Dynamique évolutive des éléments transposables de type séquence d'insertion dans les génomes des bactéries endosymbiotiques Wolbachia

Nicolas Cerveau

Université de Poitiers

THESE

pour l'obtention du Grade de

DOCTEUR DE L'UNIVERSITE DE POITIERS

(Diplôme National – Arrêté du 7 août 2006)

Spécialité : **Biologie de l'environnement, des populations, écologie**

préparée au laboratoire **Écologie Évolution Symbiose**

dans le cadre de l'École Doctorale **Sciences pour l'environnement Gay Lussac**

présentée et soutenue publiquement par

Nicolas CERVEAU

le 6 décembre 2011

Dynamique évolutive des éléments transposables de type séquence d'insertion dans les génomes des bactéries endosymbiotiques *Wolbachia*

Directeur de Thèse : **Richard CORDAUX**

Co-Directeur de Thèse : **Didier BOUCHON**

Jury

Emmanuelle LERAT	Chargé de Recherche CNRS	Rapporteur
Michael CHANDLER	Directeur de Recherche CNRS	Rapporteur
Pierre CAPY	Professeur des Universités	Examinateur
Richard CORDAUX	Chargé de Recherche CNRS	Examinateur
Didier BOUCHON	Professeur des Universités	Examinateur

Toutes les vérités sont faciles à comprendre une fois découvertes,

à nous de les découvrir.

Galilée

Résumé

Les éléments transposables (ET) sont l'une des forces majeures ayant participé à l'évolution des génomes, c'est pourquoi leur dynamique évolutive est beaucoup étudiée. Les ET ont été particulièrement bien caractérisés chez les eucaryotes, notamment grâce à la présence de nombreuses copies dégradées dans les génomes qui sont les fossiles de leur activité passée. A contrario, chez les procaryotes, les ET détectés sont considérés comme très récemment insérés dans les génomes, ce qui permet plus difficilement d'étudier leur dynamique à long terme. Les modèles de la dynamique des ET procaryotes prédisent un turn-over très rapide qui débute par l'arrivée d'ET dans un génome via un transfert horizontal, suivie d'une forte augmentation du nombre de copies et de leur perte rapide. Mes travaux de thèse avaient pour but de caractériser la dynamique évolutive des séquences d'insertion (IS) dans les génomes des bactéries endosymbiotiques *Wolbachia*. Les IS sont le type d'ET avec la structure la plus simple et le plus fréquent chez les procaryotes. Les modèles prédisent une très faible abondance voire une absence d'IS dans les génomes des bactéries endosymbiotiques. Néanmoins, les génomes de *Wolbachia* font partie des génomes bactériens qui présentent la plus forte abondance en IS. *Wolbachia* est une bactérie intracellulaire stricte présente chez de nombreuses espèces d'arthropodes et de nématodes. Les souches de *Wolbachia* sont généralement parasites du sexe chez les arthropodes (entraînant de nombreuses altérations de la reproduction) et mutualistes chez les nématodes.

Nous avons combiné à la fois une approche *in silico*, avec l'analyse de 5 génomes de *Wolbachia*, et une approche expérimentale de biologie moléculaire. Nous avons ainsi pu obtenir des confirmations empiriques des modèles de dynamique évolutive des IS en caractérisant leur dynamique passée (à court et à long terme) dans les génomes de *Wolbachia*.

Nous avons mis en évidence que les génomes de *Wolbachia* contiennent de 52 à 372 copies d'IS. La couverture des IS représente jusqu'à 14,2% du génome, ce qui est une des

plus fortes couvertures rapportées à ce jour chez les procaryotes. De plus, nous avons déterminé qu'une proportion exceptionnelle des copies identifiées était non fonctionnelle (>70%). Cette importante quantité de copies non fonctionnelles nous a permis d'étudier les mécanismes de dégradation des copies d'IS. Nos résultats montrent que la dégradation des IS s'est produite selon au moins deux modes : le premier entraîne une lente dégradation des copies par accumulation de substitutions et de petites délétions alors que le second, moins fréquent, entraîne une dégradation très rapide des copies par accumulation de grandes délétions.

De plus, l'étude de la divergence nucléotidique et de la taille relative des copies d'IS a mis en évidence que l'activité de transposition des IS n'est pas constante au cours du temps. Les deux phases de forte activité des IS observées dans les génomes de *Wolbachia*, une ancienne et une récente (peut-être toujours en cours), sont séparées par une phase d'activité plus réduite. A notre échelle de résolution, les deux phases de forte activité sont quasiment synchronisées dans les différents génomes et pourraient coïncider avec des évènements de diversification de souches de *Wolbachia*. Des simulations suggèrent que les phases de forte activité des IS nécessitent d'être précédées par l'arrivée indépendante de nouveaux IS via des transferts horizontaux. La recherche par PCR de 17 groupes d'IS dans 22 souches de *Wolbachia* a mis en évidence un grand nombre de transferts horizontaux, en cohérence avec la phase d'activité récente observée dans les génomes étudiés. Cette phase d'activité récente semble s'être déroulée de façon indépendante dans les génomes étudiés car les groupes d'IS ayant subi une forte expansion ne sont pas les mêmes dans chacun des génomes.

Nous nous sommes focalisés sur l'expression des IS du génome de *w*VulC. L'ensemble des séquences des groupes d'IS étudiés présente des signaux d'initiation de la transcription. Néanmoins, la plupart sont localisés après le codon d'initiation du gène de transposase (protéine permettant la transposition). Il ne pourrait donc y avoir que transcription

partiel du gène codant la transposase Ceci suggère que l'environnement génomique des IS joue un rôle important dans la transcription complète des IS et donc dans le mécanisme de transposition. L'analyse de l'expression de groupes d'IS de *w*VulC a mis en évidence que la majorité des groupes d'IS sont transcrits quelque soit le statut de fonctionnalité des copies qui les composent (potentiellement fonctionnelles ou non). L'étude de l'expression de toutes les copies d'un groupe d'IS a mis en évidence que les régions flanquantes en amont des copies sont co-transcrites avec les copies. Ceci suggère que la transcription pourrait être initiée dans les régions flanquantes et confirme donc l'importance de l'environnement génomique. De plus, nous avons montré que les copies d'IS d'un groupe ne sont pas toutes transcrites.

En conclusion, notre travail a permis de mieux comprendre la dynamique évolutive des IS dans les génomes de *Wolbachia*. Il montre que : (1) la dégradation des IS peut se faire selon un mode lent ou un autre plus rapide (mais moins fréquent), (2) l'activité de transposition des IS a varié au cours du temps, (3) les phases de forte activité de transposition nécessitent l'apport de nouvelles copies d'IS via des transferts horizontaux qui ont été très fréquents dans un passé proche, (4) les séquences des IS possèdent les capacités de promouvoir leur propre transcription mais dans la plupart des cas les ARNm produites sont incomplets, ce qui suggère l'importance de l'environnement génomique des copies et (5) les groupes d'IS sont transcrits quelque soit le statut fonctionnel des copies. Ces phénomènes mis en évidence chez *Wolbachia* pourraient expliquer les fortes variations d'abondance d'IS parfois observées entre des souches de bactéries phylogénétiquement proches, la présence majoritaire de copies récentes dans les génomes bactériens et pourraient être applicables, au moins partiellement, à d'autres génomes bactériens.

Mots clés : éléments transposables, séquences d'insertions, dynamique évolutive, procaryotes, *Wolbachia*

Abstract

Transposable elements (TE) are one of the major driving forces of genome evolution and so their evolutionary dynamics is particularly studied. Long-term TE evolution can readily be reconstructed in eukaryotes thanks to many degraded copies constituting genomic fossil records of past TE proliferations. By contrast, in prokaryotes, TE are considered to be recently acquired, which complicates the study of their evolutionary dynamics. Dynamic models on prokaryotic TE predict a high turn-over that starts by TE acquisition by horizontal transfers, followed by copy number increase and subsequent elimination. My Ph.D. thesis work aimed to characterise the evolutionary dynamics of insertion sequences (IS) in the genomes of the bacterial endosymbiont *Wolbachia*. IS are the simplest and the most abundant TE in prokaryotic genomes. Models predict that bacterial endosymbiont genomes should contain few or no IS. However, *Wolbachia* genomes are among those that contain the highest IS abundance. *Wolbachia* is an intracellular bacterium infecting numerous arthropod and nematode species. *Wolbachia* strains are generally reproductive parasites in arthropods (inducing numerous reproduction alterations) and mutualists in nematodes.

We combined *in silico* descriptions, with the analysis of five *Wolbachia* genomes, and molecular biology experiments. Thus, we obtained empirical confirmation of the models of IS evolutionary dynamics by characterising their past dynamics in *Wolbachia* genomes.

We showed that *Wolbachia* genomes contain between 52 and 372 IS copies. IS coverage represents up to 14.2% of the genome, which is one of the highest coverage described at this time among prokaryotes. In addition, we found that a surprising proportion of copies were non functional and degraded (>70%). This abundance of non functional copies allowed us to study IS degradation processes. Our results showed that IS degradation follows at least two distinct pathways. The first process consists of slow copy degradation by

accumulation of small substitutions and degradations over time. The second process, less frequent, consists of rapid copy degradation by accumulation of large deletions.

In addition, study of nucleotide divergence and relative IS copy size showed that IS transpositional activity was not constant during time. Two phases of high IS activity were detected in *Wolbachia* genomes, one old and one more recent (and maybe ongoing). These two phases were separated by a phase of relative quiescence. At our resolution, the two phases of high IS activity seem to be synchronised in the studied genomes. In addition, these activity phases might coincide with diversification phases of *Wolbachia* strains. Simulations suggest that IS activity phases need to be preceded by new functional IS incoming with horizontal transfers. PCR detection of 17 IS groups in a sample of 22 diverse *Wolbachia* strains showed a high number of horizontal transfers according to the recent phase of high IS activity observed in the studied genomes. This recent phase of IS activity seems to occur independently in the studied genomes, because IS groups which have expanded during this phase of activity are not the same in the different genomes.

We next focused our work on the study of IS expression in the *Wolbachia w*VulC genome. We showed that all IS sequences of the studied IS groups contain promoters. However, the majority of them are localised downstream of the start codon of the transposase gene (protein allowing transposition of the IS). Thus, it could only promote partial transcription of the transposase coding gene. This suggests that the genomic environment of IS copies plays an important role in the transposition mechanism. Experimental study of IS group expression showed that the majority of the IS groups are transcribed, whatever the functional status of their IS copies (potentially functional or not). Expression study of all the copies from the IS4wA_1 group showed that flanking regions of IS copies are co-transcribed with the IS copies. This suggests that IS transcription could be initiated in flanking regions

and confirms the importance of the genomic environment. In addition, we showed that all copies of the IS4wA_1 group are not transcribed.

To conclude, our work allowed us to better understand IS evolutionary dynamics in *Wolbachia* genomes. We showed that : (1) IS copy degradation follows two different pathways (one slow and one more rapid but less frequent), (2) transpositional activity of IS varied during time, (3) high activity phases require acquisition of new IS copies by horizontal transfers, which were frequent in a recent past of *Wolbachia* evolution, (4) IS sequences contain the coding capability to induce their own transposition but, in the majority of cases, the produced mRNA is non complete, highlighting the importance of IS copy genomic environment and (5) IS groups are transcribed whatever the functional status of the IS copies. All mechanisms described in *Wolbachia* genomes could explain the high abundance variation that was observed between closely related bacterial strains, the predominant presence of recent IS copies in bacterial genomes and could be applicable, at least partially, to other bacterial genomes.

Key words: transposable elements, insertion sequences, evolutionary dynamics, prokaryotes, *Wolbachia*

Remerciements

Tout d'abord, merci à l'ensemble des membres de mon jury, Emmanuelle LERAT, Michael CHANDLER et Pierre CAPY d'avoir accepté de juger ce travail. Je tiens à remercier en particulier Pierre CAPY et Mylène WEILL, qui n'a malheureusement pas pu participer à l'évaluation de mon travail, pour avoir pris part à mes comités de thèse et m'avoir conseillé.

Je voudrais remercier très chaleureusement mes deux directeurs de thèse, Richard et Didier, sans qui ce travail n'aurait pu être aussi bien mené. Je voudrais les remercier pour leur patience et leur conseil tout au long du déroulement de ma thèse, ce qui a permis mon plein épanouissement. Merci à Richard pour ton organisation qui m'a permis de réaliser et d'écrire ma thèse dans la sérénité et avec un minimum de stress. Merci Didier, pour toutes les discussions qui nous avons eu autour de notre petit café du matin. Enfin, merci à vous deux pour de cette thèse qui s'est déroulée sans accroc.

Merci à l'ensemble des membres du laboratoire qui m'ont supporté pendant mes 3 années de thèse mais aussi avant pendant mes stages de Master première et deuxième année. Merci à tous le personnel technique pour l'aide qu'ils m'ont apporté tant au niveau théorique que pratique. Merci à Christelle, notre secrétaire, qui nous facilite la vie tous les jours que ce soit pour les commandes ou toutes autres formalités administratives. Enfin, merci à l'ensemble des chercheurs, enseignants-chercheurs, post-doc et autre personnel précaire avec qui j'ai pu longuement discuter dans la bonne humeur.

Merci aux habitants passés et présents de la salle des étudiants et aux jeunes non permanents expérimentés : Vincent, Sam, Gaël, Bérénice, Winka, Lise, Jessica, Isabelle, Sébastien L., Sébastien V., Lenka, Sandrine, Hajer, Mehdia, Dalila, Mauricio, Gipo, Romain

et tous les autres que j'aurai pu oublier. Enfin, merci à toutes les filles qui ont supporté mes remarques et mes sarcasmes sans m'avoir passé par la fenêtre.

Merci à mes parents et à ma sœur, sans qui tout cela n'aurait pu être possible à tout point de vue. Merci de m'avoir soutenu et désolé de ne pas toujours avoir été disponible.

Enfin, merci à Johanna pour m'avoir apporté tant de choses, pour ton amour et ta patience. Merci pour ton écoute et ce même la nuit quand dans ton sommeil je te réveille pour te parler de mes « éclairs de génie » avant de me rendormir aussitôt.

Sommaire

Résumé ..... *III*

Abstract ..... *VI*

Remerciements ..... *IX*

Sommaire ..... *XI*

Liste des figures ..... *XIII*

Liste des tableaux ..... *XV*

Liste des abréviations ..... *XVII*

1. Préambule *1*

2. Introduction *4*

2.1. Classification et modes de transposition des ET **4**
2.1.1. Les ET de classe 1 5
2.1.2. Les ET de classe 2 8
2.1.3. Les ET non autonomes 11

2.2. Dynamique évolutive des ET **12**
2.2.1. Impacts des ET 12
2.2.2. Contrôle de l'activité de transposition 16
2.2.3. Dynamique des ET chez les eucaryotes 19
2.2.4. Dynamique des ET chez les procaryotes 21

2.3. La symbiose à Wolbachia **24**
2.3.1. Diversité de souches de Wolbachia 24
2.3.2. Mode de vie 26
2.3.3. Localisation tissulaire 26
2.3.4. Les génomes de Wolbachia 27

3. Détection et caractérisation des IS de Wolbachia *31*

3.1. Contexte **31**

3.2. Méthodes **32**
3.2.1. Détection 32
3.2.2. Classification 33
3.2.3. Etude de la distance génomique entre copies d'IS 34
3.2.4. Importance des MGE dans les variations de taille de génome de Wolbachia 35
3.2.5. Analyses de diversité 37
3.2.6. Analyse structurelle 40

3.3. Résultats et discussion **40**
3.3.1. Abondance des IS chez Wolbachia 40
3.3.2. Distribution des IS dans les génomes de Wolbachia 42
3.3.3. Relation entre MGE et taille de génome 48
3.3.4. Prévalence des familles d'IS chez Wolbachia 53
3.3.5. Distribution des IS dans les familles 56

3.3.6. Structure des copies 60
3.4. Conclusion 61
4. Dynamique évolutive des IS 64
4.1. Contexte 64
4.2. Méthodes 65
4.2.1. Divergence nucléotidique et dégradation des copies d'IS 65
4.2.2. Etude de la dynamique évolutive des IS 66
4.2.3. Simulations de dynamique évolutive des IS 67
4.2.4. Recherche de copies orthologues 70
4.2.5. Détection expérimentale des groupes d'IS dans divers génomes de Wolbachia 70
4.3. Résultats et discussion 74
4.3.1. Mécanismes de dégradations des IS 76
4.3.2. Dynamique à long terme des IS dans les génomes de Wolbachia. 79
4.3.3. Déroulement d'une phase d'activité d'IS 89
4.3.4. Recherche de signes d'activité pour d'autres groupes d'IS 96
4.3.5. Quelle origine pour les copies potentiellement fonctionnelles? 96
4.3.6. D'autres sources de transferts? 105
4.4. Conclusion 108
5. Etude de l'expression des IS du génome de wVulC 110
5.1. Contexte 110
5.2. Méthodes 113
5.2.1. Choix des groupes d'IS 113
5.2.2. Analyses bio-informatiques 114
5.2.3. Analyse in vivo 116
5.3. Résultats et discussion 120
5.3.1. Analyses de l'expression des groupes d'IS 120
5.3.2. Analyse de l'expression de chaque copie d'un groupe d'IS 127
5.4. Conclusion 131
6. Discussion générale 134
Bibliographie 142
Annexe 1 : article de synthèse publié dans l'ouvrage « Evolutionary Biology – Concepts, Biodiversity, Macroevolution and Genome Evolution » 155
Annexe 2: épreuves non corrigées de l'article de recherche accepté pour publication en septembre 2011 dans la revue Genome Biology and Evolution 178
Annexe 3: Scénarios de dynamique évolutive à long terme des IS dans les génomes de Wolbachia 191
Annexe 4 : Nombre de promoteurs prédits dans le sens négatif pour chaque groupe d'IS étudiés. 192

Liste des figures

Figure 2-2 : Structure type d'un ET de classe 2. 7

Figure 2-3 : Mécanisme de transposition en « couper-coller ». 9

Figure 2-4 : Détection de *Wolbachia* en hybridation *in situ* fluorescente (FISH) chez le crustacé isopode terrestre *Armadillidium vulgare*. 25

Figure 2-5 : Phylogénie des 5 souches de *Wolbachia* étudiées 28

Figure 3-1 : Distribution de la fréquence des distances entre copies pour chaque génome. 41

Figure 3-2 : Corrélation entre la densité en IS et la médiane de la distribution des distances entre copies. 43

Figure 3-3 : Représentation graphique de la distribution des IS dans les génomes de *Wolbachia* 45

Figure 3-4: Corrélation entre taille de génome de *Wolbachia* et nombre de copies d'IS. 47

Figure 3-5 : Relation entre taille de génome de *Wolbachia* et couverture totale en MGE..... 47

Figure 3-6: Corrélation entre taille de génome de *Wolbachia* et couverture de 3 types de MGE. 50

Figure 3-7: Proportion des 3 types de MGE dans les variations de taille de génome entre souches de *Wolbachia* 50

Figure 3-8: Fréquence des familles d'IS identifiées dans les cinq génomes de *Wolbachia*. 55

Figure 3-9 : Dendrogramme basé sur l'abondance et la distribution des groupes d'IS des cinq génomes de *Wolbachia*. 57

Figure 3-10: Proportion de copies d'IS en fonction de leur statut de fonctionnalité. 59

Figure 4-1: Détection par PCR de l'infection par *Wolbachia* de trois individus de l'espèce d'isopode terrestre *Cylisticus convexus* 71

Figure 4-2 : Corrélation entre la divergence nucléotidique et la taille relative des copies d'IS des génomes de *w*Mel, *w*Ri, *w*Pel et *w*VulC. 75

Figure 4-3 : Distribution des fréquences (a) de la divergence nucléotidique et (b) de la taille relative des copies d'IS pour chacun des génomes de *Wolbachia*. 80

Figure 4-4 : Distribution des divergences nucléotides obtenue dans le cadre de la simulation de l'évolution des IS dans un génome bactérien. 83

Figure 4-5 : Scénario de dynamique évolutive à long terme des IS dans les génomes de *Wolbachia* 86

Figure 4-6 : Expansion du groupe ISWen2 dans le génome de *w*Ri. 93

Figure 4-7 : Détection par PCR de la présence du groupe ISWpi11 dans les 11 souches des supergroupes de diversité de *Wolbachia* B et G. 97

Figure 4-8 : Historique des acquisitions et pertes de 17 groupes d'IS dans 22 souches de *Wolbachia*. 100

Figure 4-9 : Distribution de la divergence nucléotidique entre paire de séquences obtenues lors de l'analyse de la distribution des 17 groupes d'IS dans les 22 souches de *Wolbachia*. 101

Figure 5-1 : Gel de vérification de la qualité des ARN extraits. 115

Figure 5-2 : Description des amorces utilisées pour réaliser les tests d'expression *in vivo*... 117

Figure 5-3 : Localisation des promoteurs potentiels des 4 groupes d'IS avec des copies potentiellement fonctionnelles 121

Figure 5-4 : Détection par PCR de l'expression de groupes d'IS présents dans le génome de *w*VulC. 125

Figure 5-5 : Détection par PCR de l'expression de copies du groupes IS4wA_1 présentes dans le génome de *w*VulC. 128

Liste des tableaux

Tableau 2-1 : Caractéristiques générales des 5 génomes de *Wolbachia*. 28

Tableau 3-1 : Analogies entre notion écologique et ET 36

Tableau 3-2: Caractéristiques générales des IS dans les cinq génomes de *Wolbachia*. 39

Tableau 3-3: Nombre de copies de chacune des 12 familles d'IS dans les cinq génomes de *Wolbachia*. 52

Tableau 3-4: Indices de diversités et d'équitabilité calculés à partir de la distribution des groupes d'IS des cinq génomes de *Wolbachia*. 55

Tableau 4-1 : Liste des souches de *Wolbachia* utilisées pour la détection des groupes d'IS. .. 69

Tableau 4-2 : Caractéristiques des amorces utilisées pour la recherche des 17 groupes d'IS dans les souches de *Wolbachia*. ..... **Erreur ! Signet non défini.**

Tableau 4-2 : Caractéristiques des amorces utilisées pour la recherche des 17 groupes d'IS dans les souches de *Wolbachia*. 73

Tableau 4-3 : Coefficient de corrélation entre taille relative et divergence nucléotidique et p-value pour les différents jeux de données 77

Tableau 4-4: Valeurs des tests de comparaisons de distribution de groupe d'IS pour la classe de taille [100-80%[..... 88

Tableau 4-5 : Distribution des copies potentiellement fonctionnelles présentes dans les cinq génomes de *Wolbachia*. 88

Tableau 4-6: Caractéristiques des groupes d'IS utilisés pour construire des réseaux d'haplotypes 90

Tableau 4-7 : Polymorphisme de présence/absence d'insertion de 44 copies d'IS dans les génomes de *w*Mel et *w*Ri. 95

Tableau 4-8 : Distribution des copies d'IS insérées dans les prophages de 4 génomes de *Wolbachia*. 104

Tableau 4-9 : Recherche de similarité des IS de *Wolbachia* dans la base de données non redondante de NCBI datant du 1er octobre 2011 106

Tableau 5-1 : Liste et caractéristiques des groupes d'IS du génome de *w*VulC choisis pour l'analyse d'expression. 112

Tableau 5-2 : Caractéristiques des amorces utilisées pour étudier l'expression des groupes d'IS. 117

Tableau 5-3 : Caractéristiques des amorces utilisées pour étudier l'expression des copies du groupe IS4wA_1. 119

Tableau 5-4 : Nombre de promoteurs prédits dans le sens positif pour chaque groupe d'IS étudiés. 119

Tableau 5-5: Résultats des RT-PCR (40 cycles) réalisées pour les 12 groupes d'IS 125

Tableau 5-6: Résultats des RT-PCR réalisées pour les 7 copies du groupe IS4wA_1 128

Liste des abréviations

ADN	Acide désoxyribonucléique
ADNc	ADN complémentaire
AMPc	Adénosine monophosphate cyclique
ARN	Acide ribonucléique
ARNm	ARN messager
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
EST	Expressed sequence tag
ET	Elément transposable
FISH	Fluorescent in situ hybridization
IS	Séquence d'insertion
Kb	Kilobase (1000 paires de base)
LTR	Long terminal repeat
Mb	Mégabase (1 million de paires de base)
MGE	Eléments génétiques mobiles
miRNA	Micro interacting RNA
MITE	Miniature inverted repeat transposable element
ORF	Open reading frame
pb	Paire de base
PCR	Polymerase chain reaction
piRNA	Piwi interacting RNA
RNAi	interférence ARN
RT	Reverse transcription
RTase	Reverse transcriptase
SINE	Short interspread nuclear element

siRNA	Small interacting RNA
TIR	Terminal inverted repeat
TSD	Target site duplication

1. Préambule

Les éléments transposables (ET) sont des fragments d'ADN qui font partie de la grande famille des éléments génétiques mobiles (MGE). Ils sont composés d'un ou plusieurs modules (domaines codant des protéines ou gènes) leur conférant la capacité de se déplacer et dans la plupart des cas, de se multiplier au sein d'un génome procaryote ou eucaryote (Toussaint and Merlin 2002). Ils ont été identifiés pour la première fois en 1950 par Barbara Mc Clintock chez le maïs (McClintock 1950). Cette découverte, d'abord passée inaperçue ou considérée comme absurde par la communauté scientifique de l'époque, lui a valu de recevoir en 1983 le prix Nobel de médecine. La découverte de Barbara Mc Clintock, comme certaines grandes découvertes, se fit par hasard. Elle travaillait sur la variabilité des souches de maïs lorsqu'elle identifia deux loci en association : Dissociator (*Ds*) et activator (*Ac*). Elle démontra que le locus *Ds* était notamment à l'origine de modifications chromosomiques (duplication, perte de fragments…) lorsqu'il était en présence du locus *Ac*. Plus tard elle identifia un troisième élément qu'elle appela *Suppressor-mutator* (*Spm*) et présentant un comportement plus complexe que les deux premiers. Des études postérieures ont montré que ces 3 loci étaient les premiers ET identifiés. Barbara Mc Clintock arrêta ses recherches sur les ET du maïs faute de pouvoir faire passer ses idées dans la communauté scientifique. A cette époque, l'ensemble de la communauté scientifique était persuadé que les gènes avaient une position unique dans le génome. Ainsi, la plupart des chercheurs pensaient que la découverte de Barbara Mc Clintock était un cas particulier isolé. Ce n'est que dans les années 60 que Barbara Mc Clintock réussit à faire connaître son travail (McClintock 1961). Depuis, elle est perçue comme un précurseur et de nombreuses études se focalisent sur les ET et leurs implications. On sait notamment que les ET peuvent représenter une part importante des génomes de certains organismes, comme l'homme (Lander, et al. 2001) ou encore le maïs

(Schnable, et al. 2009). Il est maintenant couramment admis que les ET sont présents dans la quasi-totalité des organismes vivants (Bennetzen 2000; Hua-Van, et al. 2011; Hua-Van, et al. 2005; Siguier, et al. 2006a). Ainsi, une étude menée sur 10 millions de gènes codant issus de l'ensemble des branches du monde vivant a permis de mettre en évidence que les transposases sont les gènes les plus présents dans la nature (Aziz, et al. 2010). En l'espace de 60 ans les ET sont sortis de l'anonymat et se pose la question de leur impressionnante réussite évolutive. Comment des éléments génétiques mobiles parfois considérés comme des parasites génomiques (Orgel and Crick 1980) ont pu coloniser les génomes de presque toutes les espèces du monde vivant ? Cette étonnante réussite pourrait être expliquée par un ensemble de paramètres lié aux modes de transposition des ET, à leur mode de transfert entre génomes hôtes, à la régulation de l'activité de transposition et à leur impact sur les génomes.

