFLP profiles obtained with both enzymes were analyzed using GENESCAN version 3.1 software (Applied Biosystems). Data sets were constructed by using peak height threshold. Data were normalized by dividing each peak height by the total signal intensity. Only peaks above 1% of total signal intensity were considered. A new normalization was then computed. To take into account the variability among PCR experiments, triplicate PCRs for each sample were analyzed separately and data from the three runs were compiled by retaining peaks presented in at least two out of the three PCR runs.

Gene cassette profiling

ICC1 primer and fluorescently labeled (FAM) ICC4 primer (Table 1) were used to amplify gene cassettes from metagenomic DNA samples. PCRs were performed in triplicate for each sample consisting of 0.2 μM of each primer, 25 μL of Master Mix Amplitaq Gold 360 (Applied Biosystems) and 1 μL of DNA template in a final volume of 50 μL. PCR conditions were as follows: 10 min at 95°C, 35 cycles of 45 s at 95°C, 45 s at 55°C and 1 min at 72°C, and finally 10 min at 72°C. Gene cassette sizes were then analyzed by electrophoresis on an ABI Prism 3130xl Genetic Analyzer (Applied Biosystems) as described above using GeneScan 600 LIZ Size Standard. Only cassettes with size ranging from 100 to 700 bp were considered for analysis. The data set was normalized as described for T-RFLP.

intI and gene cassette libraries

PCRs were performed in a final volume of 50 μL containing 1 μL of template DNA, 0.4 μM of each primer (see Table 1) except intI948R primer (4 μM), 0.1 mM dNTPs and 25 μL of Master Mix Amplitaq Gold 360 2X (Applied Biosystems). Amplification conditions were as follows: 10 min at 95°C, 35 cycles of 45 s at 95°C, 45 s at annealing temperature and 1 min at 72°C, and finally 10 min at 72°C. PCR products were gel-purified using GFX DNA and Gel Band Purification kit (GE Healthcare). Purified products were cloned by using pGEM-T easy Vector System II (Promega). Inserts of positive clones were PCR-amplified. PCR products were sequenced by GATC Biotech (Sanger sequencing; kit BigDye Terminator v3.1 on ABI 3730xl, Applied Biosystems). All sequences were analyzed with Sequencher® software. For gene cassette analysis, first the presence of both primer sequences was confirmed, and then ORFs were determined using Sequencher® software. Cassette ORFs and *intI* sequences were translated to protein sequences with BioEdit® software. Then a BlastP search (http://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE=Proteins) was conducted to determine protein sequence similarities with sequences deposited in databanks. Cassette ORF deduced proteins were also entered into online protein prediction programs InterProScan sequence search, Psortb and SignalP, located within ExPASy proteomics server (http://www.expasy.org/tools/#secondary), in order to predict protein functional sites, cellular localization and presence of signal peptides.

To statistically compare IntI libraries, all amino acid sequences were aligned with BioEdit® software using ClustalW Multiple alignment tool (Needleman and Wunsch 1970). Distances were then calculated with Phylip protdist tool on Mobyle Pasteur Portal (http://mobyle.pasteur.fr/cgi-bin/portal.py#forms::protdist). DOTUR (Schloss and Handelsman 2005) was then used to obtain clusters, using a 98% amino-acid identity threshold, to construct a pivot table and calculate a Shannon index.

These sequence data were deposited in the DNA database under the Accession No.: LN774972- LN775293 for integrases and LN775294- LN775354 for gene cassettes.

Multivariable analyses

Non-metric multidimensional scaling (nMDS) analysis implemented in the Vegan packages of R was used to analyze T-RFLP, gene cassettes profiling and IntI sequences data. For all ordinations, the Bray–Curtis distance was used to generate dissimilarity matrices. The cluster analysis based on these dissimilarities was computed with Vegan packages of R.

intI1 quantification

Quantitative PCRs (qPCRs) were performed in duplicate on metagenomic DNA. Standard curves, constructed with plasmid-carrying cloned genes, were used to quantify *intI1* and 16S rRNA gene, respectively. The reaction mixture contained 1 μL template diluted at 1/10, 0.2 μM of each primer (Table 1) and 12.5 μL of Power SYBR® free PCR master mix (Applied Biosystem) for a final volume of 25 μL. All real-time PCRs were performed on a MX3005P (Stratagene). Amplification parameters were as follow: 5 min at 95°C, 40 cycles of 15 s at 95°C, 15 s at 62°C (*intI1*) or 55°C (16S rRNA gene), and 20 s at 72°C. After amplification, a melting curve was carried out to confirm the amplification of specific products.

RESULTS

Site classification according to contamination level

In order to determine the contamination level for each site, the metal concentrations were determined. According to these concentrations, two groups were established: (i) weakly impacted sites (Gave2, Gave3, Osm and Lou; Fig. 2) showing Zn and organotin contaminations corresponding to levels found in the area (Lespes *et al.* 2005); within this group, Lou site is located outside the Gave de Pau basin and is not influenced by urban zones. (ii) Strongly impacted sites (Gave1, Effluent, Mar and SS; Fig. 2, Table 2) showing in particular high content levels of Cd, Pb and Cu. The highest metal concentrations were found in Mar, the site impacted by an industrial discharge, whereas the other three sites were impacted by the WWTP.

Influence of contamination on bacterial communities

Bacterial community diversity was estimated using the Shannon index based on T-RFLP profiles (Table 3). It ranged from 3.11 to 3.64, with the lowest diversity in SS and Effluent sites, and the highest in Mar. It is likely that the level of contamination did not affect the bacterial diversity.

To further compare the bacterial communities, an nMDS analysis of T-RFLP data was performed (Fig. 3A). The analysis showed that Lou and Mar bacterial communities were both distinct to those from the other sites, themselves arranged in two clusters, each grouping bacterial communities with more than 30% similarity. One cluster contained the bacterial communities from the three weakly contaminated sites from *Gave de Pau* basin: Osm, Gave2 and Gave3. In this cluster, the bacterial communities of Gave2 and Gave3 presented about 70% similarity. The second cluster contained the bacterial communities from strongly contaminated sites (SS, Effluent and Gave1) that are located in the continuum of WWTP's influence. In this cluster, bacterial community structures from Effluent and Gave1 sites showed about 40% similarity.

Influence of contamination on IntI diversity

For each site, a library of approximately 800 bp *intI* fragments was constructed from the sediment metagenome. A total of 684 IntI amino-acid sequences were obtained and, considering a 98% amino-acid identity threshold as suggested by Hall (2012), were grouped into 322 previously undescribed integron classes.

The IntI diversity was assessed with the Shannon index. The four contaminated sites (Gave1, Effluent, SS and Mar) showed a lower ($H < 3.5$) Shannon diversity index than the less contaminated sites (Lou, Gave2, Gave3 and Osm; $H > 3.5$; Table 4). It is likely that contamination level influenced IntI sequence diversity.

To further compare the IntI protein pool from each sample, we performed an nMDS analysis (Fig. 3B). The IntI pools of the four strongly contaminated sites formed a cluster with 10% similarity whereas those of the weakly polluted sites were spread apart. It is important to notice that within the cluster containing the libraries from the most polluted sites, Effluent and Gave1 libraries grouped together with 20% similarity, and apart from those of the highest polluted sites, SS and Mar. These observations were in accordance with the trend of the Shannon index suggesting an effect of the pollution level. It is likely that the lesser polluted sites not only presented a larger IntI diversity but also divergent IntI pools.

No IntI1 sequences were found in the IntI libraries. However, class 1 integrons have been found in various freshwater environments (Wright *et al.* 2008; Rosewarne *et al.* 2010; Gaze *et al.* 2011; Moura *et al.* 2012; Nardelli *et al.* 2012; Stalder *et al.* 2014). This observation suggested that the primers used efficiently amplify a wide diversity of IntI sequences but do not allow IntI1 detection. In order to take class 1 integron into account in the analysis, this gene was specifically quantified by qPCR on metagenomic DNA (Fig. 4). The relative *intI1* abundance was higher in the metagenomes from contaminated sites. Finally, no *intI1* was detected in sediment from Osm, a weakly impacted site.

Influence of contamination on gene cassette pool

In order to assess part of the integron gene cassette pool, the gene cassette size was analyzed. The gene cassette diversity was estimated with the Shannon index (Table 3). It ranged from 3.24 (Gave1) to 4.06 (SS). No link between the Shannon index and the level of contamination could be established.

The structure of the gene cassette pool analyzed by nMDS (Fig. 3C) showed that Lou and Mar presented distinct cassette pools from the other samples, in the same way as the bacterial community structure. The other samples belonging to *Gave de Pau* basin were grouped together, but two distinct clusters were observed. In the first cluster, Gave2, Gave3 and Osm showed about 40% similarity. In the second cluster, Gave1, Effluent and SS were grouped with about 40% similarity between Gave1 and Effluent. These results showed that the gene cassette composition in the weakly impacted sediments of *Gave de Pau* basin was distinct from that of the strongly impacted sites under the WWTP influence.

Two libraries of gene cassettes were constructed from metagenomic DNA of Mar and Effluent, the most contaminated natural sites, in order to detect genes potentially involved in pollutants adaptation/mitigation. Out of a total of 96 sequences, 60 different gene cassettes with predicted ORFs were identified. The majority of ORFs had unknown functions and unknown functional motifs (76% and 81% of each library, respectively). Mar presented more putative transmembrane or

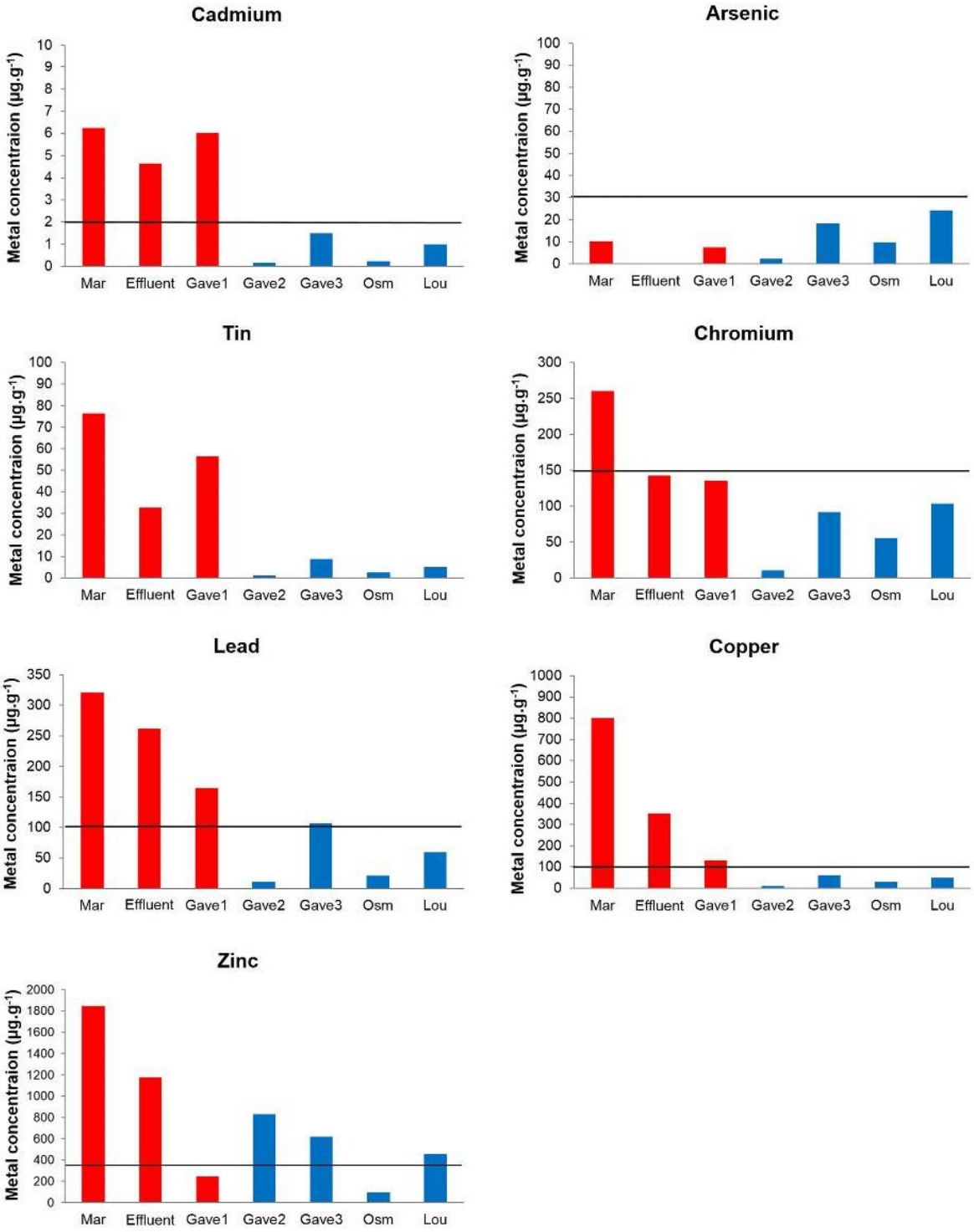


Figure 2. Metal concentrations in sediment (mg kg^{-1}) from each sample. The bold line represents quality threshold for each metal in freshwater sediment according to the French decree of 9 August 2006. In blue, the weakly impacted sites and in red, the most impacted sites.

lipoproteins than the other site (Fig. 5). Two ORFs encoded proteins with similarity to proteins of known function: a plasmid maintenance system killer protein domain and a glyoxalase/bleomycin resistance/dihydroxybiphenyl 1, 2-dioxygenase family protein. The sequence alignment of the latter showed that it was not a bleomycin resistance protein. Among 17 gene cassettes similar to gene cassettes previously detected in contaminated sites, 14 sequences showed significant homology with gene cassettes found in Halifax harbor sediment (Canada, Koenig *et al.* 2008) and 3 with sequences from Sydney Tar Ponds (Canada, Koenig *et al.* 2009).

Table 2. Metal concentrations (mg kg^{-1}, dry weight) in SS from this study and from the literature.

Metal	SS sample (this study)	Scancar *et al.* (2000)	Contin, Goi and De Nobili (2012)	Directive 86/278/EEC[a]
As	2.8	NA	Na	NA
Cd	5.5	2.8	1.31	20–40
Cr	98.1	856	76.7	1000
Cu	493.4	433	520	1000–1750
Pb	81.9	126	66.4	750–1200
Sn	32.8	NA	NA	NA
Zn	720.1	2032	1027	2500–4000

[a]European recommendations of maximum permissible concentrations of heavy metals in SS for agricultural uses.

Table 3. Shannon diversity based on T-RFLP[a] of 16S rRNA gene and gene cassette size profiles.

	SS	Effluent	Gave1	Gave2	Gave3	Mar	Osm	Lou
16S rRNA gene	3.11	3.25	3.58	3.5	3.59	3.64	3.4	3.46
Gene cassettes	4.06	3.84	3.24	3.58	3.44	3.47	3.54	3.86

[a]T-RFLP data include 5′ patterns obtained from 16S rDNA digested with *AluI* or *HhaI*.

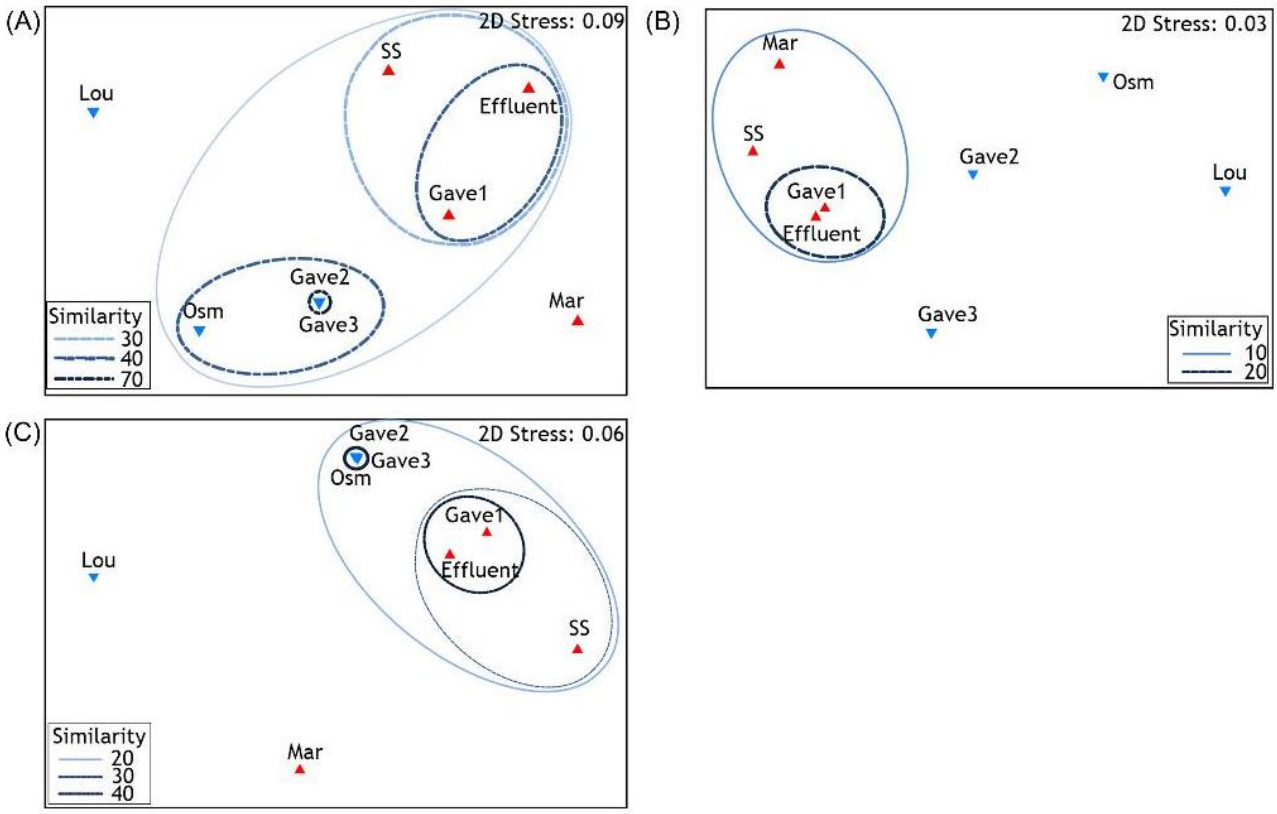


Figure 3. nMDS plots of (A) bacterial community structure assessed by 16S rRNA gene T-RFLP; (B) IntI sequences libraries; and (C) gene cassette pool assessed by size profiles, based on a distance matrix calculated from Bray–Curtis similarity. T-RFLP data include 5′ patterns obtained by 16S rRNA gene amplicons digested by *AluI* or *HhaI*. In blue, the weakly impacted sites and in red, the most impacted sites.

DISCUSSION

A wide diversity of IntI in natural environments

In this study, the integron integrase diversity from sediments of a freshwater environment was investigated. Previous studies focused on IntI sequences from environments other than clinical, but few focused on IntI sequences other than IntI1, 2 or 3. Six studies, respectively from soil (Nield *et al.* 2001), heavy-metal-contaminated-mine tailings (Nemergut, Martin and Schmidt 2004), deep-sea hydrothermal vents (Elsaied *et al.* 2007), polluted marine sediments (Elsaied *et al.* 2011), Cascadian submarine gas hydrate bearing core (Elsaied *et al.* 2014) and various terrestrial, marine and deep-sea environments (Rodríguez-Minguela *et al.* 2009) allowed the description of 3, 6, 11, 33, 37 and 169 IntI sequences not already deposited in databases. In the present study, the primers used allowed the successful amplification of

Table 4. IntI library coverage and Shannon diversity (considering a 98% amino-acid identity threshold).

	SS	Effluent	Gave1	Gave2	Gave3	Mar	Osm	Lou
Number of IntI sequences per library	79	76	108	88	85	85	86	77
Number of different IntI sequences	30	41	52	66	66	36	49	57
IntI library coverage (%)	77	66	67	41	35	73	58	40
Shannon index	2.84	3.42	3.49	4.06	4.06	3.06	3.55	3.89

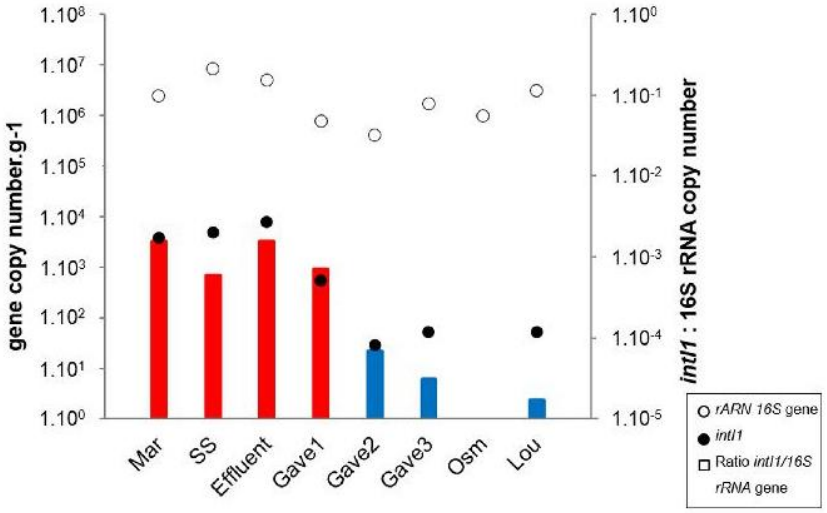


Figure 4. Class 1 integrase (*intI1*) copy number, 16S rRNA gene copy number and 16S rRNA gene/*intI1* copy number ratio for the various sites as assessed by quantitative PCR. In blue, the weakly impacted sites and in red, the most impacted sites.

a wider diversity of *intI*. Considering a 98% amino-acid identity threshold as suggested by Hall (2012), all IntI sequences were new and grouped in 322 classes. It is the first time that such a wide diversity has been identified by direct amplification from total genomic DNA extracted from environmental samples. The primer pair used was also efficient for the amplification of a wide diversity of *intI* from other samples including marine and coastal sediments (Abella *et al*. 2015). Therefore, integrons show an important diversity in environments other than clinical, indicating that high integron diversity is not specific to the samples studied here.

The high level of diversity of IntI revealed in environmental studies questions the relevance of the current classification and the identity threshold (98%) applied to establish it. When IntI sequences from the present study were analyzed considering either 80% or 60% amino-acid identity threshold [since IntI1 has 59% identity with IntI3 (Collis *et al*. 2002), a threshold of 60% still distinguishes class 1 and class 3 integrons], the number of sequences classified as different fell from 322 to 207 or 138, respectively. Thus, the observed IntI diversity remained high, even if a homology was found between several sequences from this study and sequences from databases: 21 and 112, considering a threshold of 80% and 60%, respectively. Therefore, modifying the threshold reduced the number of new sequences that represented 90% and 19%, respectively. Nonetheless, the nMDS analysis of the IntI pool at 80% and 60% threshold indicated that the IntI diversity in the weakly impacted sites, Lou and Osm, was still different from the other sites (Supplementary data 1). These observations strongly suggest that the IntI classification could be simplified if it is revised to an 80% identity threshold, allowing a better view of integron distribution in environmental samples without losing information about impact of environmental conditions on the IntI diversity.

A global pool of gene cassettes

Two gene cassette libraries were constructed from two impacted samples. An advanced search was carried out to identify patterns and cellular localization in order to characterize the gene cassettes associated with the integrons as accurately as possible. Out of a total of 60 different sequences identified as gene cassettes, 22% of ORFs encoded proteins with at least hypothetical general functions. This is consistent with previous surveys (Koenig *et al*. 2008, 2009; Wright *et al*. 2008). Only seven sequences presented known patterns or domains. The identified functions were not associated with known metal or antibiotic resistance genes, as observed by Wright *et al*. (2008) who studied metal and antibiotic contaminated samples. It is likely that the number of total sequences was too low to allow the identification of gene cassettes related to environmental contamination. Such relation was done with a big effort of sequencing in case of hydrocarbon pollution (Koenig *et al*. 2009). Moreover, with the exception of the most extreme environments, even when a large amount of sequence data has been generated, the rarefaction curves were always unsaturated, demonstrating a wide gene cassette diversity outside clinical environments (Koenig *et al*. 2008, 2009; Wright *et al*. 2008; Elsaied *et al*. 2011, 2014). Thus, this observation underlines the importance of sequencing a large amount of gene cassettes from non-clinical environments. However, this sequencing approach omits the first gene cassettes, precisely those that reflect the most recent response to selection pressure (Huang

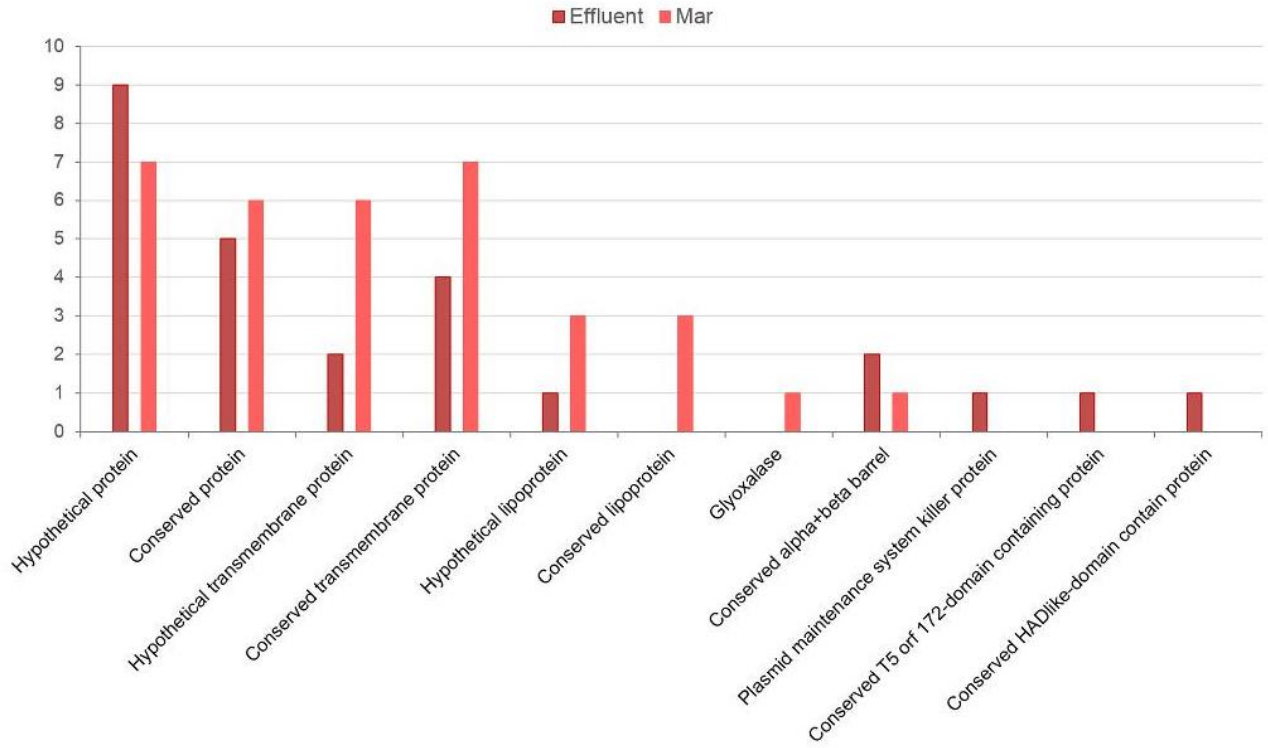


Figure 5. Characterization of ORFs in gene cassettes sequenced from Mar and Effluent metagenomes.

et al. 2009; Elsaied *et al.* 2011), a good target when searching for adaptive functions to cope with contamination.

Several gene cassette ORFs showed significant homologies with gene cassettes previously recovered in multicontaminated sites [Halifax Harbour, Canada (Koenig *et al.* 2008) and Sydney Tar Pond, Canada (Koenig *et al.* 2009)]. Elsaied *et al.* (2011) proposed the hypothesis of a 'global marine biogeographic distribution of gene cassette encoding proteins'. Interestingly, some gene cassettes found here in the freshwater environments studied have also been found in marine environments. Thus, we assume that gene cassette distribution occurs more globally, gene cassettes being spread not only by ocean currents but also by vectors such as human activities (Gaze *et al.* 2011; Stalder *et al.* 2014), wild animal microbiota (Power, Emery and Gillings 2013; Sarria-Guzmán *et al.* 2014; Shah *et al.* 2014) and atmosphere (Ling *et al.* 2013).

Integron diversity is impacted by contamination

The main purpose of the present study was to determine the diversity of integrons in sediments of a freshwater basin according to contamination level. In the weakly contaminated sites (Gave2, Gave3 and Osm), the level of IntI diversity was greater than in the strongly contaminated sites (Table 4). However, while the pool of IntI sequences was dissimilar between these three sites (Fig. 3B), their bacterial community structures were grouped together (Fig. 3A). Thus, the absence of a strong selection pressure allowed a wide variety of IntI sequences in the bacterial community. While IntI diversity could differ from one site to another, the gene cassette pool was similar (Fig. 3C). Thus, despite a similar community structure, the weakly impacted sediments presented a large diversity of integrons that shared a common cassette pool. Interestingly, while the level of IntI diversity was similar between Lou (the weakly impacted site from another basin) and the less impacted samples of the Gave de Pau River (Table 4), the IntI and gene cassette pools from Lou were distinct, as was Lou community structure (Fig. 3). This observation suggested that both bacterial community structure and integron diversity were strongly influenced by environmental conditions.

The strongly contaminated sites exhibited the lowest IntI diversity (Table 4), especially Mar and SS. The pool of IntI sequences in Mar, SS, Effluent and Gave1 was also grouped within 10% similarity in the nMDS analysis (Fig. 3B). Because the industrial waste effluent (Mar) presented a different nature of contamination than the activated SS and the effluent area of the WWTP (Effluent and Gave1), it is likely that the IntI diversity was more influenced by the contamination level than that by the contamination type. In contrast, the bacterial community structure and the cassette pool structure from the three sites impacted by the WWTP were similar, but far from Mar and the weakly impacted sites (Fig. 3A and B). Since the IntI diversity was similar, the differences between the gene cassette pools could not be explained solely by the community structure but probably by the type of the contaminant pressure. It is known that the community structure of the WWTP effluent could be impacted by the WWTP discharges, especially by antibiotic-resistant bacteria (Reinthaler *et al.* 2003; Drury, Rosi-Marshall and Kelly 2013; Marti, Jofre and Balcazar 2013; Marti and Balcázar 2014). As Gave1 and Effluent samples are close to the WWTP discharges, the bacterial communities may be composed of resistance-integron-carrying bacteria, also present in SS. The results presented in this study show a greater *intI1* abundance in these sites than in the other *Gave de Pau* River sites, confirming the proposition to use *intI1* as a proxy for anthropogenic pollution (Gillings *et al.* 2015).

The gene cassette pool of Mar was different from those of the other contaminated sites (Fig. 3). The Mar site is located in an industrial effluent highly contaminated with metals, without any input from urban nor farm discharge. The fact that the contamination type was different than that observed in the other sites

attested that gene cassette diversity could be influenced by environmental conditions. Interestingly, Mar site showed an important *intI1* relative abundance. It is in accordance with a previous study showing an increase of class 1 integron in a context of a high-level metal contamination (Wright *et al.* 2008). In addition, the IntI pools of Mar and the other contaminated sites were similar, suggesting that other IntI sequences than IntI1 could be also favored by both types of contamination.

An ecotype of gene cassettes

Having observed similarities between gene cassette pools from nearby environments, Michael *et al.* (2004) suggested a co-assorting group of gene cassette structure. In addition, an extensive sequencing effort allowed Koenig *et al.* (2008) to show that similar gene cassettes were observed in two similar ecosystems, both contaminated by raw sewage discharges, but geographically distant. They hypothesized about the existence of an integron gene cassette ecotype. Wright *et al.* (2008) came to a compatible conclusion, under which different selection pressures could lead to different gene cassette pools. Our results showed that gene cassette diversity could be influenced by contamination type. Hence, all these observations argue for the possible existence of ecotypical cassettes.

Thus, selection pressure results in gene cassette selection. Since it has also been shown that genes with functions related to contaminant type were found in the gene cassette pool (Koenig *et al.* 2009), it is likely that integrons contribute to the mobilization of adaptive genes in response to selection pressure. Our results add to these previous observations and provide further arguments for the involvement of the non-clinical integron in the adaptation of bacterial communities as a response to the presence of contaminants in natural environments.

SUPPLEMENTARY DATA

Supplementary data are available at FEMSEC online.

ACKNOWLEDGEMENTS

We thank Dr Sandra Barantal for her good advice, Dr Gilles Bareille and Pr Béatrice Lauga for their help in sampling. We also thank Dr Thomas Bersinger for metal analysis and Viviane Jean for technical assistance. We thank the Direction Générale de l'Armement (DGA) and the Aquitaine Regional Government Council (France) for financial support to J. Abella. We acknowledge the Regional Platform for Environmental Microbiology PREMICE supported by the Aquitaine Regional Government Council (France) and the urban community of Pau-Pyrénées (France). We thank anonymous reviewers for improving comments and suggestions on the first version of the manuscript.

FUNDING

The study was supported by a grant from CG64 (Conseil Général 64) and ANR DECAPAGE (project ANR-CESA-2011-006 01).

Conflict of interest. None declared.

REFERENCES

Abella J, Bielen A, Huang L, *et al.* Integron diversity in marine environments. *Environ Sci Pollut Res* 2015;**22**:15360–9.

Boucher Y, Labbate M, Koenig JE, *et al.* Integrons: mobilizable platforms that promote genetic diversity in bacteria. *Trends Microbiol* 2007;**15**:301–9.

Byrne-Bailey KG, Gaze WH, Zhang L, *et al.* Integron prevalence and diversity in manured soil. *Appl Environ Microb* 2011;**77**:684–7.

Cambray G, Guerout A-M, Mazel D. Integrons. *Annu Rev Genet* 2010;**44**:141–66.

Collis CM, Hall RM. Expression of antibiotic resistance genes in the integrated cassettes of integrons. *Antimicrob Agents Ch* 1995;**39**:155–62.

Collis CM, Kim M-J, Stokes HW, *et al.* Integron-encoded IntI integrases preferentially recognize the adjacent cognate attI site in recombination with a 59-be site. *Mol Microbiol* 2002;**46**:1415–27.

Contin M, Goi D, De Nobili M. Land application of aerobic sewage sludge does not impair methane oxidation rates of soils. *Sci Total Environ* 2012;**441**:10–8.

Demarre G, Frumerie C, Gopaul DN, *et al.* Identification of key structural determinants of the IntI1 integron integrase that influence *attC* × *attI1* recombination efficiency. *Nucleic Acids Res* 2007;**35**:6475–89.

Drury B, Rosi-Marshall E, Kelly JJ. Wastewater treatment effluent reduces the abundance and diversity of benthic bacterial communities in urban and suburban rivers. *Appl Environ Microb* 2013;**79**:1897–905.

Elsaied H, Stokes HW, Kitamura K, *et al.* Marine integrons containing novel integrase genes, attachment sites, *attI*, and associated gene cassettes in polluted sediments from Suez and Tokyo Bays. *ISME J* 2011;**5**:1162–77.

Elsaied H, Stokes HW, Nakamura T, *et al.* Novel and diverse integron integrase genes and integron-like gene cassettes are prevalent in deep-sea hydrothermal vents. *Environ Microbiol* 2007;**9**:2298–312.

Elsaied H, Stokes HW, Yoshioka H, *et al.* Novel integrons and gene cassettes from a Cascadian submarine gas-hydrate-bearing core. *FEMS Microbiol Ecol* 2014;**87**:343–56.

Gaze WH, Zhang L, Abdouslam NA, *et al.* Impacts of anthropogenic activity on the ecology of class 1 integrons and integron-associated genes in the environment. *ISME J* 2011;**5**:1253–61.

Gestal AM, Liew EF, Coleman NV. Natural transformation with synthetic gene cassettes: new tools for integron research and biotechnology. *Microbiology* 2011;**157**:3349–60.

Gillings MR. Integrons: past, present and future. *Microbiol Mol Biol R* 2014;**78**:257–27.

Gillings MR, Gaze WH, Pruden A, *et al.* Using the class 1 integron-integrase gene as a proxy for anthropogenic pollution. *ISME J* 2015;**9**:1269–79.

Gillings MR, Holley MP, Stokes HW, *et al.* Integrons in *Xanthomonas*: a source of species genome diversity. *P Natl Acad Sci USA* 2005;**102**:4419–24.

Gillings MR, Krishnan S, Worden PJ, *et al.* Recovery of diverse genes for class 1 integron-integrases from environmental DNA samples. *FEMS Microbiol Lett* 2008;**287**:56–62.

Hall RM. Mobile gene cassettes and integrons: moving antibiotic resistance genes in Gram-negative bacteria. *CIBA Found Symp* 1997;**207**:192–205.

Hall RM. Integrons and gene cassettes: hotspots of diversity in bacterial genomes. *Ann NY Acad Sci* 2012;**1267**:71–8.

Hall RM, Collis CM. Mobile gene cassettes and integrons: capture and spread of genes by site-specific recombination. *Mol Microbiol* 1995;**15**:593–600.

Hardwick SA, Stokes HW, Findlay S, *et al.* Quantification of class 1 integron abundance in natural environments using real-time quantitative PCR. *FEMS Microbiol Lett* 2008;**278**:207–12.

Henriques IS, Fonseca F, Alves A, *et al.* Occurrence and diversity of integrons and β-lactamase genes among ampicillin-resistant isolates from estuarine waters. *Res Microbiol* 2006;**157**:938–47.

Holmes AJ, Gillings MR, Nield BS, *et al.* The gene cassette metagenome is a basic resource for bacterial genome evolution. *Environ Microbiol* 2003;**5**:383–94.

Huang L, Cagnon C, Caumette P, *et al.* First gene cassettes of integrons as targets in finding adaptive genes in metagenomes. *Appl Environ Microb* 2009;**75**:3823–5.

Koenig JE, Boucher Y, Charlebois RL, *et al.* Integron-associated gene cassettes in Halifax Harbour: assessment of a mobile gene pool in marine sediments. *Environ Microbiol* 2008;**10**:1024–38.

Koenig JE, Sharp C, Dlutek M, *et al.* Integron gene cassettes and degradation of compounds associated with industrial waste: the case of the SydneyTar Ponds. *PLoS One* 2009;**4**:e5276.

Labbate M, Case RJ, Stokes HW. The integron/gene cassette system: an active player in bacterial adaptation. *Methods Mol Biol* 2009;**532**:103–25.

Lane DJ. 16S/23S rRNA sequencing. Nucleic acid techniques in bacterial systematics. In: Stackebrandt E, Goodfellow M (eds). *Sequencing Techniques in Bacterial Systematics*. Chichester: *John Wiley & Sons Ltd*, 1991, 115–74.

Laroche E, Pawlak B, Berthe T, *et al.* Occurrence of antibiotic resistance and class 1, 2 and 3 integrons in *Escherichia coli* isolated from a densely populated estuary (Seine, France). *FEMS Microbiol Ecol* 2009;**68**:118–30.

Lespes G, Bancon-Montigny C, Aguerre S, *et al.* Organotin speciation in waters and sediments in the Adour-Garonne basin. *Rev Sci Eau* 2005;**18**:47–63.

Levesque C, Brassard S, Lapointe J, *et al.* Diversity and relative strength of tandem promoters for the antibiotic resistance genes of several integrons. *Gene* 1994;**142**:49–54.

Ling AL, Pace NR, Hernandez MT, *et al.* Tetracycline resistance and class 1 integron genes associated with indoor and outdoor aerosols. *Environ Sci Technol* 2013;**47**:4046–52.

Marchesi JR, Sato T, Weightman AJ, *et al.* Design and evaluation of useful bacterium-specific PCR primers that amplify genes coding for bacterial 16S rRNA. *Appl Environ Microb* 1998;**64**:795–9.

Marti E, Balcázar JL. Use of pyrosequencing to explore the benthic bacterial community structure in a river impacted by wastewater treatment plant discharges. *Res Microbiol* 2014;**165**:468–71.

Marti E, Jofre J, Balcazar JL. Prevalence of antibiotic resistance genes and bacterial community composition in a river influenced by a wastewater treatment plant. *PLoS One* 2013;**8**:e78906.

Mazel D. Integrons: agents of bacterial evolution. *Nat Rev Microbiol* 2006;**4**:608–20.

Mazel D, Dychinco B, Webb VA, *et al.* A distinctive class of integron in the *Vibrio cholerae* genome. *Science* 1998;**280**:605–8.

Michael CA, Andrew NR. Co-assortment in integron-associated gene cassette assemblages in environmental DNA samples. *BMC Genet* 2010;**11**:75.

Michael CA, Gillings MR, Holmes AJ, *et al.* Mobile gene cassettes: a fundamental resource for bacterial evolution. *Am Nat* 2004;**164**:1–12.

Michael CA, Labbate M. Gene cassette transcription in a large integron-associated array. *BMC Genetics* 2010;**11**:82.

Moura A, Henriques I, Ribeiro R, *et al.* Prevalence and characterization of integrons from bacteria isolated from a slaughterhouse wastewater treatment plant. *J Antimicrob Chemoth* 2007;**60**:1243–50.

Moura A, Henriques I, Smalla K, *et al.* Wastewater bacterial communities bring together broad-host range plasmids, integrons and a wide diversity of uncharacterized gene cassettes. *Res Microbiol* 2010;**161**:58–66.

Moura A, Jové T, Ploy M-C, *et al.* Diversity of gene cassette promoters in class 1 integrons from wastewater environments. *Appl Environ Microb* 2012;**78**:5413–6.

Nardelli M, Scalzo PM, Ramírez MS, *et al.* Class 1 integrons in environments with different degrees of urbanization. *PLoS One* 2012;**7**:e39223.

Needleman SB, Wunsch CD. A general method applicable to the search for similarities in the amino acid sequence of two proteins. *J Mol Biol* 1970;**48**:443–53.

Nemergut DR, Martin AP, Schmidt SK. Integron diversity in heavy-metal-contaminated mine tailings and inferences about integron evolution. *Appl Environ Microb* 2004;**70**:1160–8.

Nemergut DR, Robeson MS, Kysela RF, *et al.* Insights and inferences about integron evolution from genomic data. *BMC Genomics* 2008;**9**:261.