2. Introduction

Ce chapitre s'appuie en partie sur l'article de synthèse que nous avons publié dans l'ouvrage « Evolutionary Biology – Concepts, Biodiversity, Macroevolution and Genome Evolution », présenté en annexe 1 (Cerveau et al 2011a).

2.1. Classification et modes de transposition des ET

La classification des ET a été et est encore sujette à controverses et varie en fonction des critères considérés par les auteurs. Ils peuvent être séparés en 5 classes si l'on considère le type de protéine permettant la transposition (Curcio and Derbyshire 2003), en 3 classes en fonction de l'acide nucléique intermédiaire de transposition et de la présence de longues répétitions terminales (LTR) (Eickbush and Eickbush 2005) ou en deux grandes classes en fonction de la nature de l'acide nucléique intermédiaire de la transposition (Finnegan 1989). C'est cette dernière méthode de classification qui est la plus couramment utilisée. Cette classification sépare d'une part, les ET de classe 1 qui transposent via un intermédiaire à ARN et d'autre part les ET de classe 2 qui transposent via un intermédiaire à ADN (Finnegan 1989) (Figure2-1).

Ces deux classes ont été identifiées aussi bien chez des organismes eucaryotes (copia, L1 pour la classe et Tc1-Mariner, P pour la classe 2) que procaryotes (intron de groupe II pour la classe 1 et séquence d'insertion pour la classe 2) (Daboussi and Capy 2003; Feschotte and Pritham 2007b; Hua-Van, et al. 2005; Lambowitz and Zimmerly 2010b; Leclercq, et al. 2011; Siguier, et al. 2006a). A l'intérieur des classes 1 et 2, des regroupements ont été créés en fonction des domaines protéiques présents ou absents dans les ET et de leur similarité de séquence (Wicker, et al. 2007). Cette classification en deux groupes a été affinée afin d'intégrer les ET non autonomes (ex : Miniature Inverted Transposable Element ou MITE et

Short interspread nuclear element ou SINE) capables de transposer en utilisant la machinerie d’une autre copie fonctionnelle

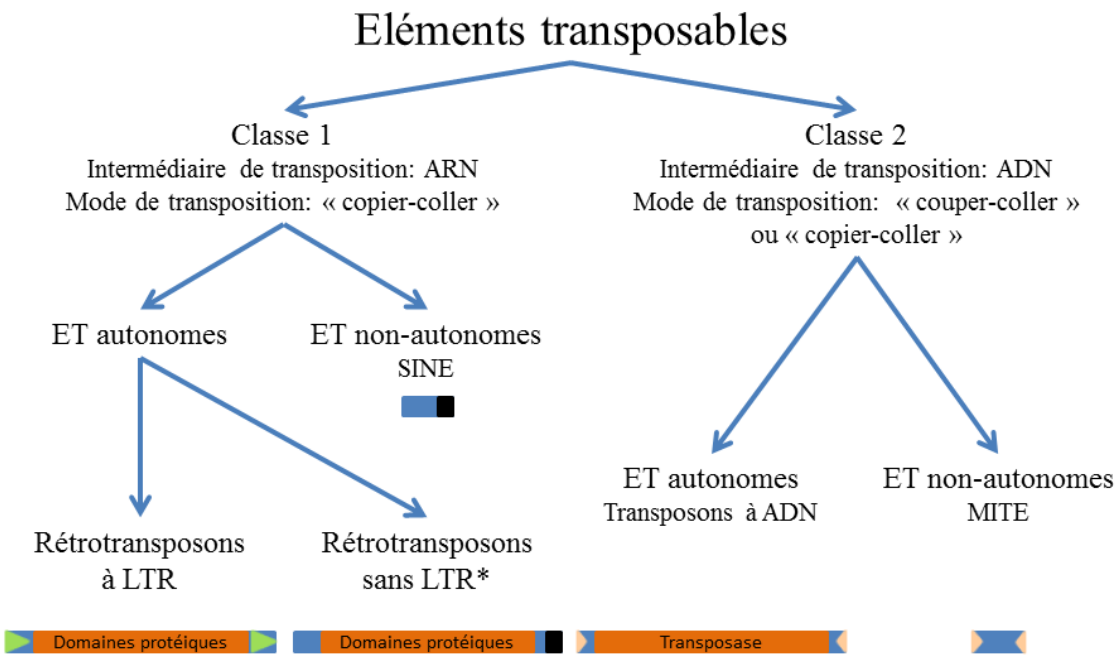

Figure 2-1 : Représentation simplifiée de la classification de ET.

Pour chacun des types d'ET autonomes ou non, une représentation schématique de la structure a été réalisée. Les rectangles bleus représentent la séquence non codante de l'ET et les rectangles rouges la zone codant les domaines protéiques ou la transposase. Les triangles verts représentent les longues répétitions terminales (LTR) des rétrotransposons à LTR. Les triangles jaunes des éléments de classe 2 représentent les répétitions terminales inversées (TIR). Les rectangles noirs représentent la queue poly(A) des rétrotransposons sans LTR. * Dans cette classification les introns de groupe II procaryotes peuvent être considérés comme des rétrotransposons sans LTR.

présente dans le génome (Wicker, et al. 2007). La différence majeure entre les ET de classes 1 et 2, en plus de la nature de l'intermédiaire de transposition, est le mécanisme de transposition (Figure 2-1). En effet, les ET de classe 1 transposent selon un mode réplicatif en « copier-coller » alors que ceux de classe 2 transposent selon un mode conservatif en « couper-coller » (Finnegan 1989) ou réplicatif en « copier-coller » (Curcio and Derbyshire 2003) en fonction des familles. Le mode de transposition des ET a d'importantes conséquences sur leur dynamique et sur leur capacité à se propager dans un génome. Dans les deux sections suivantes les caractéristiques et les modes de transposition des deux classes d'ET ont été décrits.

2.1.1. Les ET de classe 1

Les ET de classe 1 ou rétrotransposons possèdent différentes caractéristiques qui leur sont propres comme par exemple la transposition en « copier-coller » via un intermédiaire à ARN. Chaque cycle de transposition permet de générer une nouvelle copie. Du fait du passage par un intermédiaire à ARN, l'ensemble des ET de classe 1 partage une enzyme commune : la transcriptase inverse (RTase). Cette enzyme peut synthétiser un brin d'ADN à partir d'une matrice d'ARN (Eickbush and Jamburuthugoda 2008; Wicker, et al. 2007). Cette RTase ainsi que d'autres domaines codés par les ET sont aussi retrouvés dans les virus à ARN de vertébrés (rétrovirus). Ces virus utilisent la RTase pour passer d'une forme ARN à une forme ADN avant de s'insérer dans le génome de la cellule qu'ils colonisent. Il a été proposé dans la littérature que les rétrovirus de vertébrés et les ET de classe 1 pourraient partager une origine commune (Eickbush and Jamburuthugoda 2008; Mount and Rubin 1985). En plus du domaine RTase, la séquence de certains rétrotransposons code d'autres domaines protéiques ayant des homologues chez les virus comme par exemple les domaines intégrase et RNaseH (Curcio and Derbyshire 2003; Mount and Rubin 1985; Wicker, et al. 2007).

Les ET de classe 1 sont subdivisés en différents groupes, basés sur leurs caractéristiques intrinsèques (Eickbush and Eickbush 2005; Wicker, et al. 2007). Ainsi on peut distinguer les éléments possédant ou non des LTR (Eickbush and Eickbush 2005) (Figure 2-1). Un autre critère de classification des rétrotransposons est le nombre et la nature des domaines protéiques dans le ou les gènes de l'ET. Par exemple, les rétrotransposons à LTR de type *Copia* possèdent 5 domaines protéiques différents alors que les LINE de type *R2* n'en possèdent que deux (Wicker, et al. 2007). De plus, l'insertion des rétrotransposons peut entraîner la duplication du site d'insertion (target site duplication ou TSD). La présence ou non des TSD ainsi que leur taille peuvent être conservés au sein d'un groupe d'ET de classe 1, ce qui en fait un critère de classification supplémentaire de ces éléments (Wicker, et al. 2007).

La première étape du mode de transposition des ET de classe 1 nécesite un domaine protéique essentiel: la RTase qui permet de passer de la molécule d'ARN à celle d'ADN La suite de la rétrotransposition dépend de la famille considérée et de la structure de l'ET. Par exemple, les rétrotransposons à LTR catalysent leur insertion dans le génome avec une intégrase à domaine DDE (Eickbush and Malik 2002), les rétrotransposons de type DIRS pourraient utiliser une transposase à tyrosine (Curcio and Derbyshire 2003) et les rétrotransposons sans LTR utilisent les cassures déjà présentes dans l'ADN ou en génèrent de nouvelles via leur domaine endonucléase pour s'insérer (Feng, et al. 1996; Luan, et al. 1993). Chez les introns de groupe II, la traduction de l'ARN entraîne la production d'un complexe protéique qui vient se fixer sur l'élément et catalyse son propre épissage de l'ARN messager (ARNm). L'ARN ainsi épissé est ensuite reverse transcrit et inséré à un nouveau locus (Lambowitz and Zimmerly 2010a; Toro, et al. 2007).

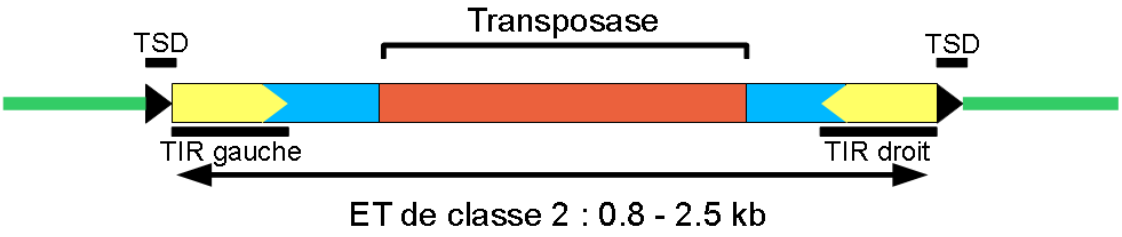

Figure 2-2 : Structure type d'un ET de classe 2.

L'ET de classe 2 est composé du gène codant la transposase (rouge) qui est entouré par des zones non codantes (en bleu) et les TIR en jaune pâle. L'ET de classe 2 est lui-même entouré de TSD en noir. La zone verte représente l'ADN du génome hôte.

2.1.2. Les ET de classe 2

L'une des caractéristiques majeures des ET de classe 2 ou transposons à ADN est la transposition généralement en « couper-coller » via un intermédiaire à ADN le plus souvent double brin (Feschotte and Pritham 2007b). Les ET de classe 2 ont été décrits aussi bien chez les procaryotes que chez les eucaryotes (Chandler and Mahillon 2002; Feschotte and Pritham 2007b). Les ET de classe 2 bactériens les plus simples et les plus abondants, qui sont par ailleurs au cœur de ma thèse, sont les séquences d'insertion (IS). Quelle que soit leur origine, les ET de classe 2 procaryotes et eucaryotes partagent des caractéristiques communes et une relation phylogénétique peut être faite entre certains IS bactériens et des transposons à ADN eucaryotes (Feschotte and Pritham 2007b). Ils sont généralement composés d'un gène unique codant la protéine permettant leur déplacement, et peuvent être délimités par des répétitions terminales inversées (TIR) de taille variable (Chandler and Mahillon 2002; Feschotte and Pritham 2007b) (Figure 2-2). Les TIR sont les sites de reconnaissance et de liaison de la transposase. Pour certaines familles d'ET de classe 2, l'insertion d'une copie entraîne comme pour les ET de classe 1 la création de TSD (Figure 2-2).

Les ET de classe 2 sont divisés en différentes familles ou groupes en fonction de leur mode de transposition, de la séquence de leur transposase et de la présence ou non de TIR (Chandler and Mahillon 2002; Feschotte and Pritham 2007b). La majorité des ET de classe 2 possède au moins un cadre ouvert de lecture (ORF) qui code une transposase ou une recombinase (Curcio and Derbyshire 2003; Wicker, et al. 2007) permettant des modes de transposition différents.

Pour transposer, une copie génomique est transcrite, le plus souvent via un promoteur endogène ou semi-endogène (Nagy and Chandler 2004), puis l'ARNm est traduit en protéine. La protéine ainsi produite va se fixer sur une copie de l'ET, l'exciser puis l'insérer à un

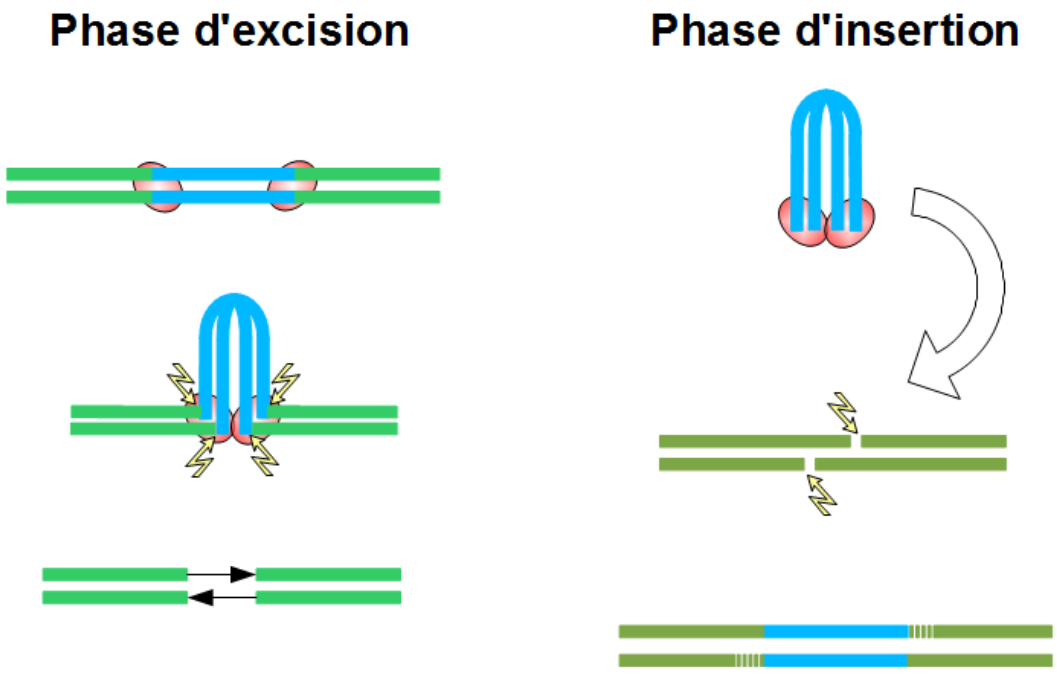

Figure 2-3 : Mécanisme de transposition en « couper-coller ».

Les zones vertes représentent les régions génomiques et la zone bleue représente le transposon à ADN. Les ovales rouges représentent la transposase et les flèches jaunes représentent les sites de cassure de l'ADN catalysé par la transposase. Le mécanisme est décomposé en deux temps, d'abord la phase d'excision du locus donneur et la phase d'insertion au locus receveur.

nouveau site dans le génome (Feschotte and Pritham 2007b; Haniford, et al. 1991; Mizuuchi 1992). Chez les procaryotes uniquement, la transposition se fait préférentiellement en *cis*, c'est-à-dire que la transposase va se fixer préférentiellement sur la copie génomique dont elle est issue ou sur une copie insérée à un locus physiquement proche dans le génome. Cette activité préférentielle en *cis* diminue la

probabilité que la transposase d'un IS donné active la transposition d'un IS similaire localisé ailleurs dans le génome (Chandler and Mahillon 2002).

Dans le cas d'une transposition en "couper-coller", l'ET de classe 2 s'excise sous forme d'ADN double-brin et en se réinsérant à un autre locus (Chandler and Mahillon 2002; Curcio and Derbyshire 2003; Feschotte and Pritham 2007b). La transposition en « couper-coller » n'entraîne pas de création de nouvelle copie génomique de l'ET considéré, et n'entraîne donc pas d'augmentation du nombre de copies dans le génome. Néanmoins, si la transposition se déroule au moment de la réplication il est possible que le nombre de copies augmente. Par exemple, si l'ET de classe 2 s'excise d'une molécule d'ADN nouvellement répliquée et s'insère dans un locus non encore répliqué, ceci peut entraîner une augmentation du nombre de copies (Cerveau, et al. 2011a).

La transposition des ET de classe 2, et plus particulièrement des IS, est catalysée majoritairement par des transposases à domaine DDE (Curcio and Derbyshire 2003). Quelle que soit l'origine de la transposase à DDE étudiée, les 3 acides aminés non contigus D, D et E sont très conservés et toujours présents (Chandler and Mahillon 2002; Wicker, et al. 2007). Après transcription de l'ET, la transposase se fixe sur une copie génomique pour catalyser sa transposition (Chandler and Mahillon 2002) (Figure2-3). Dans le cas des ET de classe 2 avec une transposase à DDE, la reconnaissance de l'ET se fait via les TIR sur lesquels la transposase se fixe pour provoquer des cassures de l'ADN et ainsi l'exciser (Chandler and Mahillon 2002; Curcio and Derbyshire 2003; Wicker, et al. 2007). Durant la phase de

déplacement vers un nouveau site d'insertion, l'ET est sous forme circulaire (Curcio and Derbyshire 2003). Lorsqu'un nouveau site d'insertion est localisé, la transposase coupe l'ADN cible et catalyse l'insertion de l'ET (Figure 2-3).

D'autres modes de transposition, cette fois-ci en « copier-coller », ont été décrits. Ainsi, certains ET de classe 2 utilisent un mécanisme de co-intégration (Schmid, et al. 1998). Ce mode de transposition peut dans certains cas particuliers être réplicatif et engendrer la production d'une nouvelle copie génomique et non pas le déplacement d'une copie déjà insérée dans le génome. Les ET de classe 2 peuvent aussi transposer via un mécanisme de « cercle roulant » (Garcillán-Barcia, et al. 2002; Mendiola and de la Cruz 1992) mais dans ce cas la transposase doit posséder deux domaines de transposase à tyrosine (Wicker, et al. 2007).

Les mécanismes de transposition décrits ci-dessus sont ceux qui sont rencontrés le plus couramment et par conséquent les mieux documentés. D'autres mécanismes plus atypiques ont été décrits.

2.1.3. Les ET non autonomes

En plus des ET autonomes décrits précédemment, de très nombreux ET non autonomes ont été identifiés à ce jour. Ils transposent via un intermédiaire de transposition à ARN (SINE : Short Interspread Nuclear Elements) ou à ADN (MITE : Miniature Inverted repeat Transposable Elements). La séquence nucléotidique des MITES, et plus particulièrement les TIR, est proche de celle des ET autonomes dont il détourne la machinerie de transposition ce qui n'est pas le cas pour les SINE. Les ET non autonomes sont de petite taille et ne possèdent pas de séquence codante. Les deux types d'ET non autonomes les plus fréquents (SINE et MITE) sont décrits brièvement.

Les SINE sont présents uniquement chez les eucaryotes. Les plus connus des SINE sont probablement les éléments *Alu* qui sont présents en très grand nombre dans le génome humain (> 1 million de copies) (Lander, et al. 2001). Les *Alu* ont été nommés ainsi car ils contiennent dans leur séquence le site de reconnaissance de l'enzyme de restriction *AluI* (Houck, et al. 1979). Un élément *Alu* a une taille d'environ 300 paires de bases (pb) et est entouré par des TSD. Les éléments *Alu* utilisent la machinerie de transposition des éléments *L1* qui sont de ET de classe 1 sans LTR pour transposer (Batzer and Deininger 2002).

Les MITES sont des éléments présents chez les eucaryotes et les procaryotes. Leur forte abondance a été mise en évidence pour la première fois chez le maïs suite à l'identification d'une copie insérée dans le gène *Waxy* codant une glucosyl-transferase (Bureau and Wessler 1992). Près de 1200 copies de MITE ont été identifiés dans le génome d'*Arabidopsis thaliana* (Feschotte, et al. 2002) et 100 000 dans celui du riz (Turcotte, et al. 2001).

2.2. Dynamique évolutive des ET

Dans les parties précédentes, j'ai décrit les grands principes de transposition des ET, cependant leur dynamique évolutive ne se résume pas à cet unique paramètre. Dans cette section, nous allons décrire deux phénomènes qui peuvent influencer la dynamique évolutive des ET : l'impact de leur insertion et les mécanismes de régulation de la transposition. Puis, nous décrirons la dynamique évolutive des eucaryotes et des procaryotes.

2.2.1. Impacts des ET

L'une des forces influençant probablement le plus l'évolution des ET dans les génomes est la sélection. En fonction de l'impact de l'insertion des ET sur la survie et la

reproduction de l'hôte, elle peut permettre soit le maintien soit la perte d'une copie. De nombreux cas d'impacts d'ET ont été mis en évidence, cependant nous ne souhaitons pas être exhaustif mais simplement illustrer l'étendue de l'influence des ET. De nombreux cas d'insertions d'ET avec un impact bénéfique ont été identifiés. Dans ce cas, la copie d'ET ou la partie induisant l'effet positif sera maintenue dans le génome. A contrario, la sélection peut aussi être à l'origine de l'élimination rapide des ET des génomes, chez les procaryotes notamment. Etant donné que les génomes procaryotes sont très denses en séquences codantes (Lawrence, et al. 2001) et que les régions intergéniques contiennent les structures de régulation des gènes, il est fortement probable que l'insertion d'un ET entraîne des effets délétères sur les génomes. La sélection pourrait alors prévenir la fixation des ET dans les génomes (Cerveau, et al. 2011a). De très nombreux effets, résultant de l'insertion des ET, ont été décrits à ce jour. Les ET peuvent bien sûr être à l'origine de phénomènes d'inactivation de gènes, mais ils peuvent aussi faire varier l'expression des gènes qui les entourent, être à l'origine de diversité génétique ou bien encore avoir une influence sur la structure du génome

L'insertion d'un ET dans une séquence peut être qualifiée de mutagénèse insertionnelle. L'impact de la mutagénèse insertionnelle le plus facile à observer est l'insertion dans une séquence codante. Une telle insertion entraîne dans la plupart des cas l'inactivation du gène dans lequel l'ET est inséré. On dénombre, par exemple, au moins 65 insertions *de novo* d'un rétrotransposon sans LTR à l'origine de maladies humaines (Belancio, et al. 2008) .Chez *Escherichia coli* dans le cadre de cette expérimentation, l'analyse des mutations d'un gène rapporteur a permis de montrer que 60% d'entre elles étaient liées à des insertions d'IS (Rodriguez, et al. 1992).

Un deuxième effet de la mutagénèse insertionnelle est possible quand les ET s'insèrent dans des régions intergéniques. Ils peuvent faire varier le niveau d'expression des gènes environnants. Ces variations de niveau d'expression des gènes environnants peuvent être liées

soit à la structure des ET soit à l'inactivation de leur système de régulation. En effet, la majorité des ET possède des structures de régulation de l'expression, comme les promoteurs et terminateurs de transcription ou encore des signaux de polyadénilation. Ce phénomène est connu depuis environ 30 ans (Charlier, et al. 1982; Jaurin and Normark 1983; Saedler, et al. 1974). Un exemple plus récent a permis de montrer que chez les bactéries l'insertion d'un ET en amont de l'opéron *glpFK* pouvait activer celui-ci sans le système de régulation habituel basé sur un intermédiaire à AMPc. Cette activation permet à la bactérie d'utiliser le glycérol comme substrat pour son développement (Zhang and Saier 2009). De plus, des rétrotransposons sans LTR possèdent des signaux de polyadénilation qui peuvent arrêter de façon prématurée la transcription et empêcher la transcription des gènes en aval s'ils sont insérés entre un gène et son promoteur (Cordaux and Batzer 2009). Une analyse bio-informatique a permis de montrer que les ET bactériens contiennent une grande quantité de terminateurs de transcription (Naville and Gautheret 2010). Les terminateurs peuvent se situer soit avant la transposase et réguler l'activité de transposition de l'ET, soit après la transposase et influer sur l'expression du gène en aval de l'ET, soit dans le milieu de l'élément auquel cas ils empêchent la production d'un ARN messager couvrant la totalité de l'élément (Dalrymple and Arber 1986).

Les ET peuvent aussi modifier l'expression de gènes adjacents en modifiant leur structure de régulation. Là encore, ce phénomène est connu depuis plus de 30 ans (Saedler, et al. 1972). Un exemple plus récent a montré que l'insertion d'un IS dans le répresseur de l'opéron codant une pompe de reflux de drogue chez *Pseudomonas putida* confère à la bactérie une meilleure résistance à des polluants comme le toluène (Wery, et al. 2001). Un mécanisme similaire a été montré sur des populations d'*E. coli* non motiles, dont la motilité a été restaurée après l'insertion d'un IS dans la région promotrice de l'opéron *flhd* qui est essentiel à la régulation du flagelle (Barker, et al. 2004).

Les ET sont très souvent considérés comme des parasites génomiques qui ne sont qu'un fardeau pour les génomes qui les hébergent, notamment à cause de l'inactivation des gènes dans lesquels ils s'insèrent. Néanmoins, la présence des ET peut aussi être une source de diversité et de production de nouveaux gènes ou de nouveaux allèles. Ainsi, les rétrotransposons peuvent changer la structure génomique d'un gène par le phénomène de transduction (Cordaux and Batzer 2009). Ce mécanisme entraîne la redistribution d'exons de régions génomiques qui ne sont pas des ET localisées en aval d'un rétrotransposon. La transduction liée au rétrotransposon SVA est ainsi à l'origine de la duplication d'environ 53 kilobases (kb) de séquences génomiques chez l'homme (Xing, et al. 2006). De plus, le détournement de la machinerie de rétrotransposition par des gènes non ET peut être à l'origine de leur duplication (Esnault, et al. 2000).

Outre les effets produits par les rétrotransposons, les ET de classe 2 peuvent aussi produire de la diversité allélique. Dans ce cas, c'est l'excision des ET, phénomène très rare, qui entraîne la production de diversité. Par exemple, lorsque les ET de classe 2 s'excisent, cette excision n'est généralement pas restreinte à l'élément uniquement, mais elle génère des marques, le plus souvent sous forme d'insertion ou de délétion de nucléotides. Lorsqu'un ET de classe 2 s'excise d'un gène, il ne rétablit donc pas nécessairement sa forme initiale. Ainsi, l'excision imprécise d'un IS inséré dans le gène *flic* de la souche P400 d'*E. coli* K12 a entraîné la création de nouveaux variants de ce gène qui est impliqué dans la production des flagelles de la bactérie et permet la colonisation des hôtes (Strauch and Beutin 2006). Le même phénomène de production de diversité a été observé dans le génome d'*Oryzias latipes*, un poisson chez lequel un gène de la pigmentation est interrompu par un ET de type *tol2*. L'excision imprécise de cet élément entraîne la production de nombreux variants phénotypiques (Koga, et al. 2006).

Les ET sont généralement des segments d'ADN d'une taille généralement comprise entre 1 et 5 Kb. Ainsi la variation de taille de génome que génère une nouvelle insertion peut-être considérée comme négligeable. Néanmoins, au vu de leur grand nombre, l'activité des ET peut tout de même avoir un impact sur la taille des génomes. Ainsi, il a été suggéré que les variations de taille de génomes observées entre des espèces de plantes pourraient être dues à des variations d'activité des ET et notamment des rétrotransposons à LTR (Bennetzen, et al. 2005). De plus, l'étude de 262 génomes procaryotes a permis de mettre en évidence que le nombre de copies d'IS est corrélé à la taille du génome (Touchon and Rocha 2007). Ainsi, les génomes les plus grands ont le plus grand nombre de copies d'IS. Cette association permet d'expliquer 40% de la variation d'abondance des IS observée dans les génomes bactériens.

En plus d'être présents en nombre souvent important dans les génomes, les copies d'ET d'une même famille peuvent avoir des séquences nucléotidiques peu divergentes, ce qui en fait de potentiels points de recombinaison ectopique. La recombinaison ectopique est un phénomène de recombinaison entre deux séquences homologues ou non présentent à deux loci génomiques différents. Ainsi, avec l'augmentation du nombre de séquences génomiques disponibles, de très nombreux réarrangements chromosomiques de taille variable ont été mis en évidence aussi bien chez les eucaryotes que chez les procaryotes. Par exemple, la comparaison de souches phylogénétiquement proches de bactéries des genres *Bordetella* et *Rickettsia* a permis de montrer que la plupart des points de cassure observés suite à des réarrangements chromosomiques étaient localisés au niveau de copies d'IS (Felsheim, et al. 2009; Parkhill, et al. 2003).

2.2.2. Contrôle de l'activité de transposition

La capacité des ET à envahir les génomes et les coûts induits par leur présence et leur activité ont entraîné la mise en place de mécanismes de régulation de la transposition. Ces

mécanismes ont certainement une influence sur la dynamique des ET dans les génomes. Ces mécanismes peuvent être liés à la structure des ET eux-mêmes ou avoir été mis en place par les cellules hôtes qu'ils colonisent. La régulation peut avoir lieu tant au niveau de la machinerie cellulaire qu'au niveau du mécanisme insertionnel (Nagy and Chandler 2004).

La séquence des ET peut permettre leur régulation transcriptionnelle et traductionnelle. Certains types d'ET possèdent leur propre site d'initiation de la transcription, leur permettant de contrôler leur propre activité de transposition. Ainsi, le promoteur des rétrotransposons avec et sans LTR et des IS se trouve dans la séquence de l'ET en amont de la partie codante. De plus, certaines familles d'IS possèdent dans leur séquence un site de décalage traductionnel ou transcriptionnel qui permet d'associer deux ORF localisées sur deux cadres de lecture différents (Baranov, et al. 2006; Cordaux 2008; Escoubas, et al. 1991; Luthi, et al. 1990; Nagy and Chandler 2004). Un décalage de cadre de lecture traductionnel peut diminuer de 50 à 99% l'activité de transposition des IS (Escoubas, et al. 1991; Vogele, et al. 1991). En outre, les sites d'initiation de la transcription ou de fixation des ribosomes peuvent être séquestrés par la formation de structures secondaires de l'ARNm induite par les TIR via des repliements (Beuzon, et al. 1999; Nagy and Chandler 2004). Un autre moyen de contrôle de l'expression des ET est la production d'anti-ARNm. La traduction d'une famille d'IS peut être très fortement réduite par la production d'un anti-ARNm venant s'hybrider à l'ARNm au niveau du site de fixation des ribosomes, empêchant ainsi l'initiation de la traduction (Ma and Simons 1990; Nagy and Chandler 2004).

Outre la structure des ET, des mécanismes de défense « actifs » contre leur prolifération ont été sélectionnés par les génomes des hôtes. Pour exemple, deux processus de défense actifs seront décrits, d'une part l'interférence ARN (RNAi) qui a été identifiée chez les eucaryotes et le système CRISPR pour « *clustered regularly interspaced short palindromic repeat* » qui a été identifié chez les procaryotes.

Le RNAi utilise des fragments d'ARNm pour diminuer l'expression des gènes cibles. La régulation des ET par ce type de mécanisme a été principalement décrit chez les eucaryotes (Aravin, et al. 2007; Brennecke, et al. 2007; Czech, et al. 2008). Il existe 3 voies différentes faisant intervenir différents types de protéines pour réguler l'expression des gènes : les *Piwi interacting RNA* (piRNA), les *micro RNA* (miRNA) et les *small interacting RNA* (siRNA) (Chapman and Carrington 2007). Chez les bactéries, des petits ARNm non codants peuvent être transcrits par des promoteurs spécifiques en réponse à des stress (Gottesman 2005). De plus, des protéines homologues aux protéines PIWI essentielles dans la voie des piRNA eucaryotes ont été identifiées dans un grand nombre de génomes bactériens et sont pour la plupart potentiellement actives (Makarova, et al. 2009). Ces observations laissent supposer que les ET bactériens pourraient eux aussi être régulés par le mécanisme de RNAi.

Chez les procaryotes, un mécanisme de défense contre l'intrusion d'ADN extra chromosomique a été mis en évidence. Il s'agit du système CRISPR qui a été identifié dans différents génomes bactériens (Jansen, et al. 2002). Les CRISPR sont des cassettes contenant des séquences répétées séparées par des espaceurs qui sont des séquences de phages ou de plasmides capturées par le génome bactérien lors de précédentes invasions (Horvath and Barrangou 2010). Il est donc fort probable que ces CRISPR puissent être une sorte de mémoire des infections passées subies par la bactérie. Lors de la première rencontre d'un phage ou d'un plasmide, un fragment de l'ADN extra-chromosomique est ajouté dans le CRISPR. Lors de la seconde rencontre du phage ou du plasmide, la cassette CRISPR est transcrite en différents petits ARN, chacun étant spécifique d'un phage ou d'un plasmide déjà rencontré. Ces petits ARN ciblent ensuite l'ADN étranger qui est inactivé et éliminé (Horvath and Barrangou 2010).

Les différents impacts que nous avons pu décrire, aussi bien chez les eucaryotes et les procaryotes, ainsi que les mécanismes de régulation de la transposition, ont probablement une influence non négligeable sur la dynamique évolutive des ET.

2.2.3. Dynamique des ET chez les eucaryotes

Chez les eucaryotes, des modèles ont été réalisés pour comprendre quels sont les mécanismes influençant l'invasion et le maintien des ET. Ainsi, lors de l'invasion d'un génome par une copie d'ET, si celle-ci transpose à un taux constant peu élevé, il est très probable qu'elle ne se maintienne pas, notamment à cause de la dérive génétique (Le Rouzic and Capy 2005). A contrario, si l'ET est trop actif, il envahit le génome, ce qui peut entraîner des dégâts irréparables allant jusqu'à la stérilité de l'organisme et donc la perte de l'ET. Dans le modèle proposé, les seuls ET qui réussissent à se maintenir dans un génome sont ceux dont la transposition est régulée. Après leur transfert, ces ET subissent une explosion de l'activité de transposition, qui sera suivie d'une forte limitation de leur activité (Le Rouzic and Capy 2005).

Une fois l'ET présent dans un génome de façon durable en une ou plusieurs copies, il est intéressant de connaitre l'évolution à long terme de ces éléments. Les modèles de génétique des populations utilisés pour prédire l'évolution à long terme des ET dans les génomes eucaryotes suggère qu'il n'existe pas un seul scénario de maintien des ET mais plusieurs qui combinent les effets de différents phénomènes (Le Rouzic, et al. 2007a).

Ainsi, la sélection négative, entraîne une élimination très rapide des ET alors que la sélection positive permet leur maintien à très long terme. Un ET avec un effet délétère sera probablement très rapidement éliminé, voire indétectable. A contrario, un ET avec un effet bénéfique peut être sélectionné et conservé, à très long terme, dans les génomes (Le Rouzic,

et al. 2007a). Dans certains cas, on parle même de domestication des ET lorsqu'ils semblent être utilisés par le génome dans lequel ils se trouvent (Feschotte and Pritham 2007b).

Un autre facteur très important dans le maintien des ET est la dérive génétique (Le Rouzic and Capy 2005). La dérive génétique est un phénomène qui entraîne la variation stochastique des fréquences alléliques dans une population. L'effet de la dérive génétique est inversement proportionnel à l'effectif efficace de la population. L'effectif efficace de la population est le nombre d'individus de la population qui peuvent participer à produire la génération suivante. C'est l'effectif total de population duquel on retire les individus trop jeunes ou trop vieux pour se reproduire. Ainsi, plus l'effectif efficace d'une population sera faible plus l'effet de la dérive génétique sera important. La dérive génétique est généralement matérialisée par l'accumulation et la fixation de mutations stochastiques non sélectionnées dans une région génomique donnée. En fonction de l'importance de cette dérive génétique et des pressions de sélection qui s'exercent, l'évolution du génome va être variable. L'évolution de régions génomiques dont la fonction est essentielle sera plus contrainte. Ainsi, seules certaines mutations seront observables, les autres étant rapidement éliminées par la sélection. Par contre, les régions moins contraintes du génome, comme les ET en général, vont évoluer plus rapidement de façon neutre. Ainsi, la dérive est une force d'évolution des ET très importante.

La sélection, la dérive génétique ou encore l'intensité de l'activité de transposition sont les facteurs qui déterminent si un ET va envahir un génome (Le Rouzic and Capy 2005) et s'y maintenir (Le Rouzic, et al. 2007a). Néanmoins, l'une des clés de la présence des ET dans les génomes est la possibilité de transfert horizontal entre génomes hôtes. En effet, ce sont ces transferts qui permettent aux ET d'arriver dans un génome. Avec le nombre croissant de séquences génomiques disponibles, de très nombreux cas de transferts horizontaux ont été identifiés à ce jour (De Setta, et al. 2011; Diao, et al. 2011; Schaack, et al. 2010; Thomas, et

al. 2011). Ainsi, il a été proposé que les relations hôtes-parasites pourraient favoriser le transfert d'ET entre eucaryotes issus de groupes phylogénétiques différents (Gilbert, et al. 2010).

Lorsque les ET arrivent à se maintenir dans un génome, ils évoluent ensuite avec la dérive et la sélection et sont dégradés. Ainsi, Les génomes eucaryotes contiennent de très nombreux ET dégradés, ce qui permet de retracer leur dynamique à long terme et de vérifier les prédictions des modèles. Dans le génome humain une chronologie de l'activité des différents types d'ET présents a pu être réalisée (Lander, et al. 2001). Les ET de classe 2 ont subi une très forte expansion durant la radiation des mammifères, au début de celle des primates il y a 60 à 150 millions d'années et sont maintenant inactifs dans le génome humain (Pace and Feschotte 2007). Les ET de classe 1 de type *Alu* et *L1* ont été très actifs durant l'évolution des primates, avant que leur activité ne décline il y a environ 20 millions d'années (Khan, et al. 2006; Xing, et al. 2004). De plus, il semble que les invasions ou bien les périodes d'activité d'ET dans un génome puissent être récurrentes. En effet, les génomes de drosophile qui ont subi une invasion fulgurante des éléments *P* en seulement 30 ans (Kidwell 1983) présentent des traces d'éléments proto*P* fossiles datés d'environ 5 millions d'années, qui sont homologues des éléments *P* actuels (Kapitonov and Jurka 2003).

La présence de copies fossiles d'ET est donc une opportunité de retracer à la fois l'histoire évolutive des ET eux-mêmes, ainsi que celle des génomes hôtes et ceci même si les ET ne sont plus actuellement actifs.

2.2.4. Dynamique des ET chez les procaryotes

Contrairement à la dynamique des ET eucaryotes, celle des ET procaryotes est beaucoup moins étudiée. Seules quelques études ont été réalisées sur la dynamique des IS (Bichsel, et al. 2010). Cependant, on peut supposer qu'une partie des modèles décrivant la

dynamique des ET eucaryotes peut être transposée aux ET procaryotes. La majorité des informations connues concernent les IS, qui sont les ET procaryotes les plus fréquents.

Les génomes de procaryotes possèdent généralement très peu d'ET. En effet, les IS ne couvrent en moyenne que 3% d'un génome (Siguier, et al. 2006a), à l'exception de quelques rares cas comme *Shigella dysenteriae* ou *Orientia tsutsugamushi* chez lesquels la couverture est supérieure à 10% (Cho, et al. 2007; Yang, et al. 2005). Cette faible abondance pourrait être expliquée par la proportion de séquences codantes qui est supérieure à 80% chez les procaryotes alors qu'elle est généralement plus faible chez les eucaryotes (Lawrence, et al. 2001). Par conséquent, les insertions d'ET chez les procaryotes ont une plus forte probabilité d'avoir un effet délétère et donc d'être éliminées par la sélection.

Il est couramment admis que les ET procaryotes sont d'origine récente, sur la base de leur répartition ou de leur divergence au sein des génomes. Par exemple, la distribution des IS dans des souches de bactéries phylogénétiquement proches est fortement variable (Parkhill, et al. 2003; Sawyer, et al. 1987; Yang, et al. 2005). De plus, la divergence très faible entre les copies de 3 familles d'IS d'*E. coli* (Lawrence, et al. 1992), ou à plus large échelle entre les copies de 20 familles d'IS de plusieurs centaines de génomes bactériens (Wagner 2006; Wagner, et al. 2007) renforce l'hypothèse d'une origine récente des ET bactériens. L'analyse des IS récemment intégrés dans les génomes bactériens suggère que leur dynamique à long terme pourrait suivre des cycles d'extinction-réinvasion dans lesquels les transferts horizontaux tiennent une place cruciale (Wagner 2006).

Le mode de vie des bactéries est un autre facteur qui influence la densité des ET dans leurs génomes et plus particulièrement celle des IS (Moran and Plague 2004). Le mode de vie des bactéries influence l'effectif efficace des populations bactériennes et fait donc varier l'effet de la dérive génétique et influence l'efficacité de la sélection. Les bactéries à vie libre ont généralement de grands effectifs efficaces de populations, ce qui implique une meilleure

efficacité de la sélection qui élimine efficacement les éléments non essentiels du génome. Par conséquent, les bactéries à vie libre ont un nombre d'ET généralement faible (Moran and Plague 2004). Lors du passage de la vie libre à la vie intracellulaire, on assiste à une diminution de l'effectif efficace de la population bactérienne qui se traduit par une baisse de l'efficacité de la sélection et une augmentation de l'effet de la dérive génétique (Dale and Moran 2006; Moran and Plague 2004; Wernegreen 2005, 2002). De plus, le passage de la vie libre vers le confinement de la vie intracellulaire implique que certains gènes ne sont plus essentiels à la survie de la bactérie (Moya, et al. 2008). Par conséquent, ces gènes peuvent subir l'effet de la dérive et être perdus par accumulation de mutations sans effet létal pour les bactéries (Moran and Plague 2004; Moya, et al. 2008; Wernegreen 2005, 2002). De plus, la diminution de l'efficacité de la sélection permet une propagation des IS dans les bactéries intracellulaires récentes (Moran and Plague 2004). Cette augmentation du nombre d'IS participe à la dégradation des génomes bactériens et à la diminution de leur taille suite à des délétions de grands fragments d'ADN causées par des évènements de recombinaison ectopique entre copies d'IS (Moya, et al. 2008; Wernegreen 2002). Ensuite, la réduction des génomes se fait de façon plus graduelle par accumulation de mutations ponctuelles (insertions, délétions, substitutions nucléotidiques) dans les gènes encore non essentiels à la survie de la bactérie, comme les ET (Moya, et al. 2008; Wernegreen 2002). Enfin, l'isolement des bactéries intracellulaires ne leur permet en principe pas d'acquérir de nouveaux gènes ni de nouveaux IS par transferts horizontaux (Dale and Moran 2006; Wernegreen 2005). Ainsi, les génomes des bactéries intracellulaires anciennes sont de petite taille et possèdent essentiellement des gènes nécessaires à la survie de la bactérie et à son interaction avec l'hôte (Dale and Moran 2006; Darby, et al. 2007; Moran, et al. 2008; Moya, et al. 2008), comme chez *Buchnera*, un endosymbiote mutualiste du puceron (Shigenobu, et al. 2000). Par conséquent, on s'attend à ce que les génomes de bactéries intracellulaires anciennes ne

contiennent que peu ou pas d'ET (Moran and Plague 2004). Néanmoins, le séquençage de génomes de bactéries endosymbiotiques a dévoilé certaines exceptions, telle que *Wolbachia pipientis* (ci-après *Wolbachia*).

2.3. La symbiose à Wolbachia

Wolbachia est une α-protéobactérie gram négative identifiée pour la première fois chez le moustique *Culex pipiens* en 1924 (Hertig and Wolbach 1924). Elle est intracellulaire obligatoire et vit dans le cytoplasme des cellules de ses hôtes. Elle est associée à ses hôtes nématodes et arthropodes depuis environ 100 millions d'années (Bandi, et al. 1998).

2.3.1. Diversité de souches de Wolbachia

Wolbachia infecte des nématodes filaires (Bandi, et al. 1998) et de très nombreux arthropodes(Bouchon, et al. 2008; Jeyaprakash and Hoy 2000; Werren and O'Neill 1997). Du fait de son large spectre d'hôtes, *Wolbachia* est probablement la bactérie endosymbiotique la plus répandue sur Terre. Une étude suggère que 66% des espèces d'insectes sont infectées par *Wolbachia* (Hilgenboecker, et al. 2008). Sur la base du séquençage de différents

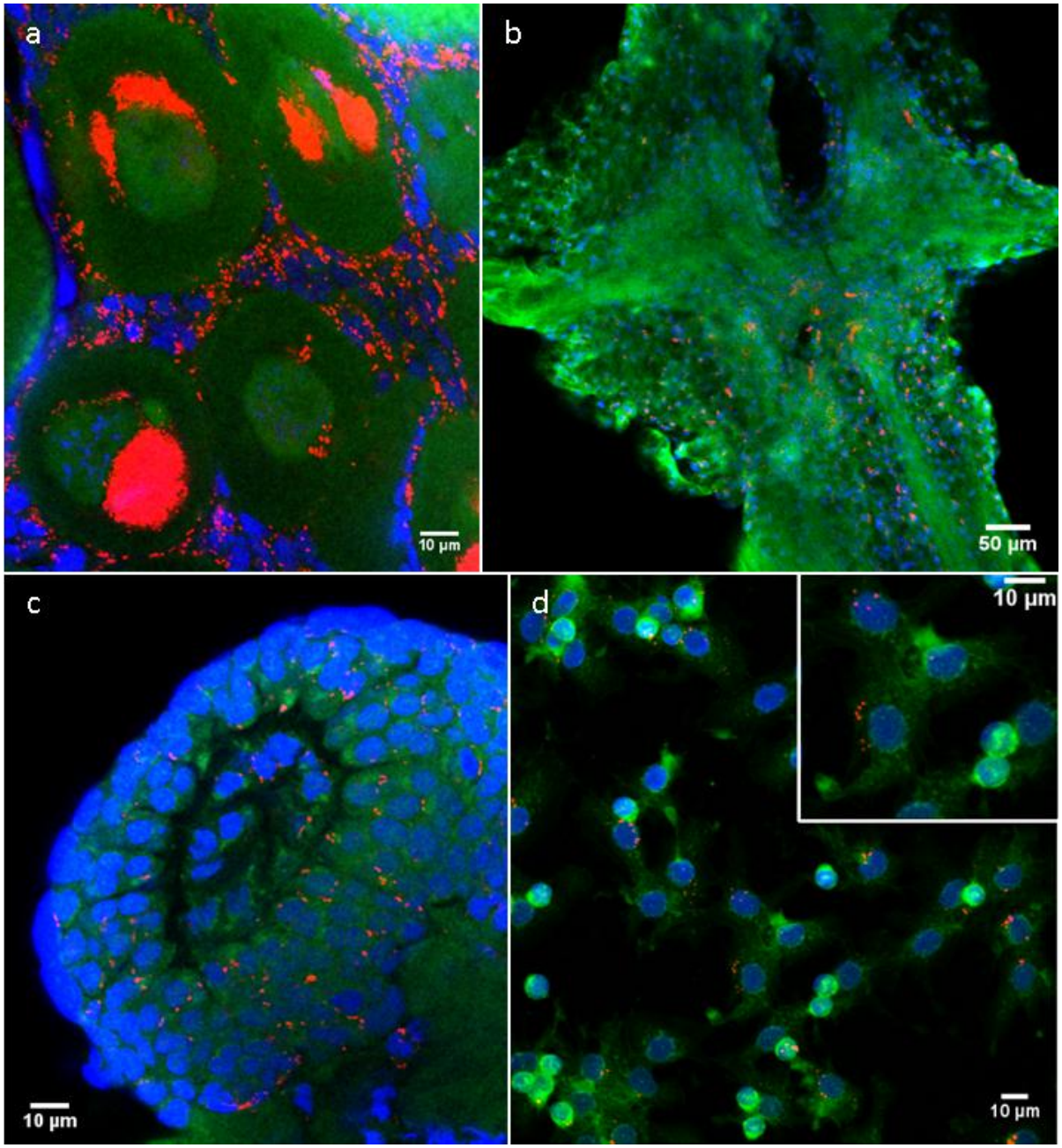

Figure 2-4 : Détection de *Wolbachia* en hybridation *in situ* fluorescente (FISH) chez le crustacé isopode terrestre *Armadillidium vulgare*.

Les noyaux des cellules sont marqués en bleu (DAPI), l'architecture cellulaire est marquée en vert via l'actine (phalloïdine FITC) et *Wolbachia* est marquée en rouge (FISH). *Wolbachia* est bien sur présente dans les ovaires (a) afin d'assurer sa transmission à la génération d'hôte suivant, mais elle est aussi présente en densité plus faible dans d'autres organes comme la chaîne nerveuse (b). Enfin, *Wolbachia* est aussi présente dans les organes hématopoïétiques (c) qui sont le centre de production des cellules du système immunitaire que sont les hémocytes (d). Crédit photographique : Joanne Bertaux LEES®

marqueurs génétiques, les souches de *Wolbachia* ont été classées en 9 supergroupes de diversité (Haegeman, et al. 2009; Lo, et al. 2007). Les supergroupes A et B contiennent la majorité des *Wolbachia* d'arthropodes, les supergroupes C et D celles des filaires, les supergroupes E à H celles de collemboles, termites et certains arthropodes et enfin le supergroupe I contient le symbiote d'un nématode parasite de plantes.

2.3.2. Mode de vie

Wolbachia entretient différents types d'interaction avec ses hôtes depuis le mutualisme jusqu'à la pathogénicité. Chez les nématodes, *Wolbachia* est mutualiste obligatoire, car elle est nécessaire au développement et à la reproduction de son hôte (Bandi, et al. 1998). Chez les arthropodes, elle est généralement symbiote facultatif et induit des altérations de la reproduction qui ont notamment pour conséquence de favoriser les femelles dans les populations hôtes et donc d'augmenter sa transmission et sa prévalence dans les populations (Cordaux, et al. 2011). Néanmoins, certains cas d'associations mutualistes obligatoires (Dedeine, et al. 2004; Hosokawa, et al. 2010) et au moins un cas de virulence (Min and Benzer 1997) entre *Wolbachia* et ses hôtes arthropodes ont été identifiés. *Wolbachia* se transmet généralement de façon verticale par voie ovocytaire de la mère aux descendants mais peut aussi se transmettre horizontalement entre différents hôtes (Cordaux, et al. 2001; Michel-Salzat, et al. 2001; Rigaud and Juchault 1995; Vavre, et al. 1999).

2.3.3. Localisation tissulaire

Chez les arthropodes, du fait de son impact sur le déterminisme du sexe et de sa transmission maternelle par voie ovocytaire, on retrouve préférentiellement *Wolbachia* dans les organes reproducteurs (Figure 2-4 A). Néanmoins, chez les isopodes terrestres, *Wolbachia*

est trouvée en densité plus faible dans différents organes comme les chaines nerveuses et les cellules du système immunitaire tels que les organes hématopoïétiques et les hémocytes (Chevalier, et al. 2011) (Figure 2-4 B, C et D). La présence de *Wolbachia* dans d'autres organes que les ovaires, et notamment dans les hémocytes circulant dans l'hémolymphe, pourrait faciliter la transmission horizontale entre des hôtes différents, si ceux-ci échangent de l'hémolymphe via des blessures par exemple (Chevalier, et al. 2011; Rigaud and Juchault 1995).

2.3.4. Les génomes de Wolbachia

A ce jour, 4 génomes provenant de différentes souches de *Wolbachia* ont été complètement séquencés et annotés. Il s'agit des endosymbiotes du nématode *Brugia malayi* (*w*Bm), des mouches *Drosophila melanogaster* (*w*Mel) et *D. simulans* (*w*Ri) ainsi que du moustique *Culex pipiens* (*w*Pel) (Tableau 2-1 – page suivante). Un cinquième génome en cours de séquençage dans notre laboratoire a été mis à notre disposition, il s'agit de celui de la souche *w*VulC infectant l'isopode terrestre *Armadillidium vulgare* (Tableau 2-1). Dans la suite du manuscrit, nous utiliserons les abréviations de chacune des souches pour les désigner. Ces 5 souches de *Wolbachia* représentent 3 supergroupes de diversité différents (Tableau 2-1 ; Figure 2-5). De plus, ces souches induisent des effets différents chez leur hôte puisqu'il y a une souche mutualiste et 4 souches parasites de la reproduction (Tableau 2-1). Les différences d'interaction entre souches de *Wolbachia* et leurs hôtes entraînent des différences dans le mode de transmission entre individus. Pour la souche mutualiste, la transmission est uniquement verticale entre la mère et les descendants alors que pour les souches parasites de la reproduction, la transmission se fait verticalement de façon prédominante, mais aussi horizontalement de façon occasionnelle (Tableau 2-1).

Tableau 2-1 : Caractéristiques générales des 5 génomes de *Wolbachia*.

Souche	*w*Bm	*w*Mel	*w*Ri	*w*Pel	*w*VulC
Hôte	*Brugia malayi* (nématode)	*Drosophila melanogaster* (mouche)	*D. simulans* (mouche)	*Culex pipiens* (moustique)	*Armadillidium vulgare* (isopode)
Interaction	Mutualiste	Parasites du sexe			
Supergroupe de diversité	D	A	A	B	B
Transmission	Verticale uniquement	Verticale (principalement) et horizontale (occasionnellement)			
Taille de génome (Mb)	1,08	1,27	1,45	1,48	1,66
% de G+C	34,2	35,2	35,2	34,2	~35%
Publication de référence	(Foster, et al. 2005)	(Wu, et al. 2004)	(Klasson, et al. 2009)	(Klasson, et al. 2008)	Non publié

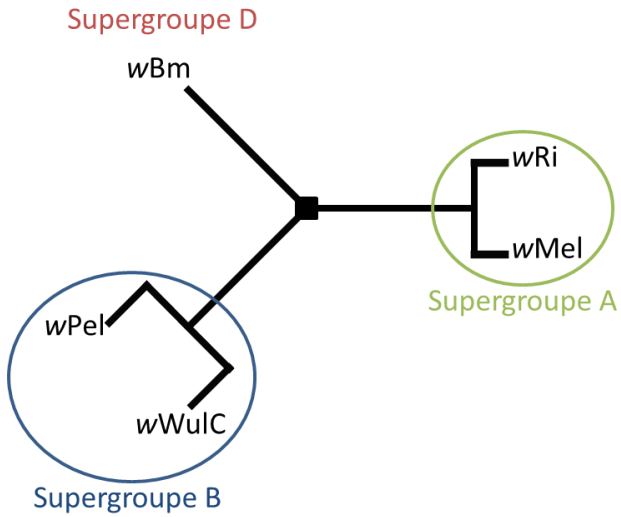

Figure 2-5 : Phylogénie des 5 souches de *Wolbachia* étudiées

La phylogénie des souches de *Wolbachia* a été réalisée à partir de la littérature (Lo, et al. 2007). La taille des branches étant arbitraire.