Nield BS, Holmes AJ, Gillings MR, *et al.* Recovery of new integron classes from environmental DNA. *FEMS Microbiol Lett* 2001;**195**:59–65.

Ploy M-C, Gassama A, Chainier D, *et al.* Integrons: an antibiotic resistance gene capture system. *Immuno-Anal Biol Spec* 2005;**20**:343–52.

Ploy M-C, Lambert T, Couty J-P, *et al.* Integrons: an antibiotic resistance gene capture and expression system. *Clin Chem Lab Med* 2000;**38**:483–7.

Point D, Bareille G, Pinaly H, *et al.* Multielemental speciation of trace elements in estuarine waters with automated on-site UV photolysis and resin chelation coupled to inductively coupled plasma mass spectrometry. *Talanta* 2007;**72**:1207–16.

Power ML, Emery S, Gillings MR. Into the Wild: dissemination of antibiotic resistance determinants via a species recovery program. *PLoS One* 2013;**8**:e63017.

Rapa RA, Labbate M. The function of integron-associated gene cassettes in *Vibrio* species: the tip of the iceberg. *Front Microbiol* 2013;**4**:385, DOI: 10.3389/fmicb.2013.00385.

Reinthaler FF, Posch J, Feierl G, *et al.* Antibiotic resistance of *E. coli* in sewage and sludge. *Water Res* 2003;**37**:1685–90.

Rodríguez-Minguela CM, Apajalahti JHA, Chai B, *et al.* Worldwide prevalence of class 2 integrases outside the clinical setting is associated with human impact. *Appl Environ Microb* 2009;**78**:5100–10.

Rosewarne CP, Pettigrove V, Stokes HW, *et al.* Class 1 integrons in benthic bacterial communities: abundance, association with Tn402-like transposition modules and evidence for coselection with heavy-metal resistance. *FEMS Microbiol Ecol* 2010;**72**:35–46.

Rowe-Magnus DA, Davies J, Mazel D. Impact of integrons and transposons on the evolution of resistance and virulence. *Curr Top Microbiol Immunol* 2002;**264**:167–88.

Rowe-Magnus DA, Mazel D. Integrons: natural tools for bacterial genome evolution. *Curr Opin Microbiol* 2001;**4**:565–9.

Sarria-Guzmán Y, López-Ramírez MP, Chávez-Romero Y, *et al.* Identification of antibiotic resistance cassettes in class 1 integrons in Aeromonas spp. strains isolated from fresh fish (*Cyprinus carpio* L.). *Curr Microbiol* 2014;**68**:581–6.

Scancar J, Milacic R, Strazar M, *et al.* Total metal concentrations and partitioning of Cd, Cr, Cu, Fe, Ni and Zn in sewage sludge. *Sci Total Environ* 2000;**250**:9–19.

Schloss PD, Handelsman J. Introducing DOTUR, a computer program for defining operational taxonomic units and estimating species richness. *Appl Environ Microb* 2005;**71**: 1501–6.

Shah SQA, Cabello FC, L'Abée-Lund TM, *et al.* Antimicrobial resistance and antimicrobial resistance genes in marine bacteria from salmon aquaculture and non-aquaculture sites. *Environ Microbiol* 2014;**6**:1310–20.

Stalder T, Barraud O, Casellas M, *et al.* Integron involvement in environmental spread of antibiotic resistance. *Front Microbiol* 2012;**3**:119, DOI: 10.3389/fmicb.2012.00119.

Stalder T, Barraud O, Jové T, *et al.* Quantitative and qualitative impact of hospital effluent on dissemination of the integron pool. *ISME J* 2014;**8**:768–77.

Stokes HW, Hall RM. A novel family of potentially mobile DNA elements encoding site-specific gene-integration functions: integrons. *Mol Microbiol* 1989;**3**:1669–83.

Stokes HW, Holmes AJ, Nield BS, *et al.* Gene cassette PCR: sequence-independent recovery of entire genes from environmental DNA. *Appl Environ Microb* 2001;**67**: 5240–6.

Wright MS, Baker-Austin C, Lindell AH, *et al.* Influence of industrial contamination on mobile genetic elements: class 1 integron abundance and gene cassette structure in aquatic bacterial communities. *ISME J* 2008;**2**:417–28.

1.2. Prédiction de la structure secondaire des sites de recombinaison *attC*

Parmi les banques de cassette de gènes réalisées dans cette étude, certaines séquences contenaient 2 cassettes de gène, séparées par un site de recombinaison *attC*. Ces sites présentent des différences dans leur séquence qui pourraient expliquer qu'ils n'aient pas été une bonne cible pour les amorces utilisées. En effet, les amorces utilisées ont été dessinées en 2001 à partir des séquences alors connues correspondant essentiellement à des sites *attC* de cassettes de gène cliniques (Stokes et al., 2001). Or la diversité de séquences des différents éléments composants les intégrons s'avère bien plus importante dans les environnements naturels, les sites de recombinaison *attC* ne faisant pas exception. De plus, les intégrases d'intégron sont capables de reconnaître une grande variété de séquences de site *attC*, la conservation de la structure secondaire du site *attC* étant plus essentielle pour l'activité de recombinaison que la séquence en elle-même. Une analyse de la structure secondaire des sites *attC* a donc été réalisée afin de vérifier que ces structures pourraient représentées un substrat pour les intégrases d'intégron (Figure III-1).

Trois des sites *attC*, $attC_{Mar2\text{-}9}$, $attC_{Mar1\_16}$ et $attC_{Mar1\_4}$ présentent des structures similaires aux sites reconnus préférentiellement par IntI1, c'est-à-dire une G et une T mésappariées séparées de 6 bases. Les sites $attC_{Effluent2\_8}$, $attC_{Effluent2\_12\_2}$, $attC_{Effluent2\_12\_3}$ et $attC_{Effluent2\_14}$ ont une structure plus atypique. Les 2 bases mésappariées sont séparées de 7 nucléotides ce qui n'est pas une distance optimale pour la reconnaissance du site *attC* par les intégrases étudiées dans la littérature. Par ailleurs, l'identité des bases mésappariées n'est pas commune aux sites *attC* jusqu'alors testés, puisqu'à l'exception du site $attC_{Effluent2\_8}$ qui présente comme bases extra-hélicales une G et une T, les autres contiennent 2 guanines pour les sites $attC_{Effluent2\_12\_2}$ et $attC_{Effluent2\_14}$ et une A et une T pour le site $attC_{Effluent2\_12\_3}$. Enfin, la structure des 2 derniers sites, $attC_{Effluent2\_12\_1}$ et $attC_{Mar1\_18}$ sont atypiques. En effet, pour le premier, ce sont 3 bases qui pourraient être mésappariées. Pour le second, il se pourrait que seule une guanine serve de reconnaissance pour l'intégrase. Il est toutefois difficile de conclure quant à la possibilité pour les intégrases d'intégron de reconnaître et de catalyser

une recombinaison au niveau de ces sites sans que cela ne soit testé dans des expériences *in vitro*.

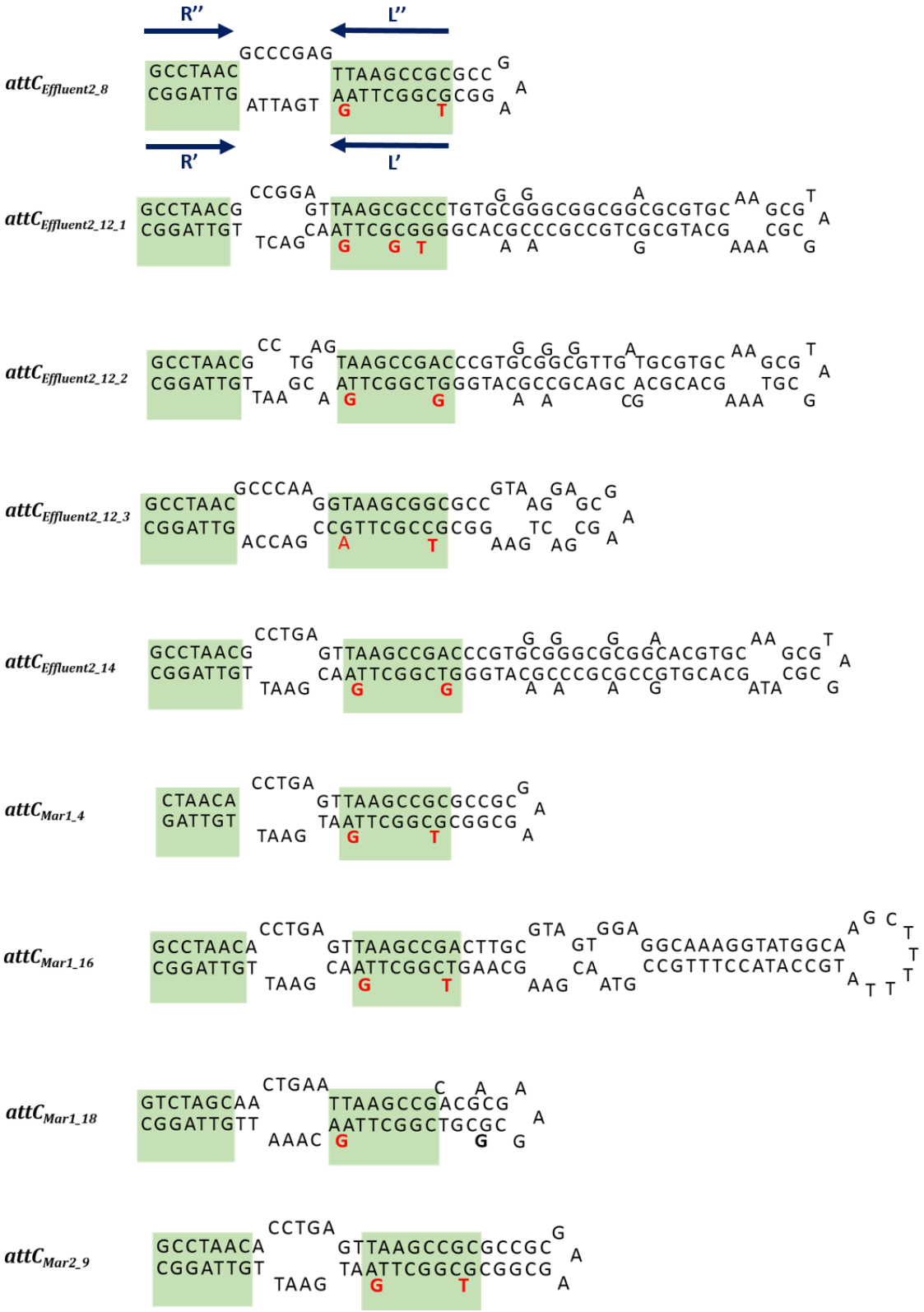


Figure III-1 : Structure secondaire prédictive des sites *attC* présents à l'intérieur des cassettes séquencées. Les séquences répétées inversées L', L'', R' et R'' sont indiquées par les flèches horizontales uniquement pour le site $attC_{Effluent2\_8}$ dans un but de lisibilité. Les box R et L sont représentées en vert pour tous les sites. Les nucléotides en gras et rouge correspondent aux bases extra-hélicales permettant la reconnaissance par les intégrases.

1.3. Analyse phylogénétique des séquences d'intégrase

Une analyse phylogénétique a été réalisée avec 126 séquences : les 31 séquences représentatives des OTUs les plus abondantes (dont l'abondance est supérieure à 0,5% du total de séquences obtenues avec toutes les banques) et 95 séquences de la base de données correspondant aux 5 séquences les plus proches des 31 séquences représentatives (après un blastP contre la banque nr).-Ainsi, plusieurs correspondances ont pu être identifiées. Une OTU représentée par la séquence *G1 3 42* présente 98% d'identité protéique avec la séquence d'intégrase d'une souche du genre *Acidovorax* (WP_024814945.1). Deux autres OTUs sont également proches de cette séquence : il s'agit des OTUs représentés par les séquences *Eff67* et *G1 31* qui partagent respectivement 97% et 96,6% d'identité protéique avec la séquence de la souche *d'Acidovorax*. L'OTU représentée par la séquence *Eff67* est retrouvée dans la majorité des sites, à l'exception de Lou et Osm. Les OTUs représentées par les séquences *G1 3 42* et *G1 31* sont moins abondantes dans l'échantillonnage et retrouvées seulement dans les sites Gave1 et Mar pour la première et Gave1 et Effluent pour la seconde.

La séquence représentative *Eff35* partage 97,6% d'identité protéique avec l'intégrase d'une *Candidatus Accumulibacter phosphatis*. Cette β-*proteobacteria* a été retrouvée dans la communauté bactérienne de stations d'épurations et fait partie des organismes capables d'accumuler les polyphosphates (PAO pour *polyphosphate-accumulating organisms*) (Seviour *et a*l., 2003). Or cette OTU est retrouvée au niveau des sites impactés par la STEP (SS, Effluent, Gave1) et dans Mar, mais est absente des autres sites moins impactés. Il est donc possible que les bactéries qui portent les séquences de cette OTU soient proches de *Candidatus Accumulibacter phosphatis*.

D'autres OTUs présentent des homologies avec des séquences de la base de données mais à un plus faible pourcentage. Il s'agit des OTUs représentées par les séquences *Mar1 18*, *Eff65*, *Mar1 35*, *Eff103* et *G1 26 19*. *Mar1 18* partage 82% et 81% d'identité protéique respectivement avec une intégrase obtenue sans isolement bactérien à partir d'un sol agricole proche d'un élevage porcin (ADP05693.1) et l'intégrase d'une souche de *Lysobacter*, isolée en Antarctique ; *Eff65* partage 83% d'identité avec une intégrase obtenue à partir d'un

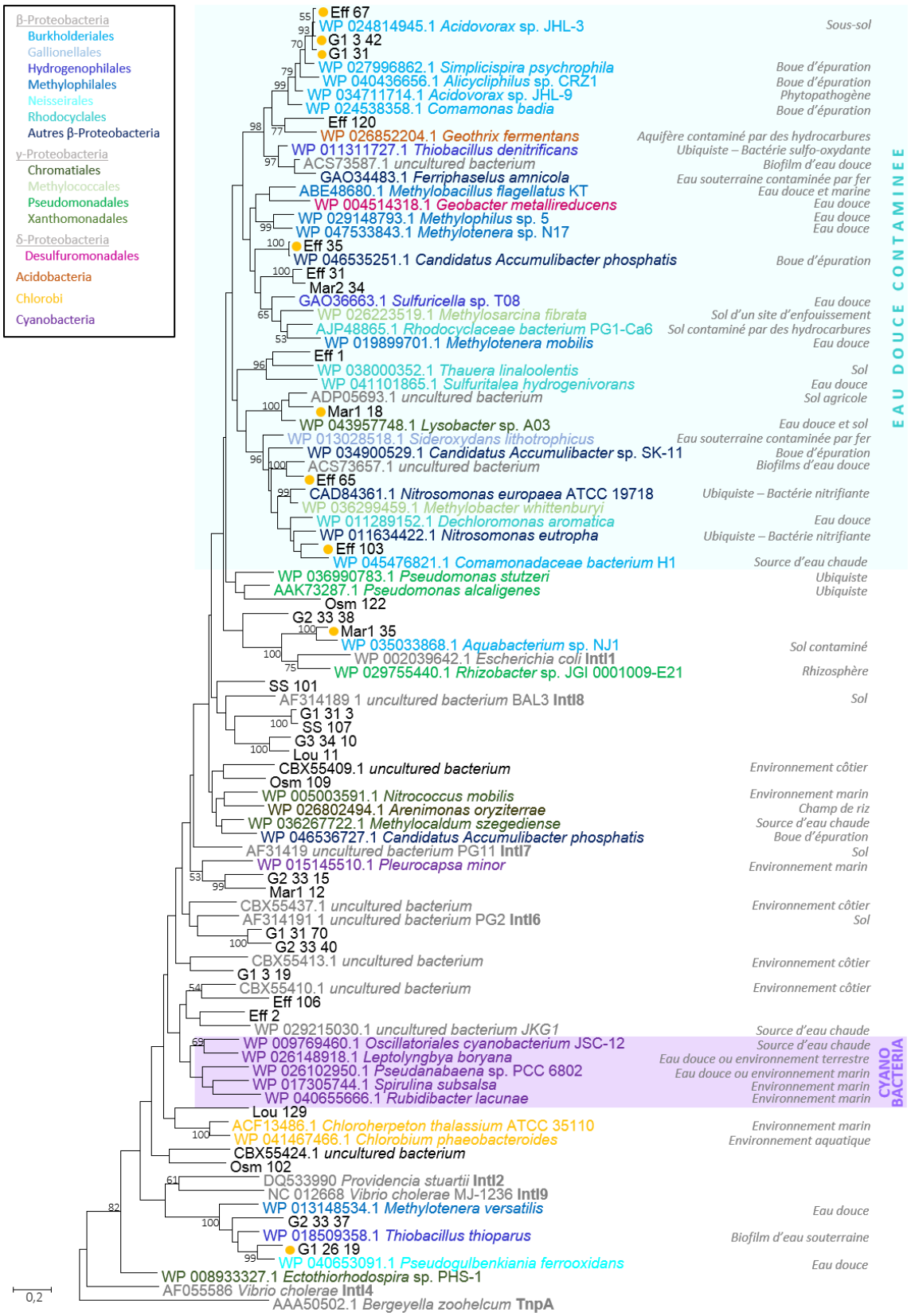


Figure III-2 : Arbre phylogénétique des OTUs les plus abondants et des 5 premières séquences obtenues après un blastP de chaque OTUs principaux (l'abondance est supérieure à 0,5% du total de séquences obtenues avec toutes les banques réalisées dans cette étude sur le Gave de Pau). ● : Séquences présentant les plus fortes homologies avec des séquences d'intégrases connues. La méthode du maximum de ressemblance a été utilisée pour calculer l'arbre à partir de l'alignement des séquences de IntI après un traitement avec Gblocks. Les valeurs de bootstrap supérieures à 50% (sur 500 itérations) sont indiquées. La séquence de TnpA de *Bergeyella zoohelcum* a été utilisée comme racine de l'arbre.

biofilm d'eau douce (ACS73585.1 ; Gillings *et al.*, 2008) ; *Mar1 35* partage 82% avec l'intégrase d'une souche *d'Aquabacterium* (WP_035033868.1), genre que l'on retrouve essentiellement dans des biofilms d'eau douce ; *Eff103* partage 81% d'identité avec l'intégrase d'une *Nitrosomonas eutropha*, une β-*proteobacteria* oxydante de l'ammonium (Stein *et al.,* 2007) ; enfin, *G1 26 19* partage 80% d'identité avec l'intégrase d'une *Pseudogulbenkiania ferrooxidans*, isolée à partir de sédiment d'un lac d'eau douce (Weber *et al.,* 2009; Byrne-Bailey *et al.*, 2012).

Toutes ces correspondances avec les séquences de la base de données sont cohérentes avec notre aire d'échantillonnage, c'est-à-dire des environnements d'eau douce plus ou moins impactés, essentiellement par le rejet d'une station d'épuration, les activités agricoles et/ou des contaminations métalliques rencontrés tout au long du cours d'eau.

2) Diversité des intégrons dans des sédiments côtiers naturels

Résumé

Les environnements marins sont sujets à un grand nombre de sources de contaminations. D'un côté, l'écoulement des eaux terrestres vers les océans transportent de nombreux contaminants provenant des activités urbaines, agricoles et industrielles. D'autre part, les environnements côtiers sont des lieux d'activités intenses industrielles et de transports maritimes qui peuvent être à l'origine de pollutions accidentelles, parfois spectaculaires telles que les marées noires. La salinité étant un des facteurs majeurs de l'environnement influençant la composition des communautés bactériennes, les communautés marines et côtières présentent des assemblages spécifiques à ce type d'environnement, différents de ceux qu'on trouve dans les écosystèmes terrestres. Il est donc important de s'intéresser aux intégrons dans ces environnements particuliers et d'évaluer les facteurs et pressions influençant leur diversité.

La publication qui suit est une compilation des résultats obtenus jusqu'à présent dans différentes études réalisées sur la diversité des intégrons dans les milieux marins et côtiers, incluant nos propres résultats sur l'analyse de la diversité des intégrases dans des environnements côtiers (séquences obtenues pendant cette thèse et par le Dr Lionel Huang pendant son doctorat (Huang, 2010).

Environ Sci Pollut Res
DOI 10.1007/s11356-015-5085-3

DECAPAGE PROJECT: HYDROCARBON DEGRADATION IN COASTAL SEDIMENTS*

Integron diversity in marine environments

Justine Abella[1] · Ana Bielen[2,3] · Lionel Huang[1,4] · Tom O. Delmont[1,5] · Dušica Vujaklija[2] · Robert Duran[1] · Christine Cagnon[1]

Received: 2 February 2015 / Accepted: 16 July 2015
© Springer-Verlag Berlin Heidelberg 2015

Abstract Integrons are bacterial genetic elements known to be active vectors of antibiotic resistance among clinical bacteria. They are also found in bacterial communities from natural environments. Although integrons have become especially efficient for bacterial adaptation in the particular context of antibiotic usage, their role in natural environments in other contexts is still unknown. Indeed, most studies have focused on integrons and the spread of antibiotic resistance in freshwater or soil impacted by anthropogenic activities, with only few on marine environments. Notably, integrons show a wider diversity of both gene cassettes and integrase gene in natural environments than in clinical environments, suggesting a general role of integrons in bacterial adaptation. This article reviews the current knowledge on integrons in marine environments. We also present conclusions of our studies on polluted and nonpolluted backgrounds.

Responsible editor: Philippe Garrigues

✉ Christine Cagnon
christine.cagnon@univ-pau.fr

[1] Equipe Environnement et Microbiologie, MELODY Group, Université de Pau et des Pays de l'Adour, IPREM UMR CNRS 5254, BP 1155, 64013 Pau, Cedex, France

[2] Laboratory for Molecular Genetics, Division of Molecular Biology, Ruđer Bošković Institute, Bijenička 54, 10000 Zagreb, Croatia

[3] Laboratory for Biology and Microbial Genetics, Department of Biochemical Engineering, Faculty of Food Technology and Biotechnology, University of Zagreb, 10000 Zagreb, Croatia

[4] Present address: Euro Engineering, Technopole Hélioparc Bât Newton, 4 rue Jules Ferry, CS N 99207, 64053 Pau, Cedex 09, France

[5] Present address: Josephine Bay Paul Center for Comparative Molecular Biology and Evolution, Marine Biology Laboratory, Woods Hole, MA, USA

Keywords Integron integrase · Gene cassettes · Contaminants · Adaptation · Bacterial communities

Introduction

Bacterial communities are able to adapt rapidly to changing conditions which help them to cope with disturbance in their environment, such as chemical contaminations. Bacterial adaptation is in part due to acquisition of new DNA sequences by horizontal gene transfer, involving genetic elements such as plasmids, phages, transposons, and integrons (Ochman et al. 2000; Smalla and Sobecky 2002). Integrons were the most recently discovered elements and, although their role in spreading antibiotic resistances is well known and well studied (Hall and Collis 1995; Ploy et al. 2000; Cambray et al. 2010; Stalder et al. 2012), their role in other contamination contexts is still in question.

Most studies on nonclinical environments have focused on "antibiotic resistance integrons" spreading, especially in freshwater and soil impacted by anthropogenic activities (Henriques et al. 2006; Moura et al. 2007; Laroche et al. 2009; Byrne-Bailey et al. 2011; Gaze et al. 2011; Stalder et al. 2014). Indeed "antibiotic resistance integrons" present a very low diversity in comparison to the variety of integrons found in natural environments (Boucher et al. 2007) suggesting that other types of integrons might be involved in bacterial adaptation in response to the large variety of possible pollutants. Few investigations tried to characterize nonclinical integrons in various natural environments: soil, sediment or water, from terrestrial, estuarine, or marine areas (e.g., Rodriguez-Minguela et al. 2009; Stokes et al. 2001).

Marine environments are subjected to a variety of pollutants from both sides: from the land, principally through agricultural practices, urban wastewaters and industrial activities

Published online: 28 July 2015

and from the sea, predominantly by crude oil from shipping and offshore mining. Because salinity is the major environmental determinant of microbial community composition (Lozupone and Knight 2007), the marine bacterial communities present specific assemblages different from those found in terrestrial ecosystems. Thus, marine environments should be taken into consideration for the full description of integron diversity.

We review knowledge on integrons in diverse marine environments including coastal zones, estuaries and the deep sea. We present the main conclusions of our studies on both polluted and non-polluted conditions.

Structure of integrons

Integron structure (Fig. 1) is characterized by (i) a gene, *intI*, encoding an integrase of the tyrosine recombinase family; (ii) a recombination site, *attI*; (iii) and a succession of gene cassettes each characterized by a coding sequence associated with a recombination site, *attC* (Stokes and Hall 1989). While *intI* and *attI* are the stable part of integrons, gene cassettes are the variable portion of these structures, showing a large sequence diversity (Holmes et al. 2003; Koenig et al. 2008) and varying in number from one to more than one hundred (Mazel 2006). Integrons can be thought of as gene cassettes reservoirs (Holmes et al. 2003; Michael et al. 2004). Integration or excision of gene cassettes is catalyzed by the integron-integrase, product of *intI*, allowing recombination between sites on the integron (Collis et al. 2002; Cambray et al. 2010) (Fig. 1). The expression of *intI* is controlled by the SOS system, for at least some integrons (Cambray et al. 2011), suggesting rearrangement of gene cassettes in response to environmental stresses. Finally, the expression of gene cassettes is ensured by the presence of one or two promoters located inside the stable part of the integron (Collis and Hall 1995; Jové et al. 2010) (Fig. 1).

Integrons in natural environments

Integrons were first discovered and largely studied in clinical contexts because of their role in spreading antibiotic resistance genes (Stokes and Hall 1989; Fluit and Schmitz 1999; Ploy et al. 2000). Three classes of integrons were described from clinical environments: classes 1, 2, and 3 according to their *intI* sequences (Ploy et al. 2000; Hall 2012). Then, other integron integrases were discovered in other environments, defining new integron classes: *intI4*, *5*, *6*, *7*, and *8* (Mazel et al. 1998; Nield et al. 2001); but today, thanks to sequencing efforts, hundreds of integron integrases have been identified from environmental samples (Rapa and Labbate 2013; Gillings 2014). The "environmental integrons" are distinct from "clinical integrons" by both integrase and gene cassette sequences (Holmes et al. 2003; Koenig et al. 2008; Gillings 2014).

The ancestors of clinical and environmental integrons are ancient, existing before the use of antibiotics (Mazel 2006; Boucher et al. 2007). Integrons are widespread in bacterial communities (Mazel et al. 1998; Stokes et al. 2006; Boucher et al. 2007; Gillings et al. 2008a; Gillings 2014). About 17 % of sequenced bacterial genomes possess integrons (Cambray et al. 2010), mostly Gram-negative Proteobacteria, and roughly in accordance with the actual ratio of sequenced genomes. Nevertheless, bacteria which do not carry such a structure can acquire an integron since these elements can be mobilized via mobile elements, such as insertion sequences, transposons, and conjugative plasmids (Mazel 2006). So far, no integrons have been described in Archaea. Culture-independent methods have detected integrons in environments as diverse as soil, sediment, biofilm, and water, irrespective of the presence of pollutants, including environments without antibiotic history (Nield et al. 2001; Stokes et al. 2001; Nemergut et al. 2004; Elsaied et al. 2007, 2011, 2014; Hardwick et al. 2008; Koenig et al. 2008, 2009; Wright et al. 2008; Rodriguez-Minguela et al. 2009; Michael and Andrew 2010; Moura et al. 2010; Gaze et al. 2011).

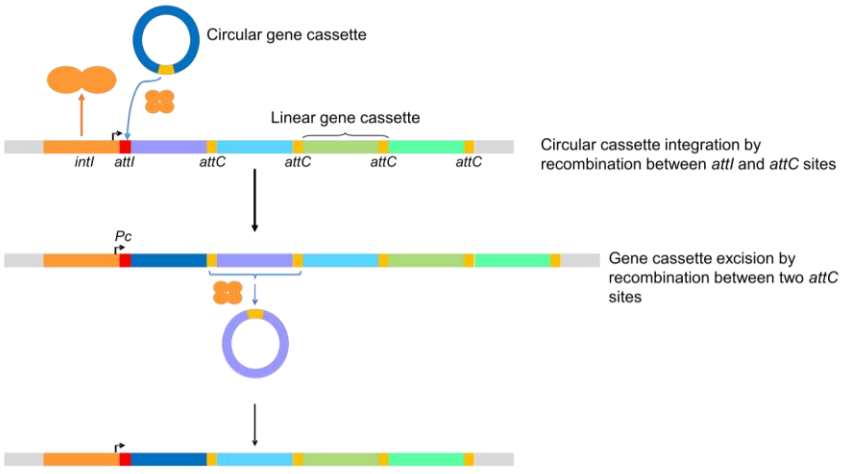


Fig. 1 General organization of integrons: scheme of structure, integration, and excision process. *intI* encodes an integrase that catalyses the integration or the excision of gene cassettes (for review, see Cambray et al. 2010)

The class 1 integron, the most abundant in clinical contexts, was detected in various natural nonpristine environments (Gillings et al. 2008b; Moura et al. 2010; Gaze et al. 2011; Drudge et al. 2012; Ling et al. 2013; Jechalke et al. 2014) with increasing abundance according to ecological perturbations (Hardwick et al. 2008; Wright et al. 2008; Rosewarne et al. 2010; Byrne-Bailey et al. 2011; Gaze et al. 2011; Gillings et al. 2015). Their presence was related to the use or the release of effluent containing antibiotic-resistant bacteria (Moura et al. 2007; Gaze et al. 2008; Byrne-Bailey et al. 2011; Wellington et al. 2013; Stalder et al. 2014). However, their presence in other contexts such as chemical contamination suggested that coselection of these integrons with other resistance elements might occur. Indeed, clinical integrons are frequently included in larger mobile elements and associated with transposons that carry metal resistance genes (Liebert et al. 1999; Wright et al. 2008).

Regarding gene cassettes, their amplification from metagenomes from various environments revealed that they can carry functions, when identified, related to the physical-chemical conditions prevailing in these environments (Nemergut et al. 2004; Koenig et al. 2009). Such observations suggested that selection of some adaptive genes from the reservoir of gene cassettes occurred in these environments. This suggests that environmental integrons contribute to the adaptation of bacterial communities in response to environmental disturbances.

Diversity of integrons in marine environments

Due to global spread of antibiotic resistances, many studies focus on clinical integrons in human impacted environments. Studies of integrons in coastal or estuarine environments were also often limited to the identification or quantification of clinical integrons (Rosser and Young 1999; Henriques et al. 2006; Gillings et al. 2008b; Laroche et al. 2009). However, most integrons from bacterial communities of nonclinical environments do not belong to integron class 1, 2, or 3. Therefore, we describe below the diversity of these environmental integrons in marine ecosystems.

Integrase diversity

Exploring the 4528 bacterial complete-genomes available at the NCBI in February 2014, 674 integron integrases were found. Among these bacteria from diverse environments, we found integron integrases in marine bacteria. An updated list of marine bacteria containing integron integrases is shown in Table 1. Most of their integrases do not belong to clinical classes and exhibit a large diversity (Heinemann et al. 2006; Boucher et al. 2007; Nemergut et al. 2008; Cambray et al. 2010). In phylogenetic analyses, they mainly group together (Boucher et al. 2007; Cambray et al. 2010). This suggests lateral transfer of integrons between marine species (Gillings 2014).

Culture-independent methods detected integrons in marine water and sediments, revealing new IntI phylotypes (11, 33, 37, and 42, respectively; Elsaied et al. 2007, 2011, 2014; Rodriguez-Minguela et al. 2009). Our investigation of *intI* diversity from coastal marine sediments collected at geographically distant sampling sites, Atlantic Ocean (Brittany, France) and the northern Adriatic Sea (Rovinj area, Croatia), identified 61 and 147 new phylotypes (according to a 98 % of amino acid identity) (Table 2). In spite of a large number of analyzed sequences, only one common sequence was found

Table 1 Marine bacteria containing integron integrases

Phyla	Genera
Bacteroidetes	*Balneola*, *Gracilimonas*
Chlorobi	*Chlorobium*, *Chloroherpeton*
Cyanobacteria	*Cyanobium*, *Leptolyngbya*, *Pseudanabaena*, *Rubidibacter*, *Spirulina*, *Synechococcus*
Lentisphaerae	*Lentisphaera*
Planctomycetes	*Blastospirellula*, *Rhodopirellula*
Proteobacteria (Beta)	*Aeromonas*, *Desulfobacula*
Proteobacteria (Delta)	*Desulfocapsa*, *Desulfuromonas*, *Geobacter*
Proteobacteria (Gamma)	*Agarivorans*, *Alcaniovorax*, *Aliivibrio*, *Alishewanella*, *Alteromonas*, *Amphritea*, *Bermanella*, *Catenovulum*, *Colwiella*, *Congregibacter*, *Cycloclasticus*, *Dasania*, *Endozoicomonas*, *Gilvimarinus*, *Glaciecola*, *Grimontia*, *Halomonas*, *Kangiella*, *Listonella*, *Marinobacter*, *Methylobacter*, *Methylomonas*, *Methylophaga*, *Microbulbifer*, *Moritella*, *Neptuniibacter*, *Nitrosococcus*, *Oceanimonas*, *Oceanobacter*, *Oceanospirillum*, *Oleispira*, *Pseudoalteromonas*, *Pseudomonas*, *Pseychromonas*, *Reinekea*, *Rheinheimera*, *Saccharophagus*, *Salinisphaera*, *Shewanella*, *Simiduia*, *Spongiibacter*, *Teredinibacter*, *Thiomicrospira*, *Thiothrix*, *Vibrio*, *Thermodesulfatator*
Verrumicrobia	*Opitutaceae*

Exploring the 4528 bacterial complete-genomes available at the NCBI in February 2014

Table 2 Integron integrase diversity in libraries from coastal sediments

	Coverage (%)	No. of total sequences	No. of phylotypes	Shannon index
Adriatic sea, Croatia				
Luka	60.2	88	47	3.35
Marina	65.6	90	38	2.57
Vani	54.9	82	45	3.39
Sredina	51.5	68	45	3.66
Atlantic Ocean, France (microcosms)				
BAL 90	62.9	62	30	2.78
BAL 270	84.9	73	18	1.65
BAL 270' (duplicate)	86.6	82	15	1.38
CTRL 270	65.2	46	20	2.45

Samples are described in Fig. 4. For more details on samples from Croatia, see Duran et al. (2015); on microcosms from French sediments, see Stauffert et al. (2013). IntI libraries were obtained as described in Fig. 2

between the French and the Croatian areas. This suggests a large diversity of integrase sequences that have not been fully explored. Further, this result was supported by the rarefaction analysis showing coverages in the range 51 to 87 % for the *intI* libraries from France and Croatia (Table 2).

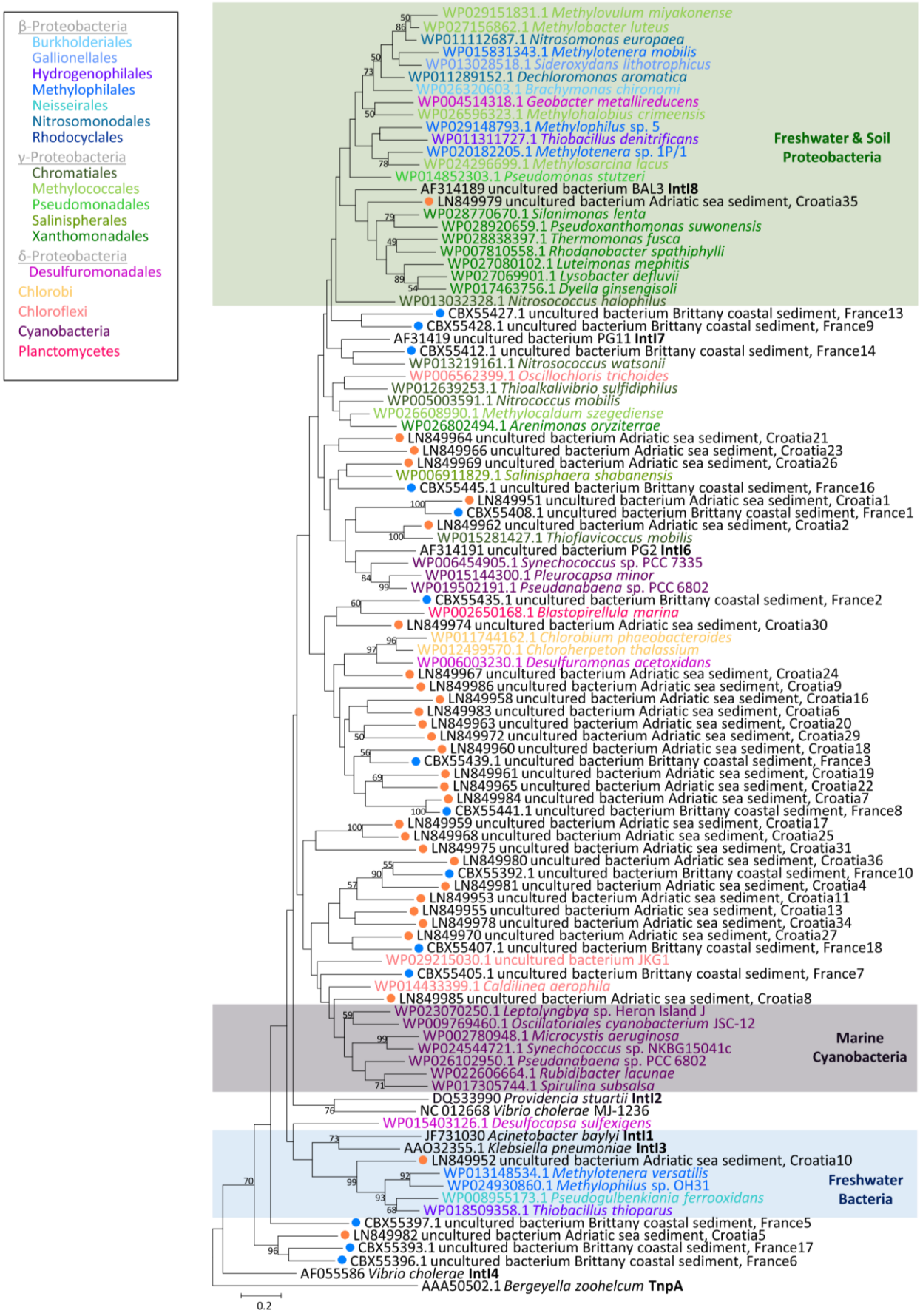
Blast searches of these new integrase sequences against bacterial sequenced genomes showed similarities mainly with integrases from freshwater and soil bacteria, but also with several integrases from marine bacteria (Fig. 2). These integrase sequence hits belonged mainly to Proteobacteria (beta and gamma) and also to Cyanobacteria, especially marine Cyanobacteria. Our phylogenetic analysis showed that the majority of these new environmental integrase sequences did not cluster with known sequences (Fig. 2), highlighting a lack of knowledge of integrase diversity from marine bacteria. Nevertheless, several of them were homologous to integrons from Proteobacteria, in particular the gamma and beta Proteobacteria (e.g., Croatia35), and several others to those from Cyanobacteria (in particular Croatia8). Phylogenetic analysis including integron integrases from other marine metagenomes did not show clustering dependent of the type or of the geographic origin of the samples (Fig. 3), but did show similarities between sequences from coastal and deep-sea environments, suggesting a global marine diversity of integron integrases.

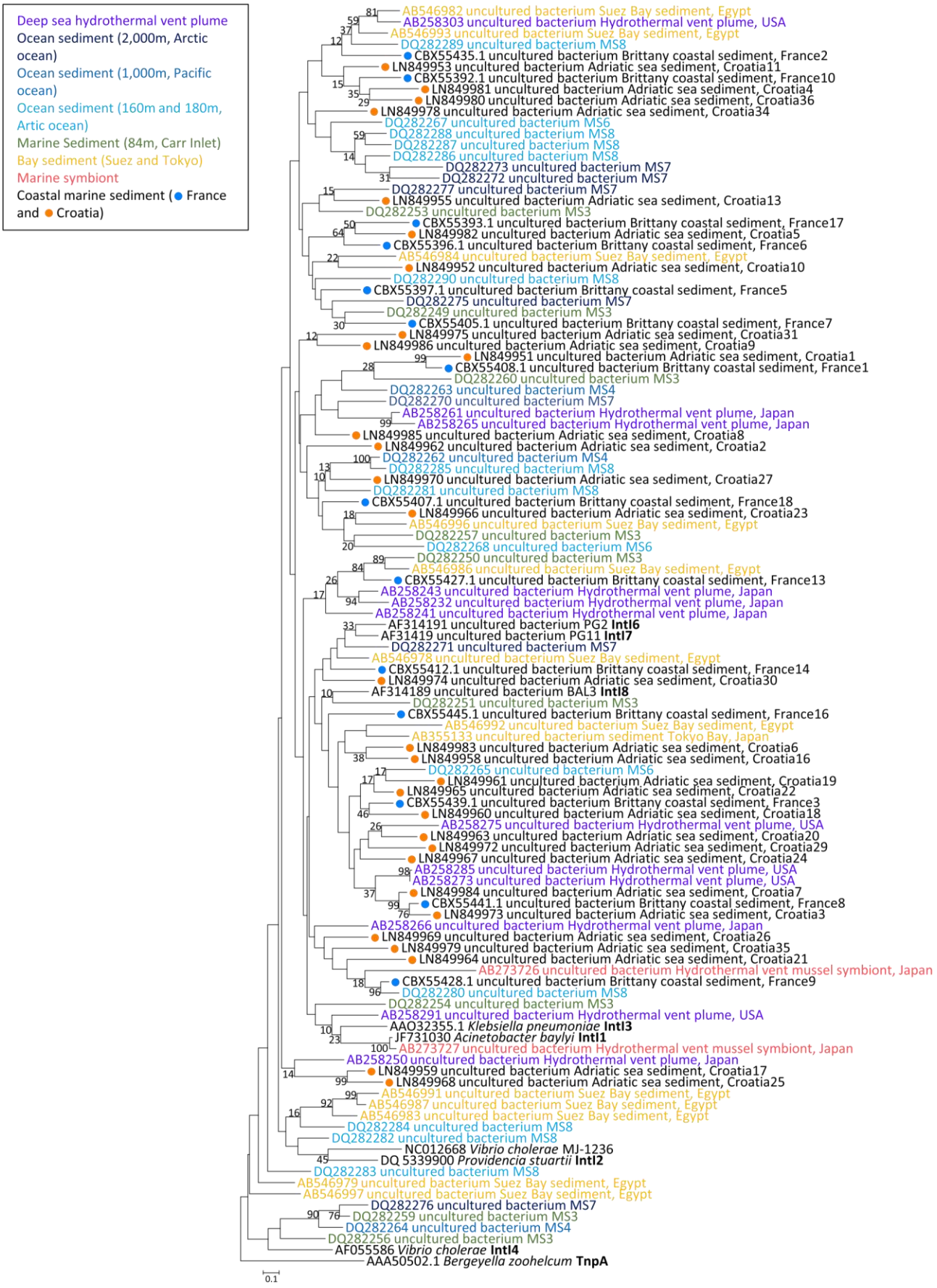
Interestingly, we could not detect class 1, 2, and 3 *intI* (*intI1*, *intI2*, and *intI3*, respectively) with PCR-based metagenomic analyses as reported in previous studies of marine water and sediments (Elsaied et al. 2007, 2011, 2014; Rodriguez-Minguela et al. 2009), while they were detected in terrestrial environments (Rodriguez-Minguela et al. 2009). However, when *intI1* was specifically targeted, we detected it in coastal sediments (Aber-Benoît, France; Fig. 5) as previously reported for estuarine water and sediments (Wright et al. 2008; Gillings et al. 2008b; Uyaguari et al. 2013). *intI2* and *intI3* were also quantified in such environments (Uyaguari et al. 2013). *intI1* and more rarely *intI2* were identified in some isolated-bacteria from various estuarine environments (Henriques et al. 2006; Wang et al. 2008; Laroche et al. 2009; Azevedo et al. 2013; Maravić et al. 2013; Pereira et al. 2013; Moura et al. 2014). To date, *intI3* has not been identified in isolated-bacteria from marine environments. Thus, these observations indicated that class 1, 2, and 3 integrons are present in coastal and estuarine environments, but the difficulties to obtain them by metagenomic analyses suggest that they are little represented.