Les génomes de *Wolbachia* possèdent à la fois des caractéristiques typiques des bactéries intracellulaires anciennes et des particularités inattendues. Ainsi, les génomes de *Wolbachia* sont de taille modeste (Tableau 2-1) par comparaison aux bactéries à vie libre dont la taille moyenne de génome est de l'ordre de 5 Mb (Wernegreen 2002). Nous avons à notre disposition 5 génomes recouvrant une gamme de taille comprise entre 1,1 et 1,7 Mb (Tableau 2-1). Le génome le plus petit correspond à la seule souche mutualiste à notre disposition, ce qui est parfaitement en accord avec les théories d'évolution génomique chez les bactéries endosymbiotiques (Wernegreen 2002). De plus, la composition des génomes de *Wolbachia* présente un biais de composition en nucléotides, avec 35% de nucléotides Cytosine et Guanine (Tableau 2-1), ce qui est là aussi caractéristique des génomes de bactéries intracellulaires (Wernegreen 2002).

Cependant, les génomes de *Wolbachia* contiennent une très grande quantité d'éléments génétiques mobiles (MGE) comme par exemple des prophages (Kent and Bordenstein 2010) mais aussi des ET (Leclercq, et al. 2011; Wu, et al. 2004). Les prophages sont des formes intégrées au génome bactérien proche de bactériophages libres qui ont été détectés dans certaines souches de *Wolbachia* (Tanaka, et al. 2009). De plus, les génomes de *Wolbachia* font partie des génomes bactériens contenant le plus d'introns du groupe II (ET de classe 1) (Leclercq, et al. 2011) et ils contiennent aussi une des plus fortes densités d'IS rapportées dans des génomes bactériens (Moran and Plague 2004). Des analyses ciblées indiquent que certains de ces IS sont anciens (Cordaux 2009), alors que d'autres présentent une forte activité récente (Cordaux, et al. 2008).

Les objectifs de ma thèse étaient d'annoter et d'analyser les IS des génomes de *Wolbachia* afin de caractériser leur dynamique évolutive à l'échelle de l'ensemble d'un génome et de tous les groupes d'IS. La variété de paramètres des génomes que nous avons étudiés nous a permis de faire des inférences évolutives sur l'importance de certains d'entre

eux dans la dynamique des ET chez *Wolbachia*. Après une description détaillée de chacun des génomes, l'analyse de la dynamique évolutive des ET a été réalisée selon une logique temporelle depuis l'activité ancienne vers la plus récente. Les inférences sur la dynamique évolutive des ET ont été faites à partir de l'analyse des 5 génomes de *Wolbachia*. De plus, nous avons utilisé une approche expérimentale pour explorer plus avant certaines parties de notre raisonnement, notamment concernant les transferts horizontaux. Nous avons ainsi pu retracer la dynamique passée des IS dans les génomes de *Wolbachia* et tenter de l'expliquer grâce à des observations faites sur l'activité récente des éléments.

3. Détection et caractérisation des IS de *Wolbachia*

Ce chapitre s'appuie en partie sur l'article de recherche ayant pour titre "Short and long-term evolutionary dynamics of bacterial insertion sequences: insights from *Wolbachia* endosymbionts" accepté pour publication en septembre 2011 dans la revue Genome Biology and Evolution et présenté en annexe 2 (Cerveau, et al. 2011b).

3.1. Contexte

Nous avons décrit dans l'introduction que les IS sont présents dans environ 80% des génomes bactériens (Siguier, et al. 2006a; Touchon and Rocha 2007) avec des variations de densité en fonction du mode de vie des bactéries (Moran and Plague 2004). Les modèles prédisent que la densité en IS devrait être plus faible chez les bactéries intracellulaires (Moran and Plague 2004; Wernegreen 2002). Néanmoins, le séquençage de génomes de bactéries intracellulaires présentant une forte densité d'IS, telles que *Orientia tsutsugamushi* ou *Wolbachia*, a remis en cause ces modèles (Nakayama, et al. 2008; Wu, et al. 2004).

Afin de mieux comprendre les raisons de la forte abondance d'IS chez *Wolbachia* nous avons caractérisé les populations d'IS de 5 génomes. Il s'agit de 4 génomes complètement séquencés et annotés (*w*Bm, *w*Mel, *w*Pel et *w*Ri) et du génome en cours de séquençage et d'assemblage dans notre laboratoire (*w*VulC) (Tableau 2-1 - page 27). Des annotations incomplètes des IS dans les génomes de *w*Mel, *w*Pel et *w*Ri ont été présentées dans les publications originales, ce qui a permis de mettre en évidence une très forte densité d'IS par rapport aux génomes d'autres bactéries endosymbiotiques. Le génome de *w*VulC en cours de séquençage et d'assemblage dans notre laboratoire n'a pas encore fait l'objet d'une annotation des IS. Concernant le génome de *w*Bm, la publication originale ne comportait pas d'annotation des IS, mais elle a été réalisée avant le début de ma thèse (Cordaux 2009). Notre

but était de décrire en détail la composition en IS des génomes à notre disposition afin de comprendre les raisons de la forte abondance des IS chez *Wolbachia*.

3.2. Méthodes

3.2.1. Détection

Les génomes complètement séquencés et annotés des souches *w*Mel, *w*Ri, *w*Pel et *w*Bm ont été consultés et téléchargés depuis le site internet de NCBI (http://www.ncbi.nlm.nih.gov/genomes/MICROBES/microbial_taxtree.html). Le séquençage du génome de *w*VulC a été coordonné au sein de notre laboratoire par Pierre Grève, Didier Bouchon et Richard Cordaux et réalisé par Roger Garrett et Chao Liu (Université de Copenhague, Danemark).

De très nombreuses méthodes d'annotations des ET dans les génomes aussi bien eucaryotes que procaryotes ont été développées (Feschotte and Pritham 2007a; Kichenaradja, et al. 2010; Lerat 2010). Chacune des méthodes ayant des avantages et des inconvénients, nous avons décidé d'utiliser 3 méthodes différentes. D'une part, une annotation *de novo* a été réalisée avec le logiciel Repeatscout en utilisant le paramètre l-mers=15 pb (Price, et al. 2005). Cette méthode à l'avantage de ne pas faire appel à une base de données de référence mais elle ne permet d'identifier que les éléments répétés ayant une fréquence supérieure ou égale à 3. Des fragments de 15 pb sont générés à partir de la séquence génomique puis comparés à celle-ci. Si le fragment trouve 3 zones de similarité ou plus, sa taille augmente tant que la similarité est maintenue. Nous avons aussi fait une recherche par similarité avec la base de données ISFinder (Siguier, et al. 2006b) via l'interface internet ISsaga (Varani, et al. 2011). Cette méthode permet la détection de l'ensemble des IS présentant une similarité avec

au moins un élément de la base ISFinder, ce qui nécessite qu'il y ait des séquences similaires dans la base de données. Enfin, nous avons fait une recherche avec les mots clés « transposase » et « transposon » dans l'annotation originale de chaque génome.

La combinaison de ces 3 méthodes nous a permis de créer une première bibliothèque de copies d'IS de grande taille qui a ensuite été utilisée pour rechercher des fragments d'IS. La recherche de fragments a été réalisée au niveau nucléotidique par des recherches de similarité via BLASTN avec les paramètres par défaut, en retirant le filtre sur les *low complexity region*. La taille minimum des fragments a été fixée à 40 pb, avec une similarité minimum de 75%, une *e-value* inférieure à 5%, une valeur de *reward* de 2 et une valeur de *penalty* de 3. Afin d'éviter les redondances, les positions de chaque copie ont été relevées et comparées à celles des autres copies du génome considéré. Ceci a notamment permis d'identifier 38 cas de copies d'IS coupées en deux par l'insertion d'un autre IS. Dans ce cas, chacune des copies coupées a été comptée comme un seul élément (3 dans *w*Mel, 6 dans *w*Ri, 9 dans *w*Pel et 20 dans *w*VulC). Enfin, nous avons retiré du jeu de données des copies qui avaient été identifiées comme des faux positifs (par exemple, le gène *DnaA* a été identifié par erreur comme un IS de la famille IS21). Ainsi, le jeu de données final basé sur les 5 génomes contient 870 copies d'IS.

3.2.2. Classification

Une famille a été assignée à chaque copie d'IS par recherche de similarité de la séquence requête traduite en protéine contre la base de données nucléique ISFinder, elle aussi traduite en protéine (TBLASTX) (Siguier, et al. 2006b). Les séquences de chaque famille ont ensuite été alignées avec l'algorithme ClustalW implémenté dans Bioedit ver 7.0 (Hall 1999) et les alignements ont été vérifiés manuellement. La diversité des séquences au sein d'une famille ne nous a pas permis d'aligner toutes les séquences d'une même famille entre elles,

donc nous avons fait des regroupements de séquences alignables. Par conséquent, nous avons défini des groupes à l'intérieur des familles comme des sous-ensembles de séquences nucléotidiques qui peuvent s'aligner entre elles, mais qui ne peuvent pas s'aligner avec celles des autres groupes. Comme pour les familles, nous avons assigné des noms aux groupes d'IS par recherche de similarité en TBLASTX avec la base de données ISFinder. Les groupes d'IS n'ayant pas d'homologues dans ISFinder et sans copie potentiellement fonctionnelle (non déposable dans ISFinder) ont été nommés : IS« nom de la famille »-w (pour *Wolbachia*) suivie d'une lettre par groupe d'IS différent (ex : le groupe IS110wA).

3.2.3. Etude de la distance génomique entre copies d'IS

Afin d'analyser la distribution des copies dans les 5 génomes de *Wolbachia*, nous avons étudié la distance génomique qui existe entre les différentes copies de chacun des génomes. Pour cela, dans chaque génome, nous avons classé les copies dans l'ordre croissant en fonction de leurs coordonnées génomiques. Puis, nous avons comparé les positions de chaque copie avec celles de la suivante pour calculer la distance génomique minimale entre les deux copies. Dans le cas où une copie est insérée dans une autre, la distance entre les deux copies a été considérée comme nulle. Les distances génomiques entre copies ont ensuite été regroupées dans des classes de 1 Kb entre 0 et 60 Kb.

De plus, nous avons caractérisé la relation qui existe entre la densité en IS et la médiane de la distribution des distances entre copies en faisant des régressions linéaires et non linéaires. Les régressions ont été réalisées et testées statistiquement avec le logiciel R. Nous avons calculé pour chacune des régressions le critère d'Akaike (AIC) qui permet de discriminer les modèles, le modèle avec la plus faible valeur étant celui qui correspond le mieux aux données.

De plus, pour chacun des génomes une distribution aléatoire de distance entre copies a été générée. Pour cela, nous avons considéré pour chaque génome, la taille *t* qui est la taille du génome de laquelle nous avons retiré la taille totale couverte par les IS et le nombre *n* de copies d'IS du génome considéré. Nous avons ensuite simulé pour chaque génome *n* insertions d'IS dans une taille *t*. Puis nous avons classé par ordre croissant la position de chaque insertion et nous avons calculé la distance entre les positions de deux insertions contiguës. Enfin, nous avons comparé par un test de Wilcoxon les distances entre copies observées et théoriques pour chacun des génomes.

3.2.4. Importance des MGE dans les variations de taille de génome de Wolbachia

Afin d'évaluer l'importance des MGE dans les variations de taille de génomes observées entre souches de *Wolbachia,* nous avons analysé les variations de la taille totale couverte par les 3 types de MGE présents dans les génomes de *Wolbachia* : les prophages, les introns de groupe II et les IS. Nous avons tout d'abord calculé la couverture totale de chaque type de MGE dans chacun des génomes de *Wolbachia* en faisant la somme de la taille de tous les éléments identifiés et annotés. Les données concernant les prophages ont été obtenues soit en cherchant dans les annotations originales des génomes pour *w*Mel et *w*Pel soit par recherche de similarité en utilisant les séquences présentes dans les génomes de *w*Mel et *w*Pel. L'annotation des introns de groupe II dans les génomes de *Wolbachia* a été réalisée par Dr. Sébastien Leclercq (Leclercq et al 2011) et l'annotation des IS est issue directement de notre travail.

Nous avons ensuite calculé d'une part les variations de taille entre les génomes de *Wolbachia* et d'autre part les variations de taille totale de chaque type de MGE entre les différents génomes de *Wolbachia*. Dans certains cas, les valeurs étaient négatives. Ceci

Tableau 3-1 : Analogies entre notion écologique et ET

Notion	Définition écologique	Analogies avec les ET
Écosystème	Milieu de vie hétérogène composé de nombreux habitats ayant des caractéristiques variables	Génome
Espèce	Ensemble d'individus se ressemblant et interféconds	Ensemble de copies se ressemblant et appartenant à la même famille ou au même groupe d'ET
Individu	Un membre d'une espèce	Une copies d'ET
Population	Ensemble des individus d'une même espèce présents dans l'environnement	Ensemble de copies d'une même famille ou groupe d'ET présent dans le génome
Communauté	Ensemble d'individus de différentes espèces présents dans l'écosystème	Ensemble de copies d'ET de différentes familles ou groupes présent dans le génome
Richesse spécifique	Nombre d'espèces identifiées dans un écosystème	Nombre de familles ou groupes d'ET présents dans un génome
Abondance spécifique	Nombre d'individus d'une espèce	Nombre de copies d'une famille ou d'un groupe d'ET
Niche écologique	Conditions nécessaires pour qu'une espèce puisse s'établir dans un habitat	Paramètres permettant à l'ET d'être présent dans un génome
Taux de naissance	Nombre de naissance à chaque génération	Taux de transposition des copies d'ET
Taux de mortalité	Nombre de mort à chaque génération	Taux de délétion ou de dégradation des copies d'ET
Migration	Déplacement d'individus entre écosystèmes	Transferts horizontaux de copies d'ET entre génomes
Compétition	Mise en concurrence des individus d'une même espèce ou non pour une ressource de l'écosystème	Mise en concurrence des copies d'ET pour une ressource du génome (ex : les sites d'insertions libres)

L'ensemble des informations contenues dans ce tableau est issu de la littérature (Le Rouzic, et al. 2007b; Venner, et al. 2009).

signifie que le génome qui a servi de référence, avait une couverture totale, en un type de MGE, plus faible que celui avec lequel on l'a comparé. Dans ce cas, nous avons ajouté la valeur absolue de la différence, calculée pour le type de MGE considéré, à la différence de taille totale entre les deux génomes considérés. Nous avons ensuite évalué la part de variation de taille de génome entre souches de *Wolbachia* qui est expliquée par les variations de taille totale de chaque type de MGE. D'autre part, nous avons fait des régressions linéaires entre la couverture en MGE et la taille totale des génomes pour étudier la relation entre ces deux paramètres. Les régressions ont été réalisées et testées statistiquement avec le logiciel R. Nous avons calculé pour chacune des régressions le critère d'Akaike (AIC).

3.2.5. Analyses de diversité

Afin d'analyser la composition des génomes de *Wolbachia* en IS, nous avons choisi des indices classiquement utilisés pour décrire les populations animales ou végétales. Par bien des caractéristiques les populations animales ou végétales d'un écosystème et les populations d'ET d'un génome sont proches. En effet, elles évoluent toutes les deux dans un espace fini composé de milieux hétérogènes (*i.e.* écosystème). De plus, de nombreuses notions écologiques utilisées pour décrire les populations des êtres vivants sont utilisables pour décrire les populations d'ET (Tableau 3-1). Par exemple, on peut considérer une population comme l'ensemble des individus d'une espèce présent dans un milieu mais aussi comme l'ensemble des copies d'un groupe d'ET présentes dans un génome. Etant donné ce parallèle, il est concevable que les indices de diversité, utilisés classiquement pour caractériser les communautés d'espèces d'un écosystème, soient utilisables pour caractériser la diversité des ET d'un génome (Brookfield 2005; Le Rouzic, et al. 2007b; Venner, et al. 2009). Nous avons appliqué ces concepts d'écologie aux IS présents dans les génomes de *Wolbachia* en considérant chaque génome comme un écosystème donné, les groupes d'IS comme des

espèces, le nombre de copies d'IS dans un groupe donné comme leur abondance et le nombre de groupe par génome comme la richesse spécifique.

L'ensemble des indices écologiques a été calculé avec le package BiodiversityR implémenté dans le programme R (http://www.r-project.org/). Différents indices de diversité existent. Nous avons choisi d'utiliser les deux plus connus : l'indice de Shannon (noté H') et l'indice de Simpson (noté D).

L'indice de Shannon se calcule de la façon suivante: $H' = \sum_{i=1}^{S} \left(\frac{N_i}{N} \times \log_2 \frac{N_i}{N} \right)$ où Ni représente l'abondance du groupe d'IS i et N l'abondance cumulée de l'ensemble des groupes d'IS d'un génome. Il est sensible aux espèces dites rares, c'est-à-dire peu abondantes. La diversité est minimale si tous les individus d'un milieu font partie de la même espèce et maximale l'abondance de toutes les espèces est égale. L'indice de diversité de Shannon est accompagné de l'indice d'équitabilité ou d'équirépartition de Pilou noté J. Cet indice qui se calcule de la façon suivant: $J = \frac{H'}{\log(S)}$ où H' est l'indice de Shannon et S est la richesse spécifique du milieu considéré, c'est-à-dire le nombre total de groupe d'IS présents dans le génome. Il varie entre 0 et 1 et permet de caractériser la structure d'un peuplement. Il est maximal quand toutes les espèces ont une abondance identique et minimale quand une espèce prédomine toutes les autres.

L'indice de Simpson se calcule de la façon suivante: $D = \sum_{i=1}^{S} \left(\frac{N_i \times (N_i - 1)}{N \times (N - 1)} \right)$ où Ni représente l'abondance du groupe d'IS i et N l'abondance de l'ensemble des groupes d'IS d'un génome. Cet indice est sensible aux espèces abondantes. Il aura une valeur de 0 pour indiquer le maximum de diversité, et une valeur de 1 pour indiquer le minimum de diversité.

Enfin pour comparer les communautés d'IS présentes dans les 5 génomes, nous avons réalisé un dendrogramme sur la base de la présence/absence des groupes d'IS et de leur

Tableau 3-2: Caractéristiques générales des IS dans les cinq génomes de *Wolbachia*.

	*w*Bm	*w*Mel	*w*Pel	*w*Ri	*w*VulC
Taille de génome (Mb)	1,08	1,27	1,48	1,45	1,66
Nombre total de copies d'IS	52	105	170	171	372
Densité en IS (copies/Mb)	48	83	115	118	224
Couverture génomique en IS (pb)	28 188	76 886	123 823	159 767	235 507
Proportion génomique en IS	2,6%	6,1%	8,4%	11,0%	14,2%

abondance dans les différents génomes. Nous avons utilisé la méthode d'analyse ward qui s'approche de la méthode neighbor joining de calcul pour la création d'arbre phylogénétique.

3.2.6. Analyse structurelle

Les séquences génomiques ont été étendues aux 500 pb en amont et en aval de chacune des copies d'IS des différentes familles ou groupes. Ceci nous a permis de clairement identifier les bornes de chaque copie et de savoir si elles étaient complètes ou non grâce à l'identification des composants typiques des IS que sont les TIR et les TSD (s'ils étaient présents) ou par comparaison des séquences flanquantes des copies lorsqu'il y en a plusieurs.

De plus, les gènes codant les transposases ont été recherchés et identifiés sur chacune des copies complètes à l'aide des logiciels ORFfinder (http://www.ncbi.nlm.nih.gov/projects/gorf/) et FSfinder (Moon, et al. 2004). Ce dernier permet notamment de détecter les décalages de cadre de lecture (Baranov, et al. 2006) observés pour certains groupes d'IS (Chandler and Mahillon 2002; Cordaux 2008).

3.3. Résultats et discussion

3.3.1. Abondance des IS chez Wolbachia

Nous avons utilisé 3 méthodes de détection complémentaires pour identifier les copies d'IS présentes dans les génomes de *w*Bm, *w*Mel, *w*Pel, *w*Ri et *w*VulC. Cette analyse nous a permis de mettre en évidence que ces génomes contenaient entre 52 et 372 copies d'IS d'une taille supérieure à 40 pb (Tableau 3-2). Ces abondances sont considérables car les génomes de bactéries, quel que soit leur mode de vie, contiennent généralement peu de copies d'IS. En effet, l'étude de 262 génomes procaryotes a mis en évidence un nombre médian de 12 copies

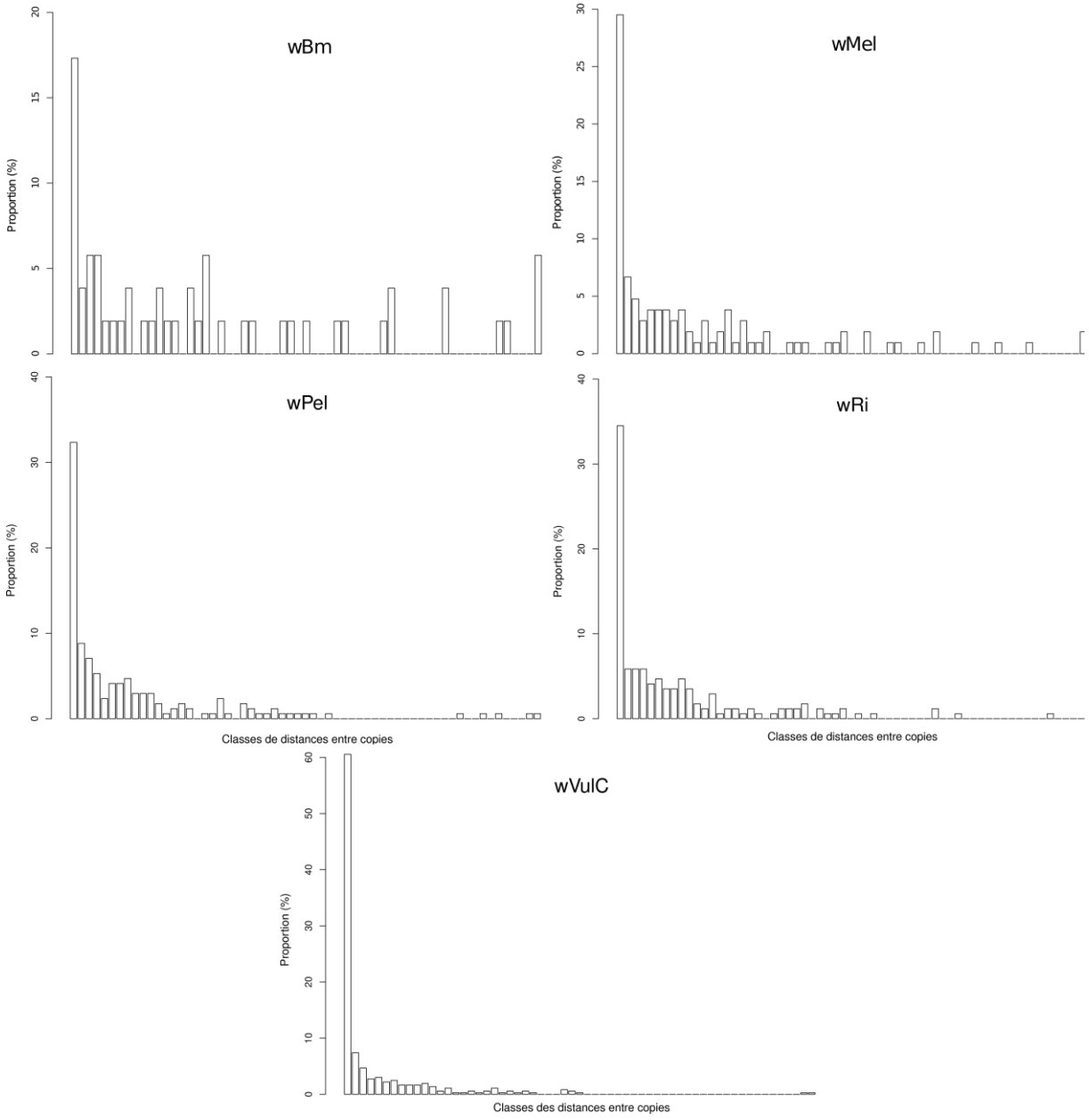

Figure 3-1 : Distribution de la fréquence des distances entre copies pour chaque génome.

Les distances génomiques entre copies d'IS ont été classées en 60 classes de 1Kb chacune, entre 0 et 60 Kb plus une classe pour les distances supérieures à 60Kb.

d'IS (avec des valeurs comprises entre 0 et 320) par génome (Touchon and Rocha 2007). Les génomes de *Wolbachia*, notamment celui de *w*VulC, sont probablement parmi les génomes bactériens connus à ce jour ayant le plus de copies d'IS. Les IS couvrent jusqu'à 14,2% des génomes de *Wolbachia* (Tableau 3-2), ce qui dépasse de très loin la couverture couramment observée dans les génomes procaryotes. En effet, la couverture des IS dans les génomes bactériens est généralement inférieure à 3% (Siguier, et al. 2006a), à l'exception de quelques rares cas comme *Shigella dysenteriae*, *Orientia tsutsugamushi* ou *Sulfolobus solfataricus* dont la couverture est supérieure à 10% (Brugger, et al. 2004; Cho, et al. 2007; Filee, et al. 2007; Nakayama, et al. 2008; Yang, et al. 2005).

La forte abondance des IS dans les génomes de *Wolbachia* pourrait être liée au schéma d'insertion des s copies. En effet, certains ET ciblent d'autres ET (Chillon, et al. 2010) ou des régions génomiques répétées, ce qui diminue leur impact sur le génome et augmente leur chance de se propager dans celui-ci (Mohr, et al. 2010). De plus, certains introns de groupe II sont connus pour s'insérer dans des régions faiblement sélectionnées comme par exemple en aval de terminateurs de transcription (Robart, et al. 2007). Enfin, certains ET de classe 2 sont connus pour s'insérer préférentiellement dans des terminateurs de transcription bactériens, ce qui leur assure probablement d'être transcrits au moins passivement à la suite du gène qui se trouve en amont (Tenzen, et al. 1990). Toutefois, dans les génomes de *Wolbachia*, seules 0 à 5,4% des copies d'IS par génome sont interrompues par d'autres copies d'IS, ce qui ne peut que partiellement expliquer la forte abondance des IS.

3.3.2. Distribution des IS dans les génomes de Wolbachia

Afin de mieux caractériser la répartition des copies d'IS dans les génomes de *Wolbachia*, nous avons analysé la distance séparant chacune des copies de la suivante la plus proche, quel que soit leur groupe (Figure 3-1). Les histogrammes de fréquences des distances

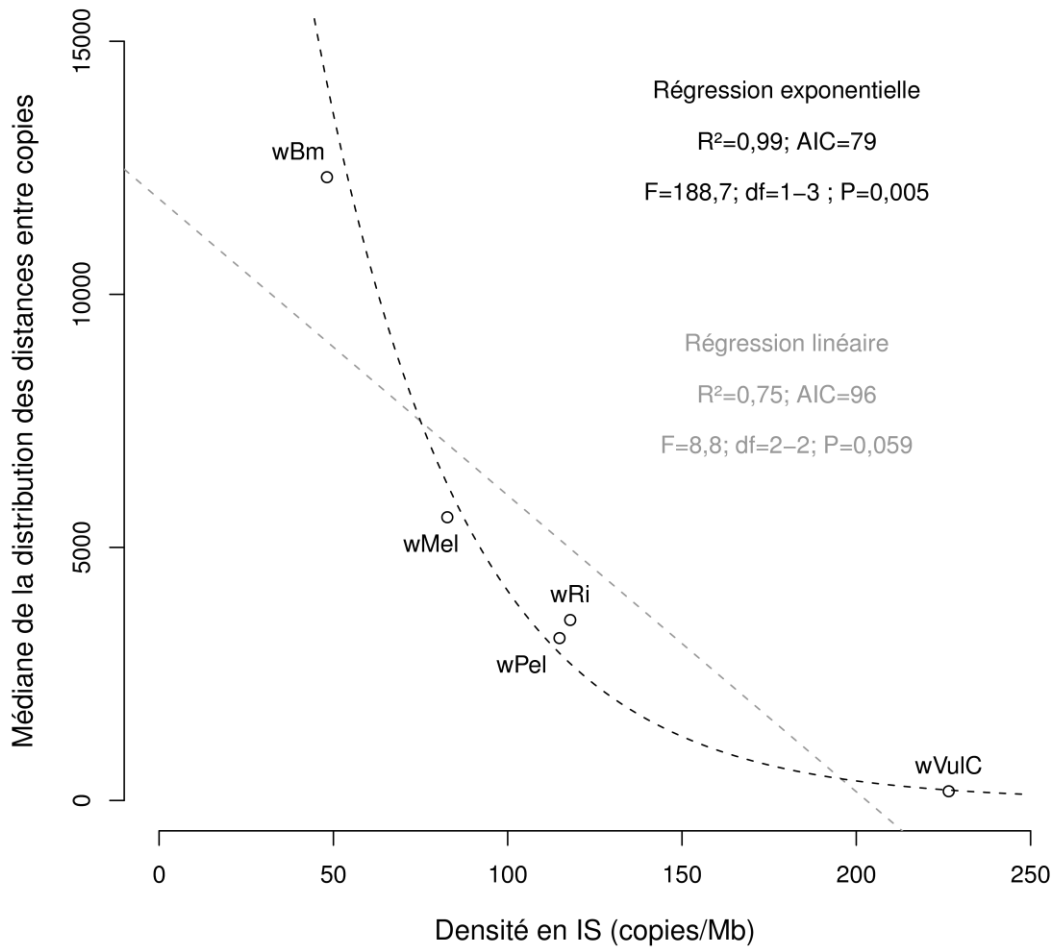

Figure 3-2 : Corrélation entre la densité en IS et la médiane de la distribution des distances entre copies.

Les régressions linéaires et exponentielles ont été générées et testées statistiquement avec le programme R.

entre copies permet de voir que, dans les 5 génomes, la classe de distances entre copies avec la fréquence la plus élevée est [0-1[Kb (Figure 3-1). Afin de savoir si le schéma de distribution des IS est homogène entre les génomes de *Wolbachia* nous avons comparé la distribution des distances entre copies à un jeu de données généré de façon aléatoire. Pour chacun des génomes, nous avons comparé la distribution aléatoire à la distribution observée. Ainsi, la distribution des IS dans le génome de *w*Bm n'est pas statistiquement différente d'une distribution aléatoire (Test de Wilcoxon, P=0,22) alors que celles des génomes de *w*Mel, *w*Pel, *w*Ri et *w*VulC sont statistiquement différentes d'une distribution aléatoire (Test de Wilcoxon, P<0,02 pour toutes les comparaisons). Il semble que l'augmentation de la densité en IS dans un génome entraîne un changement de schéma de fixation des copies d'IS. De plus, l'étude de la relation entre densité de copies et médiane de la distribution des distances entre copies nous a permis de montrer que le modèle qui explique le mieux la relation est de type exponentiel (F=188,7; df=1-3; P=0,005; AIC=79) et non pas de type linéaire (F=8,823; df=2-2; P=0,059; AIC=89) (Figure 3-2). Ainsi, l'augmentation de la densité en IS explique à 99% la diminution exponentielle de la médiane de la distribution de la distance inter-copies (Figure 3-2). Il semble donc qu'avec l'augmentation de la densité en IS, les nouvelles insertions se fixent préférentiellement à proximité d'autres ET. Ce phénomène n'est probablement pas dû à une préférence des IS vis–à-vis de leur site d'insertion car, dans ce cas, il serait aussi observable dans le génome de *w*Bm, mais plutôt à des contraintes liées à la sélection. Au cours de l'augmentation du nombre d'IS, les copies non éliminées par la sélection sont probablement celles dont l'insertion n'a que peu ou pas d'effet sur la survie de la bactérie. Ainsi, au cours du temps, tous les sites peu ou pas contre-sélectionnés sont occupés par des copies d'IS. Avec l'augmentation de la densité, la fixation de nouvelles copies d'IS ne peut donc pas se faire dans des régions génomiques contraintes par la sélection mais dans des zones peu sélectionnées comme par exemple dans des zones déjà mutées par des ET réduisant

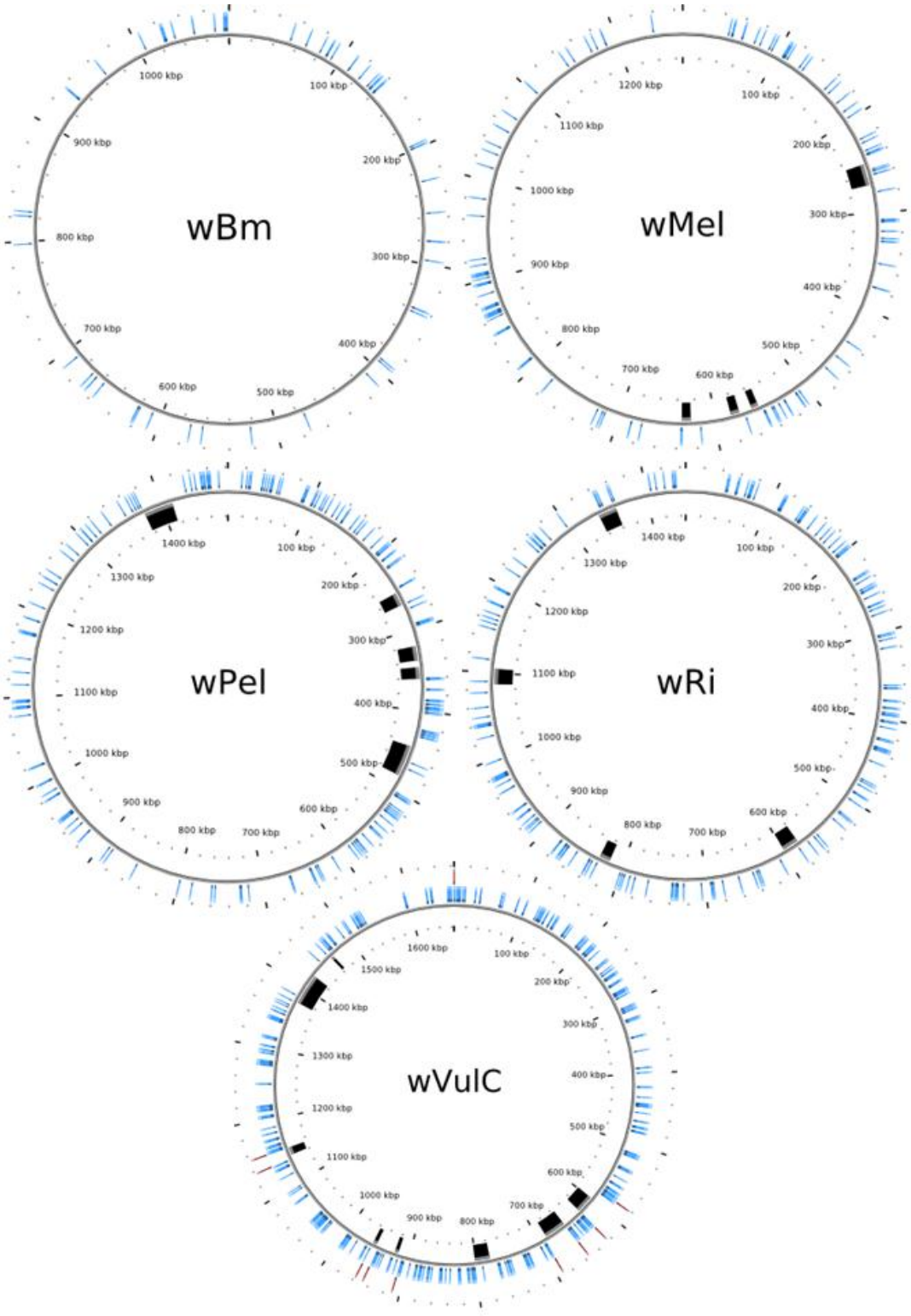

Figure 3-3 : Représentation graphique de la distribution des IS dans les génomes de *Wolbachia*

Ces représentations ont été réalisées à l'aide du logiciel cgview. Pour les cinq génomes, chaque tiret bleu indique la position d'une copie d'IS et les rectangles noirs représentent la position des prophages. Pour le génome de *w*VulC, les tirets rouges indiquent les bornes des 10 contigs artificiellement assemblés.

donc la distance entre les copies d'IS. Ainsi, l'augmentation de la densité en IS entraînerait la formation d'îlots d'ET qui devrait être visible dans les génomes de *Wolbachia*.

La représentation graphique des génomes de *Wolbachia* à l'aide du logiciel cgview a permis de visualiser la répartition des copies d'IS des 5 génomes étudiés. Nous avons ainsi identifié des zones dans lesquelles les IS sont absents (Figure 3-3). Dans les génomes de *w*Mel, *w*Pel, *w*Ri et *w*VulC, certaines de ces zones sans IS correspondent à des prophages (Figure 3-3).

Par ailleurs, la densité en IS est significativement plus faible dans les prophages que dans le reste des génomes ($Khi^2=123,42$; df=3; $P<10^{-25}$). Cette densité plus faible pourrait être due soit à une forte pression de sélection s'exerçant sur les régions prophagiques, soit au fait que les bactériophages aient été actifs et se soient insérés dans les génomes de *Wolbachia* après la phase d'activité des IS. Même si les phages sont des éléments égoïstes tout comme les ET, il est possible que leur séquence soit maintenue sous sélection purifiante afin de conserver leur capacité de déplacement. En effet, les bactériophages actifs ne contiennent que peu d'IS, probablement parce que la probabilité d'une insertion délétère est forte (Leclercq and Cordaux 2011). Par ailleurs, il a été montré qu'un bactériophage présent chez *Hamiltonella defensa* était à l'origine de la protection du puceron *Acyrthosiphon pisum* contre la guêpe parasitoïde *Aphidius ervi* (Oliver, et al. 2009). Dans cette interaction, le bactériophage produit des toxines qui tuent les larves de la guêpe parasitoïde en développement dans le corps du puceron. De plus, chez certains eucaryotes, des virus intégrés dans le génome de leur hôte peuvent leur apporter un bénéfice en termes de survie ou de reproduction (Bezier, et al. 2009). Cependant, l'hypothèse de l'insertion récente des bactériophages dans les génomes de *Wolbachia* reste la plus probable. En effet, des particules bactériophagiques libres ayant une séquence très proche de celle des prophages des génomes de *Wolbachia* ont été isolées (Tanaka, et al. 2009). Ceci suggère que les phages de *Wolbachia* ont été actifs très récemment

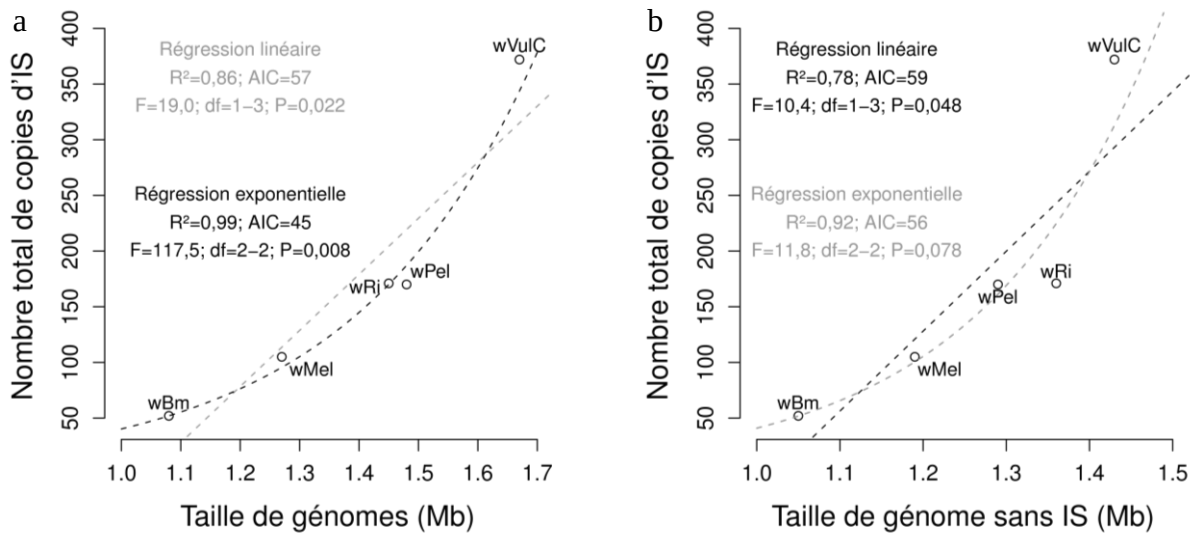

Figure 3-4: Corrélation entre taille de génome de *Wolbachia* et nombre de copies d'IS.

Ces graphiques illustrent la corrélation qu'il existe en taille de génome et nombre de copie en IS. (a) Afin de déterminer l'influence du nombre de copies d'IS sur la taille du génome nous avons étudié la relation entre nombre de copies d'IS et taille totale de génome. (b) Pour déterminer l'influence de la taille de génome sur le nombre de copies d'IS qu'il peut contenir, nous avons étudié la relation entre taille de génome sans la taille totale couverte par les IS et nombre de copies d'IS.

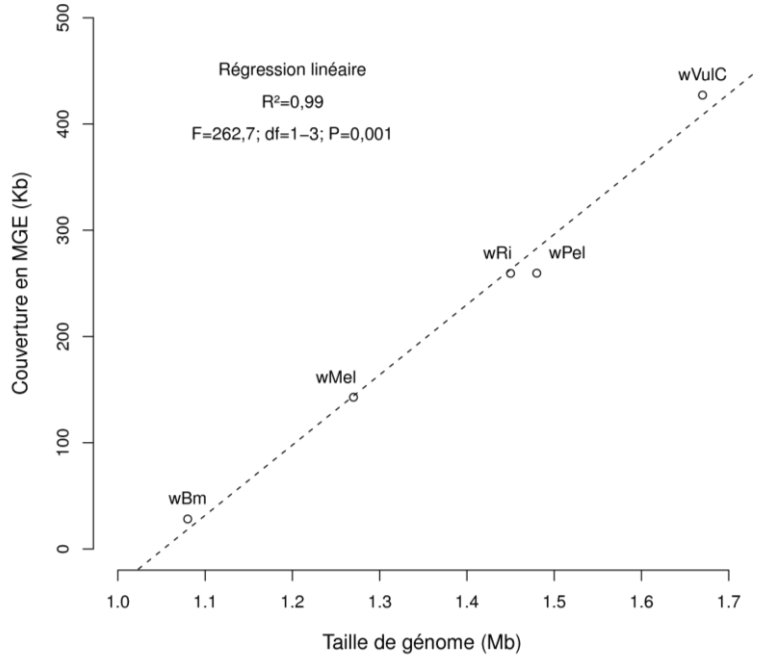

Figure 3-5 : Relation entre taille de génome de *Wolbachia* et couverture totale en MGE

Étant donné que nous voulions étudier l'impact de la couverture totale en MGE sur la taille de génome de *Wolbachia* nous n'avons pas retiré de ces derniers la couverture totale en MGE.

et/ou qu'ils le sont encore. De plus, l'analyse de la structure des IS insérés dans les prophages a montré que la plupart (12 copies sur 18) sont de grande taille et ont un gène capable de produire une transposase. Ceci suggère que l'insertion des IS dans les prophages est récente. La faible densité d'IS dans les bactériophages peut expliquer le fait que les prophages de *Wolbachia* soient des zones pauvres en IS, si les IS n'étaient que peu actifs au moment de l'insertion des bactériophages.

3.3.3. Relation entre MGE et taille de génome

Malgré la faible représentation des IS dans les génomes procaryotes (<3% en général, (Siguier, et al. 2006a)), il existe une corrélation positive entre la taille du génome et le nombre de copies d'IS (Touchon and Rocha 2007). Cette corrélation est aussi observable au sein des génomes de *Wolbachia*, que l'on retire ou non la couverture en IS de la taille du génome (Figure 3-4 a et b). La corrélation qui explique le mieux la relation entre la taille du génome avec IS et nombre de copies est de type exponentielle (F=117,5; df=2-2; *P*=0,008; AIC=45) et non pas linéaire (F=19,0; df=1-3; *P*=0,022; AIC=57). Lorsqu'on retire la taille totale couverte par des IS de la taille des génomes, seule la corrélation linéaire est statistiquement supportée. Ces corrélations entre nombre de copies et taille de génome suggèrent que plus les génomes sont grands et plus ils peuvent accueillir d'IS, mais aussi que plus les génomes accueillent d'IS plus ils sont grands. Afin de mieux cerner l'importance des IS sur la taille des génomes de *Wolbachia*, nous avons étudié la relation entre la couverture génomique de différents types de gènes et la taille des génomes de *Wolbachia*. Nous avons ainsi choisi les 3 types de MGE identifiés chez *Wolbachia*: les prophages (Kent, et al. 2011), les introns de groupe II (Leclercq, et al. 2011) et les IS.

Afin de connaître l'importance de la couverture des MGE dans la taille des génomes de *Wolbachia* nous avons étudié la relation qui existe entre les deux paramètres (Figure 3-5). Nous avons ainsi montré qu'il existe une forte corrélation linéaire entre la taille des génomes

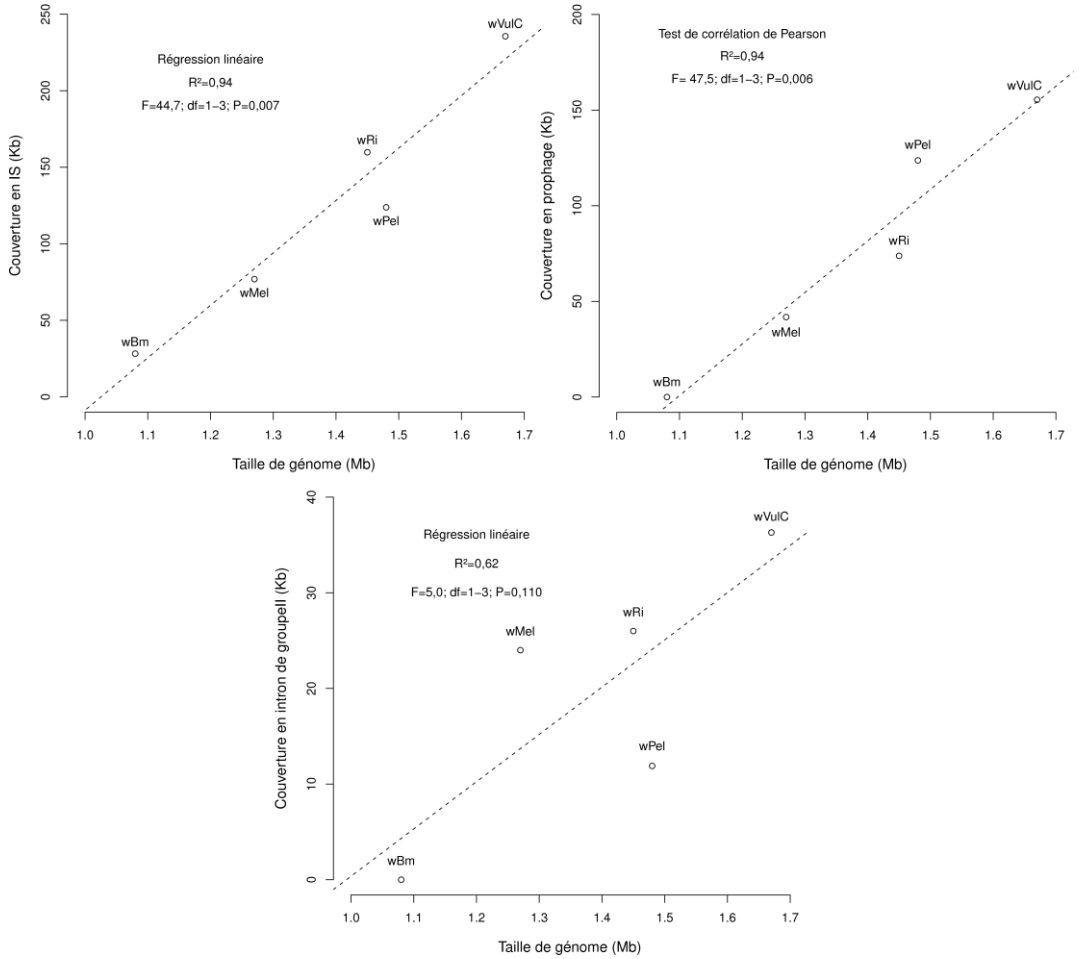

Figure 3-6: Corrélation entre taille de génome de *Wolbachia* et couverture de 3 types de MGE.

La couverture en IS (a) a été calculé à partir de l'analyse que nous avons réalisé des cinq génomes, la couverture prophagiques (b) a été calculée en se basant sur l'annotation originale des génomes et la couverture en intron de groupe II (c) a été reprise de la littérature (Leclercq, et al. 2011).

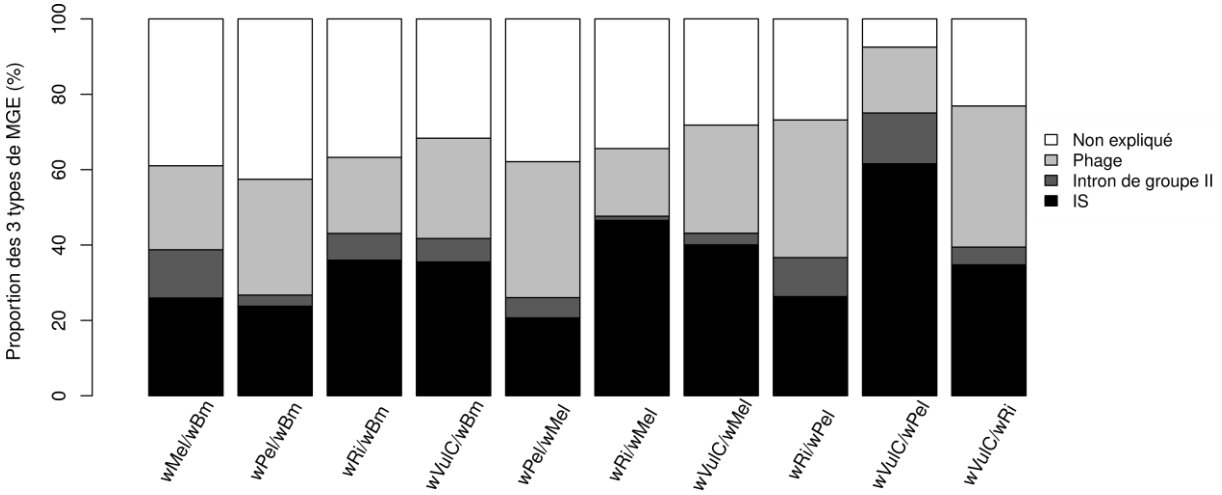

Figure 3-7: Proportion des 3 types de MGE dans les variations de taille de génome entre souches de *Wolbachia*

de *Wolbachia* et la couverture cumulée des 3 types de MGE (F=262,7 ; df=1-3 ; *P*=0,001) (Figure 3-5). Il semble donc que les MGE aient une forte influence sur les variations de taille de génome de *Wolbachia*. Pour connaître l'influence de chacun des 3 types de MGE sur la taille des génomes de *Wolbachia*, nous avons étudié la relation entre la couverture totale de chacun d'entre eux et la taille des génomes. Nous avons ainsi montré que les couvertures totales en IS et en prophages sont toutes les deux statistiquement corrélées à la taille des génomes de *Wolbachia* (respectivement F=44,6 ; df=1-3 ; *P*=0,007 et F=47,5; df=1-3 ; *P*=0,006), ce qui n'est pas le cas pour les introns de groupe II (F=5,0; df=1-3 ; *P*=0,110) (Figure 3-6). Le fait que les introns de groupe II n'influencent pas la taille des génomes de *Wolbachia* est probablement lié à leur faible couverture génomique. En effet, ils ne représentent qu'un faible pourcentage de génome (0-2%) comparativement aux prophages (0-9%) et IS (3-14%). De part leur taille, le fait que les prophages influencent fortement la taille des génomes de *Wolbachia* n'est pas surprenant. Concernant les IS, il est plus inattendu que ces éléments de petite taille, comparativement aux prophages, puissent autant influencer la taille des génomes de *Wolbachia*. Cependant, c'est le fait qu'ils soient présents en très grand nombre qui leur confère de l'importance. Il semble donc que les IS et les prophages aient une part plus importante que les introns de groupe II dans les variations de taille de génome de *Wolbachia*. Pour estimer l'impact réel des variations de couvertures en MGE sur les variations de taille de génome chez *Wolbachia*, nous avons calculé les différences de couverture totale des 3 types de MGE, que nous avons ensuite comparés aux différences de taille des 5 génomes étudiés. Notre but était de connaitre quelle proportion de la différence de taille entre génome est liée aux différences de taille totale en MGE.

Notre étude basée sur seulement 3 types de gènes permet d'expliquer en moyenne 69% de la différence de taille observée entre paires de génomes (Figure 3-7). Les différences de couverture en intron de groupe II n'expliquent en moyenne de 7% des différences de taille

Tableau 3-3: Nombre de copies de chacune des 12 familles d'IS dans les cinq génomes de *Wolbachia.*

Famille d'IS	*w*Bm	*w*Mel	*w*Pel	*w*Ri	*w*VulC
IS3		18	2	15	30
IS4	3	13	5	12	83
IS5	7	26	26	30	78
IS6			11		
IS110	21	18	6	39	61
IS200/605			5		
IS256	1	3	32	2	44
IS481	3	4	5	12	24
IS630	3	7	15	12	27
IS982	1	2	54	3	1
IS66	13	14	9	46	23
IS1182					1
Nombre total de copies d'IS	52	105	170	171	372
Nombre total de famille d'IS	8	9	11	9	10

entre les génomes de *Wolbachia*, alors que la couverture en prophage et en IS explique en moyenne respectivement 27% et 35% des variations de taille (Figure 3-7). Ces résultats confirment l'importance des prophages et des IS dans les variations de taille de génomes de *Wolbachia*.

3.3.4. Prévalence des familles d'IS chez Wolbachia

Les 870 copies d'IS que nous avons identifiées dans les 5 génomes de *Wolbachia* se répartissent en 12 familles (Tableau 3-3) parmi la vingtaine de familles d'IS connues (Chandler and Mahillon 2002; Siguier, et al. 2006a). A l'exception des familles IS6 et IS200/605 spécifiques de *w*Pel et de la famille IS1182 présente en copie unique dans *w*VulC, toutes les autres familles (9/12 soit 75%) sont présentes dans au moins 4 génomes (Tableau 3-3). On peut se demander pourquoi seules ces 12 familles sont présentes?