Gene cassette diversity

Gene cassette diversity of marine environmental integrons has been revealed to be very large by both culture-dependent and culture-independent approaches. Analysis of sequenced genomes of *Vibrio* (Boucher et al. 2007) and PCR-amplified sequences from metagenomes of various environments (contaminated coastal sediments, deep-sea hydrothermal vents, and submarine gas-hydrate-bearing core; Stokes et al. 2001; Koenig et al. 2008, 2009; Elsaied et al. 2007, 2011, 2014)

Fig. 2 Phylogenetic tree obtained from the Gblocks alignments of integron integrases (IntI) from coastal sediment metagenomes. Representative sequences of the IntI diversity from coastal sediments (*orange circles*: Adriatic sea, Rovinj area, Croatia; *blue circles*: Atlantic Ocean, Brittany, France) and the five hit sequences in BlastP analysis based on RefSeq databases. *intI1* to *8* sequences were included (except *intI5*). IntI libraries were constructed from coastal sediment metagenomes described in Duran et al. (2015) and Stauffert et al. (2013). Amplification of an 800-bp *intI* fragment was performed using as primers the inverted complementary sequence of AJH72 (Gillings et al. 2005) and *intI948R* (Elsaied et al. 2007). Purified PCR products (GFX DNA and Gel Band Purification kit, GE Healthcare) were cloned by using pGEM-T easy Vector System II (Promega) and were sequenced by GATC Biotech. Sequences were analyzed with Sequencher® software and translated with BioEdit® software. Maximum likelihood method was used for construction of the phylogenetic tree after alignment of amino acid sequence data by the MUSCLE algorithm (Edgar 2004). Bootstrap values, expressed as percentages of 500 datasets, are given for values upper than 50 % at nodes. The tree was rooted by TnpA sequence of *Bergellya zoohelcum*


β-Proteobacteria
Burkholderiales
Gallionellales
Hydrogenophilales
Methylophilales
Neisseirales
Nitrosomonodales
Rhodocyclales
γ-Proteobacteria
Chromatiales
Methylococcales
Pseudomonadales
Salinispherales
Xanthomonadales
δ-Proteobacteria
Desulfuromonadales
Chlorobi
Chloroflexi
Cyanobacteria
Planctomycetes
WP029151831.1 Methylovulum miyakonense
WP027156862.1 Methylobacter luteus
WP011112687.1 Nitrosomonas europaea
WP015831343.1 Methylotenera mobilis
WP013028518.1 Sideroxydans lithotrophicus
WP011289152.1 Dechloromonas aromatica
WP026320603.1 Brachymonas chironomi
WP004514318.1 Geobacter metallireducens
WP026596323.1 Methylohalobius crimeensis
WP029148793.1 Methylophilus sp. 5
WP011311727.1 Thiobacillus denitrificans
WP020182205.1 Methylotenera sp. 1P/1
WP024296699.1 Methylosarcina lacus
WP014852303.1 Pseudomonas stutzeri
AF314189 uncultured bacterium BAL3 IntI8
LN849979 uncultured bacterium Adriatic sea sediment, Croatia35
WP028770670.1 Silanimonas lenta
WP028920659.1 Pseudoxanthomonas suwonensis
WP028838397.1 Thermomonas fusca
WP007810558.1 Rhodanobacter spathiphylli
WP027080102.1 Luteimonas mephitis
WP027069901.1 Lysobacter defluvii
WP017463756.1 Dyella ginsengisoli
WP013032328.1 Nitrosococcus halophilus
Freshwater & Soil Proteobacteria
CBX55427.1 uncultured bacterium Brittany coastal sediment, France13
CBX55428.1 uncultured bacterium Brittany coastal sediment, France9
AF31419 uncultured bacterium PG11 IntI7
CBX55412.1 uncultured bacterium Brittany coastal sediment, France14
WP013219161.1 Nitrosococcus watsonii
WP006562399.1 Oscillochloris trichoides
WP012639253.1 Thioalkalivibrio sulfidiphilus
WP005003591.1 Nitrococcus mobilis
WP026608990.1 Methylocaldum szegediense
WP026802494.1 Arenimonas oryziterrae
LN849964 uncultured bacterium Adriatic sea sediment, Croatia21
LN849966 uncultured bacterium Adriatic sea sediment, Croatia23
LN849969 uncultured bacterium Adriatic sea sediment, Croatia26
WP006911829.1 Salinisphaera shabanensis
CBX55445.1 uncultured bacterium Brittany coastal sediment, France16
LN849951 uncultured bacterium Adriatic sea sediment, Croatia1
CBX55408.1 uncultured bacterium Brittany coastal sediment, France1
LN849962 uncultured bacterium Adriatic sea sediment, Croatia2
WP015281427.1 Thioflavicoccus mobilis
AF314191 uncultured bacterium PG2 IntI6
WP006454905.1 Synechococcus sp. PCC 7335
WP015144300.1 Pleurocapsa minor
WP019502191.1 Pseudanabaena sp. PCC 6802
CBX55435.1 uncultured bacterium Brittany coastal sediment, France2
WP002650168.1 Blastopirellula marina
LN849974 uncultured bacterium Adriatic sea sediment, Croatia30
WP011744162.1 Chlorobium phaeobacteroides
WP012499570.1 Chloroherpeton thalassium
WP006003230.1 Desulfuromonas acetoxidans
LN849967 uncultured bacterium Adriatic sea sediment, Croatia24
LN849986 uncultured bacterium Adriatic sea sediment, Croatia9
LN849958 uncultured bacterium Adriatic sea sediment, Croatia16
LN849983 uncultured bacterium Adriatic sea sediment, Croatia6
LN849963 uncultured bacterium Adriatic sea sediment, Croatia20
LN849972 uncultured bacterium Adriatic sea sediment, Croatia29
LN849960 uncultured bacterium Adriatic sea sediment, Croatia18
CBX55439.1 uncultured bacterium Brittany coastal sediment, France3
LN849961 uncultured bacterium Adriatic sea sediment, Croatia19
LN849965 uncultured bacterium Adriatic sea sediment, Croatia22
LN849984 uncultured bacterium Adriatic sea sediment, Croatia7
CBX55441.1 uncultured bacterium Brittany coastal sediment, France8
LN849959 uncultured bacterium Adriatic sea sediment, Croatia17
LN849968 uncultured bacterium Adriatic sea sediment, Croatia25
LN849975 uncultured bacterium Adriatic sea sediment, Croatia31
LN849980 uncultured bacterium Adriatic sea sediment, Croatia36
CBX55392.1 uncultured bacterium Brittany coastal sediment, France10
LN849981 uncultured bacterium Adriatic sea sediment, Croatia4
LN849953 uncultured bacterium Adriatic sea sediment, Croatia11
LN849955 uncultured bacterium Adriatic sea sediment, Croatia13
LN849978 uncultured bacterium Adriatic sea sediment, Croatia34
LN849970 uncultured bacterium Adriatic sea sediment, Croatia27
CBX55407.1 uncultured bacterium Brittany coastal sediment, France18
WP029215030.1 uncultured bacterium JKG1
CBX55405.1 uncultured bacterium Brittany coastal sediment, France7
WP014433399.1 Caldilinea aerophila
LN849985 uncultured bacterium Adriatic sea sediment, Croatia8
WP023070250.1 Leptolyngbya sp. Heron Island J
WP009769460.1 Oscillatoriales cyanobacterium JSC-12
WP002780948.1 Microcystis aeruginosa
WP024544721.1 Synechococcus sp. NKBG15041c
WP026102950.1 Pseudanabaena sp. PCC 6802
WP022606664.1 Rubidibacter lacunae
WP017305744.1 Spirulina subsalsa
Marine Cyanobacteria
DQ533990 Providencia stuartii IntI2
NC 012668 Vibrio cholerae MJ-1236
WP015403126.1 Desulfocapsa sulfexigens
JF731030 Acinetobacter baylyi IntI1
AAO32355.1 Klebsiella pneumoniae IntI3
LN849952 uncultured bacterium Adriatic sea sediment, Croatia10
WP013148534.1 Methylotenera versatilis
WP024930860.1 Methylophilus sp. OH31
WP008955173.1 Pseudogulbenkiania ferrooxidans
WP018509358.1 Thiobacillus thioparus
Freshwater Bacteria
CBX55397.1 uncultured bacterium Brittany coastal sediment, France5
LN849982 uncultured bacterium Adriatic sea sediment, Croatia5
CBX55393.1 uncultured bacterium Brittany coastal sediment, France17
CBX55396.1 uncultured bacterium Brittany coastal sediment, France6
AF055586 Vibrio cholerae IntI4
AAA50502.1 Bergeyella zoohelcum TnpA
0.2


Deep sea hydrothermal vent plume
Ocean sediment (2,000m, Arctic ocean)
Ocean sediment (1,000m, Pacific ocean)
Ocean sediment (160m and 180m, Artic ocean)
Marine Sediment (84m, Carr Inlet)
Bay sediment (Suez and Tokyo)
Marine symbiont
Coastal marine sediment (● France and ● Croatia)
AB546982 uncultured bacterium Suez Bay sediment, Egypt
AB258303 uncultured bacterium Hydrothermal vent plume, USA
AB546993 uncultured bacterium Suez Bay sediment, Egypt
DQ282289 uncultured bacterium MS8
CBX55435.1 uncultured bacterium Brittany coastal sediment, France2
LN849953 uncultured bacterium Adriatic sea sediment, Croatia11
CBX55392.1 uncultured bacterium Brittany coastal sediment, France10
LN849981 uncultured bacterium Adriatic sea sediment, Croatia4
LN849980 uncultured bacterium Adriatic sea sediment, Croatia36
LN849978 uncultured bacterium Adriatic sea sediment, Croatia34
DQ282267 uncultured bacterium MS6
DQ282288 uncultured bacterium MS8
DQ282287 uncultured bacterium MS8
DQ282286 uncultured bacterium MS8
DQ282273 uncultured bacterium MS7
DQ282272 uncultured bacterium MS7
DQ282277 uncultured bacterium MS7
LN849955 uncultured bacterium Adriatic sea sediment, Croatia13
DQ282253 uncultured bacterium MS3
CBX55393.1 uncultured bacterium Brittany coastal sediment, France17
LN849982 uncultured bacterium Adriatic sea sediment, Croatia5
CBX55396.1 uncultured bacterium Brittany coastal sediment, France6
AB546984 uncultured bacterium Suez Bay sediment, Egypt
LN849952 uncultured bacterium Adriatic sea sediment, Croatia10
DQ282290 uncultured bacterium MS8
CBX55397.1 uncultured bacterium Brittany coastal sediment, France5
DQ282275 uncultured bacterium MS7
DQ282249 uncultured bacterium MS3
CBX55405.1 uncultured bacterium Brittany coastal sediment, France7
LN849975 uncultured bacterium Adriatic sea sediment, Croatia31
LN849986 uncultured bacterium Adriatic sea sediment, Croatia9
LN849951 uncultured bacterium Adriatic sea sediment, Croatia1
CBX55408.1 uncultured bacterium Brittany coastal sediment, France1
DQ282260 uncultured bacterium MS3
DQ282263 uncultured bacterium MS4
DQ282270 uncultured bacterium MS7
AB258261 uncultured bacterium Hydrothermal vent plume, Japan
AB258265 uncultured bacterium Hydrothermal vent plume, Japan
LN849985 uncultured bacterium Adriatic sea sediment, Croatia8
LN849962 uncultured bacterium Adriatic sea sediment, Croatia2
DQ282262 uncultured bacterium MS4
DQ282285 uncultured bacterium MS8
LN849970 uncultured bacterium Adriatic sea sediment, Croatia27
DQ282281 uncultured bacterium MS8
CBX55407.1 uncultured bacterium Brittany coastal sediment, France18
LN849966 uncultured bacterium Adriatic sea sediment, Croatia23
AB546996 uncultured bacterium Suez Bay sediment, Egypt
DQ282257 uncultured bacterium MS3
DQ282268 uncultured bacterium MS6
DQ282250 uncultured bacterium MS3
AB546986 uncultured bacterium Suez Bay sediment, Egypt
CBX55427.1 uncultured bacterium Brittany coastal sediment, France13
AB258243 uncultured bacterium Hydrothermal vent plume, Japan
AB258232 uncultured bacterium Hydrothermal vent plume, Japan
AB258241 uncultured bacterium Hydrothermal vent plume, Japan
AF314191 uncultured bacterium PG2 **IntI6**
AF31419 uncultured bacterium PG11 **IntI7**
DQ282271 uncultured bacterium MS7
AB546978 uncultured bacterium Suez Bay sediment, Egypt
CBX55412.1 uncultured bacterium Brittany coastal sediment, France14
LN849974 uncultured bacterium Adriatic sea sediment, Croatia30
AF314189 uncultured bacterium BAL3 **IntI8**
DQ282251 uncultured bacterium MS3
CBX55445.1 uncultured bacterium Brittany coastal sediment, France16
AB546992 uncultured bacterium Suez Bay sediment, Egypt
AB355133 uncultured bacterium sediment Tokyo Bay, Japan
LN849983 uncultured bacterium Adriatic sea sediment, Croatia6
LN849958 uncultured bacterium Adriatic sea sediment, Croatia16
DQ282265 uncultured bacterium MS6
LN849961 uncultured bacterium Adriatic sea sediment, Croatia19
LN849965 uncultured bacterium Adriatic sea sediment, Croatia22
CBX55439.1 uncultured bacterium Brittany coastal sediment, France3
LN849960 uncultured bacterium Adriatic sea sediment, Croatia18
AB258275 uncultured bacterium Hydrothermal vent plume, USA
LN849963 uncultured bacterium Adriatic sea sediment, Croatia20
LN849972 uncultured bacterium Adriatic sea sediment, Croatia29
LN849967 uncultured bacterium Adriatic sea sediment, Croatia24
AB258285 uncultured bacterium Hydrothermal vent plume, USA
AB258273 uncultured bacterium Hydrothermal vent plume, USA
LN849984 uncultured bacterium Adriatic sea sediment, Croatia7
CBX55441.1 uncultured bacterium Brittany coastal sediment, France8
LN849973 uncultured bacterium Adriatic sea sediment, Croatia3
AB258266 uncultured bacterium Hydrothermal vent plume, Japan
LN849969 uncultured bacterium Adriatic sea sediment, Croatia26
LN849979 uncultured bacterium Adriatic sea sediment, Croatia35
LN849964 uncultured bacterium Adriatic sea sediment, Croatia21
AB273726 uncultured bacterium Hydrothermal vent mussel symbiont, Japan
CBX55428.1 uncultured bacterium Brittany coastal sediment, France9
DQ282280 uncultured bacterium MS8
DQ282254 uncultured bacterium MS3
AB258291 uncultured bacterium Hydrothermal vent plume, USA
AAO32355.1 *Klebsiella pneumoniae* **IntI3**
JF731030 *Acinetobacter baylyi* **IntI1**
AB273727 uncultured bacterium Hydrothermal vent mussel symbiont, Japan
AB258250 uncultured bacterium Hydrothermal vent plume, Japan
LN849959 uncultured bacterium Adriatic sea sediment, Croatia17
LN849968 uncultured bacterium Adriatic sea sediment, Croatia25
AB546991 uncultured bacterium Suez Bay sediment, Egypt
AB546987 uncultured bacterium Suez Bay sediment, Egypt
AB546983 uncultured bacterium Suez Bay sediment, Egypt
DQ282284 uncultured bacterium MS8
DQ282282 uncultured bacterium MS8
NC012668 *Vibrio cholerae* MJ-1236
DQ 5339900 *Providencia stuartii* **IntI2**
DQ282283 uncultured bacterium MS8
AB546979 uncultured bacterium Suez Bay sediment, Egypt
AB546997 uncultured bacterium Suez Bay sediment, Egypt
DQ282276 uncultured bacterium MS7
DQ282259 uncultured bacterium MS3
DQ282264 uncultured bacterium MS4
DQ282256 uncultured bacterium MS3
AF055586 *Vibrio cholerae* **IntI4**
AAA50502.1 *Bergeyella zoohelcum* **TnpA**
0.1

◀ **Fig. 3** Phylogenetic tree obtained from the Gblocks alignments of integron integrases (IntI) from different marine and coastal metagenomes. Representative sequences of the IntI diversity from deep sea (Washington state, USA; Arctic Ocean, Alaska, USA; western Pacific arc hydrothermal vents, Suiyo Seamount, Japan, and Pika, USA) and coastal areas (Suez Bay, Egypt; Tokyo Bay, Japan; Adriatic Sea, Croatia; Brittany, France) (Elsaied et al. 2007, 2011; Rodriguez-Minguela et al. 2009; sequences described in Fig. 2). *intI1* to *8* sequences were included (except *intI5*). Maximum likelihood method was used for construction of the phylogenetic tree after alignment of sequence data by the MUSCLE algorithm (Edgar 2004). Bootstrap values, expressed as percentages of 500 datasets, are given for values upper than 50 % at nodes. The tree was rooted by TnpA sequence of *Bergellya zoohelcum*

identified nearly 5000 distinct cassette-encoded ORFs. Since in most cases, the libraries obtained by cloning-sequencing approaches are largely unsaturated, the marine gene cassette pool is still not fully explored.

In all studies, the majority of ORFs of the gene cassette pool have no known homologues or have homologues of unknown function (70 to 80 % in general cases). Where possible, the assigned functions belong mainly to general cellular and metabolic functions (Boucher et al. 2007; Elsaied and Maruyama 2011). Except when class 1 integrons were specifically studied, few gene cassettes usually found in clinical integrons were recovered (Rosser and Young 1999; Koenig et al. 2009; Elsaied and Maruyama 2011).

Despite a rather poor knowledge of environmental gene cassettes, some gene cassettes could be found in different metagenomes from geographically distant polluted marine sediments (Suez, Egypt / Halifax, Canada; Tokyo, Japan/ Halifax, Canada), suggesting a marine global gene cassette pool (Elsaied et al. 2011).

Impact of contamination

The marine environment is subject to variable gradients of contaminants from industrial and urban discharge as well as accidental oil spills. The description of integron diversity in marine contaminated environments will provide useful information to determine the role of integrons in adaptation of bacterial communities in response to major environmental disturbances.

Integrase diversity

To evaluate the impact of contamination on integrase diversity, we performed analyses of *intI* libraries from sediments of West Istrian coast (Croatia) at sites contaminated by toxic metals, PAH and PCB (description of samples in Duran et al. 2015 and Durmiši et al. 2013) and from mesocosms of oil-contaminated mudflat sediments (Brittany, France; description of samples in Stauffert et al. 2013). MDS analysis showed more similarities between IntI pools from geographically close samples than to their pollution level (Fig. 4). Similarly, Elsaied et al. (2011) demonstrated a geographic separation of integrase sequences from two distant contaminated bays. Nevertheless, the integrase sequence pools of the two low-contaminated and the two high-contaminated Croatian sediments were grouped together, respectively (Fig. 4). In the mesocosm experiment, more similarity was also observed between integrase pools of long-time oil-contaminated duplicates than with uncontaminated control or shorter time oil-contaminated samples (Fig. 4). The decrease of the *intI* diversity in high-contaminated or long-time exposure contaminated samples (Table 2) may explain these observations.

A previous study showed that in the context of contamination, the abundance of the integrase gene of class 1 integrons increased in estuarine environments (Wright et al. 2008). Contradictory results were obtained in another estuarine area where *intI1* was in a higher abundance in a low urban-impacted site than in impacted sites, whereas *intI2* was more abundant in sites with high anthropogenic input (Uyaguari et al. 2013). These contradictory observations could be explained by the discharges feeding the second study sites, being poor in class 1 integrons, and demonstrate that clinical integrons observed in estuarine environments may come mainly from anthropogenic inputs. Nevertheless, environmental class 1 integrons are found in environmental bacteria (Gillings 2014), surely explaining the presence of *intI1* in nonanthropogenic impacted environments. As it was demonstrated that class 1 integrons increased under antibiotic and metal contaminations in freshwater environments after 7 days of contamination (Wright et al. 2008), it was interesting to determine if other contaminants, such as oil, have an impact on *intI1* abundance. Thus, we quantified *intI1* in oil-contaminated coastal sediment microcosms (Fig. 5). No significant modification of *intI1* abundance was observed between the two different microcosm conditions after 105 days, showing no impact of oil contamination on class 1 integron abundance in such environment. Thus, variations in the integron integrase pool observed in the same sediment (Fig. 4) were probably not due to the *intI1* abundance variation.

Gene cassettes in polluted environments

The estuarine samples of the Wright et al. (2008) study did not highlight the effect of contamination on cassette richness or diversity. However, Koenig et al. (2008), by an intensive gene cassette sequencing effort, demonstrated that overlap in gene cassette-type was greater between samples similar in terms of contamination than in term of localization. Gene cassette functions in relation to the selective pressure were also observed,

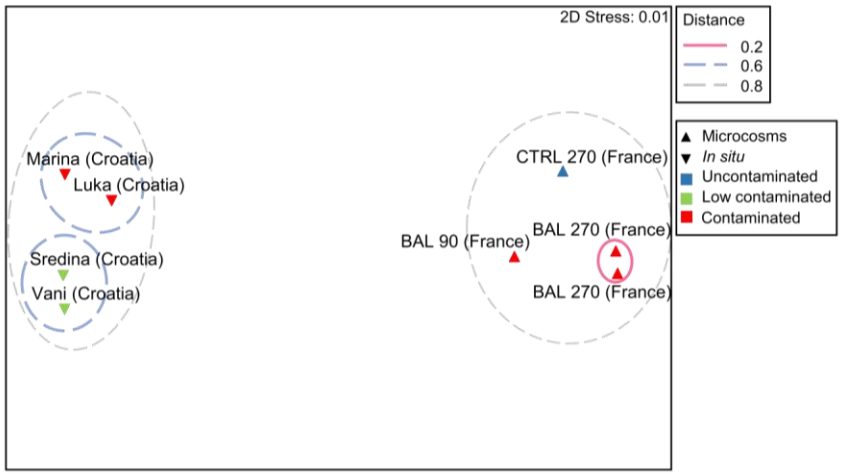


Fig. 4 MDS analysis of IntI pools from coastal sediments, based on a distance matrix calculated from Bray-Curtis dissimilarity. IntI libraries were obtained as described in Fig. 2. Sediments from West Istrian coast, Croatia: Luka and Marina sites contained the highest levels of PAHs (Total PAHs about 10 mg/kg d.m.) associated with the highest contents of toxic metals (Cd, Cu, Pb, Zn, and Hg); Vani and Sredina are less contaminated (Duran et al. 2015). Sediments from Atlantic Ocean, France, maintained in microcosms (Stauffert et al. 2013): *CTRL* control condition, *BAL* oil addition in a 2-cm layer over the sediment surface with Blend Arabian Light crude oil toped at 110 °C, *CTRL 270 CTRL* microcosm sediment sampled at day 270, *BAL 90* BAL microcosm sediments sampled at day 90, before observable oil degradation, *BAL 270* and *BAL 270'* sediments from two distinct BAL microcosms sampled at day 270 of incubation, when the oil degradation was obvious. To statistically compare IntI libraries, all amino acid sequences were aligned with BioEdit® software using ClustalW Multiple alignment tool (Needleman and Wunsch 1970). Distances were then calculated with Phylip protdist tool on Mobyle Pasteur Portal (http://mobyle.pasteur.fr/cgi-bin/portal.py#forms::protdist). DOTUR (Schloss and Handelsman 2005) was then used to obtain clusters, using a 98 % amino acid identity threshold, to construct a pivot table

particularly in the degradation of contaminants such as hydrocarbons (Koenig et al. 2009). In fact, in all marine gene cassette pools from metagenomes, coding sequences for environment-adaptive proteins were found (Elsaied and Maruyama 2011). These results are in accordance with other observations in soil (Nemergut et al. 2004; Michael and Andrew 2010) and suggest that selective pressures other than antibiotics could select specific gene cassettes in environmental integrons.

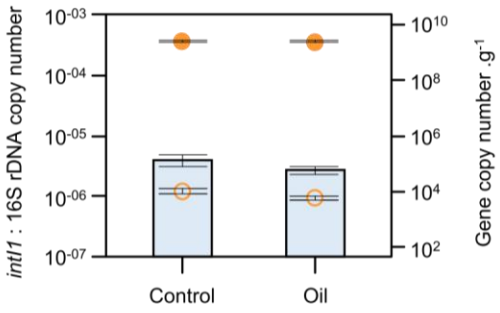


Fig. 5 Relative abundance of class 1 integron integrase gene (*intI1*) in microcosms of coastal marine sediments (Atlantic Ocean, Brittany, France). The class 1 integrase gene copy number (*open orange circles*) and 16S rRNA gene copy number (*full orange circles*) were assessed by quantitative PCR performed in duplicate on LightCycler 480 (Roche) with Light Cycler 480 SYBR green Master I kit (Roche) using qIntI1F and qIntI1R (Nemergut et al. 2004), and 338F and 518R (Lane 1991) primers, respectively. Standard curves constructed with plasmid-carrying cloned genes were used to quantify. Each value is mean of duplicate measures on five microcosm replicates. Microcosms were maintained during 105 days. Control: control microcosm without contamination; Oil: contaminated with 7g Oural oil per kilogram of sediment. No significant differences were observed between conditions (Tukey's multiple comparison, $p>0.1$)

Conclusion

In spite of few studies on the diversity of integrons present in environmental metagenomes, the best explored environments are the marine environments, both for the integrase genes and for gene cassettes. This diversity is important when considering the low diversity described in clinical environments and sequenced genomes. Cloning-sequencing approaches generate largely unsaturated libraries, and in depth knowledge of environmental integrons will require next-generation sequencing together with intensive sampling of diverse ecosystems.

Nevertheless, these first findings consolidated the hypothesis that environmental integrons might be involved in bacterial adaptive responses, especially in hydrocarbon and toxic metal contamination contexts. Thus, integrons responding to diverse environmental pressures must be characterized further, as well as the process by which they are acquired and spread in environmental bacterial communities. The marine environment is particularly vulnerable to pollutants; therefore, understanding of the adaptive mechanisms leading to the evolution of the bacterial communities in these environments is important to prospect means of protection.

Acknowledgments We acknowledge the support of the French program ANR/SEST (DHYVA project 06SEST09 and DECAPAGE project ANR450 CESA-2011-006 01), the French/Croatian bilateral program PHC-Cogito ISTRIA (Exploring and exploiting the bacterial diversity in the West Istria Sea: focus on marine Actinobacteria; project no. 25036XB), and the Croatian Ministry of Science, Education and Sports (project 098-0982913-2877). L.H. was supported by a doctoral grant from the Ministère de l'Enseignement Supérieur et de la Recherche (France) and J.A. from the Direction Générale de l'Armement (DGA) and the Aquitaine Regional Government Council (France). We thank all partners of the DHYVA project, the DECAPAGE project, the ISTRIA project, and the MELODY group for their useful discussions. We acknowledge the Regional Platform for Environmental Microbiology PREMICE supported by the Aquitaine Regional Government Council (France) and the urban community of Pau-Pyrénées (France). We acknowledge A. Betbédé-Gatas, C. Noël and C. Moya for technical assistance. We thank greatly M.R. Gillings and anonymous reviewer for very improving comments and suggestions on the first version of the manuscript.

References

Azevedo JSN, Araújo S, Oliveira CS, Correia A, Henriques I (2013) Analysis of antibiotic resistance in bacteria isolated from the surface microlayer and underlying water of an estuarine environment. Microb Drug Resist 19:64–72

Boucher Y, Labbate M, Koenig JE, Stokes HW (2007) Integrons: mobilizable platforms that promote genetic diversity in bacteria. Trends Microbiol 15:301–309

Byrne-Bailey KG, Gaze WH, Zhang L, Kay P, Boxall A, Hawkey PM, Wellington EMH (2011) Integron prevalence and diversity in manured soil. Appl Environ Microbiol 77:684–687

Cambray G, Guerout AM, Mazel D (2010) Integrons. Annu Rev Genet 44:141–166

Cambray G, Sanchez-Alberola N, Campoy S, Guerin E, Da Re S, González-Zorn B, Ploy MC, Barbé J, Mazel D, Erill I (2011) Prevalence of SOS-mediated control of integron integrase expression as an adaptive trait of chromosomal and mobile integrons. Mob DNA 2:6

Collis CM, Hall RM (1995) Expression of antibiotic resistance genes in the integrated cassettes of integrons. Antimicrob Agents Chemother 39:155–162

Collis CM, Kim MJ, Stokes HW, Hall RM (2002) Integron-encoded IntI integrases preferentially recognize the adjacent cognate attI site in recombination with a 59-be site. Mol Microbiol 46:1415–1427

Drudge CN, Elliott AVC, Plach JM, Ejim LJ, Wright GD, Droppo IG, Warrena LA (2012) Diversity of integron- and culture-associated antibiotic resistance genes in freshwater floc. Appl Environ Microbiol 78:4367–4372

Duran R, Bielen A, Paradžik T, Gassie C, Durmiši E, Cagnon C, Hamer B, Vujaklija D (2015) Exploring actinobacteria assemblages in coastal marine sediments under contrasted human influences in the West Istria Sea, Croatia. Environ Sci Pollut Res. doi:10.1007/s11356-015-4240-1

Durmiši E, Vujaklija D, Bielen A, Duran R, Hamer B (2013) Physicochemical and ecotoxicological evaluation of the Rovinj costal area sediments, Northern Adriatic, Croatia. Rapp Comm Int Mer Médit 40:330

Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 32:1792–1797

Elsaied H, Maruyama A (2011) Diversity and role of bacterial integron/gene cassette metagenome in extreme marine environments. In: de Bruijn FJ (ed) Handbook of molecular microbial ecology II: metagenomics in different habitats. Wiley-Blackwell, Weinheim, pp 309–318

Elsaied H, Stokes HW, Nakamura T, Kitamura K, Fuse H, Maruyama A (2007) Novel and diverse integron integrase genes and integron-like gene cassettes are prevalent in deep-sea hydrothermal vents. Environ Microbiol 9:2298–2312

Elsaied H, Stokes HW, Kitamura K, Kurusu Y, Kamagata Y, Maruyama A (2011) Marine integrons containing novel integrase genes, attachment sites, *attI*, and associated gene cassettes in polluted sediments from Suez and Tokyo Bays. ISME J 5:1162–1177

Elsaied H, Stokes HW, Yoshioka H, Mitani Y, Maruyama A (2014) Novel integrons and gene cassettes from a Cascadian submarine gas-hydrate-bearing core. FEMS Microbiol Ecol 87:343–356

Fluit AC, Schmitz FJ (1999) Class 1 integrons, gene cassettes, mobility, and epidemiology. Eur J Clin Microbiol Infect Dis 18:761–770

Gaze W, O'Neill C, Wellington E, Hawkey P (2008) Antibiotic resistance in the environment, with particular reference to MRSA. Adv Appl Microbiol 63:249–280

Gaze WH, Zhang L, Abdouslam NA, Hawkey PM, Calvo-Bado L, Royle J, Brown H, Davis S, Kay P, Boxall AB, Wellington EM (2011) Impacts of anthropogenic activity on the ecology of class 1 integrons and integron-associated genes in the environment. ISME J 5:1253–1261

Gillings M (2014) Integrons: past, present, and future. Microbiol Mol Biol Rev 78:257–277

Gillings MR, Holley MP, Stokes HW, Holmes AJ (2005) Integrons in Xanthomonas: a source of species genome diversity. Proc Natl Acad Sci U S A 102:4419–4424

Gillings M, Boucher Y, Labbate M, Holmes A, Krishnan S, Holley M, Stokes HW (2008a) The evolution of class 1 integrons and the rise of antibiotic resistance. J Bacteriol 190:5095–5100

Gillings MR, Krishnan S, Worden PJ, Hardwick SA (2008b) Recovery of diverse genes for class 1 integron-integrases from environmental DNA samples. FEMS Microbiol Lett 287:56–62

Gillings MR, Gaze WH, Pruden A, Smalla K, Tiedje JM, Zhu YG (2015) Using the class 1 integron-integrase gene as a proxy for anthropogenic pollution. ISME J 9:1269–1279

Hall RM (2012) Integrons and gene cassettes: hotspots of diversity in bacterial genomes. Ann N Y Acad Sci 1267:71–78

Hall RM, Collis CM (1995) Mobile gene cassettes and integrons: capture and spread of genes by site-specific recombination. Mol Microbiol 15:593–600

Hardwick SA, Stokes HW, Findlay S, Taylor M, Gillings MR (2008) Quantification of class 1 integron abundance in natural environments using real-time quantitative PCR. FEMS Microbiol Lett 278:207–212

Heinemann JA, Rosén H, Savill M, Burgos-Caraballo S, Toranzos GA (2006) Environment arrays: a possible approach for predicting changes in waterborne bacterial disease potential. Environ Sci Technol 40:7150–7156

Henriques IS, Fonseca F, Alves A, Saavedra MJ, Correia A (2006) Occurrence and diversity of integrons and β-lactamase genes among ampicillin-resistant isolates from estuarine waters. Res Microbiol 157:938–947

Holmes AJ, Gillings MR, Nield BS, Mabbutt BC, Nevalainen KMH, Stokes HW (2003) The gene cassette metagenome is a basic resource for bacterial genome evolution. Environ Microbiol 5:383–394

Jechalke S, Schreiter S, Olters B, Dealtry S, Heuer H, Smalla K (2014) Widespread dissemination of class1 integron components in soils and related ecosystems as revealed by cultivation-independent analysis. Front Microbiol 4:420

Jové T, Da Re S, Denis F, Mazel D, Ploy MC (2010) Inverse correlation between promoter strength and excision activity in class 1 integrons. PLoS Genet 6:e1000793

Koenig JE, Boucher Y, Charlebois RL, Nesbo C, Zhaxybayeva O, Bapteste E, Spencer M, Joss MJ, Stokes HW, Doolittle WF (2008) Integron-associated gene cassettes in Halifax Harbour: assessment of a mobile gene pool in marine sediments. Environ Microbiol 10: 1024–1038

Koenig JE, Sharp C, Dlutek M, Curtis B, Joss M, Boucher Y, Doolittle WF (2009) Integron gene cassettes and degradation of compounds associated with industrial waste: the case of the Sydney tar ponds. PLoS One 4:e5276

Lane DJ (1991) 16S/23S rRNA sequencing. Nucleic acid techniques in bacterial systematics. In: Stackebrandt E, Goodfellow M (eds) Sequencing techniques in bacterial systematics. John Wiley and Sons Ltd, Chichester, pp 115–174

Laroche E, Pawlak B, Berthe T, Skurnik D, Petit F (2009) Occurrence of antibiotic resistance and class1, 2 and 3 integrons in *Escherichia coli* isolated from a densely populated estuary (Seine, France). FEMS Microbiol Ecol 68:118–130

Liebert C, Hall RM, Summers AO (1999) Transposon Tn*21*, flagship of the floating genome. Microbiol Mol Biol Rev 63:507–522

Ling AL, Pace NR, Hernandez MT, LaPara TM (2013) Tetracycline resistance and class 1 integron genes associated with indoor and outdoor aerosols. Environ Sci Technol 47: 4046–4052

Lozupone CA, Knight R (2007) Global patterns in bacterial diversity. Proc Natl Acad Sci U S A 104:11436–11440

Maravić A, Skočibušić M, Šamanić I, Fredotović Z, Cvjetan S, Jutronić M, Puizina J (2013) *Aeromonas* spp. simultaneously harbouring $bla_{\text{CTX-M-15}}$, $bla_{\text{SHV-12}}$, $bla_{\text{PER-1}}$ and $bla_{\text{FOX-2}}$, in wild-growing Mediterranean mussel (*Mytilus galloprovincialis*) from Adriatic Sea, Croatia. Int J Food Microbiol 166:301–308

Mazel D (2006) Integrons: agents of bacterial evolution. Nat Rev Microbiol 4:608–620

Mazel D, Dychinco B, Webb VA, Davies J (1998) A distinctive class of integron in the *Vibrio cholerae* genome. Science 280:605–608

Michael CA, Andrew NR (2010) Co-assortment in integron-associated gene cassette assemblages in environmental DNA samples. BMC Genet 11:75

Michael CA, Gillings MR, Holmes AJ, Hughes L, Andrew NR, Holley MP, Stokes HW (2004) Mobile gene cassettes: a fundamental resource for bacterial evolution. Am Nat 164:1–12

Moura A, Henriques I, Ribeiro R, Correia A (2007) Prevalence and characterization of integrons from bacteria isolated from a slaughterhouse wastewater treatment plant. J Antimicrob Chemother 60: 1243–1250

Moura A, Henriques I, Smalla K, Correia A (2010) Wastewater bacterial communities bring together broad-host range plasmids, integrons and a wide diversity of uncharacterized gene cassettes. Res Microbiol 161:58–66

Moura A, Araújo S, Alves MS, Henriques I, Pereira A, Correia ACM (2014) The contribution of *Escherichia coli* from human and animal sources to the integron gene pool in coastal waters. Front Microbiol 5:419

Needleman SB, Wunsch CD (1970) A general method applicable to the search for similarities in the amino acid sequence of two proteins. J Mol Biol 48:443–453

Nemergut DR, Martin AP, Schmidt SK (2004) Integron diversity in heavy-metal-contaminated mine tailings and inferences about integron evolution. Appl Environ Microbiol 70:1160–1168

Nemergut DR, Robeson MS, Kysela RF, Martin AP, Schmidt SK, Knight R (2008) Insights and inferences about integron evolution from genomic data. BMC Genomics 9:261

Nield BS, Holmes AJ, Gillings MR, Recchia GD, Mabbutt BC, Nevalainen KM, Stokes HW (2001) Recovery of new integron classes from environmental DNA. FEMS Microbiol Lett 195:59–65

Ochman H, Lawrence JG, Groisman EA (2000) Lateral gene transfer and the nature of bacterial innovation. Nature 405:299–304

Pereira A, Santos A, Tacão M, Alves A, Henriques I, Correia A (2013) Genetic diversity and antimicrobial resistance of *Escherichia coli* from Tagus estuary (Portugal). Sci Total Environ 461–462:65–71

Ploy MC, Lambert T, Couty JP, Denis F (2000) Integrons: an antibiotic resistance gene capture and expression system. Clin Chem Lab Med 38:483–487

Rapa RA, Labbate M (2013) The function of integron-associated gene cassettes in *Vibrio* species: the tip of the iceberg. Front Microbiol 4: 385

Rodríguez-Minguela CM, Apajalahti JHA, Chai B, Cole JR, Tiedje JM (2009) Worldwide prevalence of class 2 integrases outside the clinical setting is associated with human impact. Appl Environ Microbiol 75:5100–5110

Rosewarne CP, Pettigrove V, Stokes HW, Parsons YM (2010) Class 1 integrons in benthic bacterial communities: abundance, association with Tn402-like transposition modules and evidence for coselection with heavy-metal resistance. FEMS Microbiol Ecol 72:35–46

Rosser SJ, Young HK (1999) Identification and characterization of class 1 integrons in bacteria from an aquatic environment. J Antimicrob Chemother 44:11–18

Schloss PD, Handelsman J (2005) Introducing DOTUR, a computer program for defining operational taxonomic units and estimating species richness. Appl Environ Microbiol 71:1501–1506

Smalla K, Sobecky PA (2002) The prevalence and diversity of mobile genetic elements in bacterial communities of different environmental habitats: insights gained from different methodological approaches. FEMS Microbiol Ecol 42:165–175

Stalder T, Barraud O, Casellas M, Dagot C, Ploy MC (2012) Integron involvement in environmental spread of antibiotic resistance. Front Microbiol 3:119

Stalder T, Barraud O, Jove T, Casellas M, Gaschet M, Dagot C, Ploy MC (2014) Quantitative and qualitative impact of hospital effluent on dissemination of the integron pool. ISME J 8:768–777

Stauffert M, Cravo-Laureau C, Jézéquel R, Barantal S, Cuny P, Gilbert F, Cagnon C, Militon C, Amouroux D, Mahdaoui F, Bouyssiere B, Stora G, Merlin FX, Duran R (2013) Impact of oil on bacterial community structure in bioturbated sediments. PLoS One 8:e65347

Stokes HW, Hall RM (1989) A novel family of potentially mobile DNA elements encoding site-specific gene-integration functions: integrons. Mol Microbiol 3:1669–1683

Stokes HW, Holmes AJ, Nield BS, Holley MP, Nevalainen KMH, Mabbutt BC, Gillings MR (2001) Gene cassette PCR: sequence-independent recovery of entire genes from environmental DNA. Appl Environ Microbiol 67:5240–5246

Stokes HW, Nesbo CL, Holley M, Bahl MI, Gillings MR, Boucher Y (2006) Class 1 integrons potentially predating the association with tn402-like transposition genes are present in a sediment microbial community. J Bacteriol 188:5722–5730

Uyaguari MI, Scott GI, Norman RS (2013) Abundance of class 1-3 integrons in South Carolina estuarine ecosystems under high and low levels of anthropogenic influence. Mar Pollut Bull 76:77–84

Wang C, Dang H, Ding Y (2008) Incidence of diverse integrons and β-lactamase genes in environmental *Enterobacteriaceae* isolates from Jiaozhou Bay, China. World J Microbiol Biotechnol 24:2889–2896

Wellington EMH, Boxall ABA, Cross P, Feil EJ, Gaze WH, Hawkey PM, Johnson-Rollings AS, Jones DL, Lee NM, Otten W, Thomas CM, Williams AP (2013) The role of the natural environment in the emergence of antibiotic resistance in Gram-negative bacteria. Lancet Infect Dis 13:155–165

Wright MS, Baker-Austin C, Lindell AH, Stepanauskas R, Stokes HW, McArthur JV (2008) Influence of industrial contamination on mobile genetic elements: class 1 integron abundance and gene cassette structure in aquatic bacterial communities. ISME J 2:417–428

3) Conclusion

L'étude de l'intégrase dans des milieux aquatiques d'eau douce et côtiers a permis de montrer une importante diversité de cette protéine dans les 2 types d'environnements. Un grand nombre de nouvelles séquences IntI a pu être découvert. L'analyse de plusieurs environnements côtiers n'a permis de montrer qu'une seule séquence commune à ces environnements. De plus, les analyses MDS montrent une séparation des échantillons côtiers de Croatie et de France. Ceci suggère que la diversité des intégrases serait principalement géo-spécifique. Nous avons pu également observer des différences du niveau de diversité entre les sites d'une même aire d'échantillonnage. Dans les sites les plus contaminés, la diversité est plus faible que cela soit en milieux côtiers ou en milieu d'eau douce. Ceci suggère que la pression exercée par les contaminations du milieu pourrait influencer cette diversité. Toutefois, malgré l'important travail de clonage-séquençage, les pourcentages de recouvrement sont bas pour certaines banques. Il serait nécessaire d'obtenir un nombre plus conséquent de séquences afin d'approfondir les analyses statistiques et de vérifier nos conclusions. Face à l'importante diversité des séquences de l'intégrase, la méthode Sanger n'est pas la plus appropriée en particulier pour étudier cette diversité dans des milieux faiblement impactés. Pour cela, les méthodes de séquençage haut débit pourraient être plus adaptées, avec toutefois l'inconvénient de ne pas pouvoir travailler sur des fragments longs couvrant l'ensemble des motifs du domaine actif.