L'une des explications pourrait être que les génomes de *Wolbachia* contiennent les 12 familles d'IS généralement les plus abondantes dans les génomes procaryotes. Néanmoins, l'étude de 438 génomes bactériens a mis en évidence que la famille IS1 était la plus abondante dans les génomes étudiés (863 des 2091 copies d'IS identifiées au total (Wagner, et al. 2007)). Or, il n'y a aucune copie d'IS1 dans les génomes de *Wolbachia*. Nous n'avons par ailleurs pas détecté de copies des familles IS30 et ISL3 qui semblent être plus fréquentes dans les génomes bactériens que la famille IS982 (Touchon and Rocha 2007), que nous avons détectée dans tous les génomes de *Wolbachia*. Ainsi, il ne semble pas que nous ayons dans notre jeu de données un biais de détection vers les familles les plus abondantes.

Une autre explication de l'absence de certaines familles pourrait être liée à leurs caractéristiques intrinsèques. En effet, il existe des modes de transposition distincts faisant intervenir des enzymes dont la fonction et le mécanisme sont différents (Chandler and Mahillon 2002; Siguier, et al. 2006b). On peut donc imaginer que certaines familles ne

puissent pas transposer dans les génomes de *Wolbachia* à cause de contraintes mécanistiques. Cependant, nous avons identifié des familles dont la transposase est de type DDE (le type le plus courant chez les bactéries) mais nous avons aussi identifié la famille IS200/605 qui transpose grâce à une transposase à tyrosine (Siguier, et al. 2006b). De plus, nous avons identifié des familles d'IS possédant des TIR et d'autres non, ainsi que des familles possédant des TSD (jusqu'à 29 pb) et d'autres non. Tout ceci suggère donc que la prolifération des IS dans les génomes de *Wolbachia* n'est pas entravée par des problèmes structurels ou mécanistiques.

Une troisième explication de l'absence de certaines familles d'IS pourrait être due à une méthode de détection inappropriée qui n'aurait pas permis de trouver tous les éléments. Néanmoins, nous avons utilisé 3 méthodes de détection différentes (ISsaga, Repeatscout et recherche manuelle dans l'annotation des génomes) et complémentaires. En effet, ISsaga s'appuie sur ISFinder, une base de données contenant un grand nombre de séquences d'IS diverses et variées, RepeatScout est une méthode de détection *de novo* non contrainte par une base de données et l'annotation des génomes a été réalisée avec des logiciels tels que REPuter pour le génome de *w*Pel (Klasson, et al. 2008). Notre jeu de données est issu de l'association de ces 3 méthodes de détection, ce qui nous permet de penser que notre annotation est probablement quasi exhaustive.

Une dernière hypothèse qu'il est difficile de tester est que les génomes de *Wolbachia* n'ont jamais été en "contact" avec les familles d'IS que nous n'avons pas identifiées. Cette hypothèse est la plus simple mais difficile à justifier. Le confinement des bactéries intracellulaire rend très difficile les échanges de matériel génétique avec les bactéries à vie libre. Cependant, selon l'hypothèse de l'arène intracellulaire, des échanges de matériel génétique peuvent se produire entre bactéries intracellulaires infectant le même hôte (Bordenstein and Wernegreen 2004). Néanmoins, nous avons étudié 5 souches différentes qui

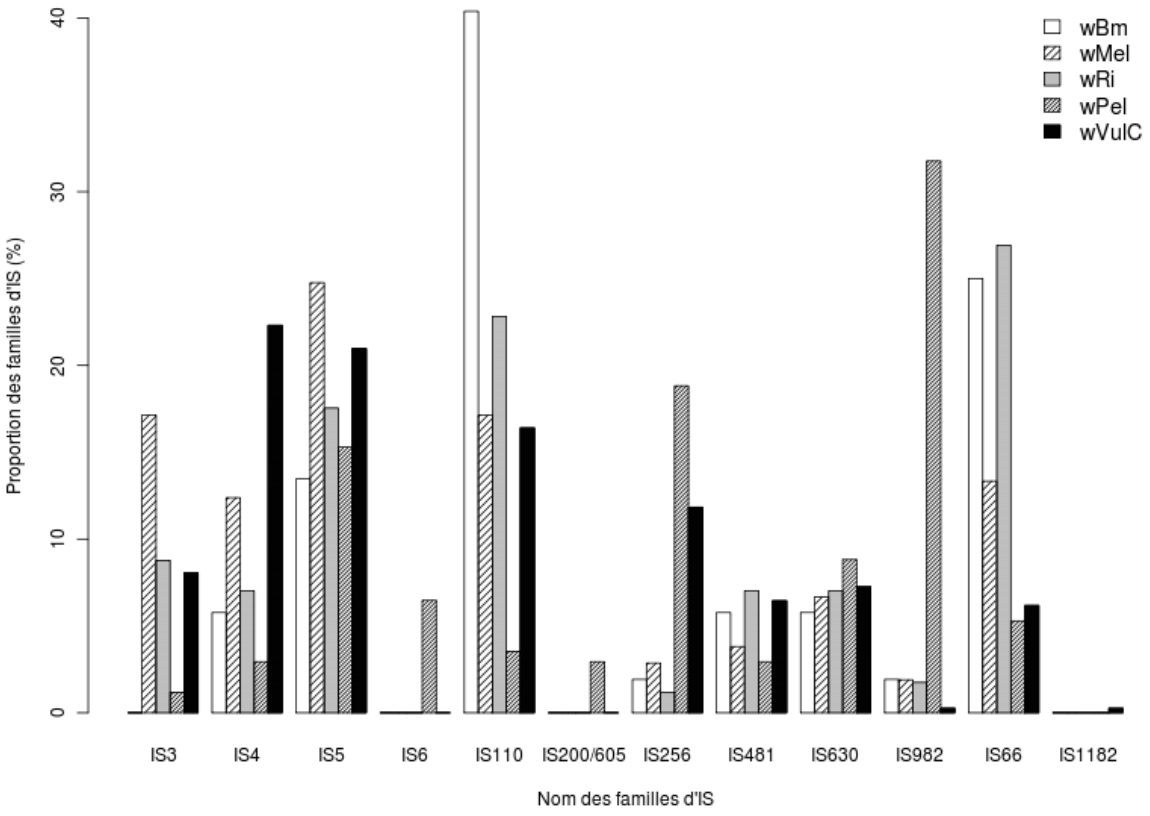

Figure 3-8: Fréquence des familles d'IS identifiées dans les cinq génomes de *Wolbachia*.

Les proportions de chaque famille ont calculé pour chaque génome séparément.

Tableau 3-4: Indices de diversités et d'équitabilité calculés à partir de la distribution des groupes d'IS des cinq génomes de *Wolbachia*.

Indices	*w*Bm	*w*Mel	*w*Pel	*w*Ri	*w*VulC
Taille de génome (Mb)	1,08	1,27	1,48	1,45	1,66
Richesse spécifique (Nombre de groupe d'IS)	19	25	21	25	35
Abondance cumulée (Nombre de copies total)	52	105	170	171	372
Indice H' de Shannon	2,34	2,79	2,31	2,56	3,13
Indice D de Simpson	0,84	0,91	0,84	0,88	0,94
Indice J d'équitabilité	0,79	0,87	0,76	0,79	0,88

sont associées à des hôtes différents vivant dans des milieux différents et donc potentiellement en contact avec des communautés bactériennes différentes (Duron, et al. 2008). Les communautés bactériennes en contact avec *Wolbachia* ne contiendraient-elles que certaines familles d'IS ou bien seules certaines familles d'IS peuvent-elles être horizontalement transmises entre bactéries ?

Aucune des explications citées ci-dessus ne peut justifier à elle seule l'homogénéité des familles d'IS présentes dans les 5 génomes de *Wolbachia* étudiés. Cependant, l'hypothèse de l'absence de contact avec les familles non identifiées semble être la plus probable.

3.3.5. Distribution des IS dans les familles

Malgré l'homogénéité de composition en familles d'IS dans les génomes de *Wolbachia*, la distribution des copies au sein des familles est très hétérogène entre les génomes (Khi^2=465,4; df=44, $P<10^{-16}$) (Figure 3-8). Comme les familles d'IS que nous avons identifiées contenaient des séquences divergentes, nous avons affiné notre analyse en classant les copies en 40 groupes d'IS différents. L'analyse des groupes d'IS a confirmé les tendances observées au niveau des familles tant au niveau de l'homogénéité de composition en groupes d'IS des génomes de *Wolbachia* (30 groupes sur 40 soit 75% sont présents dans plus d'un génome) qu'au niveau de l'hétérogénéité de distribution des IS au sein des groupes entre les génomes (Khi^2=917,23; df=156; $P<10^{-16}$).

Afin de décrire la composition en IS des génomes de *Wolbachia*, nous avons utilisé des indices écologiques permettant de caractériser les populations. Nous avons tout d'abord calculé des paramètres descriptifs comme la richesse spécifique ou encore l'abondance cumulée pour chacun des génomes (Tableau 3-4). Puis pour caractériser les populations d'IS dans les différents génomes, nous avons calculé deux indices de diversité : H' de Shannon et D de Simpson (Tableau 3-4). Nous avons calculé deux indices différents car H' peut être

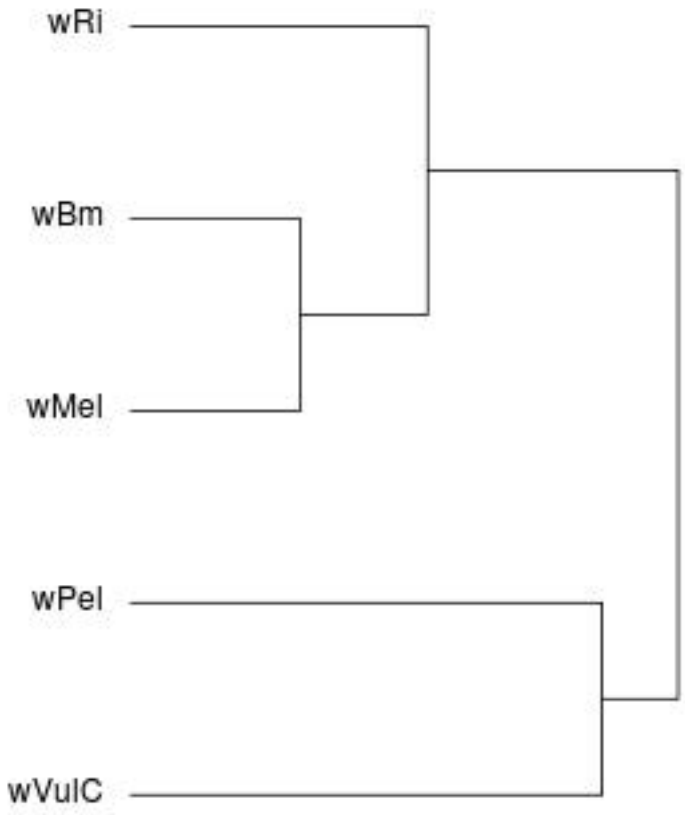

Figure 3-9 : Dendrogramme basé sur l'abondance et la distribution des groupes d'IS des cinq génomes de *Wolbachia*.

influencé par les groupes d'IS rares (c'est-à-dire en petit nombre) et D est sensible aux groupes très abondants. Étant donné que les génomes de *Wolbachia* contiennent des groupes d'IS rares et d'autres abondants, nous avons préféré calculer les deux indices pour être le plus descriptif possible. Il n'existe pas de valeur de référence permet de dire par rapport à un indice de diversité si le milieu est de bonne ou de mauvaise qualité par contre nous pouvons comparer les génomes entre eux. Ainsi, nous pouvons voir que les deux indices de diversité présentent des valeurs homogènes entre les 5 génomes de *Wolbachia* (Tableau 3-4). L'homogénéité des indices H' et D dans les 5 génomes de *Wolbachia* signifie que la diversité des IS est similaire dans chacun des génomes (Tableau 3-4). De plus, cette homogénéité des indices de diversité entre les génomes signifierait que le génome hôte n'a que peu d'influence sur la diversité en IS. La comparaison des indices d'équitabilité semble montrer qu'il existe deux types de distribution des copies d'IS différentes dans les génomes de *Wolbachia*. En effet, on peut observer que les indices d'équitabilité sont très proches entre les génomes de *w*Mel et *w*VulC de même qu'entre les génomes de *w*Bm, *w*Ri et *w*Pel (Tableau 3-4). Ceci signifie que le schéma global de distribution est le même au sein de chacun des groupes mais diffère entre les deux groupes. Cette comparaison tient compte uniquement des distributions sans nous renseigner sur la similitude de distribution des groupes d'IS présents dans deux génomes différents.

Nous avons ensuite comparé les communautés d'IS des différents génomes. Ce type d'analyse se base sur la composition en IS des différents milieux et permet de créer une classification hiérarchique des différents génomes. Ainsi, les génomes qui ont la même composition en espèces sont présents au sein d'un même groupe. Le dendrogramme que nous avons obtenu permet de visualiser deux groupes de souches de *Wolbachia*, l'un composé des souches *w*Pel et *w*VulC et l'autre composé des souches *w*Mel, *w*Ri et *w*Bm (Figure 3-9). Les regroupements réalisés avec l'analyse du dendrogramme ne peuvent pas être comparés avec

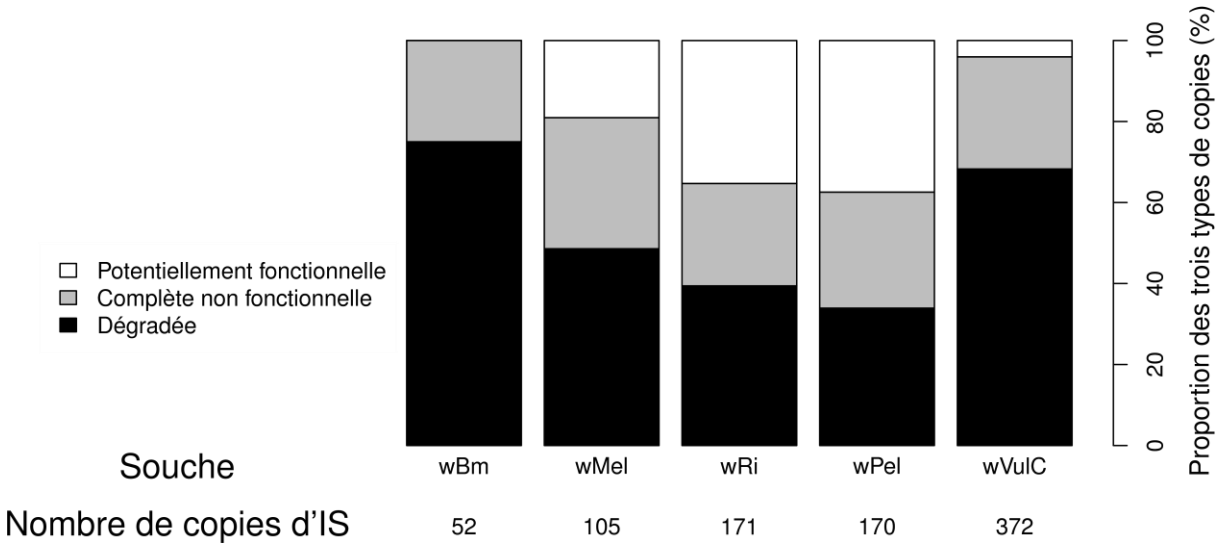

Figure 3-10: Proportion de copies d'IS en fonction de leur statut de fonctionnalité.

Les copies ont été classées selon 3 statuts: les copies complètes potentiellement fonctionnelles (entières avec une ORF de grande taille), les copies complètes non fonctionnelles (entières mais sans ORF ou avec une ORF de taille inférieure à celle attendue) et les copies dégradées (séquences délétées).

ceux fait grâce à l'indice d'équitabilité. En effet, ce ne sont pas les mêmes paramètres qui sont pris en compte. Ce dendrogramme, même si il n'est basé que sur 5 souches, comporte des points communs avec la classification des différents supergroupes de *Wolbachia* (Lo, et al. 2007) (Figure 2-5 – page 27). Par exemple, les deux souches du supergroupe B (*w*Pel et *w*VulC) et les deux du supergroupe A (*w*Mel et *w*Ri) sont regroupées. Cependant il existe aussi des différences avec la présence de *w*Bm, la souche de supergroupe D) avec *w*Mel et *w*Ri. La différence observée entre le dendrogramme de communauté d'IS de *Wolbachia* et la phylogénie des souches est facilement explicable par la grande propension des ET à transférer horizontalement. Etant donné que la répartition des groupes d'IS est très homogène entre les souches de *Wolbachia*, le hasard des acquisitions peut entrainer des rapprochements en composition de groupes dans les différents génomes. Néanmoins, on observe aussi des congruences entre la phylogénie de *Wolbachia* et le dendrogramme basé sur les communautés d'IS. On a une très franche séparation entre les *Wolbachia* du supergroupe B et les autres.

3.3.6. Structure des copies

Pour comprendre l'origine de l'hétérogénéité de distributions des copies d'IS dans les génomes de *Wolbachia*, nous avons évalué le statut fonctionnel des copies. Nous avons considéré qu'une copie est potentiellement fonctionnelle quand elle est complète (pleine taille avec les structures typiques attendues pour le groupe ou la famille considérée) et que sa séquence code une transposase couvrant au moins 70% de la taille de l'IS. Cette analyse a révélé que 4 des 5 génomes possèdent des copies potentiellement fonctionnelles (Figure 3-10). Le génome qui ne contient pas de copie potentiellement fonctionnelle est *w*Bm (Cordaux 2009), qui est par ailleurs le génome avec le moins de copies d'IS au total. Cette absence de copie potentiellement fonctionnelle pourrait être liée à la relation mutualiste stricte qui existe entre la souche de *Wolbachia* et son hôte nématode *B.malayi*. Cette relation entraîne une

transmission uniquement verticale et donc une limitation des possibilités de transferts horizontaux entre hôtes d'espèces différentes.

Les génomes de *w*Mel, *w*Pel, *w*Ri et *w*VulC, quant à eux, contiennent entre 15 et 64 copies d'IS potentiellement fonctionnelles. Ces copies ne représentent qu'une faible partie de l'ensemble des copies d'IS des génomes de *Wolbachia* (de 4 à 37% des copies identifiées dans chacun des génomes (Figure 3-10). Etant donné que les IS sont généralement considérés comme très récents dans les génomes bactériens et qu'ils sont sujets à un renouvellement rapide (Rocha 2008; Wagner 2006; Wagner, et al. 2007), la présence d'un si grand nombre de copies non fonctionnelles et plus particulièrement de copies incomplètes et dégradées dans les génomes de *Wolbachia* est très surprenante.

3.4. Conclusion

Notre analyse a permis de confirmer ce qui avait été entrevu avec l'analyse du génome de *w*Mel (Wu, et al. 2004) : les génomes de *Wolbachia* font partie des génomes bactériens dans lesquels l'abondance des IS est la plus importante. Cette forte abondance pourrait être expliquée par des transferts horizontaux de *Wolbachia* entre hôtes, altérant la stricte transmission verticale en l'exposant à de nouveaux environnements et des contacts avec d'autres bactéries, ces deux phénomènes facilitant probablement l'acquisition de matériel génétique. Néanmoins, il est aussi possible que les IS aient développé une stratégie d'insertion particulière dans les génomes de *Wolbachia* leur permettant de les envahir.

L'analyse des distances génomiques entre copies d'IS a montré que celles-ci diminuaient avec l'augmentation de la densité IS, ce qui n'est pas surprenant, mais cette diminution n'est pas linéaire. Il semble qu'avec l'augmentation de la densité en IS les nouvelles copies se fixent préférentiellement au voisinage de copies déjà présentes. Ce phénomène peut être guidé par la sélection qui n'autorise la fixation de nouvelles insertions

d'IS que dans des zones où elle est faible, c'est-à-dire des zones déjà mutées par l'insertion d'ET. On observe donc dans les génomes de *Wolbachia*, et dans celui de *w*VulC en particulier, des zones à forte densité en IS et d'autres à faible densité. Ces dernières devraient correspondre à des gènes essentiels à la survie de la bactérie. Néanmoins, plusieurs d'entre elles correspondent à des prophages présents dans les génomes de *Wolbachia*. Cette faible densité en IS dans les prophages pourrait s'expliquer soit par le fait que les bactériophages se soient insérés dans les génomes de *Wolbachia* après que les IS aient été actifs, soit par le fait qu'il y a une forte pression de sélection sur les prophages pour qu'ils restent actifs.

Contrairement à ce qui est observé chez les eucaryotes (Kidwell 2002), il existe dans les génomes bactériens une corrélation entre la taille de génome et le nombre d'ET (Touchon and Rocha 2007). Nous avons pu montrer que la taille du génome avait une influence sur la quantité d'IS qu'il pouvait accueillir et que la quantité d'IS accueilli avait un impact sur la taille du génome (Figure 3-4a et b – page 46). Nous avons aussi pu faire un lien entre variation de couverture en IS et variation de taille de génome. Les IS sont le type de MGE dont la variation de couverture explique le plus de variation de taille entre les différents génomes de *Wolbachia*.

Toutes les familles et groupes d'IS n'ont pas la même abondance dans les communautés que nous avons identifiées dans les génomes de *Wolbachia*. Malgré l'homogénéité de composition en familles et en groupes d'IS, la fréquence de chacun et chacune d'entre elles est variable parmi les 5 génomes étudiés. Cette variation pourrait être liée au statut fonctionnel des copies qui composent les différents groupes ou familles. L'analyse du statut fonctionnel des copies identifiées a révélé que les différents génomes de Wolbachia possédaient des copies potentiellement fonctionnelles. Cependant, le fait le plus marquant est que plus de 75% des copies identifiées dans les 5 génomes cumulés sont non fonctionnelles. Cette faible proportion de copies potentiellement fonctionnelles est surprenant,

car les IS sont généralement considérés comme récents dans les génomes de bactéries (Rocha 2008; Wagner 2006; Wagner, et al. 2007). Cependant, l'annotation des IS dans les génomes bactériens est rarement optimale et se borne généralement à la détection des copies de grandes taille et, le plus souvent, fonctionnelles (Varani, et al. 2011). Il est donc possible que d'autres génomes bactériens contiennent une grande proportion de copies d'IS dégradées non détectées par les méthodes d'analyse classiques.

Quoi qu'il en soit, l'abondance de copies d'IS non fonctionnelles et dégradées chez *Wolbachia* constitue une des premières opportunités d'analyser la dynamique à long terme des IS dans les génomes bactériens et de mieux comprendre leur évolution.

4. Dynamique évolutive des IS

Ce chapitre s'appuie en partie sur l'article de recherche ayant pour titre "Short and long-term evolutionary dynamics of bacterial insertion sequences: insights from *Wolbachia* endosymbionts" accepté pour publication en septembre 2011 dans la revue Genome Biology and Evolution et présenté en annexe 2 (Cerveau, et al. 2011b).

4.1. Contexte

Les ET sont des éléments d'ADN mobile dont l'importance dans l'évolution génomique des êtres vivants est capitale (Cordaux and Batzer 2009; Feschotte and Pritham 2007b). Chez les eucaryotes, certains génomes fourmillent de copies d'ET dégradées et mutées, qui sont un outil d'étude de leur dynamique évolutive passée (Kapitonov and Jurka 2003; Lander, et al. 2001). Ainsi, l'étude de la dynamique évolutive des ET humains a permis de mettre en évidence que les transposons à ADN ont cessé d'être actifs il y a ~40 millions d'années, après avoir subi une forte prolifération lors de la radiation des mammifères et au début de l'évolution de primates (Lander, et al. 2001; Pace and Feschotte 2007).

Contrairement aux génomes eucaryotes, les génomes procaryotes changent rapidement avec l'acquisition de matériel génétique via des transferts horizontaux entre génomes et des dégradations de gènes (Rocha 2008). Les ET sont un des exemples les plus frappants de cette importante variabilité génomique avec des temps de rétention très faibles (Touchon and Rocha 2007; Wagner 2006; Wagner, et al. 2007). Ainsi, de par leur forte variation d'abondance et ceci même entre souches phylogénétiquement proches et la faible divergence intragénomique entre copies d'un même groupe ou d'une même famille, les IS sont souvent considérés comme récents dans les génomes bactériens (Lawrence, et al. 1992; Parkhill, et al. 2003; Sawyer, et al. 1987). Le modèle de dynamique évolutive des IS dans les génomes

bactériens est donc basé principalement sur des observations faites sur des copies récentes. Le modèle prédit un renouvellement rapide des copies. Le cycle commence par l'arrivée via des transferts horizontaux de copies d'IS potentiellement fonctionnelles. Ces copies nouvellement arrivées sont sujettes à une forte activité de transposition et leur nombre augmente rapidement, puis les copies sont éliminées.

On peut supposer que les conditions de vie intracellulaire de *Wolbachia* avec toutes leurs implications (modification de l'efficacité de la sélection et de la dérive génétique) ont pu modifier la vitesse de déroulement des différentes phases du cycle de prolifération des IS. Ainsi, l'exceptionnelle abondance de copies non fonctionnelles et dégradées dans les génomes de *Wolbachia* constitue l'une des premières opportunités de tester empiriquement les modèles portant sur l'évolution des ET dans les génomes bactériens.

4.2. Méthodes

4.2.1. Divergence nucléotidique et dégradation des copies d'IS

Pour étudier la dégradation des copies d'IS, chaque groupe avec au moins deux copies dont une complète dans un génome a été considéré. Les copies d'IS de *w*Bm ont été retirées de cette analyse car trop peu de groupes remplissent les conditions précédemment énoncées. Pour chaque groupe sélectionné, nous avons généré une séquence consensus majoritaire à l'aide du logiciel Bioedit (Hall 1999). Pour chaque copie d'IS, nous avons calculé la divergence nucléotidique et la taille relative par rapport à la séquence consensus majoritaire de son groupe d'appartenance. La divergence a été calculée avec le logiciel Mega ver 4.0 sur la base des substitutions observées (Kumar, et al. 2008). La taille relative des copies est

exprimée comme un pourcentage de la taille de la séquence consensus majoritaire du groupe. Afin de caractériser la corrélation entre la divergence nucléotidique et la taille relative des copies nous avons utilisé la corrélation de Spearman en utilisant le programme R (http://www.r-project.org/) avec les paramètres par défaut. Afin d'étudier la relation entre la divergence nucléotidique et la taille des copies, nous avons généré et testé des régressions linéaires et non linéaires avec le logiciel R. Nous avons calculé pour chacune des régressions le critère d'Akaike (AIC) qui permet de discriminer les modèles, le modèle avec la plus faible valeur étant celui qui correspond le mieux aux données.

4.2.2. Etude de la dynamique évolutive des IS

Pour analyser la dynamique évolutive des IS, nous avons utilisé deux méthodes. D'une part, nous avons mesuré la dégradation des copies d'IS selon la taille relative des copies calculées précédemment. Nous avons regroupé les données en 5 classes de taille:]0-20%],]20-40%],]40-60%],]60-80%],]80-100%] relativement à la taille de la séquence consensus majoritaire de chacun des groupes d'IS analysés. Nous avons utilisé 5 classes car il s'agit du nombre le plus grand pour lequel toutes les classes avaient un effectif d'au moins une copie.