L'importante diversité de l'intégrase rend difficile sa classification selon le seuil de 98% d'identité protéique. En effet, en suivant cette règle, le nombre de nouvelles séquences est très important. De plus, cette règle de classification ne semble pas être suivie par tous les auteurs. Elsaied et al par exemple, se base sur un seuil de 100% d'identité protéique (Elsaied *et al.,* 2011). Nous avons réalisé une analyse de la diversité des intégrases d'eau douce en se basant sur des seuils de 80% et de 60% d'identité protéique pour déterminer les OTUs. Nous avons pu démontrer ainsi qu'avec un seuil de 80% d'identité protéique nous avions les mêmes conclusions quant aux variations de la diversité des intégrases que celles faites en se basant sur un seuil de 98% d'identité protéique. Ce seuil pourrait donc être abaissé pour pouvoir déterminer des groupes d'intégrases environnementales.

Les cassettes de gène présentent une importante diversité dans les milieux environnementaux. La majorité code des fonctions inconnues. Toutefois, il semble y avoir une corrélation entre le pool de cassettes et les contaminations du milieu comme nous avons pu le montrer dans notre étude sur les environnements d'eau douce. Certains auteurs parlent même d'écotypes de cassettes de gène (Koenig *et al.,* 2008; Michael and Andrew 2010). Le séquençage d'une trentaine de cassettes de gène à partir du métagénome de 2 sédiments d'eau douce impactés a permis de mettre en évidence des correspondances avec des séquences obtenues à partir de milieux côtiers géographiquement éloignés. Il semble donc que les 2 types d'environnements côtiers et eau douce partagent des cassettes de gène communes. Il semble donc y avoir une distribution à l'échelle mondiale des cassettes de gène, ce qui ne serait pas le cas pour les intégrases dont la diversité semble géo-spécifique.

Pour conclure, la diversité des intégrons dans l'environnement est bien plus importante que dans les milieux cliniques que cela soit au niveau de la séquence des intégrases ou des cassettes de gène. De plus, la complexité des facteurs environnementaux pouvant influencer cette diversité rend difficile l'étude des intégrons dans les milieux environnementaux. Il est donc nécessaire de travailler sur des environnements contrôlés afin de tester différents paramètres : types de contaminants, quantité de polluants, paramètres physico-chimiques,… Dans le chapitre suivant, des expérimentations en microcosmes de sédiments d'eau douce et en mésocosmes et bioréacteurs de sédiments côtiers ont été réalisés. L'impact de différents types de contaminants (hydrocarbures, métaux, antibiotiques) a été étudié.

IV- Etude de la diversité des integrons dans des environnements contrôlés

1) Dynamique de la diversité des intégrons dans des sédiments d'eau douce

1.1. Introduction

Comme nous avons pu le voir dans le chapitre précédent, la diversité des intégrons dans les environnements naturels d'eau douce est très importante. De plus, les conditions environnementales comme la présence de contaminants semblent pouvoir influencer cette diversité sans que l'on sache dans quelle mesure. Afin de mieux appréhender l'impact de ces contaminations sur la diversité des intégrons, il est nécessaire d'étudier des systèmes simplifiés. L'approche par microcosmes permet de tester séparément l'effet de différents polluants et de suivre la dynamique des intégrons au cours du temps. Actuellement, une seule étude en microcosmes a été réalisée sur des sédiments d'eau douce et seul le gène *intI1* a été suivi (Wright *et al.,* 2008). Dans l'étude qui suit, des microcosmes de sédiments d'eau douce peu impactés (échantillon Lou déjà décrit dans l'étude du paragraphe III-1.1) ont été maintenus durant une année afin de suivre la diversité des intégrons. Trois substances représentant trois catégories de polluants, métalliques, organiques et antibiotiques, ont été choisies pour contaminer les microcosmes au début de l'expérimentation : l'arsenic, le benzo[a]pyrène et le triméthoprime. Des banques de séquences IntI ont été réalisées au cours du temps (à 0 jours, 30 jours, 90 jours, 181 jours et 358 jours) pour chaque condition et pour le contrôle. La diversité des cassettes de gène a également été analysée en se basant sur la taille des cassettes. Ces données ont été comparées à la structure des communautés évaluées par T-RFLP sur le gène de l'ARNr 16S. Enfin, une analyse phylogénétique des séquences IntI a également été réalisée.

1.2. Diversité des intégrases et des cassettes de gène comparées à la structure des communautés

1.2.a. La structure des communautés

La structure des communautés a été étudiée par T-RFLP du gène de l'ARNr 16S pour chaque condition au début des incubations, au bout de 6 mois (181 jours ; T181) et à la fin des incubations (358 jours ; T358) (Figure IV-1). On observe des modifications de

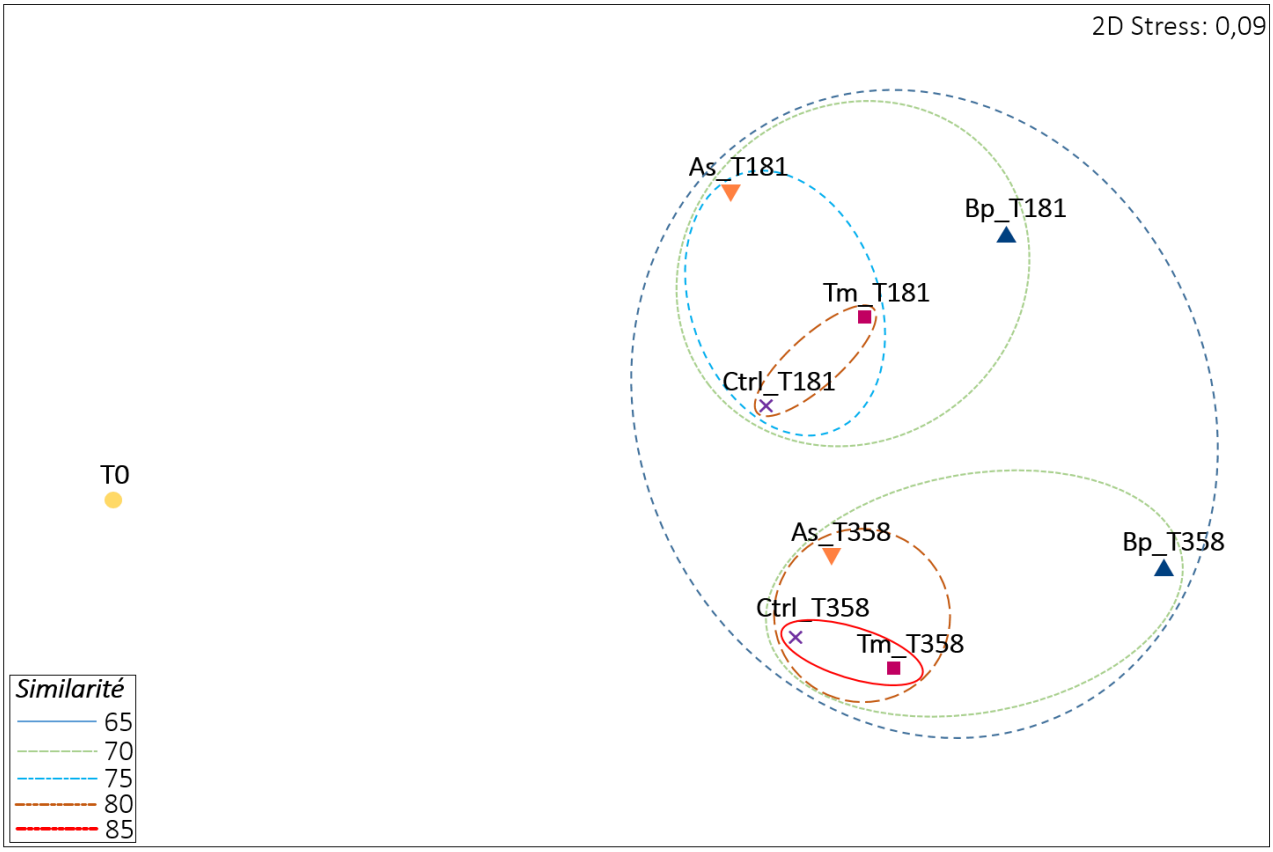


Figure IV-1 : Analyse MDS de la structure des communautés bactériennes au cours du temps selon les conditions. ● : Début de l'incubation (T0) ; × : Condition contrôle (Ctrl) ; ▼ : Condition ajout d'arsenic (As) ; ▲ : Condition ajout de benzo[a]pyrene (Bp) ; ■ : Condition ajout de trimétoprime (Tm). Analyse MDS réalisée sur les distances de Bray-Curtis calculées à partir des données de T-RFLP sur le gène de l'ARNr 16S. Les distances entre échantillons sont exprimées en pourcentage de similarité.

la structure des communautés similaires entre les 4 conditions au cours du temps. Ceci suggère un effet microcosme comme premier facteur influençant. Toutefois, en considérant les communautés à chaque temps, on peut constater quelques différences dans la structuration des communautés soumises à une contamination par du benzo[a]pyrène et dans une moindre mesure par de l'arsenic par rapport au contrôle et à la condition triméthoprime (Figure IV 1). Ces résultats montrent donc un faible effet de la contamination sur la structure des communautés bactériennes.

1.2.b. La diversité des intégrases

La diversité des séquences d'intégrase a été suivie au cours du temps par la réalisation de banques de séquences IntI. Dans un premier temps, des banques exploratoires ont été réalisées sur les premiers temps avant mon arrivée au laboratoire en séquençant 48 clones.

Des banques plus conséquentes (soit 96 clones) ont été réalisées sur le T0 et les microcosmes arrêtés à 358 jours (Tableau IV-1).

Tableau IV-1 : Tableau descriptif des banques de séquences IntI.

	Nombre de séquences	Nombre de singletons	Nombres d'OTUs	% recouvrement
T0	75	44	55	41
As_30	22	8	14	64
Bp_30	18	18	18	0.00
Ctrl_30	0	0	0	NA
Tm_30	17	11	14	35
As_90	99	52	68	47
Bp_90	23	21	22	9
Ctrl_90	0	0	0	NA
Tm_90	21	17	19	19
As_181	34	26	30	24
Bp_181	36	30	32	17
Ctrl_181	35	29	32	17
Tm_181	42	19	23	55
As_358	64	33	42	48
Bp_358	84	18	29	79
Ctrl_358	71	46	54	35
Tm_358	85	41	56	52

i. Effet d'une contamination par du benzo[a]pyrène

En condition benzo[a]pyrène, l'analyse MDS montre que le pool de séquences IntI se différencie des autres conditions à partir de 181 jours. Cette modification de la diversité des intégrases par rapport à la condition contrôle s'associe à une baisse de la diversité au bout de 358 jours comme le montre l'indice de Shannon qui est de 3,856 au début de l'incubation et de 2,811 à la fin dans cette condition (Figure IV-2).

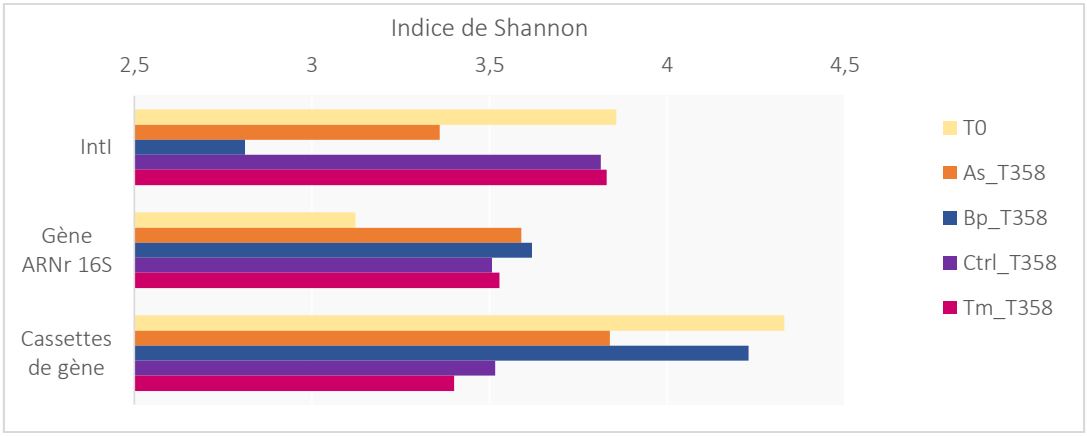


Figure IV-2 : Indice de Shannon calculé à partir des données de séquences IntI, de T-RFs du gène de l'ARNr 16S et sur les données de taille de cassettes de gène à T0 et pour les 4 conditions à T358 jours.

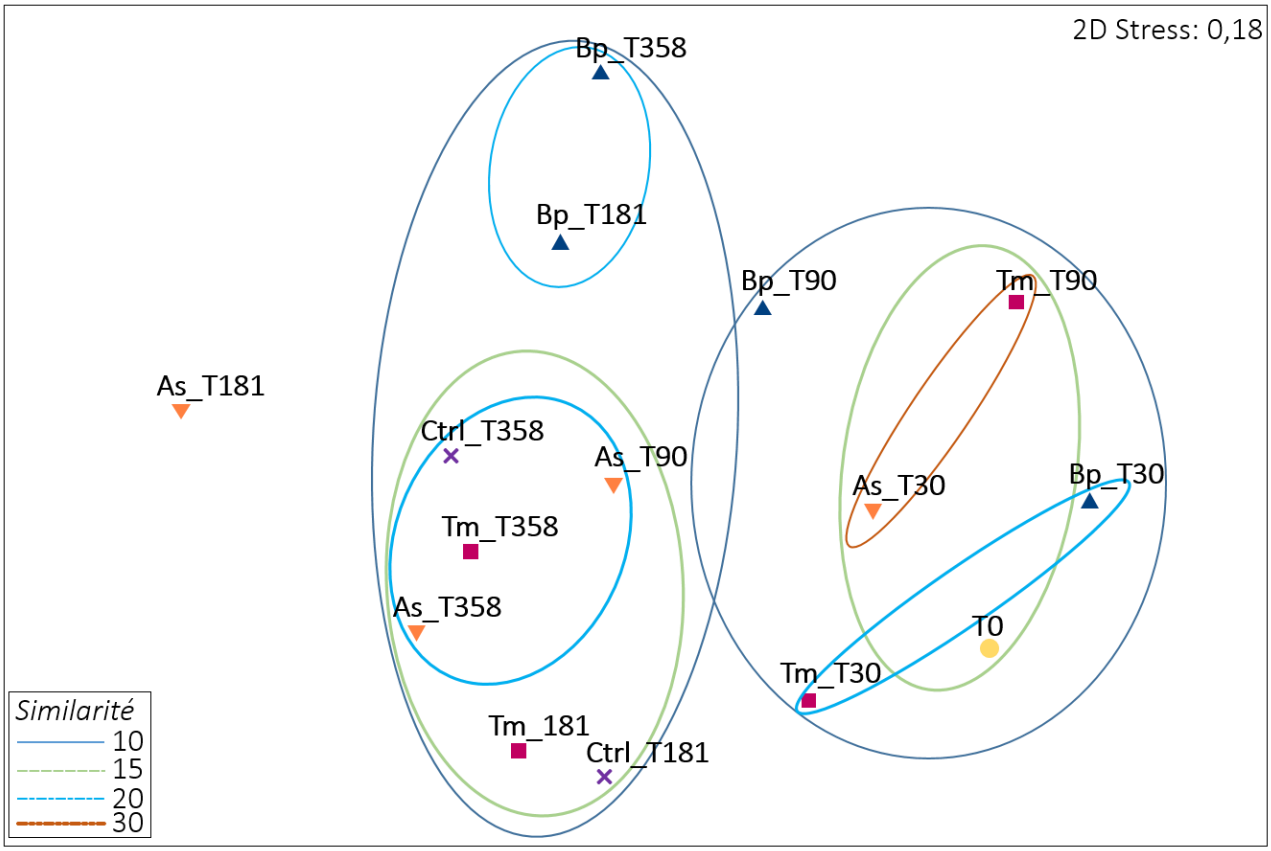


Figure IV-3 : Analyse MDS des banques de séquences d'intégrases au cours du temps selon les conditions. ● : Début de l'incubation (T0) ; × : Condition contrôle ; ▼ : Condition ajout de As ; ▲ : Condition ajout de benzo[a]pyrene ; ■ : Condition ajout de trimétoprime. Analyse MDS basée sur les distances de Bray-Curtis calculées à partir des séquences de IntI. Les distances entre les points sont exprimées en pourcentage de similarité.

Par ailleurs, trois OTUs (représentées par les séquences *BpHi358 2*, *BpHi358 9* et *BpHi358 11*) sont prépondérantes dans la banque à T358 jours. Ces OTUs représentent 13%, 15,5% et 17,9% respectivement de la banque alors qu'elles sont peu ou pas retrouvées dans les autres conditions et aux autres temps. La séquence *BpHi358 2* est proche à 73% et 72,5% d'identité

protéique des intégrases d'intégron des souches *Candidatus Accumulibacter sp*. SK-01 et SK-02 (Albertsen *et al.,* 2012). Ces souches ont été isolées à partir de boues de bioréacteurs expérimentaux pour l'étude du retraitement des eaux usées et l'élimination du phosphore. La seconde OTU représentée par *BpHi358 9* présente 61% d'identité protéique avec l'intégrase d'une souche de *Methylocaldum szegediense,* une méthanotrophe thermo-tolérante et thermophile que l'on retrouve au niveau d'un effluent de source d'eau chaude, et 60% avec une séquence provenant des sédiments côtiers pollués par du pétrole obtenue par Lionel Huang (Huang, 2010 ; paragraphe III-2). La 3e séquence *BpHi358 11* est également proche de l'intégrase d'intégron de *Candidatus Accumulibacter sp*. SK-02 à 66%. Elle présente aussi 65% d'identité avec une intégrase séquencée à partir d'un biofilm d'eau douce (Gillings, Holley and Stokes 2009). Toutes les séquences proches de ces OTUs proviennent de milieux d'eau douce, à l'exception de la séquence isolée à partir des mésocosmes de sédiments côtiers. Or, l'Aber-Benoît où a été prélevé ce sédiment reçoit des affluents d'eau douce. On peut donc conclure que ces OTUs qui représentent presque la moitié de la banque correspondent à des intégrases d'environnements d'eau douce.

ii. Effet d'une contamination par de l'arsenic

Une baisse de la diversité importante est aussi observée pour la condition arsenic à la fin des incubations comme le montre les indices de Shannon (Figure IV-2). Une OTU représentée par *AsHi358 1* est abondante à 18,75% de la banque réalisée pour la condition arsenic à T358 jours. Elle n'est retrouvée dans aucune autre banque. Cette OTU est proche à 69% d'une séquence d'intégrase appartenant à *Nitrosococcus watsonii*, une bactérie sulfureuse pourpre marine (Campbell *et al.,* 2011) et à 68% d'une des séquences isolées à partir des sédiments des mésocosmes côtiers (Huang, 2010 ; paragraphe III-2). Les 2 premières séquences les plus proches correspondent donc à des séquences côtières mais des homologies sont retrouvées également avec des séquences obtenues à partir d'environnements d'eau douce : 68,5% avec l'intégrase de *Candidatus Accumulibacter sp.* SK-02 et 68,1% avec l'intégrase de *Candidatus Accumulibacter phosphatis*. Les séquences isolées à partir de milieux côtiers pouvant provenir d'environnements d'eau douce, les résultats correspondent vraisemblablement à des intégrases d'environnements d'eau douce.

iii. Effet d'une contamination par triméthoprime

En condition triméthoprime, le pool des intégrases présente 15% de similarité avec celui de la condition contrôle à T181 jours puis 20% à 358 jours (Figure IV-3). Le niveau de diversité pour cette condition à la fin de l'expérimentation est élevé et proche de celui de la condition contrôle comme en atteste l'indice de Shannon (Figure IV-2). Toutefois à T181 jours, le pourcentage de recouvrement de la banque est de 55%, ce qui est important pour 42 séquences étudiées. Ceci est dû à la présence en abondance d'une OTU particulière, dont la séquence représentative est *TmHi181 6* (16 séquences sur les 42 séquencées au total soit 38%). Cette OTU particulière est retrouvée dans la banque réalisée à T358 jours mais en plus faible abondance : 8 séquences sur 85 (soit 9%). Cette OTU est également retrouvée dans la condition arsenic à T90 et T358 (respectivement 6,2% et 6%) et dans les autres conditions à T181 et T358 jours mais à raison de 1 ou 2 séquences. Cette OTU pourrait correspondre à une intégrase d'intégron retrouvée dans le sol. En effet, elle est proche de IntI6 et IntI8 à 67% et 66,7% d'identité protéique respectivement et de l'intégrase d'intégron d'une *Methylocaldum szegediense* avec 66% d'identité protéique (Bodrossy *et al.,* 1997). IntI6 et IntI8 ont été isolées à partir de sol (Nield *et al.,* 2001). Cette OTU étant plus abondante à 181 jours qu'à 358 jours, il serait donc intéressant d'approfondir l'analyse de la diversité des intégrases à des temps inférieurs à 181 jours afin d'étudier la dynamique de l'abondance de cette OTU au cours de 6 premiers mois de l'incubation.

iv. Suivi du gène *intI1*

Aucune séquence IntI1 n'a pu être identifiée dans les banques de même que les autres intégrases d'intégrons dits cliniques. Ceci est en accord avec nos précédentes études qui ont été réalisées au laboratoire utilisant le couple d'amorces ICC32 et ICC36. Le gène *intI1* a donc été ciblé spécifiquement par PCR quantitative et il est détecté dans le sédiment prélevé (sédiment Lou, paragraphe III-1.1 ; Erreur ! Source du renvoi introuvable.). Toutefois, il n'a pas été détecté dans les microcosmes à T0. Il semble que les bactéries porteuses de ce gène aient été perdues après le tamisage et la dilution du sédiment. Dans leur étude, Wright et ses collaborateurs n'ont pas rencontré ce problème (Wright *et al.,* 2008). Lorsque l'on compare le ratio de gènes *intI1*/ARNr 16S pour les sédiments prélevés avant la mise en microcosmes,

on peut voir que le sédiment qui a servi dans leur étude présente une abondance relative initiale de gène *intI1* bien plus élevée (entre 10^{-2} et 10^{-1} (Wright *et al.,* 2008), contre 10^{-5} dans le sédiment Lou, **Erreur ! Source du renvoi introuvable.**). Ceci est cohérent avec le fait que notre échantillon est bien faiblement impacté par les activités anthropiques.

1.2.c. La diversité des cassettes de gène

La diversité des cassettes de gène a été suivie au cours du temps en se basant sur l'analyse de la taille des cassettes. Un effet microcosme est observable sur la diversité des cassettes de gène comme pour la structure des communautés bactériennes et le pool d'intégrases (Figure IV-4).

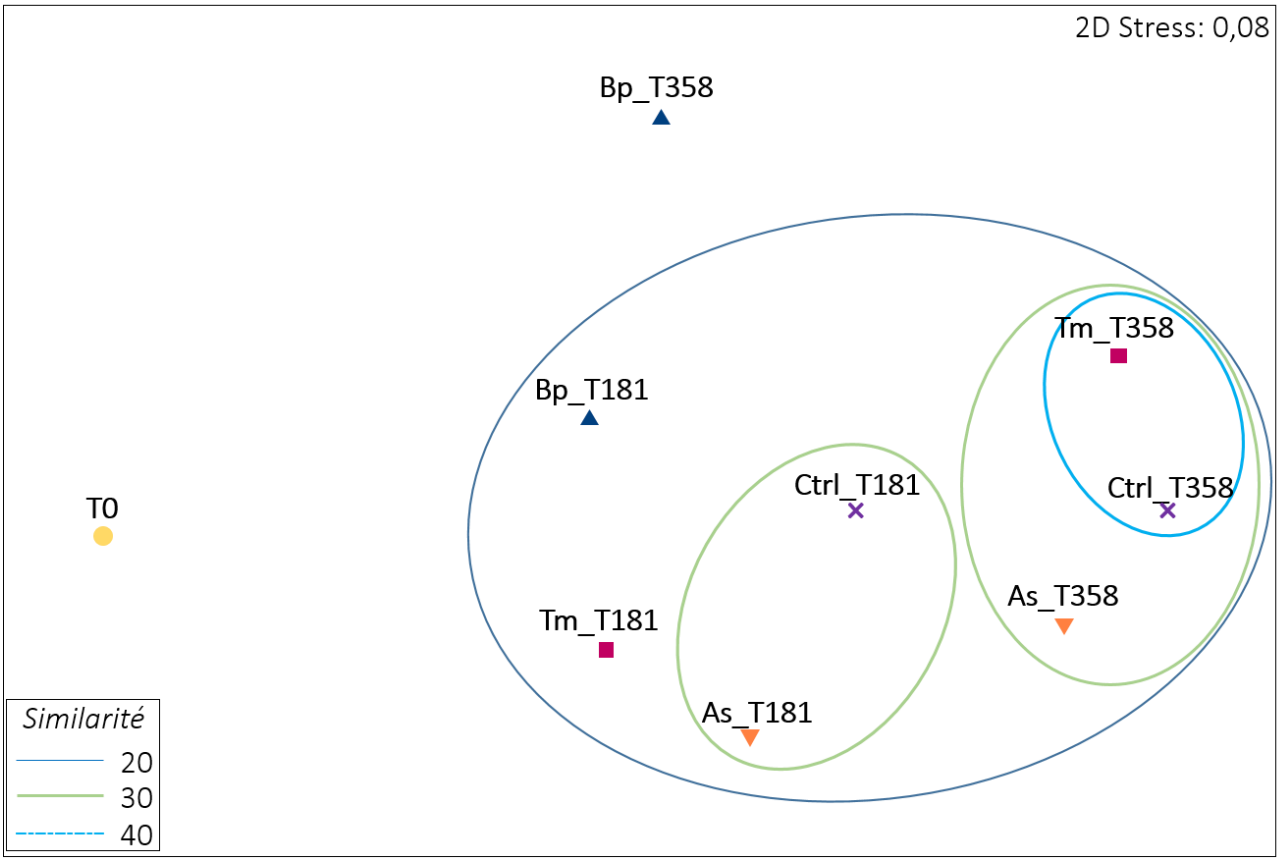


Figure IV-4 : Comparaison du pool des cassettes de gène au cours du temps selon les conditions en se basant sur l'analyse de la taille des cassettes. ● : Début de l'incubation (T0) ; × : Condition contrôle ; ▼ : Condition ajout de As ; ▲ : Condition ajout de benzo[a]pyrene ; ■ : Condition ajout de trimétoprime. Analyse MDS basée sur les distances de Bray-Curtis à partir des données de taille des cassettes de gène. Les distances entre les différents points sont exprimées en pourcentage de similarité.

Toutefois, on peut observer quelques variations en fonction des conditions : au niveau de la diversité en cassettes de gène, l'indice de Shannon chute de 4,33 au début de l'incubation à 3,5 et 3,4 respectivement pour les conditions contrôle et triméthoprime au bout de 358 jours,

ce qui suggère un effet microcosme. Cependant, pour les conditions arsenic et en particulier benzo[a]pyrène, la chute de la diversité est moindre, suggérant que ces contaminations permettraient le maintien d'un niveau de diversité en cassettes de gène important (Figure IV-2). Par ailleurs, le pool de cassette de gène dans les microcosmes soumis à une contamination par de l'arsenic ne présente que 30% de similarité avec celui de la condition contrôle, alors que celui de la condition triméthoprime est similaire à 20% avec la condition contrôle à 181 jours pour ensuite présenter 40% de similarité à 358 jours. Enfin, le pool de cassettes de gène en condition benzo[a]pyrène présente 20% de similarité avec les autres conditions à T181 jours puis se différencie nettement (Figure IV-4).

1.2.d. Diversité des intégrons dans les microcosmes d'eau douce

Ces résultats pris dans leur ensemble montrent 3 dynamiques différentes de la diversité des intégrons face aux 3 contaminants. Face à une contamination par du benzo[a]pyrène, on observe une baisse de la diversité en intégrase la plus importante à 358 jours, avec la sélection de 3 séquences particulières faiblement retrouvées dans les autres conditions. De plus, le pool de séquences IntI est le plus éloigné de la condition contrôle. Il en est de même pour la diversité en cassettes de gène. Toutefois, la même dynamique est observée pour la structure de la communauté bactérienne. Ces modifications de la diversité en intégrons pourraient donc être expliquées par la sélection de bactéries. Ceci est cohérent avec de précédentes études qui ont montré qu'il fallait 6 mois à une communauté bactérienne pour s'adapter et dégrader des hydrocarbures (Stauffert *et al.,* 2013; Terrisse 2014). Le benzo[a]pyrène est un composé organique persistant dont la demi-vie est de 57 à 530 jours (Coover and Sims 1987). Donc au bout de 358 jours, le benzo[a]pyrène est encore présent dans les microcosmes et les communautés bactériennes y sont adaptées.

Avec une contamination par de l'arsenic, on observe une baisse importante de la diversité en séquences IntI avec l'augmentation de l'abondance d'une séquence particulière au cours du temps. Tout comme la diversité de IntI, le pool de cassettes de gène est également éloigné de la condition contrôle. Le sédiment prélevé pour cette expérimentation présente naturellement un taux d'arsenic de 25 $\mu g.g^{-1}$ de sédiment, ce qui est élevé (sédiment Lou, paragraphe III-1.1 ; Erreur ! Source du renvoi introuvable.). Or, même si la concentration en

arsenic ajoutée de l'ordre de 200 mg.L^{-1} est conséquente, on peut voir que la diversité des intégrons diverge. Toutefois, cette divergence est moins notable que pour la condition benzo[a]pyrène. Il est donc possible que cette communauté soit déjà adaptée à la présence d'arsenic.

Dans le cas d'une contamination par le triméthoprime, il semblerait que la diversité en intégrase baisse à 181 jours puis remonte à 358 jours pour revenir à une diversité équivalente à celle de la condition contrôle. Ceci s'accompagne de l'augmentation à 181 jours puis la diminution à 358 jours de l'abondance d'une séquence particulière, retrouvée essentiellement dans cette condition. Par ailleurs, le pool de cassettes de gène semble s'éloigner de celui que l'on retrouve dans la condition contrôle à 181 jours, alors qu'à la fin de l'incubation elle présente plus de similarité avec la condition contrôle. La demi-vie du triméthoprime est comprise entre 20 et 100 jours dans l'environnement (Deblonde *et al*., 2012). Après 181 jours, il ne resterait donc que 25% au maximum de la concentration en triméthoprime initiale. La baisse de la pression de sélection exercée pourrait donc expliquer ce retour à la normale de la diversité des intégrons qui pourrait s'apparenter à de la résilience. Il serait donc important de caractériser cette OTU majoritaire, ne correspondant pas à une intégrase de classe 1 et qui pourrait avoir un rôle potentiellement adaptatif face à une contamination par du triméthoprime.

1.3. Analyse phylogénétique des séquences d'intégrases d'intégron

Une analyse phylogénétique a été réalisée sur 242 séquences : 119 séquences représentatives de chaque OTUs (après retrait des singletons) et 123 séquences correspondantes aux 5 premières séquences présentant le plus d'homologie après un blastP réalisé sur les 119 séquences représentatives. La Figure IV-5 présente cette analyse sous la forme d'un arbre construit à partir d'une seule séquence par cluster par obtenu.

Cette analyse montre que la majorité des séquences représentatives des OTUs sont éloignées des séquences déjà connues et donc représentent de nouvelles séquences (Figure IV-5). Toutefois, quelques séquences montrent de fortes homologies avec des intégrases déjà séquencées. C'est le cas par exemple pour l'OTU représentée par la séquence *Bp T181 10* (Figure IV-5) qui présente 92% d'identité protéique avec une séquence obtenue du métagénome d'un biofilm d'eau douce (ACS73587.1 ; Gillings *et al.*, 2008) et 84% avec l'intégrase d'une souche de *Thiobacillus denitrificans,* une bactérie ubiquiste. De même, les OTUs représentées par les séquences *Ctrl T181 10* et *Tm T181 16* (Figure IV-5n'apparaissant pas sur l'arbre) présentent environ 90% d'identité protéique avec 3 séquences obtenues à partir d'échantillons de sol (CEF48309.1, CEF49493.1 et CEF48363.1). Ces correspondances sont cohérentes avec l'origine du sédiment à partir duquel les microcosmes ont été mis en place. Par ailleurs, les intégrons étant soumis au transfert horizontal, il est difficile de conclure qu'une intégrase appartient à une espèce en se basant seulement sur le pourcentage d'identité des séquences. Par ailleurs, quelques séquences d'origines marine ou côtière présentent plus ou moins de correspondance avec des séquences IntI issues de cette analyse tout comme des séquences provenant de sol ou d'environnements d'eau douce, ce qui ne permet pas d'observer de regroupements dépendants de l'environnement.

Enfin, l'analyse phylogénétique réalisée sur l'ensemble des séquences des banques n'a pas permis de montrer de regroupements en fonction de la condition de contamination (données non présentées).

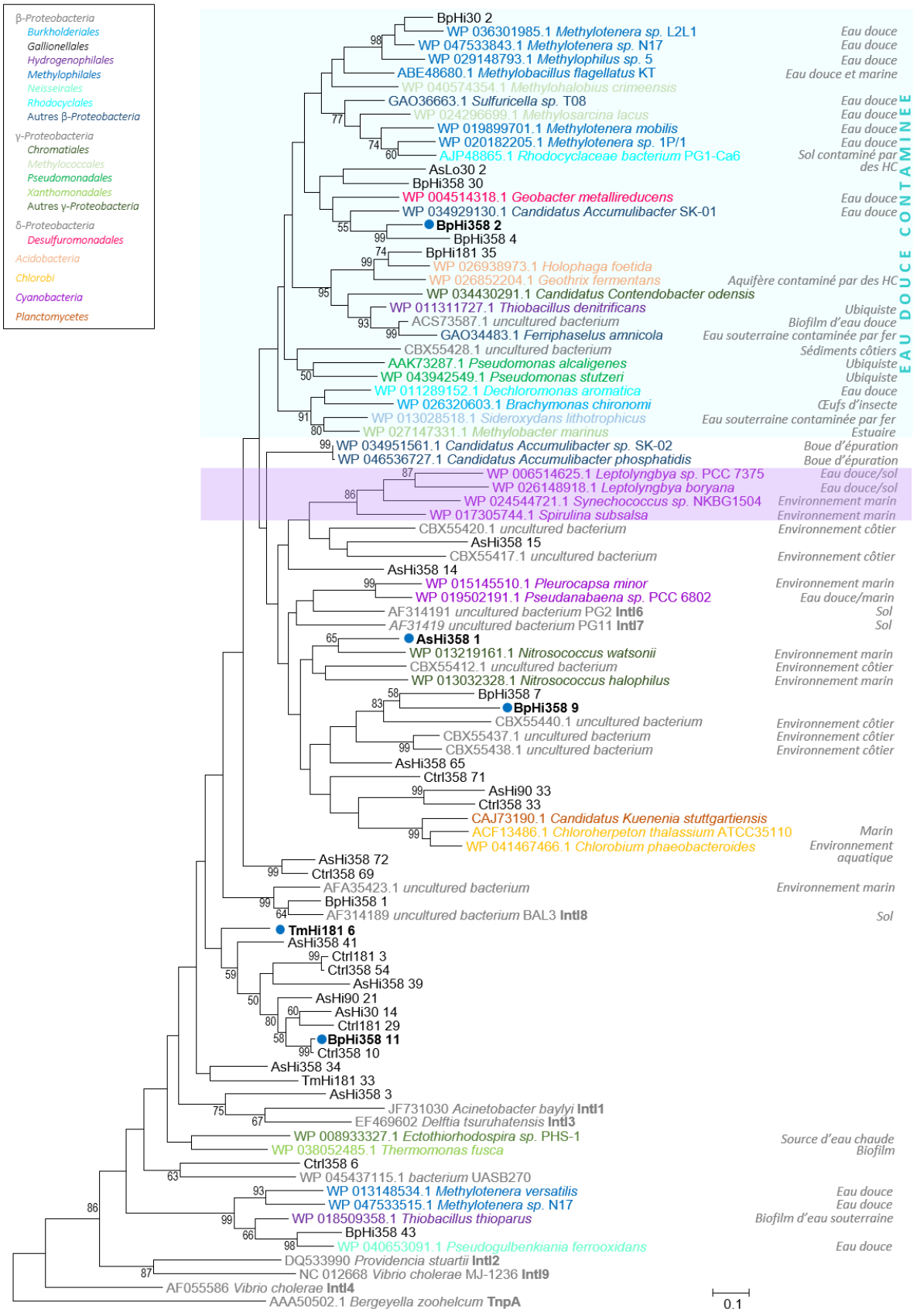


Figure IV-5 : Arbre phylogénétique des OTUs d'intégrase d'intégron obtenues représentant plus de 0.6% du total des séquences (en noir) et de séquences connues (en couleurs).L'arbre a été réalisé selon la méthode du maximum de vraisemblance sur les régions conservées des séquences, après un traitement par Gblocks sur l'alignement des séquences. Les valeurs de bootstrap supérieures à 50%, calculées sur 500 itérations sont indiquées. ● : Séquences représentatives des OTUs prépondérantes dans les banques à T358. La séquence de la protéine TnpA de *Bergeyella zoohelcum* a été utilisée comme racine de l'arbre.

1.4. Conclusion

Cette étude de la diversité des intégrons dans des communautés d'eau douce en conditions contrôlées a permis de mettre en évidence une diversité en séquences d'intégrase toujours très importante et de découvrir encore de nouvelles séquences qui viendront enrichir les bases de données. Les analyses phylogénétiques n'ont pas permis de montrer de regroupement des séquences d'intégrases en fonction de la contamination. Toutefois, une baisse de la diversité des intégrases a pu être observée en condition de contamination à la fin de l'expérimentation. Ceci est particulièrement vrai pour les conditions benzo[a]pyrène et arsenic, dont les diversités vont dans un même temps chuter et se différencier des autres conditions. Ceci confirme donc l'hypothèse émise dans le précédent chapitre que la diversité des intégrases pourrait être influencée par la contamination du milieu. Nous avions également conclu que le niveau de contamination pourrait également affecter cette diversité. Afin de confirmer cette hypothèse des microcosmes ont également été réalisés avec des concentrations plus faibles en benzo[a]pyrène, arsenic et trimétoprime. Toutefois, les données de cette analyse n'ont pas encore été exploitées.

De plus, cette étude a permis de montrer la sélection de séquences qui deviennent prépondérantes au cours du temps. Il est donc important de mettre au point une méthode de quantification de ces séquences par qPCR pour pouvoir étudier leur dynamique au cours du temps et s'assurer que leur augmentation est bien spécifique à un seul contaminant.

L'effet de la contamination sur la diversité des cassettes est principalement observable pour la condition benzo[a]pyrène dans laquelle la diversité en cassettes se différencie des autres conditions après 181 jours d'incubation. Le sédiment utilisé pour cette expérimentation présente une concentration basale en arsenic élevée due au fond géologique de la région montagneuse. Il est donc possible que dans le pool de cassettes, des gènes pourraient permettre une adaptation à cette pression environnementale. Enfin, l'analyse du pool de cassettes de gène en condition triméthoprime semble montrer une implication des intégrons jusqu'à 181 jours. Cette observation doit toutefois être confirmée par une étude plus approfondie de la diversité des cassettes à des temps antérieurs à 181 jours d'incubation.

Ainsi, il semblerait que comme nous l'avions suggéré dans le chapitre précédent, la diversité des cassettes de gène serait influencée par la nature des contaminations du milieu.

2) Dynamique de la diversité des intégrons dans des sédiments côtiers

2.1. Diversité des intégrases

Dans le cadre du projet Dhyva (projet ANR/SEST 2007-2010) des mésocosmes ont été réalisés avec des sédiments côtiers prélevés à l'Aber Benoit (Finistère, France) afin de suivre les effets d'une pollution pétrolière. Des banques de séquences d'intégrases d'intégrons ont été réalisées avant cette thèse (Huang 2010) à différents temps et ont permis de montrer une baisse de la diversité de ces séquences dans les conditions d'une pollution pétrolière (paragraphe III-2) ; Figure IV-6).

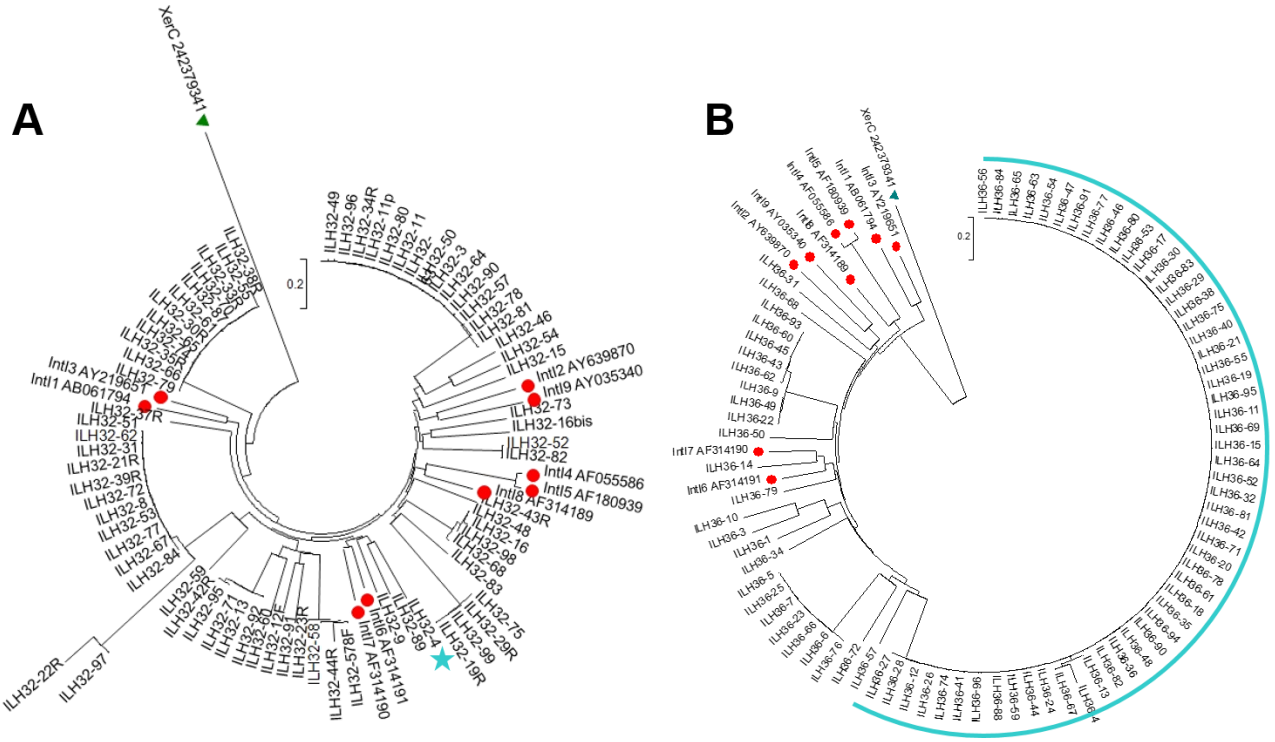


Figure IV-6 : Arbres phylogénétiques réalisés à partir de l'alignement des séquences IntI obtenues à partir des mésocosmes contaminés par du pétrole après 3 mois d'incubation (**A**) et 9 mois (**B**) (modifiée à partir de Huang et Cagnon, données personnelles). ▼ : la séquence XerC (de la famille des-tyrosine recombinases) a été utilisée pour enraciner les arbres. ● : Séquences IntI références. ★ et ↻ : Séquences IntIOPS.

Par ailleurs, l'abondance d'une séquence particulière augmente au cours du temps et devient prédominante dans la banque réalisée à 9 mois en condition de contamination par du pétrole

(Figure IV-6). Cette séquence a été appelée *intIOPS* pour *intI Oil Polluted Sediment*. L'abondance de ce gène au sein des communautés bactériennes à différents temps de prélèvements a été déterminée par qPCR et a permis de montrer une augmentation significative de l'abondance du gène intIOPS après trois mois de pollution (Figure IV-7). La prépondérance du gène intIOPS dans les sédiments étudiés a conduit Lionel Huang lors de sa thèse à émettre l'hypothèse que le (ou les) intégron(s) porteur(s) de intIOPS pourrai(en)t jouer un rôle important dans l'adaptation des communautés à la contamination au pétrole (Huang, 2010).

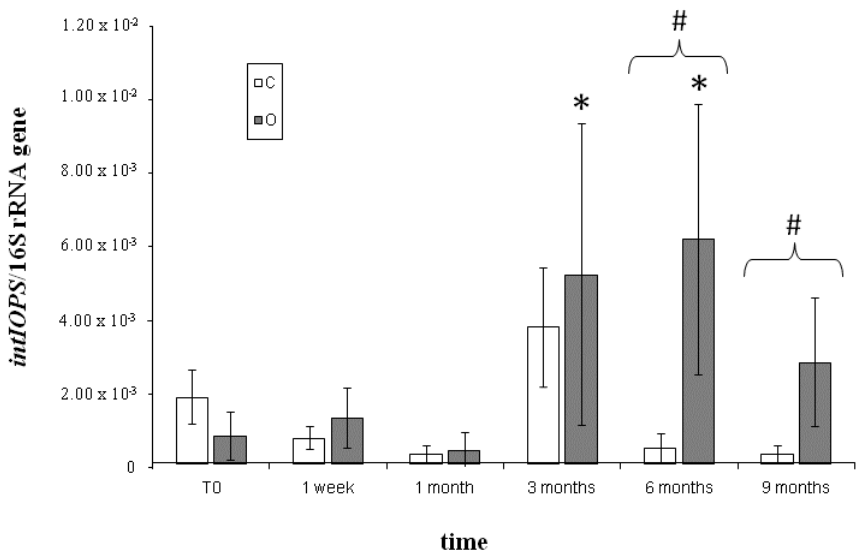


Figure IV-7 : Abondance relative de *intIOPS* dans les mésocosmes au cours du temps (d'après Huang, 2010). Le ratio *intIOPS*/gène de l'ARNr 16S a été obtenu par qPCR. Chaque valeur correspond à une moyenne de quadruplicatas d'échantillons mesurés en duplicatas. **O** : condition de contamination pas du pétrole. **C** : condition contrôle sans contamination. * : valeurs significativement différentes avec le T0 de la même condition ($p > 0.05$). # : représente une différence significative entre les 2 conditions ($p > 0.05$).

Trois ans après cette première étude, le projet ANR-CESA DECAPAGE (2011-2015) a poursuivi les travaux du projet Dhyva. Le but de ce projet est d'étudier les capacités de biodégradation des hydrocarbures par des communautés hydrocarbonoclastes en conditions oscillantes d'oxygénation. En effet, les environnements côtiers sont des écosystèmes soumis aux marées et au remaniement du sédiment par la macrofaune. De ce fait, des oscillations entre oxie et anoxie se créent et vont engendrer des variations impactant les communautés bactériennes au gré des marées et de l'action des organismes bioturbateurs. Dans ce cadre, une partie des travaux de cette présente thèse ont porté sur l'étude de la diversité des intégrons dans ces communautés.

Des sédiments côtiers prélevés à l'Aber-Benoît ont été contaminés par du pétrole, et maintenus en mésocosmes. Différents traitements ont été appliqués afin de créer des conditions d'oxygénation différentes du sédiment (sans remaniement du sédiment, avec hersage, ou avec bio-augmentation par ajout de vers bioturbateurs *Hediste diversicolor*). Les mésocosmes ont été maintenus jusqu'à observation d'une dégradation des hydrocarbures. Les communautés bactériennes hydrocarbonoclastes obtenues (communautés *Pétrole*, *Pétrole + Hersage*, *Pétrole + Hediste*) ont alors été étudiées par incubations en bioréacteurs dans différentes conditions d'oxygénation (anoxie, oxie ou oscillations oxie/anoxie) en présence de pétrole durant 15 jours (Figure II-5).

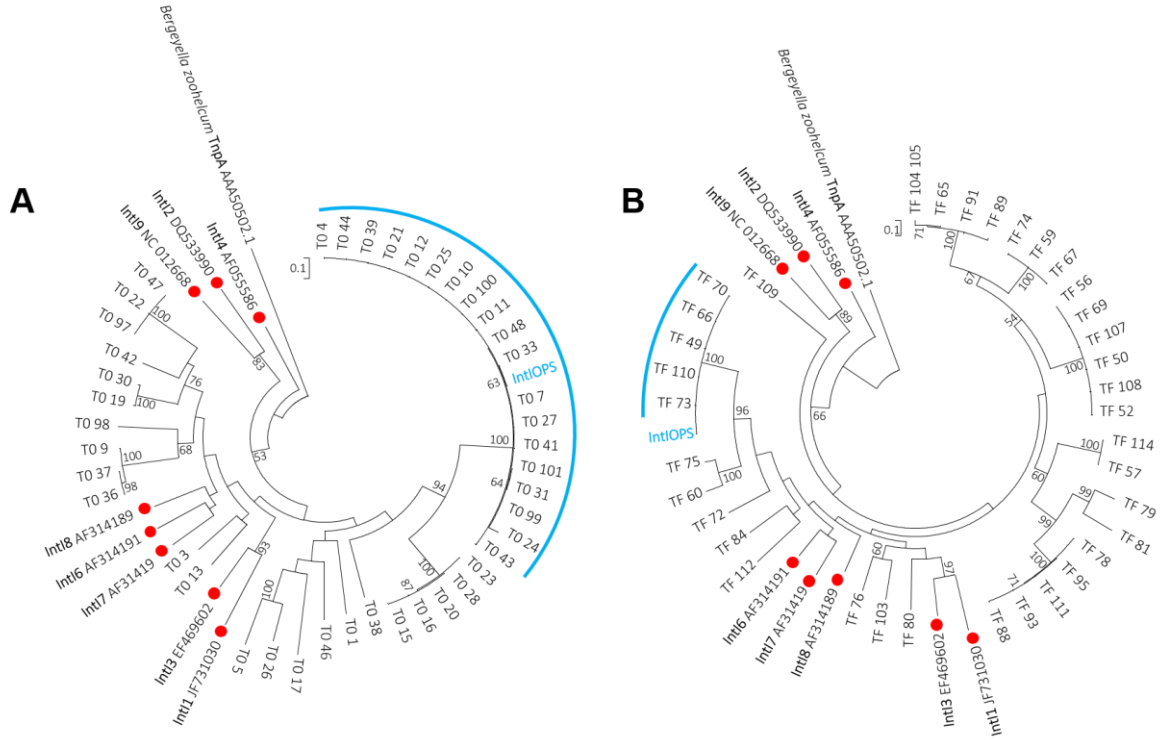


Figure IV-8 : Arbres phylogénétiques réalisés à partir de l'alignement des séquences IntI obtenues à partir des suspensions des bioréacteurs de la communauté *Pétrole* au début de l'expérimentation (A) et à la fin (B) en condition anoxie permanente en bioréacteurs. ● : Séquences références. La séquence de IntIOPS de la base de données a été intégrée aux arbres (FR718193.1) et la séquence de la protéine TnpA de *Bergeyella zoohelcum* a été utilisée comme racine de l'arbre.

Afin d'évaluer la diversité en séquences d'intégrase dans les bioréacteurs, des banques exploratoires de séquences IntI ont été réalisées à partir des suspensions de sédiment de la communauté *Pétrole* au début et à la fin de l'expérimentation en bioréacteurs en condition anoxie permanente, condition dans laquelle on s'attendait à une diversité la moins importante des intégrases. Le gène *intIOPS* a été retrouvé, représentant 45% et 15% des banques réalisées respectivement au début et à la fin des incubations (Figure IV-8). Ainsi 3 ans après, la séquence découverte par Lionel Huang était de nouveau présente dans les échantillons de cette nouvelle étude. Ceci confortait l'hypothèse émise

lors de la thèse de Lionel Huang quant au rôle potentiel que pourrait jouer l'intégron porteur de la séquence *d'intIOPS* dans l'adaptation des communautés de ce sédiment face à une contamination pétrolière.

2.2. Suivi de l'abondance du gène *intIOPS*

La quantification du gène *intIOPS* a d'abord été réalisée dans les mésocosmes au moment du prélèvement du sédiment pour la mise en place des bioréacteurs (Figure IV-9). Les résultats montrent que son abondance est plus importante dans la condition *Pétrole* par rapport au contrôle. L'abondance du gène *intIOPS* a été également suivie dans les bioréacteurs au début et à la fin de l'expérimentation et montre une baisse significative de cette abondance au cours du temps dans les conditions d'oscillations et d'oxie permanente (Figure IV-10 A). De plus, la normalisation de l'abondance relative du gène *intIOPS* à la fin de l'expérimentation par son abondance relative au début de l'expérimentation montre une réponse similaire de IntIOPS dans les 3 communautés (Figure IV-10 B). Il semble donc que d'une part sa présence est défavorisée dans les conditions d'oxygénation du sédiment en mésocosmes et dans les bioréacteurs et d'autre part que les changements de l'abondance du gène *intIOPS* sont indépendants de la communauté étudiée. Il est donc possible que l'abondance plus importante du gène *intIOPS* soit une réponse spécifique aux conditions de contamination par du pétrole en anoxie.

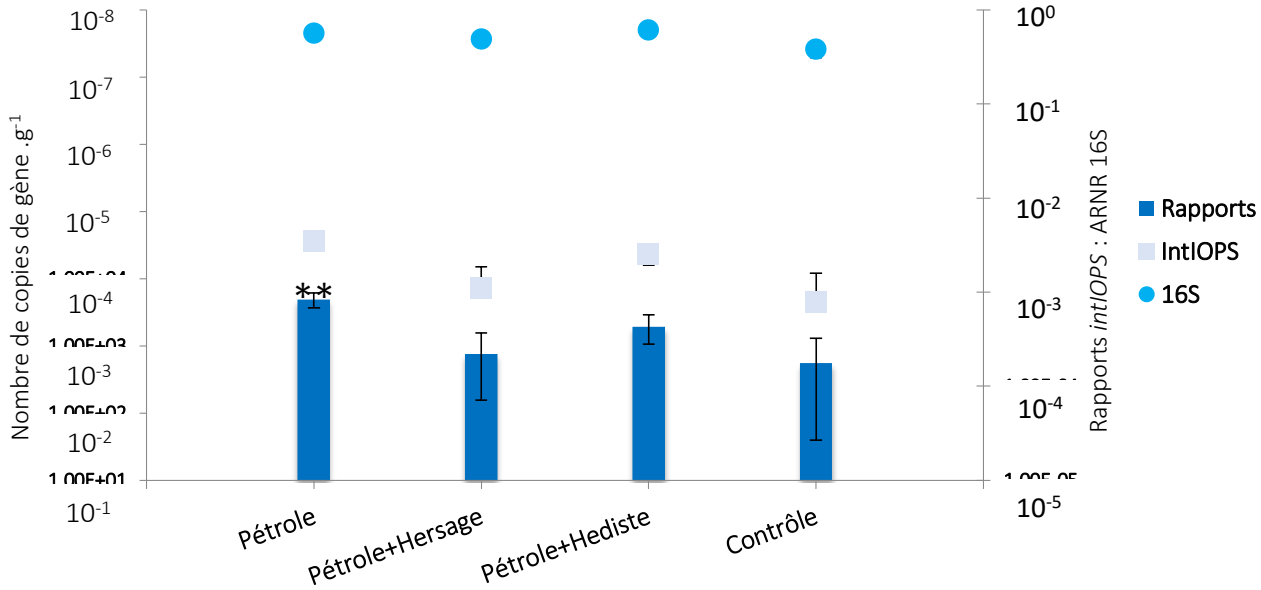


Figure IV-9 : Abondance relative du gène *intIOPS* dans les communautés bactériennes hydrocarbonoclastes au moment du prélèvement du sédiment pour la mise en place des bioréacteurs et dans le mésocosme contrôle sans contamination à la fin de l'expérimentation. * : Différence significative entre la condition *Pétrole* et les autres conditions (p>0.05 avec *Contrôle* et *Pétrole+Hersage* et p>0.1 avec *Pétrole+Hediste*).

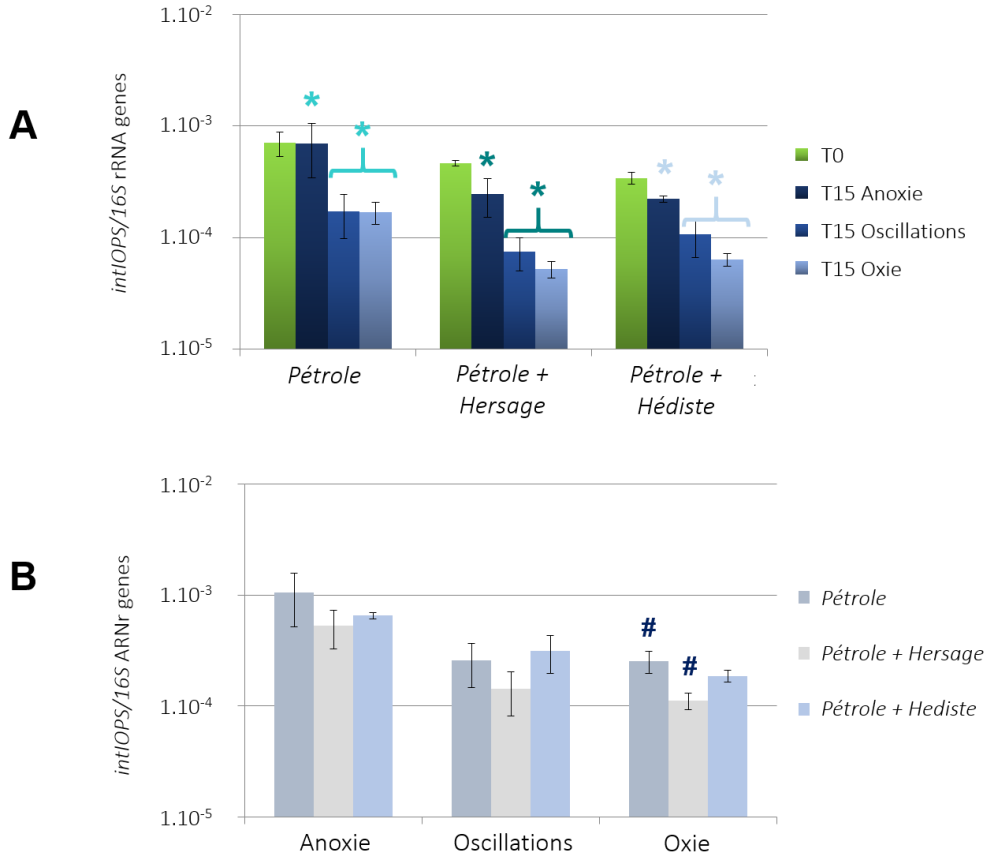


Figure IV-10 : Abondance relative du gène intIOPS dans les bioréacteurs (**A**) : Abondance dans les 3 communautés bactériennes hydrocarbonoclastes au début et à la fin de l'expérimentation en bioréacteurs pour les 3 conditions d'oxygénation. * : Différence significative entre la condition anoxie et les conditions d'oscillations et oxie (p>0.05). (**B**) : Abondance à la fin de l'expérimentation pour les 3 conditions d'oxygénation normalisée par le T0 de chaque communauté. # : Différence significative entre la communauté *Pétrole* et *Pétrole + Hersage* en condition oxie (p>0.05).

Enfin, la présence du gène *intIOPS* a été recherchée sur des consortia cultivés en conditions d'anoxie et de sulfato-réduction à partir des suspensions de sédiments issus des bioréacteurs. Sur 36 consortia analysés, 35 se sont avérés positifs pour la présence du gène *intIOPS*. Ceci confirme donc que des conditions anoxiques et de sulfato-réductions sont favorables à la présence du gène *intIOPS* suggérant qu'il pourrait être associé à des communautés de bactéries sulfato-réductrices hydrocarbonoclastes.

2.3. Etude de la dynamique des cassettes de gène dans les bioréacteurs

2.3.a. Structure de la communauté

La structure des communautés a été étudiée pour chaque condition au début et à la fin des incubations en bioréacteurs par T-RFLP sur le gène de l'ARNr 16S (Figure IV-11). La

communauté *Pétrole + Hediste* se distingue des 2 autres communautés au début de l'expérimentation et à la fin dans les conditions d'oxie et d'oscillations. Les communautés *Pétrole* et *Pétrole + Hersage* présentent 50% de similarité au début de l'expérimentation et à la fin dans les 3 conditions. Ainsi, il semble que le facteur « communauté » est le facteur principal influençant la structuration des communautés *Pétrole + Hediste* alors que pour les communautés *Pétrole* et *Pétrole + Hersage*, il s'agit du facteur « régime d'oxygénation ».

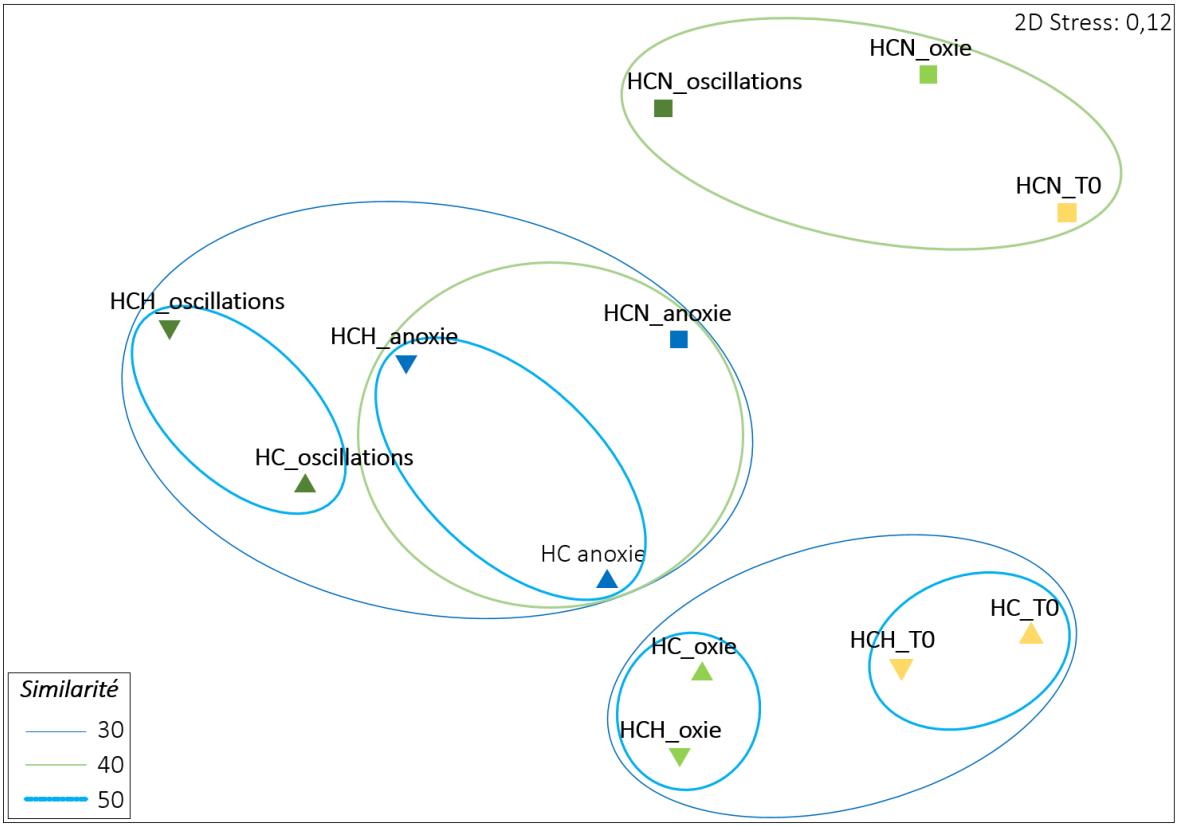


Figure IV-11 : Comparaison de la structure des 3 communautés au début et à la fin de l'expérimentation en bioréacteurs selon les conditions d'oxygénation (modifié à partir de Terrisse, 2014). ▼ : Communauté *Pétrole + Hersage* ; ▲ : Communauté *Pétrole* ; ■ : Communauté *Pétrole + Hediste*. En jaune : temps initial de l'expérimentation. En bleu : temps final en condition d'anoxie permanente. En vert clair : temps final en condition d'oxie permanente. En vert foncé : condition d'oxygénation oscillantes. Analyse MDS réalisée sur les distances de Bray-Curtis calculées à partir des données de T-RFLP sur le gène de l'ARNr 16S. Les distances entre échantillons sont exprimées en pourcentage de similarité.

2.3.b. La diversité des cassettes de gène

Le pool de cassettes de gène a été étudié à la fin de l'expérimentation en bioréacteurs pour les 3 communautés en se basant sur l'analyse de la taille des cassettes (Figure IV-12). Les 3 communautés en condition oxie présentent des diversités de cassettes proches entre elles (70% de similarité) et distantes des 2 autres conditions d'oxygénation (Figure IV-12). On observe donc une réponse similaire de la diversité des cassettes dans les 3 communautés

avec un effet structurant du régime d'oxygénation avec d'une part le pool en condition d'oxie permanente et d'autre part le pool des conditions d'anoxie permanente et d'oscillations. Il s'agit donc d'une réponse différente de celle de la structure des communautés qui montre que les modifications dans les communautés bactériennes n'expliquent pas ces différences dans la diversité des cassettes de gène.

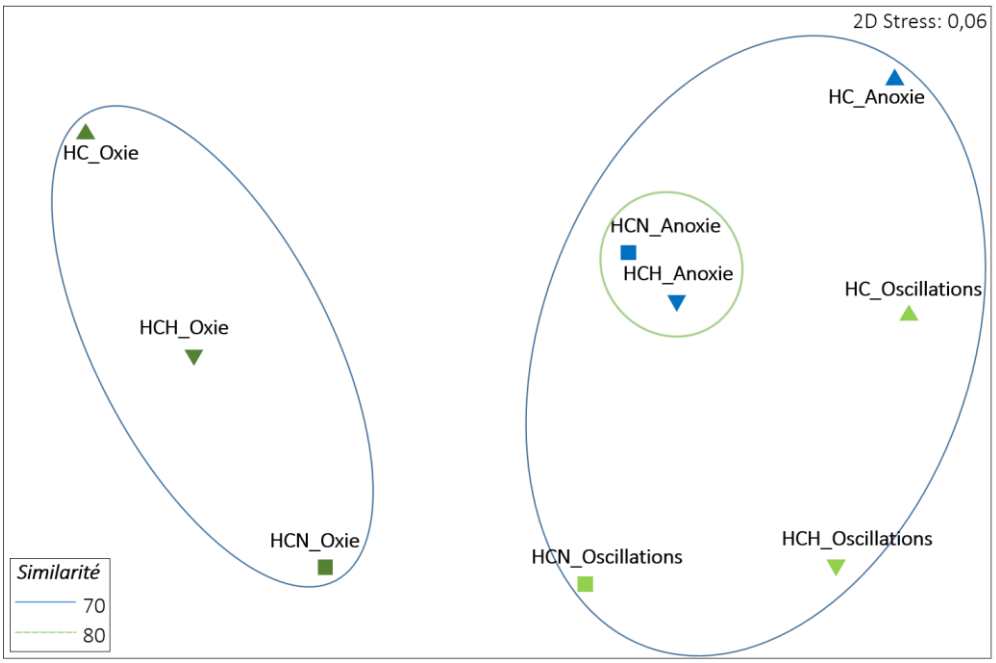


Figure IV-12 : Comparaison de la diversité des cassettes de gène dans les 3 communautés au temps final selon les conditions d'oxygénations en se basant sur l'analyse de la taille des cassettes. ▼ : Communauté *Pétrole + Hersage* ; ▲ : Communauté *Pétrole* ; ■ : Communauté *Pétrole + Hédiste*. En bleu : temps final en condition d'anoxie permanente. En vert clair : temps final en condition d'oxie permanente. En vert foncé : condition d'oxygénation oscillantes. Analyse MDS réalisée sur les distances de Bray-Curtis calculées à partir des données de T-RFLP sur le gène de l'ARNr 16S. Les distances entre échantillons sont exprimées en pourcentage de similarité.

2.4. Conclusion

Les analyses en mésocosmes ont permis de montrer qu'une contamination par du pétrole avait un impact important sur la diversité des intégrases. D'une part, une forte baisse de la diversité est observée avec une contamination pétrolière comme le montre l'indice de Shannon qui est compris entre 1,38 et 1,65 au bout de 9 mois d'incubation en condition pétrole contre 2,45 pour le mésocosme contrôle au même temps (Table 2, paragraphe III-2). Une importante baisse de la diversité avait également été observée dans les microcosmes de sédiments d'eau douce en condition de contamination par du benzo[a]pyrène (2,85 à 358 jours en condition benzo[a]pyrène contre 3,85 en condition contrôle, Figure IV-2, paragraphe IV-1.2.b.*i*). Par ailleurs, ces deux expérimentations ont montré la sélection de séquences particulières différentes. L'expérience en mésocosmes de sédiments côtiers contaminés par du pétrole a été reproduit 3 ans plus tard, et le gène *intIOPS* est toujours sélectionné, ce qui montre que ces résultats sont reproductibles et nous conforte quant à l'implication de l'intégron porteur de cette séquence dans l'adaptation de ces communautés bactériennes face à une contamination pétrolière.

L'étude de l'effet de différents régimes d'oxygénation des communautés hydrocarbonoclastes a permis de montrer une réponse similaire des 3 communautés en ce qui concerne la diversité en cassettes de gène. Or, ces 3 communautés sont distinctes. Ceci suggère que ces modifications sont indépendantes de la communauté. La diversité des cassettes répondrait non seulement aux contaminations du milieu comme nous avons pu le montrer au cours des études précédentes en milieu d'eau douce mais également aux variations en oxygène. Les variations de l'abondance du gène *intIOPS* semblent également indépendantes de la communauté, puisqu'elle varie de la même manière dans les 3 communautés hydrocarbonoclastes. Le facteur après le facteur contamination par du pétrole qui semble le plus influencer sa présence est l'oxygène. Ainsi, des conditions oxiques sont défavorables à la présence de IntIOPS. Il est également possible que ce gène soit porté par des bactéries communes aux 3 communautés, comme les bactéries sulfato-réductrices. Le genre *Desulfosarcina* par exemple est présent en abondance au début des incubations en bioréacteurs dans la communauté *Pétrole*. Son abondance ne varie pas au cours de

l'expérimentation en condition d'anoxie permanente, mais décroit de manière importante en conditions d'oscillations ou d'oxie permanente (Terrisse, 2014).

L'intégron porteur du gène *intIOPS* pourrait donc avoir un rôle dans l'adaptation des communautés bactériennes face à une pollution par du pétrole. Il était donc nécessaire de le caractériser afin de mieux comprendre quels mécanismes pourraient expliquer sa prépondérance. Le chapitre qui suit présente les résultats de cette caractérisation.

V- Caractérisation d'un intégron potentiellement adaptatif

1) Reconstitution de la séquence de l'intégron IntIOPS

Plusieurs méthodes ont été tentées dans le but de caractériser les séquences adjacentes du fragment de 800 pb connu du gène *intIOPS*. Des expérimentations de PCR inverse, de transposon display, de PCR-clonage de fragments longs réalisées avant et au cours de cette thèse n'ont pas permis d'aboutir à des résultats suffisamment concluants (Terrisse, 2011 ; Guenoc, 2013). Par ailleurs, des enrichissements à partir des sédiments dans lequel le gène *intIOPS* a pu être amplifié ont également été réalisés (paragraphe I - 2.2) dans le but d'isoler une souche porteuse de l'intégron IntIOPS. Toutefois, la diversité bactérienne présente dans les consortia était trop importante, même après plusieurs repiquages.

Nous avons alors choisi de développer une expérimentation de capture de gène qui a été réalisée en collaboration avec le groupe animé par le Pr Pierre Peyret de l'EA CIDAM 4678 de l'Université d'Auvergne et avec la contribution de la plateforme Génomique du Génopole Toulouse Midi-Pyrénées de Toulouse. Cette méthode novatrice combine la capture de fragments contenant des séquences connues du gène *intIOPS* par hybridation de sondes puis le séquençage de ces fragments capturés grâce à la technologie Illumina (Figure II-7). Elle a été réalisée à partir du métagénome de 5 échantillons : du sédiment côtier de mésocosme contaminé par du pétrole et sans remaniement (condition *Pétrole*), prélevé au même moment que le sédiment utilisé pour l'expérimentation en bioréacteurs (banque 1) ; des échantillons de suspension de sédiments prélevés dans les bioréacteurs en anoxie à la fin des incubations des 3 communautés (banque 2, banque 3 et banque 4) ; un des consortia dans lequel le gène *intIOPS* a pu être amplifié (banque D4).

Le prétraitement des séquences a permis l'élimination des séquences de taille inférieure à 100 nucléotides, de qualité inférieure à Q20 et présentant plus d'1% de N, soit au final entre 4 et 9,5% de séquences éliminées (Tableau V-1).

Tableau V-1 : Récapitulatif du nombre et de la qualité des séquences pour chaque banque avant et après prétraitement.

		Nombre de séquences avant prétraitement	GC%	Longueur séquences (nt) Q>20	Nombre de séquences après prétraitement	% de séquences filtrées
Banque 1	Read1 Read2	1 948 926 1 948 926	51 51	235 155	3 529 385	9,453
Banque 2	Read1 Read2	2 893 148 2 893 148	52 52	270 200	5 544 674	4,176
Banque 3	Read1 Read2	2 433 005 2 433 005	52 52	265 195	4 634 579	4,756
Banque 4	Read1 Read2	2 289 475 2 289 475	52 52	265 190	4 338 680	5,247
Banque D4	Read1 Read2	2 993 004 2 993 004	47 47	265 205	5 579 463	6,792

Une reconstruction de la séquence de l'intégron IntIOPS a pu être faite grâce à l'assemblage des séquences obtenues avec l'ADN des 5 échantillons (Figure II-7). Cette séquence est constituée du gène *intOPS* entier, du site *attI* associé, des régions promotrices du gène *intIOPS* et des cassettes, d'un fragment de l'environnement génétique de l'intégron IntIOPS et de 3 séquences différentes correspondant à des régions variables différentes de l'intégron IntIOPS (Figure V-1).

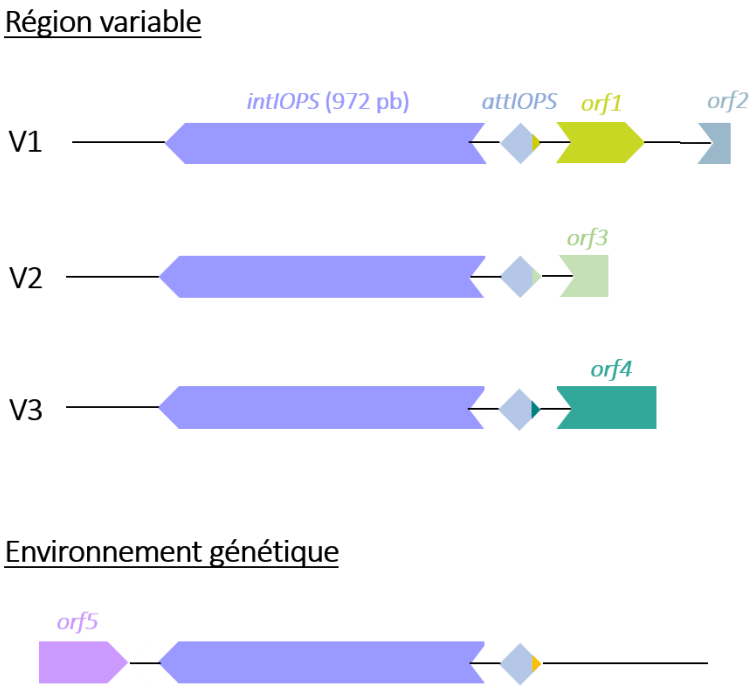


Figure V-1 : Représentation schématique des différents éléments de l'intégron IntIOPS caractérisés. En bleu, la région constante comprenant la séquence complète d'intIOPS et le site *attIOPS*. Les 3 régions variables correspondant aux 3 différentes premières cassettes identifiées sont notées V1, V2 et V3 et représentées en différentes nuances de vert. En violet, le cadre de lecture partiel identifié dans la région en amont de l'intégron.

2) Caractérisation des différents éléments de l'intégron IntIOPS

2.1. Caractérisation du gène de l'intégrase d'intégron IntIOPS

Le gène *intIOPS* entier fait 972 pb. Il est en sens inverse dans l'intégron IntIOPS par rapport aux cassettes de gène, comme c'est le cas pour la majorité des intégrons étudiés à ce jour. Après un blastX contre la base de données nr, la séquence connue la plus proche de l'intégrase IntIOPS est l'intégrase d'une *Oscillatoriales* dont le génome a été séquencé (WP_009769460.1). Les 2 intégrases partagent 62% d'identité protéique, ce qui montre que IntOPS est une nouvelle intégrase d'intégron qui n'était pas connue jusqu'alors.

Par ailleurs, la séquence obtenue de la séquence *intIOPS* présente des incertitudes au niveau de certaines positions. L'assemblage des séquences de fragments capturés ne permet pas de déterminer s'il s'agit de polymorphisme du gène *intIOPS* ou d'un biais d'expérimentation qui aurait pu intervenir notamment lors des différentes PCR ou lors du séquençage.

2.2. Le site *attI* de l'intégron IntIOPS : *attIOPS*

La région constante de l'intégron IntIOPS est bien conservée, hormis les quelques variations de séquences de *intIOPS*, jusqu'à un point se situant à 283 pb en aval du codon d'initiation du gène *intIOPS*. Ce point se situe au niveau d'un motif GTTRRRY, caractéristique du domaine R des sites *attI* (paragraphe 2.1.b ; Figure V-2). Il pourrait donc s'agir du site de recombinaison *attI* de IntIOPS et du point d'insertion des cassettes de gène. A l'exception du motif GTTRRRY, les sites *attI* sont très peu conservés et l'alignement du site putatif *attIOPS* et de 4 autres sites *attI* de référence ne montre pas d'autres motif partagé.

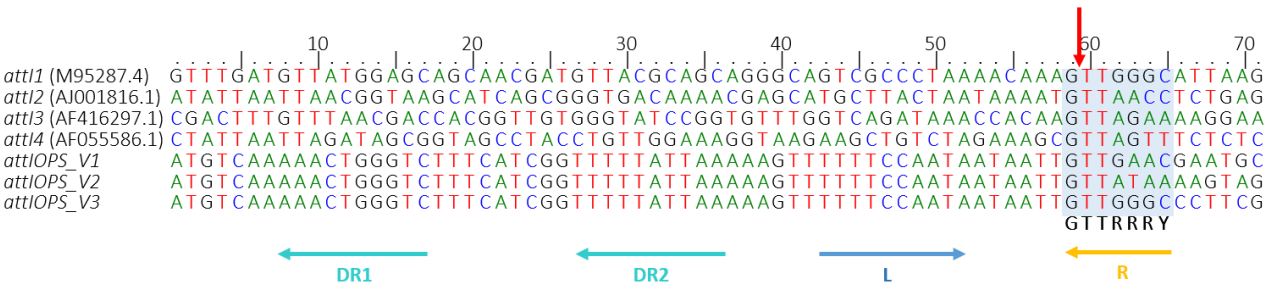


Figure V-2 : Alignement des séquences des sites *attIOPS V1, V2 et V3* de IntIOPS et des sites *attI* des intégrons de classe 1, 2 et 3 et de *Vibrio cholerae*. Les domaines DR1, DR2, L et R caractérisés chez *attI1* sont représentés par les flèches horizontales. La flèche verticale rouge représente le point d'insertion dans *attI1*.

2.3. La région promotrice de l'intégron IntIOPS

Une analyse in silico a permis de mettre à évidence une région pouvant correspondre au promoteur Pc des cassettes de gène (Figure V-3). Cette région est située en amont du site *attI* : la boite -10 putative est à 104 pb du point d'insertion des cassettes. Plusieurs motifs contenus dans cette région promotrice putative pourraient correspondre à des motifs de fixation de facteurs de régulation de l'expression : LexA, le régulateur transcriptionnel de la réponse SOS ; RpoH2, un facteur qui intervient dans la réponse au stress photooxydatif chez *Azospirillum brasilense* SP7 (Kumar et al., 2012) ; Fis, un facteur accessoire pour l'activation transcriptionnel du promoteur *mar* (pour multiple antibiotic resistance) chez *E. coli* (Martin and Rosner 1997) ; et Lrp, un facteur connu pour intervenir dans la régulation de la synthèse de la leucine mais également impliqué dans la biosynthèse d'autres acides aminés et dans leur dégradation, dans le transport de petites molécules, la synthèse de pili et dans d'autres fonctions cellulaires (Cho *et al.,* 2008).

Figure V-3 : Région située entre le début du gène *intIOPS* (en bleu) et le codon d'initiation de *l'orf1* (en vert) de la région variable V1 caractérisée. En **vert** : la région promotrice prédite des cassettes de gène. La barre verte au-dessus de la séquence symbolise les motifs de fixation des différents facteurs de régulation. En **bleu** : la région promotrice prédite du gène *intIOPS*. La barre bleue foncée en pointillée située au niveau des régions -10 correspond au site de fixation putatif de LexA. La barre bleue pleine en-dessous de la séquence symbolise les motifs de fixation de TyrR et PhoB. Les 2 boîtes - 10 sont représentées en rayé et se chevauchent partiellement. Les boîtes – 35 sont surlignées en couleur pleine. Le point d'insertion des cassettes de gène est symbolisé par la flèche rouge.