D'autre part, nous avons utilisé la méthode qui a permis d'étudier la dynamique évolutive des ET présents dans le génome humain (Lander, et al. 2001). Pour chaque groupe d'IS avec au moins deux copies dans un génome considéré nous avons calculé la divergence entre chaque paire de copies, sur la base des substitutions nucléotidiques observées, avec le logiciel Mega ver 4.0 (Kumar, et al. 2008).

Les distributions obtenues avec les deux méthodes ont été comparées séparément par des tests de Khi² calculés avec le programme R. Ce programme a aussi été utilisé pour générer les représentations graphiques des données.

4.2.3. Simulations de dynamique évolutive des IS

Les scénarios de dynamique évolutive des IS ont été élaborés en collaboration avec Dr. Sébastien Leclercq (post doctorant au laboratoire) et sont basés sur la simulation d'un jeu de données de séquences d'IS évoluant dans un génome haploïde. Les scénarios ont été créés pour permettre des comparaisons qualitatives mais ne représentent pas toute la complexité de l'évolution des séquences d'IS. 4 processus majeurs de l'évolution des IS ont été considérés: l'acquisition par transferts horizontaux, l'expansion du nombre de copies résidentes, la dégradation selon une accumulation aléatoire de substitutions et la perte par délétion des copies. Chaque séquence initiale d'IS contient une ORF potentiellement fonctionnelle représentée par une série de 300 nombres variant entre 0 et 63. Chaque nombre représente un codon hypothétique et 3 d'entre eux sont des codons stop, pour suivre le code génétique bactérien.

Chaque simulation comprend 10 000 générations et le taux de mutations a été fixé arbitrairement à $15x10^{-6}$. Ces deux critères ont été choisis pour que la divergence nucléotidique maximale soit de 30% à la fin des simulations. Si l'on utilisait un taux de mutation moins élevé, il faudrait augmenter le nombre de générations pour pouvoir obtenir la même divergence nucléotidique. Par exemple, si le taux de mutation était 10^4 fois plus faible, il faudrait augmenter le temps de 10^4 de génération pour obtenir des résultats différents entre chaque scénario, ce qui augmenterait considérablement le temps de simulation. De plus, nous ne voulons réaliser qu'une analyse qualitative et non quantitative de l'évolution des IS, c'est pourquoi le taux de mutation a pu être choisi arbitrairement. A chaque génération, chacun des codons d'une séquence peut subir une substitution qui change la valeur du codon. Si la nouvelle valeur correspond à un codon stop la copie passe irrémédiablement du statut de potentiellement fonctionnelle au statut de non fonctionnelle. Nous avons testé 4 taux de

délétions différents : égal au taux de mutations, 10 et 100 fois moins fort et égal à 0 (pas de délétion).

Deux types d'expansion de copies ont été simulés. D'une part, une expansion instantanée ou "burst" dans laquelle une copie fonctionnelle transférée horizontalement est dupliquée de nombreuses fois dans un court laps de temps. D'autre part, une expansion lente et régulière de copies potentiellement fonctionnelles déjà présentes dans le génome qui sont dupliquées à faible vitesse. Les transferts horizontaux sont simulés par addition d'une copie aléatoire potentiellement fonctionnelle provenant d'un génome source hypothétique. Les copies du génome source évoluent en parallèle de celles des autres génomes et subissent des expansions régulières pour que le réservoir soit illimité.

Cinq scénarios différents ont été simulés, pour obtenir à la fin un nombre fixe de copies *n* :

Scénario 1- Expansion unique, instantanée et ancienne : un transfert horizontal à la génération 1, immédiatement suivi par l'expansion du nombre de copies pour atteindre *n* à la génération 2.

Scénario 2- Expansion unique, instantanée et récente : un transfert horizontal à la génération 9 900 immédiatement suivi par l'expansion du nombre de copies pour atteindre *n* à la génération 9 901.

Scénario 3- Expansion lente : un transfert horizontal à la génération 1, suivi par une duplication de copie toutes les 10 000/*n* générations.

Scénario 4- Deux expansions instantanées et récentes : deux transferts horizontaux indépendants à la génération 9 900, chacun immédiatement suivi par l'expansion du nombre de copies pour atteindre *n*/2 à la génération 9 901.

Scénario 5- Une expansion instantanée et ancienne et un autre instantanée et récente : deux transferts horizontaux indépendants aux générations 1 et 9 900, chacun immédiatement suivi

Tableau 4-1 : Liste des souches de *Wolbachia* utilisées pour la détection des groupes d'IS.

Identifiant échantillons	Espèce hôte	Groupe taxonomique	Origine géographique	Supergroupe de diversité de *Wolbachia*
Abb	*Aleochara bilineata*	Insecte, Coleoptère	Canada	A
Dana	*Drosophila ananassae*	Insecte, Diptère	Rio de Janeiro, Brazil	A
Dmel	*Drosophila melanogaster* (*w*Mel)[a]	Insecte, Diptère	Antibes, France	A
wAu	*Drosophila simulans* (*w*Au)	Insecte, Diptère	Yaounde, Cameroon	A
Dsim	*D. simulans* (*w*Ri)[b]	Insecte, Diptère	Antibes, France	A
Dtri	*Drosophila triauraria*	Insecte, Diptère	Tokyo, Japan	A
wYac	*Drosophila yakuba*	Insecte, Diptère	Ogoue River, Gabon	A
Zsep	*Zaprionus sepsoides*	Insecte, Diptère	Sao Tomé	A
Atab	*Asobara tabida* (*w*Atab3)	Insecte, Hymenoptère	Antibes, France	A
Ajap	*Asobara japonica*	Insecte, Hymenoptère	Sapporo, Japan	A
Lhet	*Leptopilina heterotoma* (*w*Lhet1)	Insecte, Hymenoptère	Antibes, France	A
Zu	*Armadillidium vulgare* (*w*VulC)	Crustacé, Isopode	Saint Cyr, France	B
Bi	*A. vulgare* (*w*VulM)	Crustacé, Isopode	Méry sur Cher, France	B
Cc	*Cylisticus convexus*	Crustacé, Isopode	Avanton, France	B
Hb	*Helleria brevicornis*	Crustacé, Isopode	Bastia, France	B
Oa	*Oniscus asellus*	Crustacé, Isopode	Golbey, France	B
Slab	*Culex pipiens* (*w*Pel)[c]	Crustacé, Isopode	Montpellier, France	B
Pdp	*Porcellio dilatatus petiti*	Crustacé, Isopode	Saint Honorat, France	B
Pp	*Porcellionides pruinosus* (*w*PruIII)	Crustacé, Isopode	Nevers, France	B
Bsn	*Drosophila sechellia* (*w*Sn)	Insecte, Diptère	Archipel des Seychelles	B
Dery	*Dysdera erythrina*	Arachnide, Araneae	Saint Benoit, France	G
Mdo	*Musca domestica*	Insecte, Diptère	Poitiers, France	G

[a,b,c] utilisés comme contrôle positif pour les groupes d'IS identifiés respectivement dans les génomes de *w*Mel, *w*Ri et *w*Pel.

par l'expansion du nombre de copies pour atteindre *n*/2 aux générations 2 et 9 901.

A la fin de chaque simulation, le nombre de codon différent entre chaque paire de séquence a été calculé et exprimé en pourcentage de divergence. Chaque distribution correspond à la somme des distributions des 36 simulations, chacune d'entre elles ayant un nombre final de copies équivalent à ce qui a été observé dans l'un des 36 groupes d'IS des génomes de *Wolbachia* comprenant au moins deux copies alignables.

4.2.4. Recherche de copies orthologues

Pour identifier les copies d'IS insérées à des sites orthologues entre les génomes, nous avons réalisé des recherches par BLASTN (taille minimale 40pb, similarité minimale 75%, *e-value*<0,05, valeur de *reward* de 2 et valeur de *penalty* de 3). Nous avons utilisé comme requête 300 pb de séquences flanquantes en aval et en amont de chaque copie d'IS. Pour chaque locus contenant un IS, les régions flanquantes orthologues des différents génomes ont été alignées pour identifier des loci orthologues d'IS. Cependant, pour de nombreux loci contenant des IS, les recherches par BLASTN n'ont soit pas donné de résultats, soit des résultats partiels (c'est-à-dire que seul une petite fraction de la séquence flanquante était retrouvée dans le génome cible) ou multiples (c'est-à-dire que la séquence flanquante était retrouvée plus d'une fois dans le génome cible). Par conséquent, notre jeu de données final pour cette analyse ne contient que les loci contenant des IS pour lesquels il n'y a pas d'ambiguïté de résultats pour les deux séquences flanquantes.

4.2.5. Détection expérimentale des groupes d'IS dans divers génomes de Wolbachia

Nous avons diagnostiqué la présence de 17 groupes d'IS dans un échantillon de 22 souches de *Wolbachia* (Tableau 4-1) appartenant aux supergroupes de diversité A, B et G,

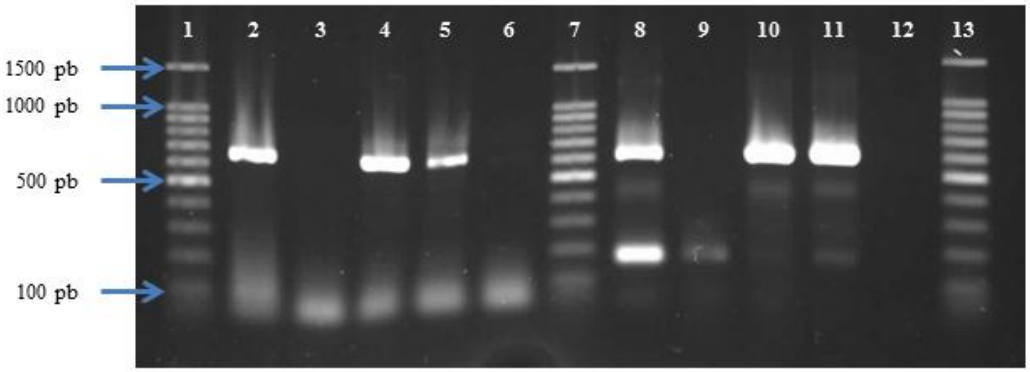

Figure 4-1: Détection par PCR de l'infection par *Wolbachia* de trois individus de l'espèce d'isopode terrestre *Cylisticus convexus*

Les marqueurs chromosomiques utilisés pour détecter la présence de *Wolbachia* sont les gènes *Wsp* (pistes 2 à 6) et *GroE* (pistes 8 à 12). Les pistes 1, 7 et 13 correspondent au marqueur de taille BenchTop (Promega). Les pistes 2 et 8 sont des témoins positifs et les pistes 6 et 12 sont des témoins négatifs. Les pistes 3 et 9; 4 et 10 ainsi que 5 et 11 correspondent respectivement aux individus 1, 2 et 3 de l'espèce d'isopode terrestre *Cylisticus convexus*. Les signaux positifs ont été séquencés en séquençage directe pour vérification. Les individus pour lesquels nous n'avons pas eu de signal n'ont pas été utilisés dans la suite des expériences.

disponibles lors d'une précédente étude (Cordaux, et al. 2008). Le statut d'infection par une souche unique de *Wolbachia* a été vérifié dans chacun des 22 échantillons par PCR (Figure 4-1) et séquençage direct de 2 ou 3 marqueurs chromosomiques de *Wolbachia* (*wsp, 16S rRNA* et *GroE*) traditionnellement utilisés pour la détection de *Wolbachia* (Cordaux, et al. 2008). Les 17 groupes d'IS ont été sélectionnés parmi ceux détectés dans les génomes de *w*Mel, *w*Ri, *w*Pel. Nous n'avons pas sélectionné de groupe d'IS du génome de *w*Bm car nous n'avions pas d'échantillon d'ADN pour tester le bon fonctionnement des amorces et nous n'avons pas sélectionné de groupe de *w*VulC car nous avons eu accès à la séquence génomique trop tardivement pour pouvoir faire cette analyse. Nous avons sélectionné les groupes ayant au moins une copie potentiellement fonctionnelle et/ou plusieurs copies de pleine taille.

Pour chaque groupe d'IS, une paire d'amorces spécifique a été proposée par le logiciel Primer3 (Rozen and Skaletsky 2000). Elles permettaient d'amplifier un fragment dont la taille est comprise entre 499 et 706 pb. Les amplifications ont été réalisées avec la GoTaq polymérase (Promega). Chaque réaction contenait 5µL de tampon 5X, 0,5µl d'un mix dNTP (8,6mM), 0,7µL de chaque amorce (10µM), 0,25µL de Taq polymérase (5 unités par µl), 1µL d'ADN matriciel et la quantité suffisante d'eau pour obtenir un volume réactionnel de 25µL. Le cycle d'amplification est initié par une dénaturation à 94°C pendant 3 minutes puis se poursuit par 35 répétitions des 3 étapes suivantes : dénaturation à 94°C pendant 30 secondes, hybridation des amorces (Tableau 4-2 page suivante) pendant 30 secondes puis élongation à 72°C pendant 1 minute par Kb de séquence à amplifier. Enfin, le cycle est terminé par une élongation à 72°C pendant 10 minutes. La visualisation des fragments amplifiés se fait par migration des produits de PCR sur gel d'agarose 1,5% puis révélation aux ultraviolets après un bain de bromure d'éthidium. Les échantillons d'ADN D.mel, D.sim et Slab correspondent respectivement aux souches *w*Mel, *w*Ri et *w*Pel et ont été utilisés comme témoins positifs. Les séquences et conditions de PCR pour chaque paire d'amorces se trouvent dans le tableau 4-2.

Tableau 4-2 : Caractéristiques des amorces utilisées pour la recherche des 17 groupes d'IS dans les souches de *Wolbachia*.

Nom	Séquence de l'amorce	Product size	Tm (°C)
IS3_wM-F1	TGTCTGGAGCATACGCATTT	646	60
IS3_wM-R1	TGTGACCCTTGATCGCTATG		
ISWosp2_wM-F	GTATGYGAYAGAGAAGCAGATA	647	60
ISWosp2_wM-R	CTTTCCATTCTTCYTCAGCTAAT		
ISWen1_wM-F	AAAGCCATGGTCTTGGGTAA	699	60
ISWen1_wM-R	TCTGGCCATAGACTGTTCTTCA		
ISWpi1-F	GATCTAAGCGAAAGGGAATGG	681	60
ISWpi1-R	CAACCCATCTTCTTGGCTGT		
ISWen2_wM-F	GCATTAGGTGGAGAAGGGAAG	697	60
ISWen2_wM-R	AAGGTCTTCCCAAGCATTTG		
ISWpi15_wM-F	CCAATTGTTCTGCCAGGATT	607	60
ISWpi15_wM-R	TGGCATGGGATACAGAGACA		
ISWpi4_wM-F	CGCCAAAACAACAGAGACAA	598	60
ISWpi4_wM-R	ATGGCGTTTTTGGTTGGTAA		
IS630wA_wM-F	AGCTACTGGCCAAGGTCTCA	633	60
IS630wA_wM-R	GGTGTGCCAACAATGCTCTA		
IS6_wP-F	TGAGCTATCGAGATTTGGAAG	569	60
IS6_wP-R	CATAGCAAAATTCTCAAAAGTAG		
ISWpi12_wP-F	TCAACCTTGGTTGCACTTTG	675	60
ISWpi12_wP-R	CAAACAGCTGCTGCAACAAT		
ISW1_wP-F	AAGTCCATCCTTTTAAGAATATAA	499	60
ISW1_wP-R	AATCCAAAGATTGAGAGAGAGCA		
ISWpi11_wP-F	TACATTATGGGAGTGGCAGC	532	60
ISWpi11_wP-R	AAATCGGTGTCAGGAACTGC		
ISWpi10_wP-F	ATAAGCGCTGTGGCAAGAAT	637	65
ISWpi10_wP-R	GCCTGCACAATCCATTACAA		
ISWpi16_wP-F	CTGTACTGTTGCGTCGAGGA	554	60
ISWpi16_wP-R	CCAAACAAAAGTCCGGTCAG		
ISWpi13_wRi-F	CAAGTTGGTAAATGCACGACA	636	60
ISWpi13_wRi-R	GCATCAAGCCCTGGAACTT		
ISWpi2_wRi-F	AGCAACAGAATTTCCAGCAT	692	60
ISWpi2_wRi-R	AACGCAACCAATGATCAACA		
ISWen3_wRi-F	TGCCTAATACTTGCGAATGC	690	60
ISWen3_wRi-R	ATTCAGTGGCCACGTTTCTC		

Les résultats du groupe ISWpi1 ont été repris de (Cordaux, et al. 2008)

Les sigles wM, wP et wRi indiquent le génome de référence dont le ou les copies ont été utilisées pour dessiner les amorces et représentent respectivement les génomes de *w*Mel, *w*Pel et *w*Ri

Tm : température utilisée pour la phase d'hybridation des amorces dans le cycle PCR.

Pour confirmer les résultats, toutes les réactions de PCR ont été réalisées au moins deux fois et les fragments obtenus ont été séquencés en séquençage direct avec le kit Big Dye Terminator. Le milieu réactionnel contient 0,5µL de BigDye, 3µL de tampon BigDye 5X, 0,5µL d'une amorce (10µM), 3µL de produit de PCR purifié et la quantité suffisante d'eau pour que le milieu réactionnel soit de 15µL. La réaction de PCR de séquençage commence par une dénaturation à 96°C pendant 30 secondes puis comprend 25 cycles des 3 étapes suivantes : dénaturation à 96°C pendant 45 secondes, hybridation de l'amorce à 55°C pendant 30 secondes et élongation à 60°C pendant 4 minutes. Pour chaque groupe d'IS, les séquences obtenues ont été alignées avec le logiciel ClustalW implémenté dans BioEdit. Puis, les alignements ont été vérifiés manuellement. La divergence nucléotidique a été calculée à l'aide du logiciel Mega ver 4.0, sur la base des substitutions observées.

Pour inférer le nombre d'acquisitions et de pertes des groupes d'IS, la distribution de chacun d'eux a été comparée à une phylogénie des 22 souches de *Wolbachia* testées (Cordaux, et al. 2001; Cordaux, et al. 2008; Lo, et al. 2007). Pour chaque groupe, nous avons privilégié le scénario le plus parcimonieux, c'est-à-dire celui qui fait intervenir le moins d'évènements d'acquisitions et de pertes pour expliquer la distribution en fonction de la phylogénie des souches de *Wolbachia*. Pour les groupes avec plus d'un scénario équi-parcimonieux, nous avons privilégié de façon conservative le scénario minimisant le nombre d'évènements d'acquisitions

4.3. Résultats et discussion

La forte abondance de copies dégradées nous a permis d'étudier les mécanismes qui régissent l'évolution des ET dans les génomes de *Wolbachia*. Nous avons tout d'abord étudié les mécanismes de dégradation des copies d'IS puis nous avons étudié leur dynamique évolutive dans un passé lointain et dans une période plus récente.

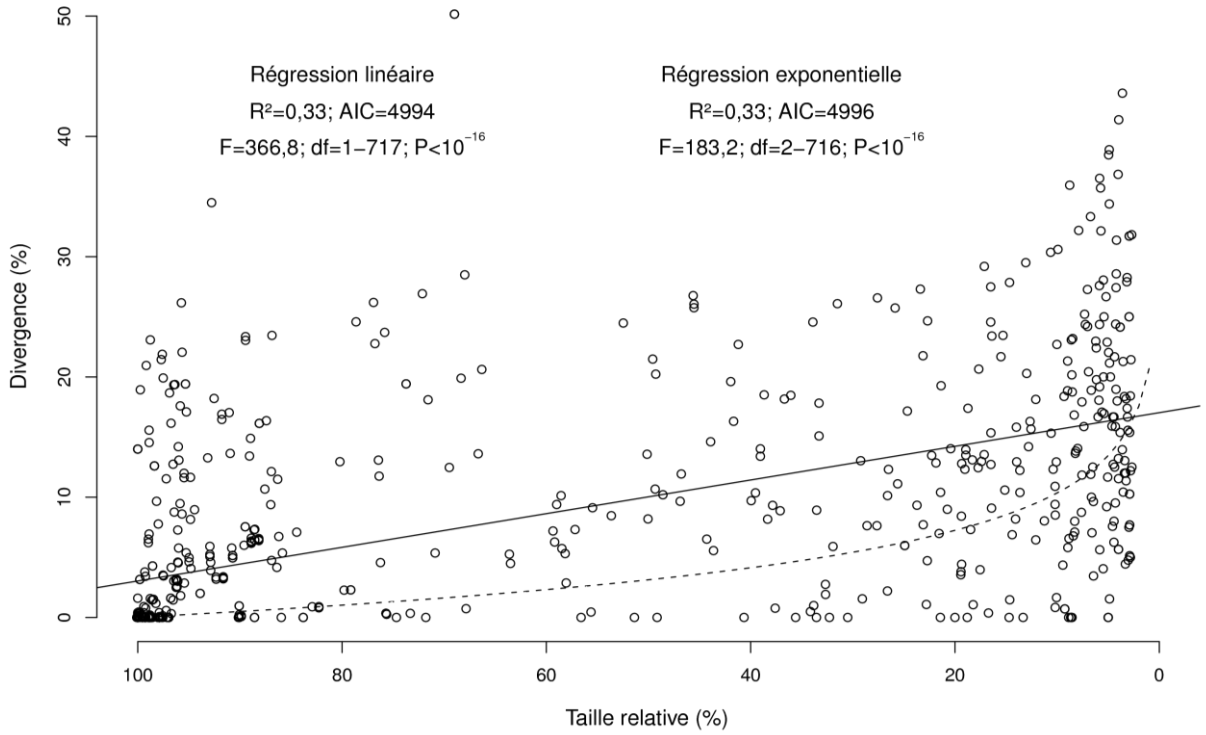

Figure 4-2 : Corrélation entre la divergence nucléotidique et la taille relative des copies d'IS des génomes de *w*Mel, *w*Ri, *w*Pel et *w*VulC.

Les copies d'IS proviennent de 28 groupes d'IS différents contenant au moins deux copies dont une complète au moins (n=719). La taille relative des copies a été calculée par rapport à la taille de la séquence consensus du groupe d'IS considéré et est exprimée en pourcentage. Seulement deux régressions sont modélisables sur notre jeu de données. Le trait plein représente la régression linéaire et le trait en pointillés représente la régression exponentielle.

4.3.1. Mécanismes de dégradations des IS

Nous sommes partis de l'hypothèse que l'érosion des IS se fait selon un modèle d'évolution neutre qui prédit une accumulation graduelle de substitutions et de délétions (Gomez-Valero, et al. 2008; Mira, et al. 2001; Moran 2003; Moran, et al. 2009; Moya, et al. 2008; Silva, et al. 2001). Ce processus prédit ainsi que la divergence nucléotidique est négativement corrélée à la taille relative des éléments. Nous avons testé cette hypothèse sur 28 groupes contenant au moins deux copies d'IS, dont au moins une entière, des génomes de *w*Mel, *w*Pel, *w*Ri et *w*VulC. Nous avons ainsi montré que la divergence et la taille relative des copies étaient corrélés (Rho^2= 0,51, $P<10^{-16}$, Figure 4-1 et Tableau 4-3 - page suivante). Cette corrélation est toujours observable même si on retire les copies complètes et/ou identiques au consensus et si les 4 génomes sont considérés séparément (Tableau 4-3 - page suivante). Ces résultats indiquent donc qu'il y a une corrélation entre l'accumulation de substitutions et l'accumulation de délétions dans les copies d'IS. Cependant, nous ne savons pas quel est le type de corrélation. En effet, contrairement au coefficient de corrélation de Pearson, le coefficient de corrélation de Spearman (Rho^2) ne définit pas uniquement une relation affine entre les deux variables testées, mais toutes les relations monotones. Nous avons donc étudié la relation entre la taille des copies et la divergence, grâce à différents types de régression linéaire et non linéaire. D'un point de vue biologique, 3 types de relation sont possibles : exponentielle (I), logarithmique (II) ou linéaire (III). Nous avons donc testé chacun de ces types de relation sur notre jeu de données (Figure 4-2).

Une relation logarithmique entre les deux variables signifierait que les copies commencent par accumuler des mutations avant d'accumuler des délétions. Ceci voudrait dire que les mutations qui touchent des copies d'IS sont plus fréquemment des substitutions que des délétions. Les données montrent certaines copies de grande taille dont le niveau de divergence nucléotidique avoisine ou dépasse les 20%, ce qui accrédite l'hypothèse d'une

Tableau 4-3 : Coefficient de corrélation entre taille relative et divergence nucléotidique et p-value pour les différents jeux de données

Génomes	Jeu de données[a]	Taille de l'échantillon	Spearman Rho²	p-value
Données des 4 génomes	A	719	0,51	<10E-16
combinées	B	617	0,37	<10E-16
	C	502	0,3	<10E-16
	D	603	0,38	<10E-16
	E	488	0,28	<10E-16
*w*Mel	A	91	0.62	<10E-16
	B	70	0.38	1.39E-8
	C	62	0.32	4.28E-6
	D	66	0.30	6.40E-7
	E	58	0.21	3.11E-4
*w*Pel	A	139	0.39	2.67E-16
	B	88	0.27	1.95E-7
	C	48	0.30	1.69E-8
	D	86	0.50	6.02E-8
	E	46	0.46	1.88E-7
*w*Ri	A	156	0.46	<10E-16
	B	129	0.37	3.12E-14
	C	88	0.39	5.03E-15
	D	124	0.51	8.85E-15
	E	83	0.43	1.52E-12
*w*VulC	A	333	0,23	<10E-16
	B	330	0,21	<10E-16
	C	304	0,21	<10E-16
	D	327	0,23	<10E-16
	E	301	0,23	<10E-16

[a] Type A: Jeu de données complet ; Type B : jeu de données sans les copies d'IS qui sont complètes et identiques au consensus ; Type C : jeu de données sans les copies d'IS qui sont identiques au consensus ; Type D : jeu de données sans les copies d'IS qui sont complètes; Type E : jeu de données sans les copies d'IS qui sont complètes ou identiques au consensus.

relation logarithmique entre divergence et taille pour ces copies. Néanmoins, de telles copies divergentes et de grande taille pourraient également résulter de transferts horizontaux indépendants de copies d'IS d'un même groupe mais provenant de sources différentes. Dans ce cas, il est tout à fait envisageable que deux copies complètes aient une divergence élevée. Cette hypothèse est illustrée par le fait que dans le groupe IS4wA présent chez *w*VulC possède deux types de copies potentiellement fonctionnelles qui ont une divergence nucléotidique de 14%. Enfin, la régression logarithmique n'est pas modélisable avec notre jeu de données, c'est pourquoi elle n'est pas représentée sur la figure 4-2.

Une relation exponentielle entre les deux variables impliquerait que les copies d'IS accumulent d'abord des délétions avant d'accumuler des substitutions. Les données montrent des copies identiques au consensus qui ne sont pas complètes, et dont certaines sont de petite taille. En effet, 26 des 217 copies identiques au consensus sont fortement tronquées, leur taille représentant de 5 à 89% de celle du consensus de leur groupe. On a donc ici un mécanisme qui entraîne des délétions de grande taille dans des copies d'IS d'origine relativement récente. Il est probable que ce phénomène fasse intervenir des mécanismes de réparation des cassures d'ADN, comme le *Non Homologuous end joining* initialement mis en évidence chez les eucaryotes (Paques and Haber 1999), mais aussi présent chez les procaryotes (Pitcher, et al. 2007), ou encore le *single strand annealing*. Ces mécanismes font appel à des homologies de séquences. L'ADN autour de la cassure sera dégradé jusqu'à ce que des séquences, pouvant s'appariées, soient libérées sur chacun des brins. Il semble donc que dans 12% des cas (26/217), la dégradation des copies d'IS se fasse selon une relation exponentielle entre la taille et la divergence. Enfin, le modèle exponentielle est statistiquement supporté sur notre jeu de données ($F=183,2$; $df=2\text{-}716$; $P<10^{-16}$ Figure 4-2). Cependant la valeur d'AIC du modèle exponentielle est supérieure à celle du modèle linéaire. C'est donc ce dernier qui explique le mieux nos données.