Le promoteur de *intIOPS* pourrait être chevauchant à la région promotrice des cassettes de gène. La boite -10 se situerait à 167 pb du codon d'initiation de la traduction de *intIOPS* (Figure V-3). Cette région promotrice contient des sites de reconnaissance pour les facteurs

de transcription PhoB et TyrR. PhoB est un régulateur positif du régulon phosphate (Bachhawat *et al.,* 2005) et TyrR peut-être un activateur ou un répresseur de la transcription (Pittard, Camakaris and Yang 2005). Le site de fixation du facteur LexA intervenant dans la réponse SOS pourrait également être présent.

Tous ces résultats ne sont que des prédictions et mériteraient une étude plus approfondie de ces séquences. Enfin, seules des expérimentations *in vitro* pourraient permettre de tester et de s'assurer que ces différentes régions sont bien des régions promotrices ou régulatrices.

2.4. Les premières cassettes associées à l'intégron IntOPS

Au cours de l'assemblage des séquences de IntIOPS, il s'est avéré que 3 régions variables différentes pourraient être associées à l'intégron IntIOPS.

La première région variable identifiée contient un cadre de lecture complet situé à 32 pb du motif GTTRRRY, nommé *orf1* (Figure V-1), suivi d'une région de 139 pb puis d'un second cadre de lecture incomplet, *orf2*. Un blastX de la séquence contenant ces 2 cadres de lecture et la région les séparant a permis de ressortir une homologie avec un fragment du génome de la δ-*proteobacteria* NaphS2, bactérie sulfato-réductrice capable de dégrader le naphtalène (DiDonato *et al.,* 2010). Le premier cadre de lecture *orf1,* d'une longueur de 213 pb, code une protéine partageant 61% d'identité protéique avec une protéine de cette souche, de fonction inconnue (WP_006425809.1). Le second cadre de lecture partiel *orf2* (55 pb) correspondrait à l'extrémité N-terminale d'une transposase de NaphS2 (80% d'identité protéique avec le fragment correspondant de la transposase WP_006425804.1) voisine du premier ORF codant la protéine homologue de *orf1*. La séquence séparant les 2 ORFs de IntIOPS et celle séparant les 2 ORFs de NaphS2 font toutes les deux 139 pb et partagent 80% d'identité nucléique (Figure V-4). Aucune séquence palindromique caractéristique des sites de recombinaison *attC* n'a pu être clairement identifiée dans cette région (Figure V-4).

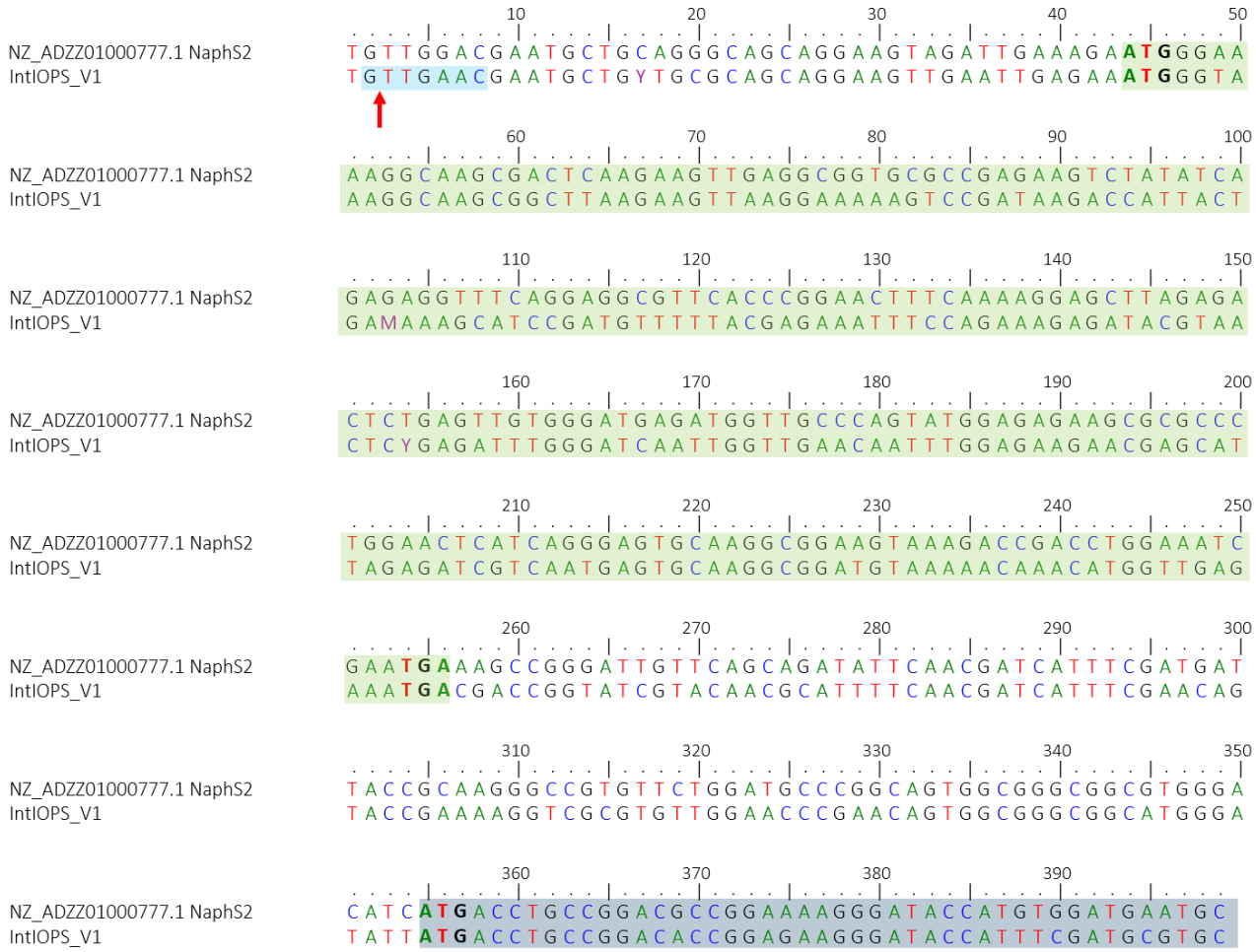


Figure V-4 : Alignement de la région variable V1 de IntIOPS et des séquences génomiques de la souche δ-*proteobacterium* NaphS2. La boîte R du site *attIOPS* est surlignée en bleu clair et le point de recombinaison est symbolisé par la flèche verticale rouge. Les gènes codant la protéine de fonction inconnue WP_006425809.1 aligné avec l'*orf1* de la région variable V1 de IntIOPS et la transposase WP_006425804.1 aligné avec l'*orf2* sont surlignées respectivement en vert et bleu foncé.

La seconde région variable identifiée contient un cadre de lecture putatif incomplet, nommé *orf3* (Figure V-1), d'une longueur de 198 pb, dont le codon d'initiation ATG est situé à 70 pb du motif GTTRRRY. Après un blastX contre la base de données nr, aucune correspondance n'a été trouvée avec des séquences connues.

La troisième région variable contient un autre cadre de lecture incomplet, nommé *orf4* (Figure V-1) d'une longueur de 273 pb et débutant 34 pb après le motif GTTRRRY. Le blastX contre la base de données nr a permis de mettre en évidence un motif conservé : il s'agit du domaine PRK11176 qui correspond à un domaine de transporteur de lipides ATP dépendant et/ou de perméase.

2.5. L'environnement génétique de IntIOPS

Une séquence de 465 pb après le codon stop du gène *intIOPS* a pu être obtenue grâce aux assemblages (Figure V 1). Dans cette séquence, un putatif cadre de lecture incomplet de 387 pb correspondant à 129 acides aminés a pu être caractérisé. Un blastX contre la base de données nr a montré qu'il pourrait correspondre à la partie C-terminale d'une transposase. Cette région présente 43% d'identité protéique avec l'extrémité C-terminale d'une transposase de *Candidatus Magnetoglobus multicellularis* str. Araruama (ETR70493.1) et 48% avec celle de la transposase d'une bactérie non isolée d'un environnement côtier (ABY56073.1). L'alignement des régions de ces 3 protéines montre plusieurs domaines bien conservés (Figure V-5).

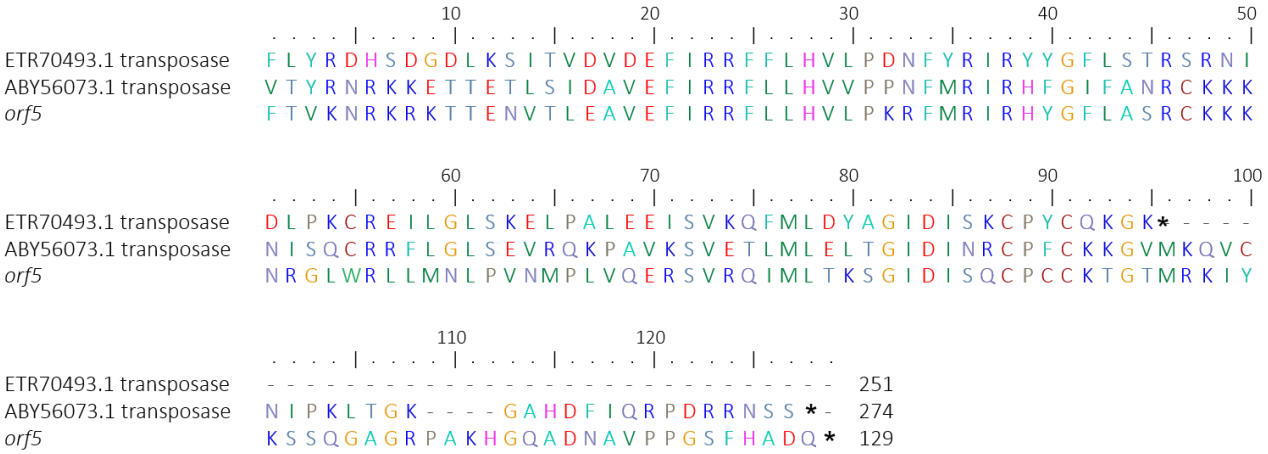


Figure V-5 : Alignement de la séquence codée par *l'orf5* de IntIOPS avec la séquence de la région C-terminale de la transposase de *Candidatus Magnetoglobus multicellularis* str. Araruama (ETR70493.1) et de la transposase de pFosPlaG (ABY56073.1).

3) Conclusion

Si l'identification des premières cassettes de gène ne permet pas de conclure pour l'instant quant au rôle potentiellement adaptatif que pourrait jouer IntIOPS dans les communautés étudiées, cela aura permis toutefois de montrer que dans les communautés étudiées l'intégron porte différentes premières cassettes de gène. Il semblerait donc qu'il y ait des mouvements de cassettes au sein de cet intégron. Ainsi, il pourrait être actif face à une pollution pétrolière. L'identification de la région promotrice du gène *intIOPS* a permis de montrer la présence de sites de liaison pour des facteurs intervenant dans des réponses aux stress comme LexA tout comme cela a été montré pour d'autres intégrons (Guerin *et al.,*

2009; Cambray *et al.,* 2011). Si les prédictions sont justes, l'expression du gène *intIOPS* pourrait être également régulée par des systèmes de régulation intervenant dans le contrôle de l'expression de fonctions du métabolisme cellulaire. Des expériences d'expression *in vitro* pourraient être réalisées pour infirmer ou confirmer ces hypothèses.

Par ailleurs, l'obtention de séquences situées en amont de l'intégron IntIOPS a permis de mettre en évidence la présence d'un gène possible codant une transposase. Il n'est pas rare de trouver ce type de gène à proximité des intégrons de manière générale, ce qui permet aux intégrons d'être mobiles. L'hypothèse de la mobilité et de la dispersion de IntIOPS au sein de la communauté bactérienne avait déjà été suggérée (Huang, 2010) pour expliquer les modifications de son abondance bien avant d'observer des modifications de la structure des communautés bactériennes face à une pollution pétrolière. La présence de cette putative transposase à proximité de l'intégron IntIOPS vient étayer cette hypothèse.

Tous ces résultats encourageants représentent une grande avancée dans la caractérisation de l'intégron IntIOPS mais ils restent toutefois encore préliminaires. De nouvelles expérimentations de captures de gène devront être réalisées avec des sondes ciblant les nouvelles séquences identifiées de IntIOPS afin d'obtenir les séquences complètes des ORFs déjà obtenus et de découvrir de nouvelles séquences plus éloignées encore du gène *intIOPS*. Des expérimentations pour étudier les régions promotrices des cassettes de gène et du gène *intIOPS* doivent également être réalisées si l'on veut pouvoir comprendre comment leur expression est régulée. Enfin, il serait intéressant de tester l'activité de recombinaison de l'intégrase IntIOPS au niveau de son site *attI* associé pour prouver que des cassettes sont bien insérées à son niveau.

Conclusions générales

Grâce aux différentes études présentées dans cette thèse, environ 800 séquences d'intégrases différentes ont été obtenues à partir de différents sédiments d'eau douce et côtiers. La quasi-totalité de ces séquences étaient encore inconnues. Par ailleurs, une analyse des génomes présents dans les bases de données réalisées en février 2014 a permis d'identifier 674 séquences d'intégrases. Le grand nombre de nouvelles séquences découvertes nous a poussé à remettre en question la classification des intégrases selon le seuil de 98% d'identité protéique. Dans l'article *Integron diversity in bacterial communities of freshwater sediments at different contamination levels* (Abella *et al*., en révision), nous avons proposé un nouveau seuil à 80% d'identité protéique qui nous semble plus adapté pour la classification des intégrons environnementaux. Même en abaissant autant ce seuil, nos conclusions restent inchangées et ce seuil ne remet pas en question la classification des intégrons de classe 1, 2, 3 et 4.

Nous avons pu confirmer par différentes études *in situ* et en microcosmes, en environnements d'eau douce et en milieux côtiers et avec différents types de polluants, que les contaminations du milieu impactent la diversité des intégrons de manière indépendante de la structure de la communauté bactérienne. En effet, nous avons pu observer en milieux naturels que la diversité des intégrases était plus faible dans les milieux les plus impactés. En microcosmes, la présence de polluant entraine une baisse, parfois très importante, de la diversité des intégrases mais également l'augmentation de l'abondance de nouvelles séquences particulières, de manière similaire aux travaux de Lionel Huang. Enfin, nous avons pu confirmer l'influence géographique sur la diversité des intégrases en montrant des diversités différentes entre des sites de même type (côtiers ou eau douce) mais éloignés géographiquement.

En ce qui concerne la diversité des cassettes de gène, nous avons pu voir que le pool de cassettes répondait bien aux conditions environnementales. En effet, le pool de cassettes semble être impacté par le type de contaminants présents dans le milieu. Ceci appuie la notion d'écotypes de cassettes proposée auparavant (Koenig et al., 2008). Enfin, les banques de cassettes réalisées lors de cette thèse ont révélé des séquences fortement similaires à d'autres cassettes obtenues à partir de sites côtiers d'autres régions du monde. Ainsi,

contrairement à l'intégrase, il pourrait y avoir une distribution mondiale des cassettes de gène.

Enfin, grâce à une méthode novatrice de capture de gène couplée à un séquençage avec la technologie Illumina, nous avons pu accéder aux séquences adjacentes du fragment du gène *intIOPS* obtenu lors de la thèse de Lionel Huang. L'intégralité du gène *intIOPS* a pu être caractérisée. Nous avons pu identifier 3 régions variables correspondant à 3 premières cassettes différentes. Ceci suggère que des mouvements de cassettes résultant de phénomènes d'insertion et d'excision ont pu intervenir au cours des incubations en mésocosmes et donc que l'intégron porteur du gène *intIOPS* pourrait être actif dans ces conditions. Enfin, un fragment d'un gène codant une transposase a pu être identifié en amont de l'intégron, venant renforcer l'hypothèse que cet intégron pourrait être mobile et se disséminer dans la communauté bactérienne.

Ainsi, les études menées lors de cette thèse ont permis de conforter un grand nombre d'observations précédemment faites et ont apporté de nouveaux éléments qui viennent soutenir notre hypothèse principale que les intégrons environnementaux seraient impliqués dans l'adaptation des communautés bactériennes en réponse à la présence de contaminants dans les milieux non cliniques. Afin d'aller plus loin, les travaux présentés dans cette thèse doivent être approfondis.

La méthode de séquençage Sanger nous a permis d'obtenir un grand nombre de séquences correspondant à un fragment du gène *intI* qui couvre l'ensemble des domaines impliqués dans l'activité de l'intégrase. Toutefois malgré l'effort de séquençage, cette méthode ne permet pas d'obtenir un bon recouvrement des banques tant la diversité des intégrases peut être importante dans certains échantillons. Or, afin de mieux comprendre comment la diversité des intégrases est influencée par les conditions environnementales, il est nécessaire d'obtenir des banques avec de bon taux de recouvrement. A ce jour, les méthodes de séquençage haut débit permettent l'obtention de données avec une profondeur suffisamment importante pour répondre à ce besoin. Actuellement, une méthode de

séquençage MiSeq du gène *intI* est en cours de développement au laboratoire et devrait permettre d'accéder à une plus grande diversité de l'intégrase dans des échantillons de la Seine (projet EC2CO DIVINT, 2015-2016).

Cette thèse a permis de caractériser une partie de l'intégron porteur du gène *intIOPS*. De nouvelles expérimentations de capture de gène pourraient être réalisées avec la synthèse de nouvelles sondes ciblant les extrémités des nouvelles de séquences, dans le but d'obtenir l'intégralité des premiers cadres de lecture identifiés et de découvrir de nouvelles séquences plus éloignées du gène *intIOPS*. Ceci permettrait d'aller plus loin dans la caractérisation des fonctions portées par cet intégron. Par ailleurs, le séquençage haut-débit a permis de révéler une certaine variabilité dans la séquence de IntIOPS et dans la région du site *attI* associé. Le clonage et le séquençage de ces régions permettrait de savoir si cette variabilité de séquences est due à un biais induit par les étapes d'amplification ou de synthèse ou bien si différents variants du gène *intIOPS* existent bien et sont associés à différentes cassettes de gène. Enfin, une première analyse des régions promotrices du gène *intIOPS* et des cassettes de gène a été réalisée en utilisant différents outils en ligne. Les premières conclusions obtenues ont besoin d'être vérifiées et approfondies. Des expérimentations *in vitro* et *in vivo* s'avèrent indispensables afin de comprendre comment sont régulées les activités de recombinaison et d'expression des cassettes de gène dans cet intégron. Ces approches nécessitent en premier lieu de cloner la région promotrice située entre le gène *intIOPS* et la première cassette de gène. Dans un second temps, la technique de gel retard pourrait être utilisée pour vérifier la fixation des différents facteurs de régulation identifiés par l'analyse *in silico* sur les régions promotrices et d'une analyse par empreinte à la DNase par exemple pour localiser la position du site de fixation. Enfin, grâce à la construction d'un plasmide dans lequel aurait été insérée la région promotrice associée à un et/ou 2 gènes de rapporteurs, la régulation de l'expression de l'intégrase et/ou des cassettes de gène par différents facteurs transcriptionnels pourrait être testée.

Cette thèse a permis de mettre en évidence des séquences d'intégrase d'intégron qui devenaient prédominantes en microcosmes d'eau douce face à une contamination par de l'arsenic, du benzo[a]pyrène ou du triméthoprime. Une mise au point de méthode de

quantification permettant le suivi de ces séquences dans les sédiments au cours du temps a été débutée mais n'a pas abouti. Il serait intéressant de poursuivre cette mise au point pour pouvoir confirmer que l'abondance de ces séquences augmente bien au cours du temps et de manière spécifique au contaminant. Par la suite, une caractérisation des séquences adjacentes grâce à la méthode de capture de gène développée pour *intIOPS* pourrait être réalisée afin de comprendre pourquoi l'abondance de ces intégrons augmentent et de déterminer si ces intégrons portent des fonctions adaptatives.

Au cours de cette thèse, une méta-analyse a été débutée sur les séquences d'intégrase contenues dans les bases de données. Cette analyse a pour but de vérifier l'existence des différents groupes d'intégrase en fonction de leur origine, marine ou terrestre, identifiés dans de précédentes études mais avec un nombre restreint de séquences. De plus, cette étude cherche également à mettre en évidence d'éventuels motifs protéiques spécifiques de ces groupes, qui pourraient être des éléments à prendre en compte pour la classification des intégrons environnementaux.

Cette méta-analyse pourrait être étendue à l'étude des sites de recombinaisons. En effet, le séquençage des cassettes de gène est rendu difficile par l'importante variabilité de séquence des sites *attC* et *attI*. Les amorces utilisées actuellement pour séquencer les cassettes en ciblant les sites *attC* ont été dessinées en 2001. Or aujourd'hui, avec le séquençage de nombreux génomes, un grand nombre de séquences de site *attC* est disponible et permettrait la création d'amorces ciblant une diversité de cassettes plus importante. La diversité du site *attI* a, quant à elle, été peu étudiée et les études se sont essentiellement limitées aux sites *attI* des intégrons de classe 1, 2 et 3 et de quelques intégrons chromosomiques. Or, une méthode plus directe pour pouvoir étudier les premières cassettes de gène dans un environnement donné, cassettes représentant les derniers évènements d'insertion et donc potentiellement adaptatives, serait de pouvoir les séquencer en ciblant d'une part le site *attC* en aval et d'autre part le site *attI* en amont de la cassette. Une méta-analyse des séquences de sites *attI* disponibles dans les bases de données permettrait d'approfondir nos connaissances sur la variabilité de ces séquences et pourrait permettre de dessiner un ou des couples d'amorces qui permettraient de séquencer la première cassette de gène.

Enfin, les différentes études réalisées sur le sujet ont montré que la diversité des cassettes de gène étaient très importantes dans l'environnement et que les protéines codées ont pour la grande majorité des fonctions encore inconnues. Afin de mieux comprendre le rôle des intégrons dans l'adaptation des bactéries dans l'environnement, il est primordial d'identifier les fonctions qu'ils portent et quel(s) avantage(s) elles peuvent apporter face aux conditions environnementales. Ainsi, au-delà de l'effort de séquençage réalisé, il faudrait également réaliser des études de protéomiques afin de caractériser les fonctions portées par les cassettes de gène.

Références bibliographiques

A

Abella J, Fahy A, Duran R, Cagnon C. Integron diversity in bacterial communities of freshwater sediments at different contamination levels. *FEMS Ecology Microbiology* 2015, 95-12

Abella J, Bielen A, Huang L *et al.* Integron diversity in marine environments. Environ Sci Pollut Res 2015.

Abremski KE, Hoess RH. Evidence for a second conserved arginine residue in the integrase family of recombination proteins. *Protein Eng* 1992;5:87–91.

Albertsen M, Hansen LBS, Saunders AM *et al.* A metagenome of a full-scale microbial community carrying out enhanced biological phosphorus removal. *ISME J* 2012;6:1094–106.

Arakawa Y, Murakami M, Suzuki K *et al.* A novel integron-like element carrying the metallo-β-lactamase gene bla(IMP). *Antimicrob Agents Chemother* 1995;39:1612–5.

Arciszewska LK, Sherratt DJ. Xer site-specific recombination in vitro. *EMBO J* 1995;14:2112–20.

Argos P, Landy A, Abremski K *et al.* The integrase family of site-specific recombinases: regional similarities and global diversity. *EMBO J* 1986;5:433–40.

B

Bachhawat P, Swapna GVT, Montelione GT *et al.* Mechanism of activation for transcription factor PhoB suggested by different modes of dimerization in the inactive and active states. *Struct Lond Engl 1993* 2005;13:1353–63.

Baharoglu Z, Bikard D, Mazel D. Conjugative DNA transfer induces the bacterial SOS response and promotes antibiotic resistance development through integron activation. *PLoS Genet* 2010;6:1–10.

Baharoglu Z, Krin E, Mazel D. Connecting environment and genome plasticity in the characterization of transformation-induced SOS regulation and carbon catabolite control of the Vibrio cholerae integron integrase. *J Bacteriol* 2012;194:1659–67.

Baharoglu Z, Mazel D. SOS, the formidable strategy of bacteria against aggressions. *FEMS Microbiol Rev* 2014;38:1126–45.

Barkovskii AL, Green C, Hurley D. The occurrence, spatial and temporal distribution, and environmental routes of tetracycline resistance and integrase genes in Crassostrea virginica beds. *Mar Pollut Bull* 2010;60:2215–24.

Barlow RS, Gobius KS. Diverse class 2 integrons in bacteria from beef cattle sources. *J Antimicrob Chemother* 2006;58:1133–8.

Barraud O, Casellas M, Dagot C *et al.* An antibiotic-resistant class 3 integron in an Enterobacter cloacae isolate from hospital effluent. *Clin Microbiol Infect* 2013;**19**:E306–8.

Beaber JW, Hochhut B, Waldor MK. SOS response promotes horizontal dissemination of antibiotic resistance genes. *Nature* 2004;**427**:72–4.

Bertin PN, Heinrich-Salmeron A, Pelletier E *et al.* Metabolic diversity among main microorganisms inside an arsenic-rich ecosystem revealed by meta-and proteo-genomics. *ISME J* 2011;**5**:1735–47.

Biskri L, Bouvier M, Guérout A-M *et al.* Comparative study of class 1 integron and Vibrio cholerae superintegron integrase activities. *J Bacteriol* 2005;**187**:1740–50.

Biskri L, Mazel D. Erythromycin esterase gene ere(A) is located in a functional gene cassette in an unusual class 2 integron. *Antimicrob Agents Chemother* 2003;**47**:3326–31.

Biyela PT, Lin J, Bezuidenhout CC. The role of aquatic ecosystems as reservoirs of antibiotic resistant bacteria and antibiotic resistance genes. *Water Sci Technol J Int Assoc Water Pollut Res* 2004;**50**:45–50.

Blázquez J, Couce A, Rodríguez-Beltrán J *et al.* Antimicrobials as promoters of genetic variation. *Curr Opin Microbiol* 2012;**15**:561–9.

Bodrossy L, Holmes EM, Holmes AJ *et al.* Analysis of 16S rRNA and methane monooxygenase gene sequences reveals a novel group of thermotolerant and thermophilic methanotrophs, Methylocaldum gen. nov. *Arch Microbiol* 1997;**168**:493–503.

Bordenstein SR, Reznikoff WS. Mobile DNA in obligate intracellular bacteria. *Nat Rev Microbiol* 2005;**3**:688–99.

Boucher Y, Labbate M, Koenig JE *et al.* Integrons: mobilizable platforms that promote genetic diversity in bacteria. *Trends Microbiol* 2007;**15**:301–9.

Bouvier M, Demarre G, Mazel D. Integron cassette insertion: A recombination process involving a folded single strand substrate. *EMBO J* 2005;**24**:4356–67.

Byrne-Bailey KG, Gaze WH, Zhang L *et al.* Integron prevalence and diversity in manured soil. *Appl Environ Microbiol* 2011;**77**:684–7.

Byrne-Bailey KG, Weber KA, Coates JD. Draft genome sequence of the anaerobic, nitrate-dependent, Fe(II)-oxidizing bacterium Pseudogulbenkiania ferrooxidans strain 2002. *J Bacteriol* 2012;**194**:2400–1.

C

Cambray G, Guerout A-M, Mazel D. Integrons. *Annu Rev Genet* 2010;**44**:141–66.

Cambray G, Sanchez-Alberola N, Campoy S *et al.* Prevalence of SOS-mediated control of integron integrase expression as an adaptive trait of chromosomal and mobile integrons. *Mob DNA* 2011;**2**, DOI: 10.1186/1759-8753-2-6.

Campbell MA, Chain PSG, Dang H *et al.* Nitrosococcus watsonii sp. nov., a new species of marine obligate ammonia-oxidizing bacteria that is not omnipresent in the world's oceans: calls to validate the names "Nitrosococcus halophilus" and "Nitrosomonas mobilis." *FEMS Microbiol Ecol* 2011;**76**:39–48.

Castresana J. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol Biol Evol* 2000;**17**:540–52.

Chao A. Nonparametric Estimation of the Number of Classes in a Population. *Scand J Stat* 1984;**11**:265–70.

Cho B-K, Barrett CL, Knight EM *et al.* Genome-scale reconstruction of the Lrp regulatory network in Escherichia coli. *Proc Natl Acad Sci* 2008;**105**:19462–7.

Clark CA, Purins L, Kaewrakon P *et al.* VCR repetitive sequence elements in the Vibrio cholerae chromosome constitute a mega-integron [1]. *Mol Microbiol* 1997;**26**:1137–8.

Clarke RK, Gorley RN. *Primer V6: User Manual - Tutorial*. Plymouth Marine Laboratory, 2006.

Coleman NV, Holmes AJ. The native Pseudomonas stutzeri strain Q chromosomal integron can capture and express cassette-associated genes. *Microbiology* 2005;**151**:1853–64.

Collis CM, Grammaticopoulos G, Briton J *et al.* Site-specific insertion of gene cassettes into integrons. *Mol Microbiol* 1993;**9**:41–52.

Collis CM, Hall RM. Expression of antibiotic resistance genes in the integrated cassettes of integrons. *Antimicrob Agents Chemother* 1995;**39**:155–62.

Collis CM, Hall RM. Comparison of the structure-activity relationships of the integron-associated recombination sites attI3 and attI1 reveals common features. *Microbiology* 2004;**150**:1591–601.

Collis CM, Kim M-J, Stokes HW *et al.* Binding of the purified integron DNA integrase IntI1 to integron- and cassette-associated recombination sites. *Mol Microbiol* 1998;**29**:477–90.

Collis CM, Kim M-J, Stokes HW *et al.* Integron-encoded IntI integrases preferentially recognize the adjacent cognate attI site in recombination with a 59-be site. *Mol Microbiol* 2002;**46**:1415–27.

Contin M, Goi D, De Nobili M. Land application of aerobic sewage sludge does not impair methane oxidation rates of soils. *Sci Total Environ* 2012;**441**:10–8.

Coover MP, Sims RC. The Effect of Temperature on Polycyclic Aromatic Hydrocarbon Persistence in an Unacclimated Agricultural Soil. *Hazard Waste Hazard Mater* 1987;**4**:69–82.

Correia M, Boavida F, Grosso F *et al.* Molecular characterization of a new class 3 integron in Klebsiella pneumoniae. *Antimicrob Agents Chemother* 2003;**47**:2838–43.

Costa SS, Mourato C, Viveiros M *et al.* Description of plasmid pSM52, harbouring the gene for the Smr efflux pump, and its involvement in resistance to biocides in a meticillin-resistant Staphylococcus aureus strain. *Int J Antimicrob Agents* 2013;**41**:490–2.

Coyne S, Rosenfeld N, Lambert T *et al.* Overexpression of resistance-nodulation-cell division pump AdeFGH confers multidrug resistance in Acinetobacter baumannii. *Antimicrob Agents Chemother* 2010;**54**:4389–93.

D

Deblonde T, Dreyer M, Hartemann P. Pharmaceuticals residues and drinking water. 2012.

Demarre G, Frumerie C, Gopaul DN *et al.* Identification of key structural determinants of the IntI1 integron integrase that influence attC × attI1 recombination efficiency. *Nucleic Acids Res* 2007;**35**:6475–89.

Deng Y, Liang Y, Cui H *et al.* Roles of integrons in the antimicrobial resistance of Gram-positive microorganisms. *Rev Med Microbiol* 2015;**26**:26–31.

DiDonato RJ Jr, Young ND, Butler JE *et al.* Genome Sequence of the Deltaproteobacterial Strain NaphS2 and Analysis of Differential Gene Expression during Anaerobic Growth on Naphthalene. *PLoS ONE* 2010;**5**:e14072.

Drury B, Rosi-Marshall E, Kelly JJ. Wastewater treatment effluent reduces the abundance and diversity of benthic bacterial communities in urban and suburban rivers. *Appl Environ Microbiol* 2013;**79**:1897–905.

Duran R, Bielen A, Paradžik T *et al.* Exploring Actinobacteria assemblages in coastal marine sediments under contrasted Human influences in the West Istria Sea, Croatia. 2015.

Van Duyne GD. A structural view of tyrosine recombinase site-specific recombination. *Mobile DNA II*. ASM Press. Washington, D.C.: N. L. Craig et al., 2002, 93–117.

E

Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 2004;**32**:1792–7.

Elsaied H, Stokes HW, Kitamura K *et al.* Marine integrons containing novel integrase genes, attachment sites, attI, and associated gene cassettes in polluted sediments from Suez and Tokyo Bays. *ISME J* 2011;**5**:1162–77.

Elsaied H, Stokes HW, Nakamura T *et al.* Novel and diverse integron integrase genes and integron-like gene cassettes are prevalent in deep-sea hydrothermal vents. *Environ Microbiol* 2007;**9**:2298–312.

Elsaied H, Stokes HW, Yoshioka H *et al.* Novel integrons and gene cassettes from a Cascadian submarine gas-hydrate-bearing core. *FEMS Microbiol Ecol* 2014;**87**:343–56.

Esposito D, Scocca JJ. The integrase family of tyrosine recombinases: Evolution of a conserved active site domain. *Nucleic Acids Res* 1997;**25**:3605–14.

van Essen-Zandbergen A, Smith H, Veldman K *et al.* Occurrence and characteristics of class 1, 2 and 3 integrons in Escherichia coli, Salmonella and Campylobacter spp. in the Netherlands. *J Antimicrob Chemother* 2007;**59**:746–50.

F

Fluit AC, Schmitz F-J. Resistance integrons and super-integrons. *Clin Microbiol Infect* 2004;**10**:272–88.

Fong KPY, Goh CBH, Tan H-M. The genes for benzene catabolism in Pseudomonas putida ML2 are flanked by two copies of the insertion element IS1489, forming a Class-I-type catabolic transposon, Tn5542. *Plasmid* 2000;**43**:103–10.

Frost LS, Leplae R, Summers AO *et al.* Mobile genetic elements: the agents of open source evolution. *Nat Rev Microbiol* 2005;**3**:722–32.

G

Gaze WH, Zhang L, Abdouslam NA *et al.* Impacts of anthropogenic activity on the ecology of class 1 integrons and integron-associated genes in the environment. *ISME J* 2011;**5**:1253–61.

Gestal AM, Liew EF, Coleman NV. Natural transformation with synthetic gene cassettes: New tools for integron research and biotechnology. *Microbiology* 2011;**157**:3349–60.

Giakkoupi P, Tryfinopoulou K, Polemis M *et al.* Circulation of a multiresistant, conjugative, IncA/C plasmid within the nosocomial Providencia stuartii population in the Athens area. *Diagn Microbiol Infect Dis* 2015;**82**:62–4.

Gillings MR. Integrons: Past, present, and future. *Microbiol Mol Biol Rev* 2014;**78**:257–77.

Gillings MR, Holley MP, Stokes HW *et al.* Integrons in Xanthomonas: A source of species genome diversity. *Proc Natl Acad Sci U S A* 2005;**102**:4419–24.

Gillings MR, Holley MP, Stokes HW. Evidence for dynamic exchange of qac gene cassettes between class 1 integrons and other integrons in freshwater biofilms. *FEMS Microbiol Lett* 2009;**296**:282–8.

Gillings MR, Krishnan S, Worden PJ *et al.* Recovery of diverse genes for class 1 integron-integrases from environmental DNA samples. *FEMS Microbiol Lett* 2008;**287**:56–62.

Good IJ. The Population Frequencies of Species and the Estimation of Population Parameters. *Biometrika* 1953;**40**:237–64.

Gravel A, Fournier B, Roy PH. DNA complexes obtained with the integron integrase Intl1 at the attl1 site. *Nucleic Acids Res* 1998;**26**:4347–55.

Grosso MD, Northwood JGE, Farrell DJ *et al.* The Macrolide Resistance Genes erm(B) and mef(E) Are Carried by Tn2010 in Dual-Gene Streptococcus pneumoniae Isolates Belonging to Clonal Complex CC271. *Antimicrob Agents Chemother* 2007;**51**:4184–6.

Guerin E, Cambray G, Sanchez-Alberola N *et al.* The SOS response controls integron recombination. *Science* 2009;**324**:1034.

Guérin E, Jové T, Tabesse A *et al.* High-level gene cassette transcription prevents integrase expression in class 1 integrons. *J Bacteriol* 2011;**193**:5675–82.

H

Hall RM. Mobile gene cassettes and integrons: Moving antibiotic resistance genes in Gram-negative bacteria. *CIBA Found Symp* 1997:192–205.

Hall RM. Integrons and gene cassettes: Hotspots of diversity in bacterial genomes. *Ann NY Acad Sci* 2012;**1267**.

Hall RM, Brookes DE, Stokes HW. Site-specific insertion of genes into integrons: Role of the 59-base element and determination of the recombination cross-over point. *Mol Microbiol* 1991;**5**:1941–59.

Hall RM, Collis CM. Mobile gene cassettes and integrons: Capture and spread of genes by site-specific recombination. *Mol Microbiol* 1995;**15**:593–600.

Hanau-Berçot B, Podglajen I, Casin I *et al.* An intrinsic control element for translational initiation in class 1 integrons. *Mol Microbiol* 2002;**44**:119–30.

Hansson K, Sundström L, Pelletier A *et al.* IntI2 integron integrase in Tn7. *J Bacteriol* 2002;**184**:1712–21.

Hardwick SA, Stokes HW, Findlay S *et al.* Quantification of class 1 integron abundance in natural environments using real-time quantitative PCR. *FEMS Microbiol Lett* 2008;**278**:207–12.

Henriques IS, Fonseca F, Alves A *et al.* Occurrence and diversity of integrons and β-lactamase genes among ampicillin-resistant isolates from estuarine waters. *Res Microbiol* 2006;**157**:938–47.

Holmes AJ, Gillings MR, Nield BS *et al.* The gene cassette metagenome is a basic resource for bacterial genome evolution. *Environ Microbiol* 2003;**5**:383–94.

Huang L. Implication des éléments génétiques mobiles dans l'adaptation des communautés bactériennes. 2010.

Huang L, Cagnon C, Caumette P *et al.* First gene cassettes of integrons as targets in finding adaptive genes in metagenomes. *Appl Environ Microbiol* 2009;**75**:3823–5.

Huang Y, Niu B, Gao Y *et al.* CD-HIT Suite: a web server for clustering and comparing biological sequences. *Bioinformatics* 2010;**26**:680–2.

I

Ilori MO, Picardal FW, Aramayo R *et al.* Catabolic plasmid specifying polychlorinated biphenyl degradation in Cupriavidus sp. strain SK-4: Mobilization and expression in a pseudomonad. *J Basic Microbiol* 2015;**55**:338–45.

J

Jacquier H, Zaoui C, Sanson-Le Pors M-J *et al.* Translation regulation of integrons gene cassette expression by the attC sites. *Mol Microbiol* 2009;**72**:1475–86.

Ji G, Silver S. Regulation and expression of the arsenic resistance operon from Staphylococcus aureus plasmid pI258. *J Bacteriol* 1992;**174**:3684–94.

Johansson C, Kamali-Moghaddam M, Sundström L. Integron integrase binds to bulged hairpin DNA. *Nucleic Acids Res* 2004;**32**:4033–43.

Johnson LL, Collier TK, Stein JE. An analysis in support of sediment quality thresholds for polycyclic aromatic hydrocarbons (PAHs) to protect estuarine fish. *Aquat Conserv Mar Freshw Ecosyst* 2002;**12**:517–38.

Jové T, Da Re S, Denis F *et al.* Inverse correlation between promoter strength and excision activity in class 1 integrons. *PLoS Genet* 2010;**6**, DOI: 10.1371/journal.pgen.1000793.

K

Koenig JE, Boucher Y, Charlebois RL *et al.* Integron-associated gene cassettes in Halifax Harbour: Assessment of a mobile gene pool in marine sediments. *Environ Microbiol* 2008;**10**:1024–38.

Koenig JE, Sharp C, Dlutek M *et al.* Integron gene cassettes and degradation of compounds associated with industrial waste: The case of the SydneyTar Ponds. *PLoS ONE* 2009;**4**, DOI: 10.1371/journal.pone.0005276.

Korona-Glowniak I, Siwiec R, Malm A. Resistance determinants and their association with different transposons in the antibiotic-resistant streptococcus pneumoniae. *BioMed Res Int* 2015;**2015**, DOI: 10.1155/2015/836496.

Kor S-B, Choo Q-C, Chew C-H. New integron gene arrays from multiresistant clinical isolates of members of the Enterobacteriaceae and Pseudomonas aeruginosa from hospitals in Malaysia. *J Med Microbiol* 2013;**62**:412–20.

Krin E, Cambray G, Mazel D. The superintegron integrase and the cassette promoters are co-regulated in Vibrio cholerae. *PLoS ONE* 2014;**9**, DOI: 10.1371/journal.pone.0091194.

L

Labbate M, Case RJ, Stokes HW. The integron/gene cassette system: an active player in bacterial adaptation. *Methods Mol Biol Clifton NJ* 2009;**532**:103–25.

Landy A. Dynamic, structural, and regulatory aspects of λ site-specific recombination. *Annu Rev Biochem* 1989;**58**:913–49.

Lane DJ. 16S/23S rRNA Sequencing. *Nucleic Acid Techniques in Bacterial Systematics*. John Wiley & Sons. New York, NY, USA: Stackebrandt E & Goodfellow M, eds, 1991, 115–75.

Laroche E, Pawlak B, Berthe T *et al.* Occurrence of antibiotic resistance and class 1, 2 and 3 integrons in Escherichia coli isolated from a densely populated estuary (Seine, France). *FEMS Microbiol Ecol* 2009;**68**:118–30.

Larouche A, Roy PH. Effect of attC structure on cassette excision by integron integrases. *Mob DNA* 2011;**2**, DOI: 10.1186/1759-8753-2-3.

Lespes G, Bancon-Montigny C, Aguerre S *et al.* Organotin speciation in waters and sediments in the Adour-Garonne basin. *Rev Sci Eau* 2005;**18**:47–63.

Lévesque C, Brassard S, Lapointe J *et al.* Diversity and relative strength of tandem promoters for the antibiotic-resistance genes of several integron. *Gene* 1994;**142**:49–54.

Levesque C, Brassard S, Lapointe J *et al.* Diversity and relative strength of tandem promoters for the antibiotic resistance genes of several integrons. *Gene* 1994;**142**:49–54.

Liebert CA, Hall RM, Summers AO. Transposon Tn21, flagship of the floating genome. *Microbiol Mol Biol Rev* 1999;**63**:507–22.

Ling AL, Pace NR, Hernandez MT *et al.* Tetracycline resistance and class 1 integron genes associated with indoor and outdoor aerosols. *Environ Sci Technol* 2013;**47**:4046–52.

Lovett ST, Hurley RL, Sutera Jr. VA *et al.* Crossing over between regions of limited homology in Escherichia coli: RecA-dependent and RecA-independent pathways. *Genetics* 2002;**160**:851–9.

M

MacDonald D, Demarre G, Bouvier M *et al.* Structural basis for broad DNA-specificity in integron recombination. *Nature* 2006;**440**:1157–62.

Maravić A, Skočibušić M, Šamanić I *et al.* Aeromonas spp. simultaneously harbouring blaCTX-M-15, blaSHV-12, blaPER-1 and blaFOX-2, in wild-growing Mediterranean mussel (Mytilus galloprovincialis) from Adriatic Sea, Croatia. *Int J Food Microbiol* 2013;**166**:301–8.

Marchesi JR, Sato T, Weightman AJ *et al.* Design and evaluation of useful bacterium-specific PCR primers that amplify genes coding for bacterial 16S rRNA. *Appl Environ Microbiol* 1998;**64**:795–9.

Márquez C, Labbate M, Ingold AJ *et al.* Recovery of a Functional Class 2 Integron from an Escherichia coli Strain Mediating a Urinary Tract Infection. *Antimicrob Agents Chemother* 2008;**52**:4153–4.

Marti E, Balcázar JL. Use of pyrosequencing to explore the benthic bacterial community structure in a river impacted by wastewater treatment plant discharges. *Res Microbiol* 2014;**165**:468–71.

Marti E, Jofre J, Balcazar JL. Prevalence of Antibiotic Resistance Genes and Bacterial Community Composition in a River Influenced by a Wastewater Treatment Plant. *PLoS ONE* 2013;**8**, DOI: 10.1371/journal.pone.0078906.

Martinez-Freijo P, Fluit AC, Schmitz F-J *et al.* Class I integrons in Gram-negative isolates from different European hospitals and association with decreased susceptibility to multiple antibiotic compounds. *J Antimicrob Chemother* 1998;**42**:689–96.

Martin RG, Rosner JL. Fis, an accessorial factor for transcriptional activation of the mar (multiple antibiotic resistance) promoter of Escherichia coli in the presence of the activator MarA, SoxS, or Rob. *J Bacteriol* 1997;**179**:7410–9.

Mazel D. Integrons: Agents of bacterial evolution. *Nat Rev Microbiol* 2006;**4**:608–20.

Mazel D, Dychinco B, Webb VA *et al.* A distinctive class of integron in the Vibrio cholerae genome. *Science* 1998;**280**:605–8.

Van der Meer JR, Zehnder AJB, De Vos WM. Identification of a novel composite transposable element, Tn5280, carrying chlorobenzene dioxygenase genes of Pseudomonas sp. strain P51. *J Bacteriol* 1991;**173**:7077–83.

Messier N, Roy PH. Integron integrases possess a unique additional domain necessary for activity. *J Bacteriol* 2001;**183**:6699–706.

Michael CA, Andrew NR. Co-assortment in integron-associated gene cassette assemblages in environmental DNA samples. 2010.

Michael CA, Gillings MR, Holmes AJ *et al.* Mobile gene cassettes: A fundamental resource for bacterial evolution. *Am Nat* 2004;**164**:1–12.

Michael CA, Labbate M. Gene cassette transcription in a large integron-associated array. *BMC Genet* 2010;**11**, DOI: 10.1186/1471-2156-11-82.

Moura A, Araújo S, Alves MS *et al.* The contribution of Escherichia coli from human and animal sources to the integron gene pool in coastal waters. *Evol Genomic Microbiol* 2014;**5**:419.

Moura A, Henriques I, Ribeiro R *et al.* Prevalence and characterization of integrons from bacteria isolated from a slaughterhouse wastewater treatment plant. *J Antimicrob Chemother* 2007;**60**:1243–50.

Moura A, Henriques I, Smalla K *et al.* Wastewater bacterial communities bring together broad-host range plasmids, integrons and a wide diversity of uncharacterized gene cassettes. *Res Microbiol* 2010;**161**:58–66.

Moura A, Jové T, Ploy M-C *et al.* Diversity of gene cassette promoters in class 1 integrons from wastewater environments. *Appl Environ Microbiol* 2012;**78**:5413–6.

N

Nakatsu C, Ng J, Singh R *et al.* Chlorobenzoate catabolic transposon Tn5271 is a composite class I element with flanking class II insertion sequences. *Proc Natl Acad Sci U S A* 1991;**88**:8312–6.

Nardelli M, Scalzo PM, Ramírez MS *et al.* Class 1 integrons in environments with different degrees of urbanization. *PLoS ONE* 2012;**7**, DOI: 10.1371/journal.pone.0039223.

Needleman SB, Wunsch CD. A general method applicable to the search for similarities in the amino acid sequence of two proteins. *J Mol Biol* 1970;**48**:443–53.

Nemergut DR, Martin AP, Schmidt SK. Integron Diversity in Heavy-Metal-Contaminated Mine Tailings and Inferences about Integron Evolution. *Appl Environ Microbiol* 2004;**70**:1160–8.

Nemergut DR, Robeson MS, Kysela RF *et al.* Insights and inferences about integron evolution from genomic data. *BMC Genomics* 2008;**9**:261.

Nield BS, Holmes AJ, Gillings MR *et al.* Recovery of new integron classes from environmental DNA. *FEMS Microbiol Lett* 2001;**195**:59–65.

Novais A, Baquero F, Machado E *et al.* International spread and persistence of TEM-24 is caused by the confluence of highly penetrating enterobacteriaceae clones and an IncA/C2 plasmid containing Tn1696::Tn1 and IS5075-Tn21. *Antimicrob Agents Chemother* 2010;**54**:825–34.

Nunes-Düby SE, Kwon HJ, Tirumalai RS *et al.* Similarities and differences among 105 members of the Int family of site-specific recombinases. *Nucleic Acids Res* 1998;**26**:391–406.

O

Oros DR, Collier RW, Simoneit BRT. The extent and significance of petroleum hydrocarbon contamination in Crater Lake, Oregon. *Hydrobiologia* 2007;**574**:85–105.

Overmann J, Fischer U, Pfennig N. A new purple sulfur bacterium from saline littoral sediments, Thiorhodovibrio winogradskyi gen. nov. and sp. nov. *Arch Microbiol* 1992;**157**:329–35.

P

Parisot N, Denonfoux J, Dugat-Bony E *et al.* KASpOD-a web service for highly specific and explorative oligonucleotide design. *Bioinformatics* 2012;**28**:3161–2.

Partridge SR, Recchia GD, Scaramuzzi C *et al.* Definition of the attl1 site of class 1 integrons. *Microbiology* 2000;**146**:2855–64.

Partridge SR, Tsafnat G, Coiera E *et al.* Gene cassettes and cassette arrays in mobile resistance integrons: Review article. *FEMS Microbiol Rev* 2009;**33**:757–84.

Peng Y, Leung HCM, Yiu SM *et al.* IDBA-UD: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. *Bioinforma Oxf Engl* 2012;**28**:1420–8.

Pfennig N, Widdel F, Trüper HG. The dissimilatory sulfate-reducing bacteria. *The Prokaryotes*. Springer. Berlin, 1981, 926–40.

Pittard J, Camakaris H, Yang J. The TyrR regulon. *Mol Microbiol* 2005;**55**:16–26.

Ploy M-C, Gassama A, Chainier D *et al.* Integrons: An antibiotic resistance gene capture system. *Immuno-Anal Biol Spec* 2005;**20**:343–52.

Ploy M-C, Lambert T, Couty J-P *et al.* Integrons: An antibiotic resistance gene capture and expression system. *Clin Chem Lab Med* 2000;**38**:483–7.

Point D, Bareille G, Pinaly H *et al.* Multielemental speciation of trace elements in estuarine waters with automated on-site UV photolysis and resin chelation coupled to inductively coupled plasma mass spectrometry. *Talanta* 2007;**72**:1207–16.

Poirel L, Carattoli A, Bernabeu S *et al.* A novel IncQ plasmid type harbouring a class 3 integron from Escherichia coli. *J Antimicrob Chemother* 2010;**65**:1594–8.

Power ML, Emery S, Gillings MR. Into the Wild: Dissemination of Antibiotic Resistance Determinants via a Species Recovery Program. *PLoS ONE* 2013;**8**, DOI: 10.1371/journal.pone.0063017.

R

Ramírez MS, Piñeiro S, Group AIS *et al.* Novel Insights about Class 2 Integrons from Experimental and Genomic Epidemiology. *Antimicrob Agents Chemother* 2010;**54**:699–706.

Rapa RA, Labbate M. The function of integron-associated gene cassettes in Vibrio species: The tip of the iceberg. *Front Microbiol* 2013;**4**, DOI: 10.3389/fmicb.2013.00385.

Recchia GD, Stokes HW, Hall RM. Characterisation of specific and secondary recombination sites recognised by the integron DNA integrase. *Nucleic Acids Res* 1994;**22**:2071–8.

Reinthaler FF, Posch J, Feierl G *et al.* Antibiotic resistance of E. coli in sewage and sludge. *Water Res* 2003;**37**:1685–90.

Rodríguez-Minguela CM, Apajalahti JHA, Chai B *et al.* Worldwide Prevalence of Class 2 Integrases outside the Clinical Setting Is Associated with Human Impact. *Appl Environ Microbiol* 2009;**75**:5100–10.

Romine MF, Stillwell LC, Wong K-K *et al.* Complete sequence of a 184-kilobase catabolic plasmid from Sphingomonas aromaticivorans F199. *J Bacteriol* 1999;**181**:1585–602.

Rosewarne CP, Pettigrove V, Stokes HW *et al.* Class 1 integrons in benthic bacterial communities: Abundance, association with Tn402-like transposition modules and evidence for coselection with heavy-metal resistance. *FEMS Microbiol Ecol* 2010;**72**:35–46.

Rosser SJ, Young H-K. Identification and characterization of class 1 integrons in bacteria from an aquatic environment. *J Antimicrob Chemother* 1999;**44**:11–8.

Rotkin-Ellman M, Solomon G, Gonzales CR *et al.* Arsenic contamination in New Orleans soil: Temporal changes associated with flooding. *Environ Res* 2010;**110**:19–25.

Rowe-Magnus DA, Davies J, Mazel D. Impact of integrons and transposons on the evolution of resistance and virulence. *Curr Top Microbiol Immunol* 2002;**264**:167–88.

Rowe-Magnus DA, Mazel D. Integrons: Natural tools for bacterial genome evolution. *Curr Opin Microbiol* 2001;**4**:565–9.

S

Sadowski PD. The Flp recombinase of the 2-mum plasmid of Saccharomyces cerevisiae. *Prog Nucleic Acid Res Mol Biol* 1995;**51**:53–91.

Sarria-Guzmán Y, López-Ramírez MP, Chávez-Romero Y *et al.* Identification of antibiotic resistance cassettes in class 1 integrons in Aeromonas spp. strains isolated from fresh fish (Cyprinus Carpio L.). *Curr Microbiol* 2014;**68**:581–6.

Scancar J, Milacic R, Strazar M *et al.* Total metal concentrations and partitioning of Cd, Cr, Cu, Fe, Ni and Zn in sewage sludge. *Sci Total Environ* 2000;**250**:9–19.

Schloss PD, Handelsman J. Introducing DOTUR, a computer program for defining operational taxonomic units and estimating species richness. *Appl Environ Microbiol* 2005;**71**:1501–6.

Seviour RJ, Mino T, Onuki M. The microbiology of biological phosphorus removal in activated sludge systems. *FEMS Microbiol Rev* 2003;**27**:99–127.

Shah SQA, Cabello FC, L'Abée-Lund TM *et al.* Antimicrobial resistance and antimicrobial resistance genes in marine bacteria from salmon aquaculture and non-aquaculture sites. *Environ Microbiol* 2014;**16**:1310–20.

Shannon CE. A Mathematical Theory of Communication. *Bell Syst Tech J* 1948;**27**:379–423.

Shibata N, Doi Y, Yamane K *et al.* PCR Typing of Genetic Determinants for Metallo-β-Lactamases and Integrases Carried by Gram-Negative Bacteria Isolated in Japan, with Focus on the Class 3 Integron. *J Clin Microbiol* 2003;**41**:5407–13.

Simberloff D. USE OF RAREFACTION AND RELATED METHODS IN ECOLOGY. *ASTM Spec Tech Publ* 1978:150–65.

Smith CJ, Danilowicz BS, Clear AK *et al.* T-Align, a web-based tool for comparison of multiple terminal restriction fragment length polymorphism profiles. *FEMS Microbiol Ecol* 2005;**54**:375–80.

Springael D, Top EM. Horizontal gene transfer and microbial adaptation to xenobiotics: New types of mobile genetic elements and lessons from ecological studies. *Trends Microbiol* 2004;**12**:53–8.

Stalder T, Barraud O, Casellas M *et al.* Integron involvement in environmental spread of antibiotic resistance. *Front Microbiol* 2012;**3**, DOI: 10.3389/fmicb.2012.00119.

Stalder T, Barraud O, Jové T *et al.* Quantitative and qualitative impact of hospital effluent on dissemination of the integron pool. *ISME J* 2014;**8**:768–77.

Stauffert M, Cravo-Laureau C, Jézéquel R *et al.* Impact of Oil on Bacterial Community Structure in Bioturbated Sediments. *PLoS ONE* 2013;**8**, DOI: 10.1371/journal.pone.0065347.

Stein LY, Arp DJ, Berube PM *et al.* Whole-genome analysis of the ammonia-oxidizing bacterium, Nitrosomonas eutropha C91: implications for niche adaptation. *Environ Microbiol* 2007;**9**:2993–3007.

Stokes HW, Hall RM. A novel family of potentially mobile DNA elements encoding site-specific gene-integration functions: integrons. *Mol Microbiol* 1989;**3**:1669–83.

Stokes HW, Hall RM. Sequence analysis of the inducible chloramphenicol resistance determinant in the TN1696 integron suggests regulation by translational attenuation. *Plasmid* 1991;**26**:10–9.

Stokes HW, Holmes AJ, Nield BS *et al.* Gene Cassette PCR: Sequence-Independent Recovery of Entire Genes from Environmental DNA. *Appl Environ Microbiol* 2001;**67**:5240–6.

Stokes HW, Nesbø CL, Holley M *et al.* Class 1 integrons potentially predating the association with tn402-like transposition genes are present in a sediment microbial community. *J Bacteriol* 2006;**188**:5722–30.

Stokes HW, O'Gorman DB, Recchia GD *et al.* Structure and function of 59-base element recombination sites associated with mobile gene cassettes. *Mol Microbiol* 1997;**26**:731–45.

T

Terrisse F. Influence des oscillations anoxie/oxie sur des communautés microbiennes hydrocarbonoclastes de sédiments intertidaux. 2014.

Terrisse F, Duran R, Vitte I *et al.* Simulation of anoxic/oxic oscillations in crude oil-degrading bioreactors. *Hydrocarbon and Lipid Microbiology Protocols*. Humana Press, 2014.

Thomas CM, Nielsen KM. Mechanisms of, and barriers to, horizontal gene transfer between bacteria. *Nat Rev Microbiol* 2005;**3**:711–21.

Tuffin IM, de Groot P, Deane SM *et al.* An unusual Tn21-like transposon containing an ars operon is present in highly arsenic-resistant strains of the biomining bacterium Acidithiobacillus caldus. *Microbiology* 2005;**151**:3027–39.

U

Uyaguari MI, Scott GI, Norman RS. Abundance of class 1-3 integrons in South Carolina estuarine ecosystems under high and low levels of anthropogenic influence. *Mar Pollut Bull* 2013;**76**:77–84.

V

Vazquez D, Mansoori GA. Identification and measurement of petroleum precipitates. *J Pet Sci Eng* 2000;**26**:49–55.

Vickers AA, Potter NJ, Fishwick CWG *et al.* Analysis of mutational resistance to trimethoprim in Staphylococcus aureus by genetic and structural modelling techniques. *J Antimicrob Chemother* 2009;**63**:1112–7.

Vinué L, Jové T, Torres C *et al.* Diversity of class 1 integron gene cassette Pc promoter variants in clinical Escherichia coli strains and description of a new P2 promoter variant. *Int J Antimicrob Agents* 2011;**38**:526–9.

W

Walker GC. General principles of DNA repair in microorganisms and implications for future research. *Birth Defects Orig Artic Ser* 1989;**25**:45–59.

Wang C, Dang H, Ding Y. Incidence of diverse integrons and b -lactamase genes in environmental Enterobacteriaceae isolates from Jiaozhou Bay, China. *World J Microbiol Biotechnol* 2008;**24**:2889–96.

Weber KA, Hedrick DB, Peacock AD *et al.* Physiological and taxonomic description of the novel autotrophic, metal oxidizing bacterium, Pseudogulbenkiania sp. strain 2002. *Appl Microbiol Biotechnol* 2009;**83**:555–65.

Widdel F, Bak F. Gram-Negative Mesophilic Sulfate-Reducing Bacteria. In: Balows A, Trüper HG, Dworkin M, et al. (eds.). *The Prokaryotes*. Springer New York, 1992, 3352–78.

Wright MS, Baker-Austin C, Lindell AH *et al.* Influence of industrial contamination on mobile genetic elements: Class 1 integron abundance and gene cassette structure in aquatic bacterial communities. *ISME J* 2008;**2**:417–28.

X

Xu H, Davies J, Miao V. Molecular characterization of class 3 integrons from Delftia spp. *J Bacteriol* 2007;**189**:6276–83.

Y

Yano H, Garruto CE, Sota M *et al.* Complete Sequence Determination Combined with Analysis of Transposition/Site-specific Recombination Events to Explain Genetic Organization

of IncP-7 TOL Plasmid pWW53 and Related Mobile Genetic Elements. *J Mol Biol* 2007;**369**:11–26.

Z

Zhu JY, Duan GC, Yang HY *et al.* Atypical class 1 integron coexists with class 1 and class 2 integrons in multi-drug resistant Shigella flexneri isolates from China. *Curr Microbiol* 2011;**62**:802–6.

Annexes

Environ Sci Pollut Res
DOI 10.1007/s11356-015-4373-2

DECAPAGE PROJECT: HYDROCARBON DEGRADATION IN COASTAL SEDIMENTS*

Effect of physical sediments reworking on hydrocarbon degradation and bacterial community structure in marine coastal sediments

Robert Duran[1,6] · Patricia Bonin[2] · Ronan Jezequel[3] · Karine Dubosc[3] · Claire Gassie[1] · Fanny Terrisse[1] · Justine Abella[1] · Christine Cagnon[1] · Cecile Militon[2] · Valérie Michotey[2] · Franck Gilbert[4,5] · Philippe Cuny[2] · Cristiana Cravo-Laureau[1]

Received: 22 November 2014 / Accepted: 12 March 2015
© Springer-Verlag Berlin Heidelberg 2015

Abstract The present study aimed to examine whether the physical reworking of sediments by harrowing would be suitable for favouring the hydrocarbon degradation in coastal marine sediments. Mudflat sediments were maintained in mesocosms under conditions as closer as possible to those prevailing in natural environments with tidal cycles. Sediments were contaminated with Ural blend crude oil, and in half of them, harrowing treatment was applied in order to mimic physical reworking of surface sediments. Hydrocarbon distribution within the sediment and its removal was followed during 286 days. The harrowing treatment allowed hydrocarbon compounds to penetrate the first 6 cm of the sediments, and biodegradation indexes (such as *n*-C_{18}/phytane) indicated that biodegradation started 90 days before that observed in untreated control mesocosms. However, the harrowing treatment had a severe impact on benthic organisms reducing drastically the macrofaunal abundance and diversity. In the harrowing-treated mesocosms, the bacterial abundance, determined by 16S rRNA gene Q-PCR, was slightly increased; and terminal restriction fragment length polymorphism (T-RFLP) analyses of 16S rRNA genes showed distinct and specific bacterial community structure. Co-occurrence network and canonical correspondence analyses (CCA) based on T-RFLP data indicated the main correlations between bacterial operational taxonomic units (OTUs) as well as the associations between OTUs and hydrocarbon compound contents further supported by clustered correlation (ClusCor) analysis. The analyses highlighted the OTUs constituting the network structural bases involved in hydrocarbon degradation. Negative correlations indicated the possible shifts in bacterial communities that occurred during the ecological succession.

Keywords Bacteria · Bioremediation · Oil pollution · Harrowing · Mesocosms · T-RFLP · Co-occurrence network · Community structure

Responsible editor: Philippe Garrigues

✉ Robert Duran
Robert.duran@univ-pau.fr

[1] Equipe Environnement et Microbiologie, MELODY Group, Université de Pau et des Pays de l'Adour, IPREM UMR CNRS 5254, BP 1155, 64013 Pau Cedex, France

[2] Aix Marseille Université, CNRS, IRD, MIO UM 110, 13288 Marseille, France

[3] Centre de Documentation, de Recherche et d'Expérimentations sur les pollutions accidentelles des Eaux, 715 rue Alain Colas, CS 41836 Brest, France

[4] Université de Toulouse; INP, UPS; EcoLab (Laboratoire écologie fonctionnelle et environnement), 118 Route de Narbonne, 31062 Toulouse, France

[5] CNRS; EcoLab, 31062 Toulouse, France

[6] Université de Pau et des Pays de l'Adour, Bâtiment IBEAS, BP 1155, 64013 Pau Cedex, France

Introduction

Oil spills are spectacular accidents threatening marine ecosystems as evidenced by the recent *Deepwater Horizon* (Gulf Mexico, 2010) disaster with approximately 700 Ktons of crude oil released. But, since the *Deepwater Horizon* catastrophe, the *Cedre* (Centre de documentation, de recherche et d'expérimentations sur les pollutions accidentelles des eaux; http://www.cedre.fr/), French institution in charge of the response to accidental water pollution, has recorded 22 less publicised marine accidents over the world. Oil spills are

Published online: 07 April 2015

estimated to release more than 1.3 million tons of hydrocarbons per year in the sea, revealing the extent of the threat. Appropriate responses, which effectiveness is depending on the pollutant type (crude oil, heavy fuel oil, diesel…), must be implemented according to the environment (sea, beaches, mudflats…) in order to mitigate the ecological impact. Most cleaning methods are well adapted to sandy and rocky beaches, but they are very difficult to transpose for cleaning complex ecosystems characterised by fine, sandy-muddy sediments such as mudflats and estuarine areas, which are highly productive areas. In such areas, physical removal is impossible and oil spills are thus difficult to control (Duran and Goñi-Urriza 2010). Biological treatments attempting to manage microorganisms, main actors in hydrocarbon biodegradation (McGenity 2014; Timmis et al. 2010), have been developed and appears to be friendly adapted methods for such areas. Bio-stimulation and bio-augmentation strategies aim to enhance the microbial activities by adding nutrients and hydrocarbon-degrading microorganism, respectively (Goni-Urriza et al. 2013). Alternative bioremediation approaches should take into account the characteristic conditions constraining the ecosystem functioning and thus use them as advantages, particularly the oxic/anoxic oscillations due to tide and sediment reworking by burrowing organisms (Cravo-Laureau and Duran 2014). The bioturbation affects the sediment matrix and the inhabiting microbial communities both via burrow construction modifying the physical structure of sediments and via ventilation modifying the chemical parameters, oxygen supply and nutrients availability (Adamek and Marsalek 2013; Kristensen et al. 2012). We recently demonstrated that applying anoxic/oxic oscillations to oiled sediments resulted in the acceleration of hydrocarbon degradation when restoring oxic conditions after an anoxic period (Vitte et al. 2013; Vitte et al. 2011). Similarly, increasing the bioturbation activity by the addition of burrowing organisms such as *Hediste diversicolor* did not affect the hydrocarbon degradation (Stauffert et al. 2013) modifying the bacteria and archaea communities' structures (Cravo-Laureau and Duran 2014; Stauffert et al. 2014), suggesting functional redundancy in hydrocarbon degradation. We hypothesised that physical sediments reworking such as harrowing will affect the microbial community structure in oil-contaminated sediments, which would be a strategy to manage microbial communities' structures and activities for, at the applied point of view, increasing hydrocarbon degradation in contaminated areas. At the academic point of view, the controlled modification of microbial communities' structures is an opportunity to gain information to decipher the rules of microbial assemblages in oil-contaminated sediments (Cravo-Laureau and Duran 2014; Morales and Holben 2011).

In the present study, mudflat sediments maintained under tidal cycles in order to ensure conditions close to those prevailing in the natural environment, were contaminated by crude oil and subjected or not to a harrowing treatment. The effects of this physical treatment on hydrocarbon degradation and on microbial communities' structure were compared to those observed in oiled sediments without harrowing treatment, used as control.

Materials and methods

Sampling and experimental setup

Field site and experimental devices have been previously described (Stauffert et al. 2013). Muddy-sand sediments were collected at the tidal basin Aber-Benoît (Treglonou, Brittany (France), 48° 33′ 12.40″ N; 4° 32′ 8.69″ W). In order to maintain the sediment structure, sediments were collected carefully with a core collector and then transferred over geotextile membrane on mesocosm-boxes (65 cm×50 cm×41 cm; approximately 30 L of sediment). Six mesocosm-boxes were set up at room temperature and connected with a device supplying natural sand-filtered and UV-treated seawater from the Oceanopolis aquarium (Brest, France). Tidal cycles (rise/fall 12 h—6 h high tide and 6 h low tide) were simulated using a water level control consisting of an up and down drainage system. Seawater was renewed at each tidal cycle (20 L of water per box). At high tide, the water level was maintained at about 5 cm above the sediment surface using a faucet ballcock. In order to determine the effect of harrowing treatment on oiled sediments, two conditions were applied in triplicate (N=6) as follows: (i) control condition (C) corresponding to oil-contaminated sediments and (ii) harrowing treatment (H) corresponding to oil-contaminated sediments in which harrowing by a rake was performed twice a week. In order to ensure a uniform contamination of sediments, the upper 2-cm layer of each mesocosm was sampled, pooled, mixed with crude oil (final concentration 24±4 mg/g of wet sediment) and homogenised. This mixture was spread over the sediment surface (2-cm layer). The oil used for the experimentation was the Russian export blend crude oil (REBCO), an Ural type crude oil, distilled at 110 °C to eliminate the more volatile hydrocarbon compounds. This oil contained approximately 60 % of saturated hydrocarbons (including 6.2 % alkanes and 53.7 % unresolved complex mixture), 25 % of aromatic compounds (including 0.8 % polyaromatic hydrocarbons (PAHs) and 24 % unresolved complex mixture), 10 % of resins and 5 % of asphaltenes. The major metals are nickel (50 mg/kg) and vanadium (100 mg/kg). The tidal cycle was initiated 2 h after reloading sediments.

Sediment subsamples were collected after 63, 151 and 286 days of incubation for chemical and molecular analyses. Four core samples (0–10 cm; falcon tube, 10 mL) were collected from each mesocosm: one for chemical analyses (triplicate conditions), and three for molecular analyses (sampling

replicates×triplicate conditions: nine replicates per condition at each sampling time). The subsamples were collected 10 min after reaching the low tide. The sampling hole was loaded with a plastic tube in order to maintain the sediment's structure. Samples were immediately frozen in liquid nitrogen after sampling and stored at −80 °C until analyses. The effect of the harrowing on sediment reworking was estimated by using fluorescent particulate tracers (luminophores) as previously described (Stauffert et al. 2013). A suspension of 100 g of orange luminophores (63–125 mm) was homogeneously spread over the sediment surface of each mesocosm 1 h after reloading the sediments surface. At the end of the experiment, two cores samples were collected (PVC corers; diameter 10 cm, length 20 cm) and sliced into 1-cm-thick sediment layers for the analysis of the macrobenthic community densities and reworking activity (bioturbation).

Macrobenthic communities' densities and bioturbation

Each 1-cm-thick layer obtained from each corer sampled at the end of the experiment was fixed with 4 % formaldehyde. The sediments were first sieved with a 500-μm mesh in order to retain macroorganisms and then with a 44-μm mesh to collect the sediment fraction containing the luminophores particles. Each sediment layer retained on the 500-μm mesh was preserved in 70 % alcohol. The macroorganisms were identified to the major taxonomic level possible and counted using a stereomicroscope at magnification ×40 (Duport et al. 2007). Each sediment layer retained on the 44-μm mesh was freeze-dried, gently crushed to powder, homogenised and luminophores were counted using a microplate reader (Biotek Synergy Mx; Majdi et al. 2014) at λex/λem 450/585 nm. This resulted in a particulate tracer distribution profile over depth, for each sediment core. The sediment reworking (i.e. production of a biodiffusion-like coefficient Db describing particle transport in the whole sedimentary column) was quantified as previously described (Stauffert et al. 2013).

Chemical analyses

Prior to analyses, each sediment core was divided into two layers (0–6 and 6–10 cm). Each layer was then split for the determination of moisture content (50 °C for 24 h), and for chemical analyses (total petroleum hydrocarbon, TPH, concentrations and alkanes/aromatics distributions).

Sediment samples were extracted with 30 mL methylene chloride at 100 °C and 2000 psi for 14 min in a Dionex ASE 350 accelerated solvent extractor. Extractions of sediment were performed twice for each sample. The organic extracts were dried over Na_2SO_4 (activated at 400 °C for 4 h) and concentrated to approximately 2 mL using a Syncore (Büchi, Germany). Extracts were fractionated using a solid phase extraction (SPE) cartridge (silica/cyanopropyl (SiO_2/C3-CN; 1.0/0.5 g, 6 mL; Interchim, France) (Alzaga et al. 2004). Saturate and aromatic fractions were eluted simultaneously with 8 mL of methylene chloride/pentane (80:20, *v*/*v*) and concentrated to approximately 2 mL.

Extracts were injected in a gas chromatography Agilent 6890N equipped with a split/splitless injector (splitless time 1 min, flow 50 mL/min) coupled to an Agilent 5975 mass selective detector (MSD) (electronic impact 70 eV, voltage 1200 V). The injector temperature was maintained at 300 °C. The interface temperature was 300 °C. The GC/MS temperature gradient was from 50 °C (1 min) to 300 °C (20 min) at 5 °C/min. The carrier gas was helium at a constant flow of 1 mL/min. The capillary column used was an HP-5 MS 30 m×0.25 mm ID×0.25 μm film thickness.

TPH contents were analysed according to the standard NF EN ISO 9377-2 protocol using the GCMS in Scan mode. Different concentrations of the oil were diluted with methylene chloride and used to calibrate the Scan mode response and then to quantify the TPH extracted from the sediment samples.

n-alkanes and PAHs semi-quantifications were performed using Single Ion Monitoring (SIM) mode with the most representative fragment (saturates) or the molecular ion (PAHs) of each compound at a minimum of 1.4 cycles/s. *n*-alkane and PAH degradation was followed with respect to 17α(H), 21β(H)-hopane (m/z=191) used as a conserved internal biomarker during the analysis (Prince et al. 1994).

Bacterial community analyses

The sediment cores were separated in two horizons (0–6 and 6–10 cm). The PowerSoil DNA isolation kit (MoBio Laboratories) was used for metagenomic DNA extraction following the manufacturer's recommendations. Briefly, sediments were centrifuged at 5000*g* for 15 min, to remove the water phase and the initial step of horizontal vortexing was performed for 30 min. DNA was eluted in 100 μL water and stored at −20 °C until use.

The primer couple 63F (5′-CAGGCCTAACACATGCAAGTC-3′), 1387R (5′-GGGCGGWGTGTAACAAGGC-3′), specific for bacteria (Marchesi et al. 1998), was used to amplify the 16S rRNA gene. The forward primer was labelled at the 5′ end with phosphoramidite fluorochrome carboxyfluorescein (FAM). The PCR conditions were as previously described (Duran et al. 2008). Briefly: initial denaturation (95 °C for 10 min) followed by 35 cycles of denaturation (95 °C for 45 s), annealing (58 °C for 45 s), and extension (72 °C for 60 s) and a terminal extension at 72 °C for 10 min. The reaction mixture (25 μL final volume) contained 12.5 μL of Amplitaq Gold 360 Master Mix 2x (Life Technologies), 0.2 μM of each primer and 1 μL of DNA template. PCR products were visualised by agarose gel electrophoresis and purified with the PCR purification kit (GE Healthcare).

The purified 16S rRNA amplified fragments were digested by 3.5 U of *Alu*I (Takara) at 37 °C in a final volume of 10 μL

for 3 h. Then, digested products (1 μL) were mixed with deionised formamide (10 μL) and 0.25 μL of internal size standard (GS-500 Rox, Applied Biosystems). Fluorescently labelled fragments were separated and detected with an ABI Prism 3130XL capillary sequencer (Applied Biosystems) run in Genetic Analyser version 3.1. Injection was performed electrokinetically at 1.2 kV for 23 s, and the runs at 15 kV were completed within 20 min. Terminal restriction fragment length polymorphism (T-RFLP) profiles were analysed using GENEMAPPER version 4.1 software (Applied Biosystems). Only terminal fragments whose size ranged from 35 to 500 bp and whose height was greater than 20 fluorescence units were considered for analysis. T-RFLP profiles were aligned by identifying and grouping homologous fragments and normalised by calculating relative abundances of each terminal-restriction fragment (T-RFs) from height fluorescence intensity, using T-REX software (Culman et al. 2009). Shannon and Evenness indexes were calculated from T-RFLP patterns using the MVSP software (Multi-Variate Statistical Package 3.12d, Kovach Computing Services, 1985–2001, UK).

Bacterial abundance

The bacterial abundance was assessed by the determination of 16S rRNA gene copy numbers. It was performed by real time PCR with Sso Advanced SYBR Green Supermix using a CFX96 Real Time System (C1000 Thermal Cycler, Bio-Rad Laboratories, 170 CA, USA). The primer set GML5F-Uni516R (Michotey et al. 2012) was used. The real time PCR cycles consisted of a denaturation step of 5 s, a hybridization step of 20 s at 56 °C with an elongation step of 40 s. Gene abundance in mesocosm DNA samples was determined in relation with a calibration standard corresponding to dilution of plasmids harbouring Gammaproteobacterial SSU rRNA gene fragment (from 9.10^6 to 9.10^3 copies). The resulting conditions lead to a Q-PCR efficiency of 97.1 % (R^2=0.994). At the end of the PCR reaction, the specificity of the amplification was checked from the first derivative of their melting curves.

Statistics and multivariate analyses

Data statistical analyses were performed with STATISTICA software, version 5.5 (StatSoft, USA). The analysis of variance (ANOVA) was performed in order to investigate the effects of treatments on bacterial abundances and TPH content. The comparisons of the biodiffusion coefficients (Queiros, #23) and of the macroorganisms' densities were performed to assess differences between the experimental conditions by means of the non-parametric Kruskal-Wallis one-way analysis of variance by ranks (Kruskal and Wallis 1952).

Bacterial communities' comparisons

In order to compare bacterial communities (T-RFLP patterns data) according to treatment and time, two dimension non-metric multidimensional scaling ordination (NMDS) based on Bray–Curtis distances was performed using the function metaMDS from the "Vegan" package in R (Oksanen et al. 2011). The significant differences between groups were tested with an analysis of similarities (ANOSIM) performed with Primer 6 software (Primer E, Plymouth, UK).

In order to determine the relationships between bacterial communities (T-RFLP patterns data) and environmental parameters (hydrocarbon concentrations and bacterial abundance), a canonical correspondence analysis (CCA) was carried out using MVSP software (Multi-Variate Statistical Package 3.12d, Kovach Computing Services, 1985–2001, UK). Clustered correlation (ClusCor) analysis was described previously (Weinstein et al. 1997). A programme for making clustered correlation is available (http://discover.nci.nih.gov.gate1.inist.fr).

Co-occurrence network for detecting microbial interactions

In order to determine the bacterial co-occurrence and mutual-exclusion relationships, a bacterial network was built by a Cytoscape plugin CoNet 1.0b6 (Co-occurrence Network inference; http://psbweb05.psb.ugent.be/conet/; Faust et al. 2012). The input matrix consisted of T-RFLP data from all the samples (triplicates of both treatments and horizons). The analysis was performed with two correlation methods (Spearman and Pearson), edge scores for the randomization with 1000 bootstraps iterations. The resulting networks were merged applying Fichers'Z P value threshold 0.05 and Bonferroni multiple-test corrections. The co-occurrence bacterial network was visualised with Cytoscape (Smoot et al. 2011).

Results and discussion

Harrowing effect on macrobenthic communities' densities and bioturbation

Because the volume of sediments required for the analyses is rather large (about 1.2 L per core), the structure and the densities of macroorganisms as well as the luminophores distribution with depth were only determined at the end of the experiment (286 days). Although there is no significant statistical difference (chi-square=2.4; p=0.1213), the harrowing treatment had a negative effect on the macrofauna community, by decreasing both the number of individuals and the species richness in comparison to controls (Table 1). This could be

Table 1 Species composition, mean densities of macrobenthic organisms (ind.m^{-2}) and biodiffusion coefficients (Db, cm^2 y^{-1}) observed in the cores (n=2 or 3) sampled at the end of the experiment (286 days) in the (C) control (oil-contaminated sediments) and in the (H) harrowing (harrowed oil-contaminated sediments) treatments

Treatments	C			H	
	Core 1	Core 2	Core 3	Core 1	Core 2
Species					
Hydrobia ulvae	11	6	11	0	0
Streblospio shrubsolii	0	0	3	0	0
Cirratulus cirratus	2	0	2	0	0
Oligochaeta sp1	51	45	22	24	14
Oligochaeta sp2	0	0	1	0	0
Oligochaeta sp3	18	30	1	26	13
Hediste diversicolor	10	14	2	2	7
Cirratulus sp.	1	0	0	0	0
Terebellidae ind.	0	0	0	1	0
Spionidae sp.	1	0	1	0	0
Heteromastus filiformis	1	0	1	0	0
Total (ind.core^{-1})	92	95	45	53	34
Mean density (ind.m^{-2})	11,384.30			6403.60	
Db (cm^2 y^{-1})	4.53±1.4			4.13±0.93	

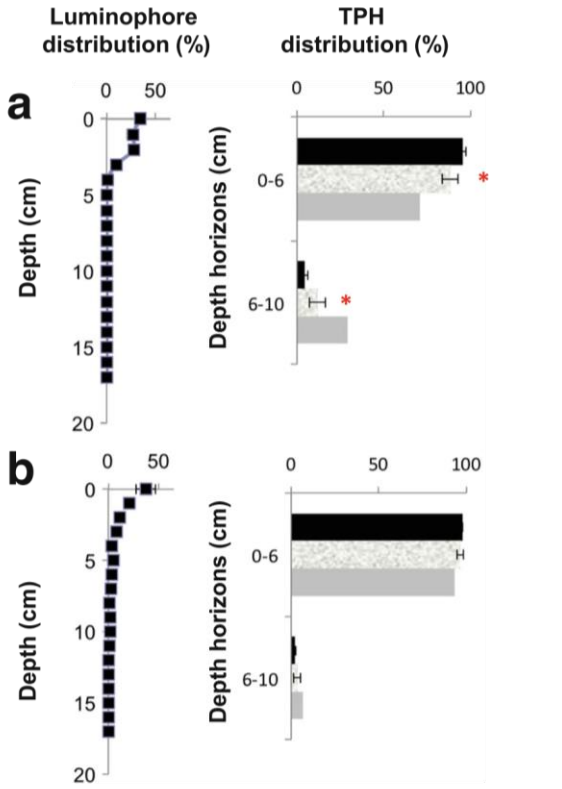


Fig. 1 Effect of the harrowing treatment (**a**) compared to the control (**b**) on sediment reworking (luminophore distribution) and total petroleum distribution (*TPH*). Luminophore distribution was determined at the end of the experiment (286 days). TPH distribution was determined at 63 days (*black*), 151 days (*white*) and 286 days (*grey*). *Error bars* indicate the standard error (n=3). *Asterisk* show significant differences between treatments (ANOVA p=0.05)

explained by the direct killing of individuals and by the destruction of the macrofaunal habitats due to the mechanical effect of harrowing. This is further supported by the absence in the harrowed sediments of *Hydrobia ulvae* (Gastropoda) and *Cirratulus cirratus* (Polychaeta) individuals, which sediment reworking and mobility traits correspond to surficial modifiers with limited movements' capacities (Queiros et al. 2013). These species are therefore most likely sensitive to perturbations of the sediment surface. By contrast, no significant difference (chi-square=0.02; p=0.8845) in the sediment mixing intensity between control (4.53±1.4 cm2 y^{-1}; mean± SD, n=3; Table 1) and harrowed treatments (4.13± 0.93 cm2 y^{-1}; mean±SD, n=3; Table 1) was observable. The luminophore distribution, however, indicated that the harrowing treatment resulted in the greater homogenization of the first 4 cm of the surface sediments in comparison to controls (Fig. 1).

Harrowing effect on hydrocarbon degradation

The mechanical mixing of the surface sediments induced the hydrocarbon compounds to be introduced deeper in the sediments in comparison to the untreated control (ANOVA p= 0.05) as shown by the distribution of total petroleum hydrocarbon (TPH) (Fig. 1). But more than 70 % of TPH were retained within the first 6 cm depth all along the experiment although the TPH proportion increased with time on the deeper horizon (6–10 cm, Fig. 1a). We thus considered the horizons 0–6 and 6–10 cm for further study.

The hopane units, normalisation using the 17α(H),21β(H)-hopane as internal standard (Prince et al. 1994), which serve as proxy to estimate petroleum biodegradation in the environment was used to follow the hydrocarbon degradation (Fig. 2a). At the end of the experimentation (286 days), the hydrocarbon (n-alkanes and PAHs) biodegradation reached 80 % in both harrowing and control conditions (Fig. 2a). Obviously, the degradation started faster in the harrowing condition with 50 % hydrocarbon degradation at 63 days whilst at the same incubation time the degradation was 23 % in the control without harrowing (Fig. 2a; ANOVA p=0.05). Similarly, the decrease of the n-C_{18}/phytane ratio (Fig. 2b) indicated a faster biodegradation in the harrowing condition than in the control but the highest difference was observed at day 151 (ANOVA p=0.05). Interestingly, at 63 days, the biodegradation of PAHs was similar to that of n-alkanes (65 %) in the harrowing condition whilst it was estimated to 49 % for n-alkanes and 37 % for PAHs in the control condition (data not shown). However, at day 286, the biodegradation of n-alkanes reaching 91 % was higher than that of PAHs estimated at

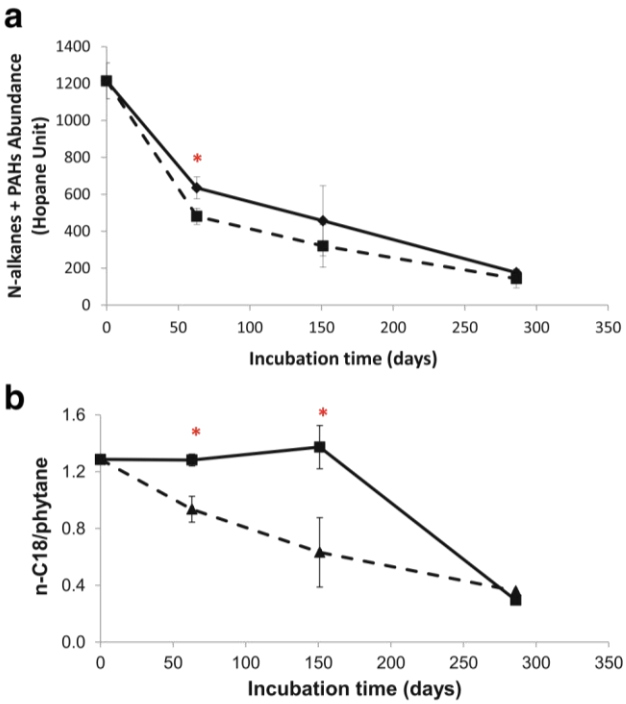


Fig. 2 Biodegradation of hydrocarbon compounds indicated by hopane (**a**) and phytane (**b**) biomarkers during the incubation under harrowing (*dashed line*) and control (*full line*) conditions on the 0–6 cm horizon. The hydrocarbon content was determined at 63, 151 and 286 days. The *Error bars* indicate standard deviation (n=3). *Asterisk* show significant differences between treatments (ANOVA p=0.05)

68 %. This observation indicated that harrowing, by homogenising the surface of sediments, facilitated the hydrocarbon biodegradation probably enabling the sediments' oxygenation and thus the oxygen attack of molecules, the first step of the aerobic hydrocarbon catabolism. We previously demonstrated in a bioreactor study that oxygenation was a critical step in hydrocarbon biodegradation (Vitte et al. 2013; Vitte et al. 2011), which is consistent with the aerobic biodegradation pathways so far described (Cerniglia 1992; McGenity 2014; Timmis et al. 2010). However, the harrowing treatment not only had a negative effect on the macrofauna, by decreasing the number of individuals from 77.3±28.0 to 43.5± 13.4 ind.core^{-1} in control and harrowed sediments, respectively, but also by reducing the species richness from 10 (control) to 4 (harrowing) at the end of experiment. This was probably due to the mechanical effect of harrowing by direct killing of individuals and by the destruction of the macrofaunal habitats.

Harrowing effect on bacterial community structure in oiled sediments

In order to determine the effect of the harrowing treatment on oil-degrading bacterial communities, microbiological analyses were performed at the latest stages of incubation (151 and 286 days) when the biodegradation of n-alkanes+PAHs was over 55 % in all conditions (Fig. 2a) but at distinct levels according to n-C_{18}/phytane ratio as for the 151 days (Fig. 2b; ANOVA p=0.05) allowing the comparison of contrasted situations. First, the bacterial abundances were determined by estimating the 16S rRNA gene copy numbers by Q-PCR. At the beginning of the experiment, the bacterial abundance was estimated to $6.5.10^{10}$±5.10^{9} 16S rRNA gene copies per gram of sediment for the 0–6 cm horizon. At 151 days, the bacterial abundance was found below 4.10^{10} 16S rRNA gene copies per gram of sediment for each condition (0–6 cm horizon; Fig. 3) showing the effect of crude oil on bacterial growth. The bacterial abundance remained stable for the control condition whilst it reached the initial level for the harrowing condition at the end of the experiment (286 days; Fig. 3a). This observation suggested that the harrowing favoured bacterial growth probably by increasing the oxygenation in the 0–6 cm horizon of the sediment. The effect of physical perturbation in marine sediments has been previously discussed (Widdicombe and Austen 2001), and the increase of diffusive oxygen uptake rate after harrowing demonstrated (Soltwedel et al. 2008). It is important to notice that in the 6–10-cm horizon, the bacterial abundance remained stable around 4.10^{10} 16S rRNA gene copies per gram of sediment until the end of experiment in the control condition (Fig. 3b) whereas it was found lower in the harrowing condition increasing slightly at 286 days (Fig. 3a). These observations indicated that harrowing influenced the bacterial communities not only at the surface (0–6 cm) of sediments but also at lower depth (6–10 cm horizon), disturbing probably bacterial populations living in anaerobic zones by introducing both hydrocarbon

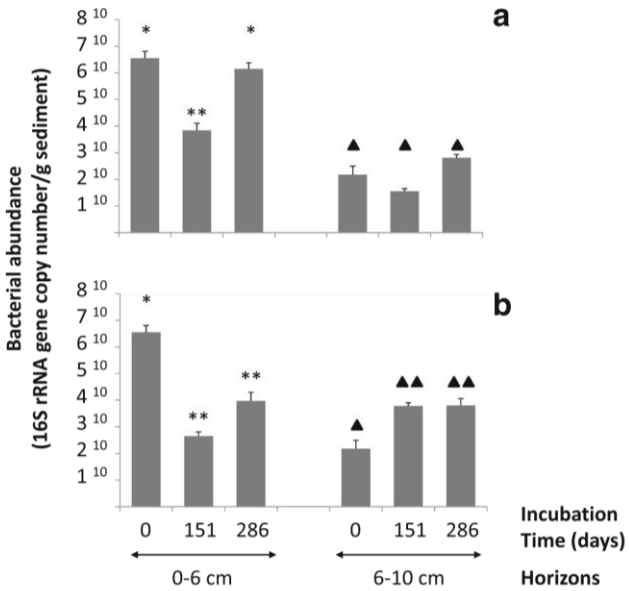


Fig. 3 Bacterial abundances in harrowed (**a**) and control (**b**) sediments. *Error bars* indicate the standard error (n=3). *Symbols* refer to homogenous groups according to ANOVA test

compounds and oxygen in deeper horizons whilst modifying the sediment structure as previously shown (Widdicombe and Austen 2001).

The bacterial community structures were then determined by T-RFLP based on 16S rRNA genes analysis. The T-RFLP patterns showed an average of 26.5±1.7 T-RFs (terminal restriction fragments; or operational taxonomic units (OTUs)). The Shannon and Evenness were almost similar (no significant differences) in all conditions with average 2.94±0.2 and 0.7±0.1 respectively.

The bacterial communities' structures from the different treatments were compared by NMDS. At 151 days, the bacterial communities' structures were similar except those from the first 0–6-cm horizon in the harrowing treatment that grouped apart (Fig. 4a; R=0.1, p=0.12) showing the effect of the harrowing treatment whereas the bacterial communities' structures were found different at 286 days (Fig. 4b; R=0.65, p=0.002) indicating that each treatment and each horizon was characterised by a specific bacterial community. Thus, considering the 0–6 cm horizon where biodegradation occurred, this observation suggested that the bacterial communities of the harrowing treatment at 286 days were distinct to those from the control treatment although the biodegradation rates were found similar. We have reported comparable observation with sediments in which bioturbation reworking was enhanced by the addition of *H. diversicolor* (Cravo-Laureau and Duran 2014; Stauffert et al. 2013). Such observation also indicated that the mechanical reworking by harrowing imposed a selection above the presence of oil, resulting in a distinct bacterial community structure. It is thus likely that the harrowing was a stronger driver for the bacterial community than oil, because the effect of oil was hidden by the harrowing effect. In response to environmental perturbations, the bacterial communities may be resistant, resilient or functionally redundant (Allison and Martiny 2008; Yachi and Loreau 1999). The results reported here could be explained either by the fact that the specific degraders group was not affected or by functional redundancy in hydrocarbon degradation. The manipulation of bacterial community structure and composition, for instance by harrowing as reported here, will provide the opportunity to decipher the mechanisms underlying functional redundancy by combining in depth characterisation of bacterial composition, by high throughput sequencing, with the analysis of appropriate genes involved in hydrocarbon degradation.

The observed changes in the bacterial T-RFLP fingerprints in the harrowed mesocosms may be also explained in part by an indirect effect as harrowing modified the structure of the macrobenthic community and the density of macroorganisms (Table 1). Indeed, previous studies demonstrated that the absence/presence of specific macroorganisms like polychaetes may influence the bacterial communities' structure not only in uncontaminated sediments (Pischedda et al. 2011) but also in oil-contaminated sediments (Cuny et al. 2007; Stauffert et al. 2013). Differences in mechanical and biological mixing in the sediments may have also induced different environmental conditions for bacterial communities.

Additionally, NMDS analyses showed a clear temporal shift of bacterial community structure between 151 and 286 days in control treatment (Fig. 4c; R=0.44, p=0.003) and harrowing treatment (Fig. 4d; R=0.47, p=0.004), which resulted to the separation of the bacterial communities of the surface (0–6 cm) from those of the deeper horizon (6–10 cm) without affecting the bacterial abundance (Fig. 3b). This observation is in accordance with ecological succession theory that hypothesises a succession of microorganisms following

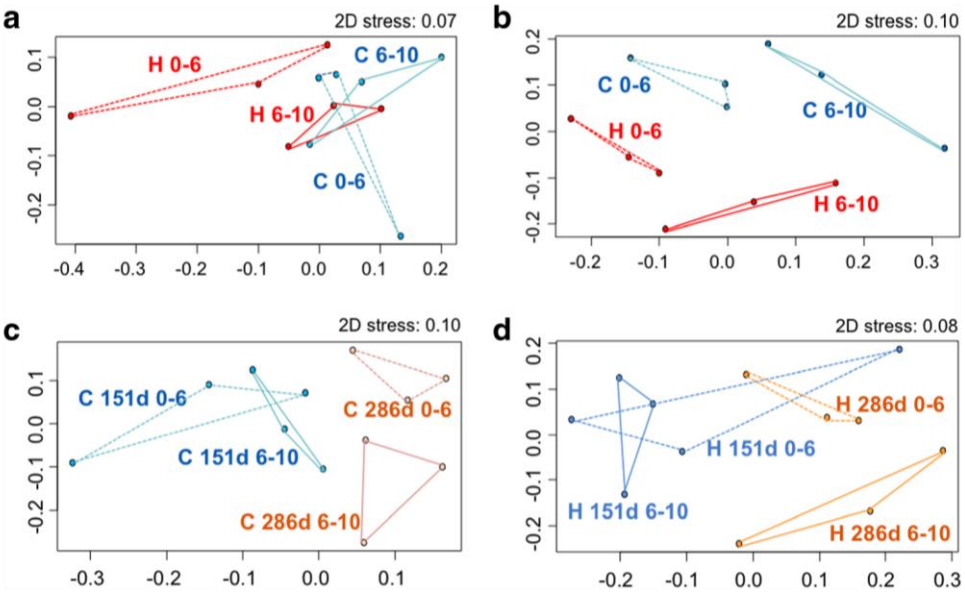


Fig. 4 Comparison of bacterial community structures by non-metric multidimensional scaling (*NMDS*) analyses based on T-RFLP 16S rRNA gene patterns. **a** and **b** effect of treatment: comparisons of bacterial community structures at 151 (**a**) and 286 (**b**) days from harrowing treatment (*H*, *red*) and control (*C*, *blue*) for the horizons 0–6 cm (*dashed lines*) and 6–10 cm (*full lines*). **c** and **d** temporal shifts: bacterial community structure shifts between 151 (*blue*) and 286(*orange*)days in sediments from control (**c**) and harrowing (**d**) treatments. Analyses were performed in triplicates, which are linked by *lines*

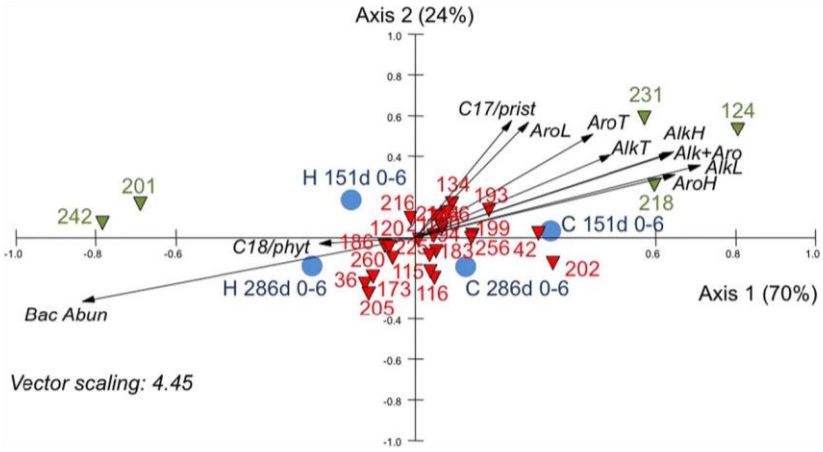


Fig. 5 Canonical correspondence analysis (*CCA*) correlating bacterial populations (*triangles*, T-RFLP OTUs) and communities (*blue circles*, T-RFLP patterns) with the hydrocarbon content (*Alk* and *Aro*, see below), biodegradation indexes *(C17/prist* and *C18/phyt*) and bacterial abundance (*Bac Abun*). The analysis is based on mean of triplicate samples for T-RFLP profiles, bacterial abundance and hydrocarbon contents. The specific OTUs (T-RFs from T-RFLP profiles) are indicated in *green* and the common OTUs in *red. Blue circles* show the bacterial communities in harrowed (*H*) and control (*C*) sediments at 151 (*151d*) and 286(*286d*) days for the horizon 0–6 cm (*0–6*). Hydrocarbon content: *AlkT* total alkane compounds, *AlkL* light molecular weight alkanes (<20 C atoms), *AlkH* high molecular weight alkanes (>20 C atoms), *AroT* total polyaromatic hydrocarbons (*PAHs*), *AroL* light molecular weight PAHs (<3 benzene rings), *AroH* high molecular weight PAHs (>3 benzene rings), *Alk+Aro* total of alkanes and PAHs compounds. Biodegradation indexes correspond to the ratio of *n*-C17 alkane/pristane (*C17/prist*) and *n*-C18 alkane/phytane (*C18/phyt*). The bacterial abundance (*Bac Abun*) corresponds to 16S rRNA gene copy numbers assessed by Q-PCR

the succession of metabolites during the degradation (Head et al. 2006). Such modifications have been observed in sediments from different habitats including highly saline (Bordenave et al. 2004a; Bordenave et al. 2004b) and marine (Chronopoulou et al. 2013; Coulon et al. 2012) ecosystems contaminated by crude oil. However, it is important to notice that the bacterial community structures of the sediments surface (0–6 cm) at 286 days were still slightly related to those at 151 days in the harrowing-treated mesocosms (Fig. 4d), endorsing that the harrowing was strongly affecting the bacterial communities.

Bacterial assemblage in oiled sediments

In order to determine the bacterial populations (OTUs) associated with the degradation capacities, we performed

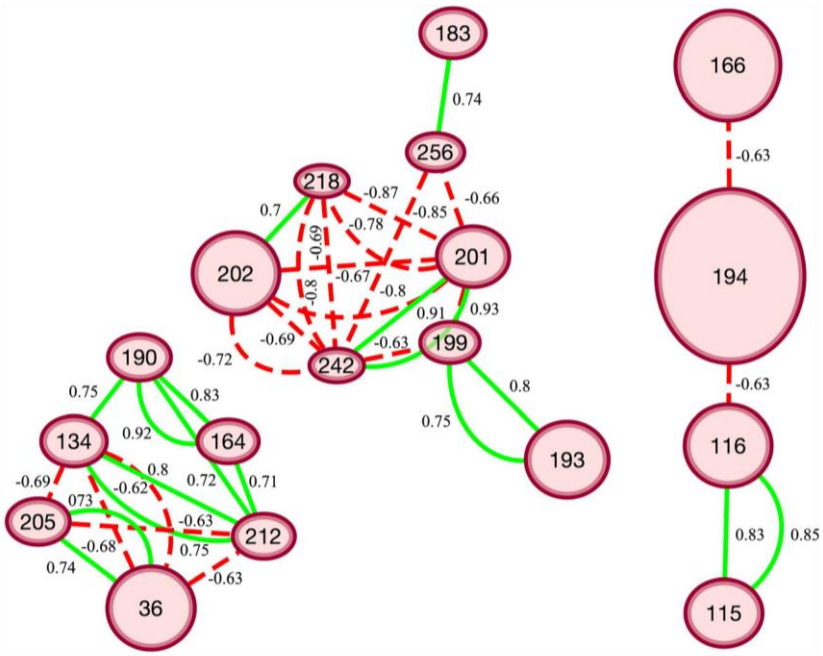


Fig. 6 Co-occurrence bacterial network showing the relationships between the OTUs. The network was inferred from T-RFLP patterns applying Pearson (*straight lines*) and Spearman (*curved lines*) correlations with 1000 bootstrap iterations, Fichers'Z *P* value threshold 0.05 and Bonferroni multiple-test corrections. The positive correlations are indicated by *green lines* and mutual exclusions by *red dashed lines. Node circles* correspond to OTUs, the size is proportional to the relative abundances considering all T-RFLP profiles. *Numbers at the edge* correspond to correlation weight

canonical correspondence analyses (CCA) correlating T-RFLP patterns with the hydrocarbon content, degradation indexes and bacterial abundance. The objective being to observe the correlations between OTUs and hydrocarbons, the CCA was performed with the samples in which biodegradation occurred, i.e. the 0–6 cm horizon of both treatments (Fig. 5). The CCA, explaining 94 % of the data distribution, showed that hydrocarbon content and bacterial abundances were in opposite directions along the axis 1 (70 % of the distribution), indicating that the presence of hydrocarbons probably limited the bacterial growth. Another explanation could be that more degradation occurred where bacteria was more abundant. Three OTUs (124, 218 and 231) were associated with high hydrocarbon content; they determine the specificity of the control bacterial community (C 151d 0–6) at 151 days when the biodegradation was low in the control mesocosms (Fig. 2), as indicated by the correlation with the n-C_{17}/pristane ratio in the CCA. These three OTUs may be tolerant to the presence of crude oil but their role in the biodegradation must be limited. In contrast, two OTUs (201 and 242) were associated with high bacterial abundance, explaining the difference of the harrowing-treated communities at 286 days (H 286d 0–6). It is likely that these two OTUs became predominant once the negative effect of hydrocarbon was attenuated; they probably play a minor role in hydrocarbon degradation. These observations are in accordance with previous studies demonstrating that some populations in hydrocarbon-degrading consortia play an auxiliary role in the bacterial community such as utilising metabolites (Dastgheib et al. 2012; Hii et al. 2009), increasing the bioavailability of the molecules (Kanaly et al. 2002), surfactant production (Ranchou-Peyruse et al. 2004) and syntrophic interactions (Hirschler-Réa et al. 2012; Kleinsteuber et al. 2008).

In the CCA, most OTUs were grouped in the centre (red in Fig. 5), suggesting that they play a pivotal role in the bacterial community organisation. In order to test the hypothesis that the central OTUs in the CCA constitute the framework of the bacterial assemblage in oiled sediments, we conducted a co-occurrence network analysis. Because the objective was to determine the correlations between OTUs in oiled sediments, all the data set (both treatments and horizons) was included in the analysis in order to exclude the relationships that may occur in non-degrading conditions. Interestingly, the co-occurrence network analysis confirmed that most of these OTUs had a pivotal role in the microbial assemblage (Fig. 6), revealing three sub-networks.

The first sub-network formed by the most abundant OTUs (194, 166) shows mutual exclusion with no positive relation with other OTUs. It is likely that these OTUs, although dominant, did not play a crucial role in the bacterial community as indicated by the CCA (Fig. 5). The second sub-network was formed by four couples (183/256, 193/199, 218/202 and 201/242) with strong positive correlations (Pearson >0.7; p value< 0.05). These couples excluded each other, suggesting ecological succession since negative associations are considered as structural shifts (Faust et al. 2012). Among them, it is important to notice that the couple formed by OTUs 201 and 242 were found in the harrowing treatment at 286 days as highlighted in the CCA (Fig. 5). Furthermore, these couples showed different correlation with hydrocarbon compounds as revealed by the clustered correlation (ClusCor) analysis (Fig. 7), suggesting different degradation capacities and thus supporting the ecological succession. The couples 183/256 and 201/242 showed no or weak correlation (−0.17<Pearson>0.2) and even exclusion (Pearson<−0.3) with hydrocarbon compounds, respectively. In contrast, the couple 193/199

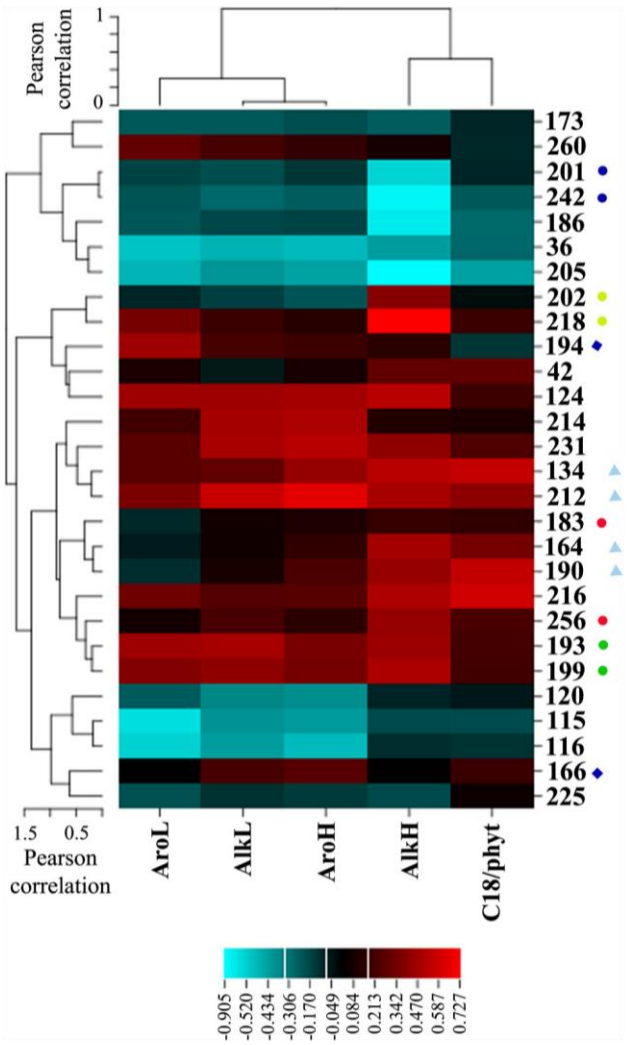


Fig. 7 Clustered correlation (*ClusCor*) between T-RFLP OTUs and chemical data. Colour legend shows Pearson correlation, *red* indicating high positive correlation and *blue* high negative correlation (exclusion). Clustering was performed applying Pearson correlation and complete linkage as cluster method. Abbreviations of the environmental parameters are those indicated in Fig. 5. *Diamonds* indicate OTUs of the sub-network 1, *coloured circles* the different OTUs couples of the sub-network 2 and *triangles* the OTUs of the sub-network 3

showed correlation with all hydrocarbons (Pearson>0.47) whereas the couple 218/202 was strongly correlated to high molecular weight (MW) alkanes (Pearson>0.7) and low MW aromatic compounds (0.47<Pearson>0.58; Fig. 7).

The co-occurrence network analysis did not found correlations for OTUs 124 and 231, which were associated to the hydrocarbon content in the CCA and the ClusCor analysis (Figs. 5 and 7). This observation confirms their limited role in the bacterial assemblage. It is likely that these OTUs were present transiently in the mesocosms. The third sub-network showed the central role of four OTUs (134, 164, 190 and 212) ensuring most associations, supported by positive Pearson correlations (>0.7; p value<0.05), constituting thus the network structural bases. The ClusCor analysis showed that the OTUs 212 and 134 were correlated with all types of hydrocarbon compounds whilst OTUs 164 and 190 were correlated with high MW alkanes and excluded low MW aromatics (Fig. 7). Noteworthy, the OTU 36 was associated with the OTU 205 but showed mutual exclusion with OTUs 134 and 212, major OTUs in the frame of the network, the OTUs 36 and 205 showing mutual exclusion with hydrocarbon compounds (Fig. 7). This observation is in accordance with the NMDS and CCA analyses (Figs. 4 and 5), suggesting that this negative association may correspond to a structural shift due to the effect of petroleum. The bacterial community modification has been demonstrated in many cases after petroleum contamination and resulting in ecological succession (Bordenave et al. 2004a; Bordenave et al. 2007; Head et al. 2006; Païssé et al. 2008). Similar modifications have been also reported for other perturbations such as biotic reworking activity by polychaetes (Christensen et al. 2002; Gilbert et al. 1996; Grossi et al. 2002; Stauffert et al. 2013), crustaceans (Bertics and Ziebis 2009), meiofauna (Louati et al. 2013) and natural perturbations (Paerl et al. 2002). Although the network was constructed with low-resolution fingerprinting method comparing to high throughput sequencing methods, it demonstrates the complex interactions between bacteria that constitute bacterial communities. Further studies are required to decipher the bacterial interactions and their role in hydrocarbon degradation. The in-depth characterisation of microbial communities by high throughput sequencing technology is a promising approach (Cravo-Laureau and Duran 2014) to describe microbial assemblage that may provide useful information to manage microbial degradation capacities.

Conclusion

The control of bacterial assemblages and activities is the major challenge for the implementation of biological processes for mitigating the effect of pollutants. Several approaches including bio-stimulation and bio-augmentation have been developed to favour and enhance bacterial activities involved in biodegradation (Goni-Urriza et al. 2013). In this study, we demonstrated that the mechanical sediment reworking by harrowing favoured the initiation of bacterial hydrocarbon degradation activities. However, the intensive harrowing treatment performed in this study, applied twice a week, had a damaging effect on the macrofauna that may be deleterious for the whole ecosystem functioning. The usefulness of such harrowing treatment should thus be carefully examined before to envisage applying it for bioremediation processes. But overall, the harrowing treatment resulted in a distinct bacterial assemblage than that observed in the control. At the academic point of view, studying the modification of bacterial communities is a useful way to understand the assemblage rules driving bacterial communities' organisation; harrowing is an easy way to manipulate bacterial communities.

Acknowledgments We acknowledge the support of the French programme ANR DECAPAGE (project ANR-CESA-2011-006 01). We would like to thank all partners of the DECAPAGE project and MELODY group for their useful discussions. We acknowledge the Regional Platform for Environmental Microbiology PREMICE supported by the Aquitaine Regional Government Council (France) and the urban community of Pau-Pyrénées (France). We thank the Direction Générale de l'Armement (Edgar) and the Aquitaine Regional Government Council (France) for financial support to JA. We acknowledge B. Bouyssiere, S. Guasco, M.S. Goñi-Urriza and J. Gury for technical assistance during sampling campaign.

References

Adamek Z, Marsalek B (2013) Bioturbation of sediments by benthic macroinvertebrates and fish and its implication for pond ecosystems: a review. Aquacult Int 21:1–17

Allison SD, Martiny JBH (2008) Resistance, resilience, and redundancy in microbial communities. Proc Natl Acad Sci U S A 105:11512–11519

Alzaga R, Montuori P, Ortiz L, Bayona JM, Albaigés J (2004) Fast solid-phase extraction-gas chromatography-mass spectrometry procedure for oil fingerprinting: application to the prestige oil spill. J Chromatogr 1025:133–138

Bertics VJ, Ziebis W (2009) Biodiversity of benthic microbial communities in bioturbated coastal sediments is controlled by geochemical microniches. ISME J 3:1269–1285

Bordenave S, Fourçans A, Blanchard S, Goni-Urriza MS, Caumette P, Duran R (2004a) Structure and functional analyses of bacterial communities changes in microbial mats following petroleum exposure. Ophelia 58:195–203

Bordenave S, Jézéquel R, Fourçans A, Budzinski H, Merlin FX, Fourel T, Goñi-Urriza M, Guyoneaud R, Grimaud R, Caumette P, Duran R (2004b) Degradation of the "Erika" oil. Aquat Living Resour 17: 261–267

Bordenave S, Goñi-Urriza MS, Caumette P, Duran R (2007) Effects of heavy fuel oil on the bacterial community structure of a pristine microbial mat. Appl Environ Microbiol 73:6089–6097

Cerniglia CE (1992) Biodegradation of polycyclic aromatic hydrocarbons. Biodegradation 3:351–368

Christensen M, Banta GT, Andersen O (2002) Effects of the polychaetes Nereis diversicolor and Arenicola marina on the fate and distribution of pyrene in sediments. Mar Ecol Prog Ser 237:159–172

Chronopoulou PM, Fahy A, Coulon F, Païssé S, Goñi Urriza M, Acuña Alvarez L, McKew BA, Lawson T, Timmis KN, Duran R, Underwood GJC, McGenity TJ (2013) Impact of a simulated oil spill on benthic phototrophs and nitrogen-fixing bacteria in mudflat mesocosms. Environ Microbiol 15:242–252

Coulon F, Chronopoulou P-M, Fahy A, Sandrine P, Goñi-Urriza M, Peperzak L, Acuña Alvarez L, McKew BA, Brussaard CPD, Underwood GJC, Timmis KN, Duran R, McGenity TJ (2012) Central role of dynamic tidal biofilms dominated by aerobic hydrocarbonoclastic bacteria and diatoms in the biodegradation of hydrocarbons in coastal mudflats. Appl Environ Microbiol 78: 3638–3648

Cravo-Laureau C, Duran R (2014) Marine coastal sediments microbial hydrocarbon degradation processes: contribution of experimental ecology in the omics'era. Front Microbiol 5:39

Culman SW, Bukowski R, Gauch HG, Cadillo-Quiroz H, Buckley DH (2009) T-REX: Software for the processing and analysis of T-RFLP data. BMC Bioinformatics 10

Cuny P, Miralles G, Cornet-Barthaux V, Acquaviva M, Stora G, Grossi V, Gilbert F (2007) Influence of bioturbation by the polychaete Nereis diversicolor on the structure of bacterial communities in oil contaminated coastal sediments. Mar Pollut Bull 54: 452–459

Dastgheib SMM, Amoozegar MA, Khajeh K, Shavandi M, Ventosa A (2012) Biodegradation of polycyclic aromatic hydrocarbons by a halophilic microbial consortium. Appl Microbiol Biotechnol 95: 789–798

Duport E, Gilbert F, Poggiale JC, Dedieu K, Rabouille C, Stora G (2007) Benthic macrofauna and sediment reworking quantification in contrasted environments in the Thau Lagoon. Estuar Coast Shelf Sci 72:522–533

Duran R, Goñi-Urriza MS (2010) Impact of pollution on microbial mats. In: Timmis K (ed) Handbook of Hydrocarbon and Lipid Microbiology. Springer Berlin Heidelberg, pp 2339-2348. doi:10.1007/978-3-540-77587-4_170

Duran R, Ranchou-Peyruse M, Menuet V, Monperrus M, Bareille G, Goñi MS, Salvado JC, Amouroux D, Guyoneaud R, Donard OFX, Caumette P (2008) Mercury methylation by a microbial community from sediments of the Adour estuary (Bay of Biscay, France). Environ Pollut 156:951–958

Faust K, Sathirapongsasuti JF, Izard J, Segata N, Gevers D, Raes J, Huttenhower C (2012) Microbial co-occurrence relationships in the human microbiome. PLoS Comp Biol 8

Gilbert F, Stora G, Bertrand JC (1996) In situ bioturbation and hydrocarbon fate in an experimental contaminated Mediterranean coastal ecosystem. Chemosphere 33:1449–1458

Goni-Urriza M, Cravo-Laureau C, Duran R (2013) Microbial bioremediation of aquatic environments. In: Férard J, Blaise C (eds) Encyclopedia of aquatic ecotoxicology, vol 2. Springer, Berlin, pp 709–720

Grossi V, Massias D, Stora G, Bertrand JC (2002) Burial, exportation and degradation of acyclic petroleum hydrocarbons following a simulated oil spill in bioturbated Mediterranean coastal sediments. Chemosphere 48:947–954

Head IM, Jones DM, Röling WF (2006) Marine microorganisms make a meal of oil. Nat Rev Microbiol 4:173–182

Hii YS, Law AT, Shazili NAM, Abdul-Rashid MK, Lee CW (2009) Biodegradation of Tapis blended crude oil in marine sediment by a consortium of symbiotic bacteria. Int Biodeterior Biodegrad 63: 142–150

Hirschler-Réa A, Cravo-Laureau C, Casalot L, Matheron R (2012) Methanogenic octadecene degradation by syntrophic enrichment culture from brackish sediments. Curr Microbiol 65:561–567

Kanaly RA, Harayama S, Watanabe K (2002) Rhodanobacter sp strain BPC1 in a benzo a pyrene-mineralizing bacterial consortium. Appl Environ Microbiol 68:5826–5833

Kleinsteuber S, Schleinitz KM, Breitfeld J, Harms H, Richnow HH, Vogt C (2008) Molecular characterization of bacterial communities mineralizing benzene under sulfate-reducing conditions. FEMS Microbiol Ecol 66:143–157

Kristensen E, Penha-Lopes G, Delefosse M, Valdemarsen T, Quintana CO, Banta GT (2012) What is bioturbation? The need for a precise definition for fauna in aquatic sciences. Mar Ecol Prog Ser 446:285–302

Kruskal WH, Wallis WA (1952) Use of ranks in one-criterion variance analysis. J Am Stat Assoc 47:583–621

Louati H, Ben Said O, Soltani A, Got P, Mahmoudi E, Cravo-Laureau C, Duran R, Aissa P, Pringault O (2013) The roles of biological interactions and pollutant contamination in shaping microbial benthic community structure. Chemosphere 93:2535–2546

Majdi N, Bardon L, Gilbert F (2014) Quantification of sediment reworking by the Asiatic clam Corbicula fluminea Muller, 1774. Hydrobiologia 732:85–92

Marchesi JR, Sato T, Weightman AJ, Martin TA, Fry JC, Hiom SJ, Wade WG (1998) Design and evaluation of useful bacterium-specific PCR primers that amplify genes coding for bacterial 16S rRNA. Appl Environ Microbiol 64:795–799

McGenity TJ (2014) Hydrocarbon biodegradation in intertidal wetland sediments. Curr Opin Biotechnol 27:46–54

Michotey V, Guasco S, Boeuf D, Morezzi N, Durieux B, Charpy L, Bonin P (2012) Spatio-temporal diversity of free-living and particle-attached prokaryotes in the tropical lagoon of Ahe atoll (Tuamotu Archipelago) and its surrounding oceanic waters. Mar Pollut Bull 65:525–537

Morales SE, Holben WE (2011) Linking bacterial identities and ecosystem processes: can 'omic' analyses be more than the sum of their parts? FEMS Microbiol Ecol 75:2–16

Oksanen J, Blanchet FG, Kindt R, Legendre P, Minchin PR, O'Hara RB, Simpson GL, Solymos P, Henry M, Stevens H, Wagner H (2011) vegan: Community ecology package. R package version 2.0-2. http://CRAN.R-project.org/package=vegan

Paerl HW, Dyble J, Twomey L, Pinckney JL, Nelson J, Kerkhof L (2002) Characterizing man-made and natural modifications of microbial diversity and activity in coastal ecosystems. Antonie Van Leeuwenhoek 81:487–507

Païssé S, Coulon F, Goñi Urriza M, Peperzak L, McGenity TJ, Duran R (2008) Structure of bacterial communities along a hydrocarbon contamination gradient in a coastal sediment. FEMS Microbiol Ecol 66: 295–305

Pischedda L, Militon C, Gilbert F, Cuny P (2011) Characterization of specificity of bacterial community structure within the burrow environment of the marine polychaete Hediste (Nereis) diversicolor. Res Microbiol 162:1033–1042

Prince RC, Elmendorf DL, Lute JR, Hsu CS, Haith CE, Senius JD, Dechert GJ, Douglas GS, Butler EL (1994) 17-Alpha(H),21-beta(H)-hopane as a conserved internal marker for estimating the biodegradation of crude-oil. Environ Sci Technol 28:142–145

Queiros AM, Birchenough SNR, Bremner J, Godbold JA, Parker RE, Romero-Ramirez A, Reiss H, Solan M, Somerfield PJ, Van Colen C, Van Hoey G, Widdicombe S (2013) A bioturbation classification of European marine infaunal invertebrates. Ecol Evol 3:3958–3985

Ranchou-Peyruse A, Moppert X, Hourcade E, Hernandez-Raquet G, Caumette P, Guyoneaud R (2004) Characterization of brackish anaerobic bacteria involved in hydrocarbon degradation: a combination of molecular and culture-based approaches. Ophelia 58:255–262

Smoot ME, Ono K, Ruscheinski J, Wang PL, Ideker T (2011) Cytoscape 2.8: new features for data integration and network visualization. Bioinformatics 27:431–432

Soltwedel T, Lansard B, Gilbert F, Hasemann C, Bell E, Sablotny B, Eagle M, Kershaw P, Rabouille C (2008) An "Integrated Sediment Disturber" (ISD) to study the impact of repeated physical perturbations on sediment geochemistry and the small benthic biota. Limnol Oceanogr Methods 6:307–318

Stauffert M, Cravo-Laureau C, Jézéquel R, Barantal S, Cuny P, Gilbert F, Cagnon C, Militon C, Amouroux D, Mahdaoui F, Bouyssiere B, Stora G, Merlin FX, Duran R (2013) Impact of oil on bacterial community structure in bioturbated sediments. PLoS ONE 8

Stauffert M, Duran R, Gassie C, Cravo-Laureau C (2014) Response of archaeal communities to oil spill in bioturbated mudflat sediments. Microb Ecol 67:108–119

Timmis KN, McGenity TJ, Meer JR, Lorenzo V (2010) Handbook of hydrocarbon and lipid microbiology. Springer, Berlin

Vitte I, Duran R, Jezequel R, Caumette P, Cravo-Laureau C (2011) Effect of oxic/anoxic switches on bacterial communities and PAH biodegradation in an oil-contaminated sludge. Environ Sci Pollut Res 18: 1022–1032

Vitte I, Duran R, Hernandez-Raquet G, Mounier J, Jézéquel R, Bellet V, Balaguer P, Caumette P, Cravo-Laureau C (2013) Dynamics of metabolically active bacterial communities involved in PAH and toxicity elimination from oil-contaminated sludge during anoxic/oxic oscillations. Appl Microbiol Biotechnol 97:4199–4211

Weinstein JN, Myers TG, Oconnor PM, Friend SH, Fornace AJ, Kohn KW, Fojo T, Bates SE, Rubinstein LV, Anderson NL, Buolamwini JK, vanOsdol WW, Monks AP, Scudiero DA, Sausville EA, Zaharevitz DW, Bunow B, Viswanadhan VN, Johnson GS, Wittes RE, Paull KD (1997) An information-intensive approach to the molecular pharmacology of cancer. Science 275:343–349

Widdicombe S, Austen MC (2001) The interaction between physical disturbance and organic enrichment: an important element in structuring benthic communities. Limnol Oceanogr 46:1720–1733

Yachi S, Loreau M (1999) Biodiversity and ecosystem productivity in a fluctuating environment: the insurance hypothesis. Proc Natl Acad Sci U S A 96:1463–1468

s

RESUME

Les intégrons sont des éléments génétiques bactériens. Découverts récemment dans le contexte clinique, ils sont présents dans le génome d'un certain nombre de bactéries provenant d'environnements très variés. Ils sont composés d'un gène codant une intégrase et d'une succession de cassettes de gènes. L'activité de l'intégrase permet l'acquisition, la perte ou le réarrangement des cassettes. Par ailleurs un promoteur permet l'expression des cadres de lecture contenus dans les cassettes de gène. Ainsi les intégrons sont à la fois des réservoirs de gènes et des systèmes d'expression de ces gènes. Dans le contexte clinique, ils sont connus pour être impliqués dans l'adaptation des bactéries pathogènes. Ils sont en effet capables d'acquérir et de diffuser des gènes conférant un avantage sélectif face à la pression exercée par l'usage des antibiotiques et des biocides, et par ailleurs d'être mobilisés afin d'être transférés horizontalement.

Quelques études ont porté sur les intégrons en dehors des environnements cliniques. Elles ont permis de caractériser de nombreuses cassettes de gènes, sans toutefois en atteindre toute la diversité, à partir de bactéries ou de communautés bactériennes issues d'environnements soumis à différents niveaux de contaminations. Cependant, la diversité de l'intégrase a été peu étudiée, car le plus souvent les études se sont limitées aux séquences d'intégrons cliniques. Ainsi, les intégrons environnementaux sont encore mal connus et mal caractérisés. Les objectifs de ma thèse étaient de caractériser la diversité des intégrons environnementaux, avec un focus particulier sur les intégrases, à partir d'environnements soumis à des contaminations chimiques variables, dans le but d'évaluer le rôle possible des intégrons dans l'adaptation des bactéries face à des perturbations environnementales.

Au cours de ces travaux, environ 800 séquences d'intégrases différentes, pour la plupart encore inconnues, ont été obtenues à partir de différents sédiments d'eau douce et côtiers. Des études *in situ* et en microcosmes, d'environnements d'eau douce ou de milieux côtiers, et avec différents types et niveaux de polluants, ont permis de mettre en évidence un impact des contaminations du milieu sur la diversité des intégrons, sur l'intégrase comme sur les cassettes de gène, de manière indépendante à la structure de la communauté bactérienne. Enfin, lors de cette thèse a été réalisée la caractérisation d'un intégron potentiellement adaptatif face à une pollution pétrolière, porteur de la séquence *intIOPS* mise en évidence et nommée par Lionel Huang lors de sa thèse.

Finalement, les résultats obtenus lors de cette thèse ont apporté de nouveaux éléments qui viennent soutenir notre hypothèse principale que les intégrons environnementaux seraient impliqués dans l'adaptation des communautés bactériennes en réponse à la présence de contaminants dans les milieux non cliniques.

MOTS-CLES

Intégrons environnementaux, diversité, intégrase, cassettes de gène, contaminations chimiques, pollution pétrolière, contamination aux métaux