Le dernier type de relation entre la taille et la divergence qui s'applique le mieux à notre jeu de données est la relation linéaire. Dans ce cas, les copies d'IS accumulent de façon concomitante des substitutions et des délétions de taille modeste au cours du temps. Dans le cas où la taille des substitutions est de grande taille, on peut dans certains cas revenir à une relation de type exponentielle si la délétion est de très grande taille, ce qui semble être peu fréquent (<12%).

4.3.2. Dynamique à long terme des IS dans les génomes de Wolbachia.

Les copies d'IS accumulent de façon neutre des mutations (substitutions et délétions) au cours du temps, c'est pourquoi la divergence nucléotidique et la taille relative sont deux outils intéressants pour analyser la dynamique évolutive des IS dans les génomes de *Wolbachia*.

L'analyse de la divergence est une méthode qui a déjà fait ses preuves, par exemple pour étudier la dynamique évolutive des ET dans le génome humain (Lander, et al. 2001). Nous avons analysé 813 copies de 28 groupes d'IS des 5 génomes ayant au moins deux copies alignables. Le jeu de données utilisé pour cette analyse est donc plus large que celui utilisé pour étudier la dégradation des copies d'IS et contient des données provenant du génome de *w*Bm. Nous avons inclus les données issues de copies d'IS de *w*Bm car le nombre de groupes d'IS éligible nous permet d'obtenir un nombre de données suffisant pour faire des comparaisons avec les autres génomes. Il ne sera cependant pas possible de comparer l'évolution des copies d'IS de *w*Bm avec celles des autres génomes car la conversion génique modifie l'échelle temporelle en homogénéisant les séquences (Cordaux 2009). Pour chaque groupe d'IS nous avons calculé la divergence nucléotidique entre chaque paire de copies dans

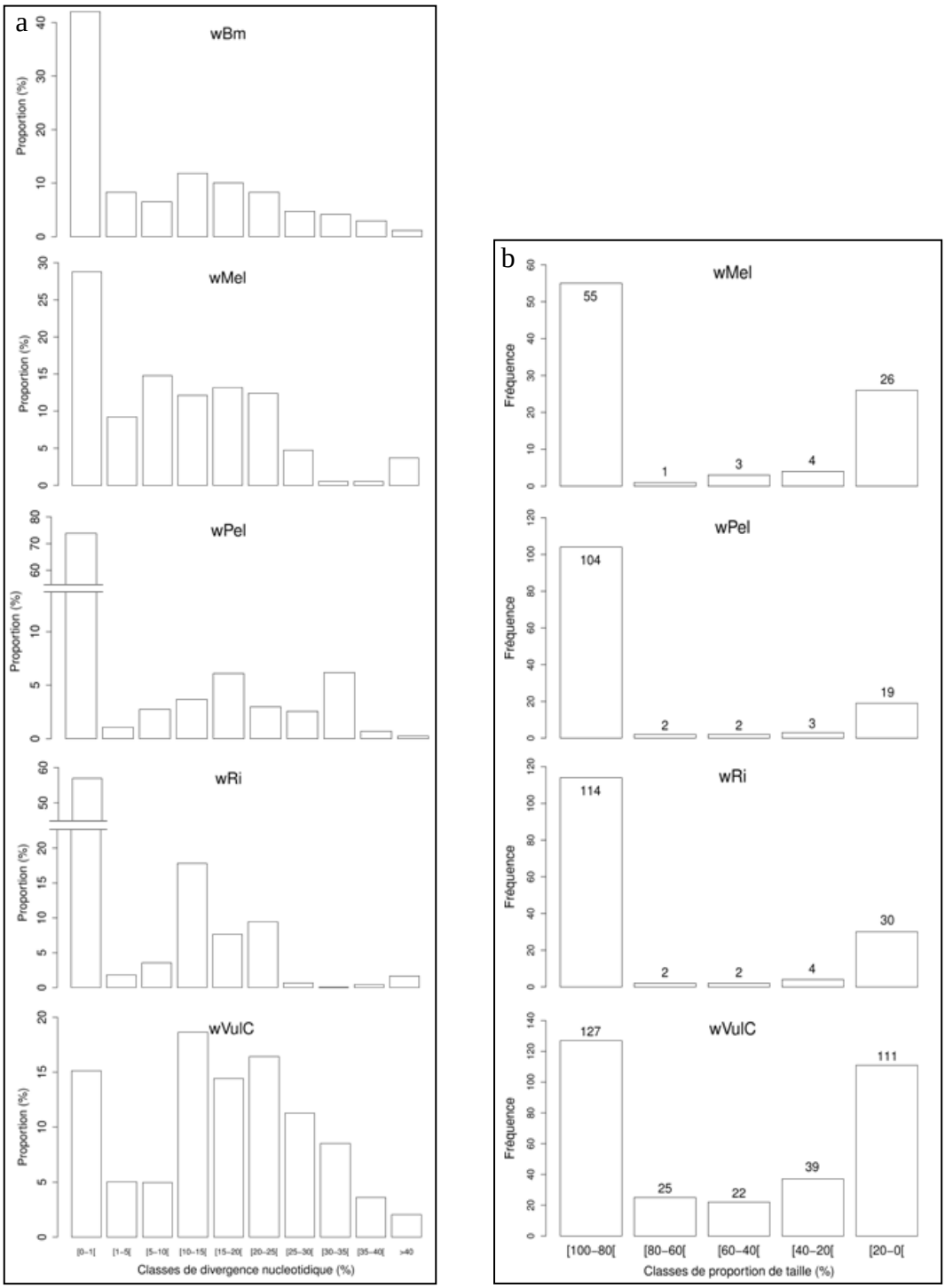

Figure 4-3 : Distribution des fréquences (a) de la divergence nucléotidique et (b) de la taille relative des copies d'IS pour chacun des génomes de *Wolbachia*.

(a) Les analyses ont été réalisées sur 6 groupes d'IS de *w*Bm (n=38 copies), 17 de *w*Mel (n=95), 12 de *w*Pel (n=153), 18 de *w*Ri (n=168), 28 de *w*VulC (n=359) ayant au moins de séquences alignables. La distribution est basée sur 169 comparaisons pour *w*Bm, 379 pour *w*Mel, 2152 pour *w*Pel, 1612 pour *w*Ri et 3087 pour *w*VulC. 10 classes de divergences ont été faites : [0-1[; [1-5[; [5-10[; [10-15[; [15-20[; [20-25[; [25-30[; [30-35[; [35-40[et >40% de divergence nucléotidique.

(b) Les analyses ont été réalisées sur 16 groupes d'IS de *w*Mel (n=89 copies), 9 de *w*Pel (n=130), 15 de *w*Ri (n=152), 22 de *w*VulC (n=324) ayant au moins de deux copies, dont une complète au moins. La taille des copies est exprimée en pourcentage de la taille de la séquence consensus

les différents génomes.

L'étude de la dynamique évolutive des IS a aussi été réalisée en se basant sur la taille relative des copies par rapport au consensus majoritaire des groupes considérés. Cette analyse a été faite avec le même jeu de données que celui utilisé pour étudier la dégradation des copies d'IS. Cependant, nous avons retiré les copies identiques au consensus mais fortement dégradées (correspondant plus à un modèle de dégradation exponentiel) pour ne pas altérer notre distribution. La taille de chaque copie a été comparée au consensus majoritaire de son groupe d'IS et est exprimée en pourcentage de la taille de celui-ci. Nous avons ensuite classé les copies des génomes de *w*Mel, *w*Pel, *w*Ri et *w*VulC dans 5 classes de tailles d'importance équivalentes:]0-20%],]20-40%],]40-60%],]60-80%] et]80-100%]. Nous avons utilisé 5 classes, car c'est le nombre maximal pour lequel chaque classe présente un effectif d'au moins une copie. De plus, contrairement à la divergence, la distribution est bornée par deux valeurs (0 et 100% de la taille de la séquence consensus du groupe), ce qui supprime les problèmes d'artéfacts de classification. En effet, il est possible de créer artificiellement un mode dans une distribution, en regroupant toutes les données situées à la fin de la distribution dans une même classe.

Les deux méthodes d'analyses ont mis en évidence un même schéma global observable dans tous les génomes avec la présence de deux modes dans les distributions (Figure 4-3). Le premier mode est généralement constitué de la majorité des copies qui présentent une divergence nucléotidique très faible (<1%) (Figure 4-3a) ou une grande taille (>80%) (Figure 4-3b). Ceci suggère une forte prolifération récente de copies d'IS induisant un grand nombre de comparaisons avec peu de divergence nucléotidique et une grande taille. De plus, on peut observer un second mode qui est pour l'analyse de divergence plus ou moins marqué en fonction des génomes avec des divergences nucléotidiques supérieures à 10% (Figure 4-3a) ou une faible taille (<25%) (Figure 4-3b). Ceci suggère que les génomes de *Wolbachia*

contiennent une part importante de copies d'IS anciennes qui sont le résultat d'une activité passée.

Ainsi, nous avons montré que la distribution de la divergence des copies d'IS dans les génomes de *Wolbachia* est bimodale. Le schéma observé peut être expliqué soit (1) par une variation du taux de délétion et de substitution en fonction de la taille des copies, soit (2) par des variations dans l'activité de transposition en fonction du temps. Les variations des taux de délétions et de substitutions sont peu probables. En effet pour la taille relative, il faudrait que le taux de délétion soit plus fort dans les 3 classes de taille intermédiaires de façon à ce que les copies se retrouvent très rapidement dans la classe [20-0[. L'analyse des copies identiques aux consensus nous a montré que le phénomène de grande délétion n'arrivait que très peu fréquemment. De plus, seulement 9% des 186 copies comprises dans la classe de taille [0-20%[ont une divergence nucléotidique inférieure à 5% par rapport au consensus du groupe. En outre, il faudrait que l'accélération du taux de délétion soit accompagnée d'une accélération du taux de mutation car on observe le même schéma pour les deux variables et elles sont corrélées linéairement. L'hypothèse la plus probable est donc celle d'une variation d'activité des IS au cours du temps. Ce même schéma est observable plus ou moins clairement dans les 5 génomes étudiés, ce qui suggère que l'évolution de l'activité de transposition suit un schéma analogue dans les différents génomes. Néanmoins, il existe des différences entre les génomes et notamment dans le génome de *w*VulC (Figure 4-3a et b). Dans la distribution des fréquences de divergence nucléotidique des IS de *w*VulC, c'est le second mode qui est majoritaire et non pas le premier. Ceci suggère donc que la phase d'activité récente des IS a été moins forte dans ce génome ou qu'elle n'en est qu'à ses débuts. De plus, la distribution de la taille relative des copies du génome de *w*VulC présente deux modes de fréquences quasi égales et est différente de celles observées dans les 3 autres génomes (test du khi²; $p<0,001$ pour toutes les comparaisons). Ceci suggère que la dynamique des IS de *w*VulC est analogue

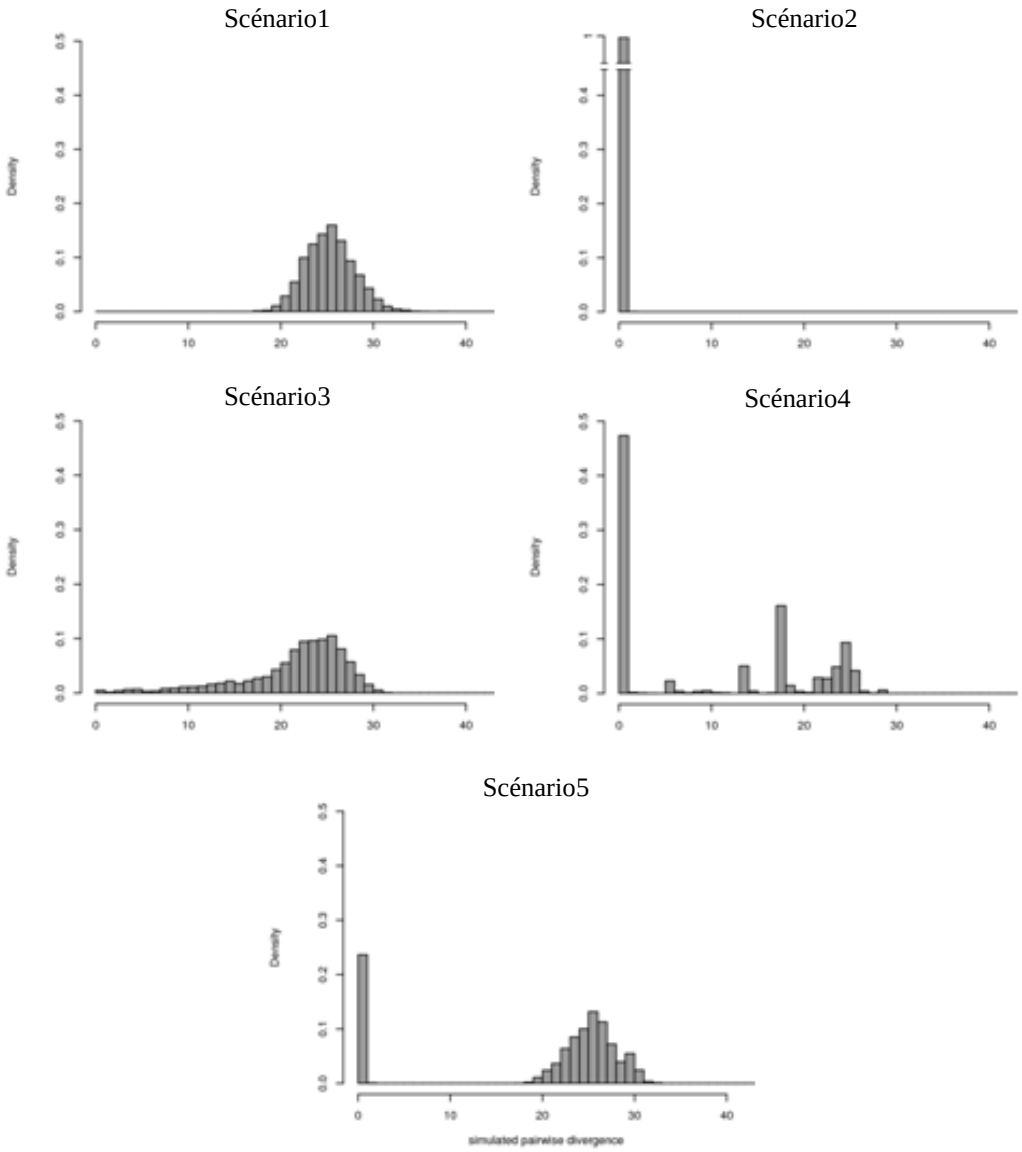

Figure 4-4 : Distribution des divergences nucléotides obtenue dans le cadre de la simulation de l'évolution des IS dans un génome bactérien.

Scénario 1 : un seul pic de transposition ancien ; scénario 2 : un seul pic de transposition récent ; scénario 3 : transposition à un faible taux ; scénario 4 : deux pics de transposition récents et indépendants et scénario 5 : un pic de transposition récent et un autre ancien.

Chaque distribution est un pool des distributions de 36 simulations différentes, chacune d'entre elles avec un nombre de copies d'IS équivalent à celui d'un des groupes d'IS de *Wolbachia*.

mais tout de même différente de celles observées dans les autres génomes.

Pour analyser plus en détail l'origine du schéma de distribution bimodale observée précédemment, nous avons simulé l'évolution de la divergence nucléotidique des copies d'IS dans un génome bactérien hypothétique, selon différents scénarios. Ceux-ci font intervenir des pics d'activité de transposition ayant une intensité variable, des transferts horizontaux, l'accumulation de substitutions et la perte de copies.

Les deux premiers scénarios simulent l'acquisition d'une unique copie d'IS, suivie d'une forte expansion au début (scénario 1) ou à la fin (scénario 2) de la simulation (Figure 4-4). Le résultat est un mode unique avec soit peu de divergence dans le cas de l'expansion récente, soit beaucoup de divergence dans le cas de l'expansion ancienne. Le scénario 3 simule la transposition peu fréquente mais constante d'une copie d'IS présente au début de la simulation (Figure 4-4). Le résultat est une distribution avec un mode unique de divergence supérieure à 20% avec un aplatissement de la distribution vers les faibles divergences. Les deux derniers scénarios (4 et 5) simulent chacun deux acquisitions indépendantes d'IS et deux forts évènements de transpositions (Figure 4-3). Dans le scénario 4, les deux phénomènes sont récents (ce qui correspond au scénario 2 répété deux fois) alors que dans le scénario 5, il y a un évènement récent et un autre ancien (ce qui correspond aux scénarios 1 et 2 cumulés). Ce dernier scénario est analogue au modèle de dynamique évolutive des IS dans les génomes bactériens proposé dans la littérature (Wagner 2006; Wagner, et al. 2007). Le schéma de distribution résultant des deux fortes expansions récentes (scénario 4) présente un mode avec une faible divergence et de nombreux pics à des niveaux de divergence variable. Le scénario 5 est le seul qui présente une distribution clairement bimodale avec un mode pour les fortes divergences (correspondant à l'ancien transfert et son burst) et un second pour les très faibles divergences (correspondant au récent transfert et son burst), comme observé dans les génomes de *Wolbachia* (Figure 4-4). Il est à noter que les variations de taux de délétions des copies

entraînent la diminution du nombre de copies anciennes diminuant la hauteur du mode concernant les copies anciennes sans pour autant qu'il ne disparaisse (Annexe 3).

En conclusion, la distribution bimodale des divergences nucléotidiques et de la taille relative des copies d'IS suggère que l'activité de transposition dans les génomes de *Wolbachia* n'est pas constante au cours du temps. De plus, il semble que la dynamique de transposition des IS dans les 5 génomes de *Wolbachia* soit analogue, même si on peut observer des différences entre génomes, notamment pour *w*VulC. Pour ce dernier, les distributions de la divergence nucléotidique et de la taille relative des copies suggèrent que les IS de *w*VulC ne sont pas dans la même phase de dynamique d'activité que ceux des autres génomes de *Wolbachia*. Dans les 5 génomes, la plupart des copies d'IS semble avoir été produite dans au moins deux périodes majeures d'activité de transposition, une ancienne (divergence > 10% ou taille relative < 25%) et une récente (divergence très faible ou taille > 80%) qui est peut-être toujours en cours. Enfin, nos simulations de l'évolution nucléotidique des copies d'IS confirment que la dynamique évolutive des IS de *Wolbachia* nécessite des transferts horizontaux suivis de pic d'activité de transposition pour expliquer les distributions de divergences nucléotidiques observées.

Afin de comparer les phases d'activité d'IS dans les génomes de *Wolbachia*, nous avons affiné notre analyse de la dynamique des IS en étudiant la composition en groupes d'IS des deux classes de taille relatives extrêmes (c'est-à-dire [100-80%[et [20-0%[). Les compositions en groupes d'IS de la classe de taille [20-0%[ne sont pas statistiquement différentes entre *w*Mel et *w*Ri (Khi^2=12,4; df=13, *P*=0,49), mais elles diffèrent significativement des compositions observées pour *w*Pel et *w*VulC qui sont différentes entre elles (khi^2=56,95; df=18; $P=6.10^{-6}$). Ceci suggère qu'une vague de prolifération ancienne s'est déroulée chez l'ancêtre commun de *w*Mel et *w*Ri, ce qui parait plausible car ces deux souches sont phylogénétiquement proches (Klasson, et al. 2009; Wu, et al. 2004). De plus, cette vague

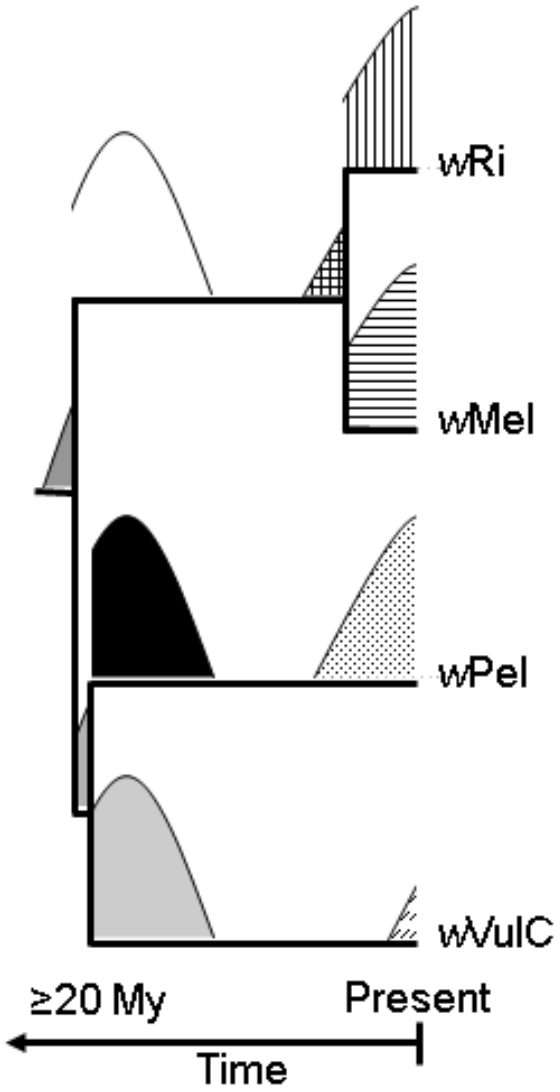

Figure 4-5 : Scénario de dynamique évolutive à long terme des IS dans les génomes de *Wolbachia*

La taille des branches de l'arbre phylogénétique est arbitraire, il nous sert uniquement de support pour pouvoir inférer des vagues de prolifération d'IS. Les vagues d'expansion se déroulent de façon indépendante dans différents souches comme en témoignent les couleurs de remplissage différentes. Nous avons pu mettre en évidence deux périodes de transposition. La première qui a probablement démarré très récemment et qui est peut-être toujours en cours et la seconde, plus ancienne, qui a probablement démarré dans l'ancêtre commun des 4 souches. En effet, nous avons trouvé une copie d'IS dégradée insérée à un site orthologue dans les génomes de *w*Mel, *w*Pel et *w*Ri ; de plus, les génomes de *w*Pel et *w*VulC présentent aussi des copies d'IS dégradées insérées à des sites orthologues, ce qui suggère une origine phylogénétique commune à ces deux souches. Ceci indique que notre analyse de la dynamique évolutive des IS de *Wolbachia* couvre une période d'au moins 20 millions et au plus 60 millions d'années d'évolution (estimation basée sur la divergence estimée des lignées *w*Mel/*w*Ri et *w*Pel (Kuo and Ochman 2009; Werren, et al. 1995).

ancienne s'est probablement déroulée au moment de la séparation entre *w*VulC et *w*Pel qui sont deux souches de supergroupe de diversité B (Figure 4-5) car leur distribution de groupe d'IS sont statistiquement différentes dans la classe de taille [20-0%[. La recherche de copies d'IS insérées dans des sites orthologues dans les génomes de *w*Mel, *w*Pel, *w*Ri et *w*VulC a permis de mettre en évidence seulement une copie fragmentaire d'ISWpi16 qui est insérée dans un site orthologue dans les génomes de *w*Pel, *w*Mel et *w*Ri. Dans le génome de *w*VulC, le site orthologue a été retrouvé. Il présente une séquence de 20 pb qui ne peut pas être reliée au groupe ISWpi16. Néanmoins, 4 fragments de copies d'IS sont insérés dans des sites orthologues entre les génomes de *w*VulC et *w*Pel, mais qui ne sont pas présents dans les génomes de *w*Mel et *w*Ri. On peut donc supposer que l'expansion ancienne a commencé un peu avant la divergence des lignées *w*Pel/*w*VulC et *w*Mel/*w*Ri et qu'elle s'est prolongée ensuite, le fragment d'ISWpi16 ayant probablement été dégradé plus rapidement dans le génome de *w*VulC que dans les autres. De plus, le peu de fragments de copies d'IS insérées dans des sites orthologues entre *w*Pel et *w*VulC suggère que ces deux souches ont divergé rapidement après la séparation de la lignée *w*Mel/*w*Ri. Ceci est d'ailleurs supporté par le fait que les distributions des groupes d'IS dans la classe de taille]20-0%] sont statistiquement différentes entre les génomes *w*Pel et *w*VulC (Figure 4-5).

Concernant la classe de taille]100-80%], toutes les comparaisons réalisées montrent que les distributions sont statistiquement différentes les unes des autres (Tableau 4-4 - page suivante). Ceci suggère donc que l'histoire évolutive récente des IS est spécifique à chacun des génomes étudiés (Figure 4-5). Concernant le génome de *w*VulC, nous avons suggéré que l'activité des IS n'est pas dans la même phase que les autres génomes mais nous ne savons pas si elle se situe en début ou en fin d'une phase d'expansion. Nous avons pris le parti de représenter la phase récente avec une activité de plus faible intensité mais nous ne pouvons savoir si les IS sont en début, en fin ou dans une phase de forte activité.

Tableau 4-4: Valeurs des tests de comparaisons de distribution de groupe d'IS pour la classe de taille [100-80%[

Génomes comparés	Valeur de khi²	df	*P*
*w*Mel/*w*Ri	63,2	18	6.10^{-7}
*w*Mel/*w*Pel	117,8	18	$< 2.10^{-16}$
*w*Mel/*w*VulC	98,9	25	9.10^{-11}
*w*Ri/*w*Pel	194,6	16	$< 2.10^{-16}$
*w*Ri/*w*VulC	154,4	24	$< 2.10^{-16}$
*w*Pel/*w*VulC	132,5	24	$< 2.10^{-16}$

Tableau 4-5 : Distribution des copies potentiellement fonctionnelles présentes dans les cinq génomes de *Wolbachia*.

Famille d'IS	Groupe d'IS	*w*Bm	*w*Mel	*w*Ri	*w*Pel	*w*VulC
IS3	IS3wB					6
IS4	ISWen1		3			
	IS4wA					6
IS5	ISWpi1		13	20		
IS110	ISWen2		1	16		3
	ISWpi12				3	
	ISWpi13			1		
	ISWpi14		1			
IS200/605	ISW1				2	
IS256	ISWpi15		1		6	
IS481	ISWpi2			5		
	ISWpi4		1	1		
IS630	ISWpi11				4	
	ISWpi10				1	
IS982	ISWpi16				44	
IS66	ISWen3			21		
Nombre de copies potentiellement fonctionnelles		0	20	64	60	15
Proportion		0%	19%	37%	35%	4%
Nombre de copies non fonctionnelles		52	85	107	110	357
Proportion		100%	81%	63%	65%	96%

4.3.3. Déroulement d'une phase d'activité d'IS

La structure bimodale des distributions de divergence nucléotidique et de taille suppose l'idée de phase d'activité de transposition intense entrecoupée de phases de quiescence, ce qui est confirmé par nos simulations. Ceci pose la question du déroulement d'une phase d'expansion, comment se comportent les groupes ayant des copies potentiellement fonctionnelles? D'où proviennent les copies actives? Sont-elles présentes de longue date dans le génome ou arrivent-elles par des transferts horizontaux?

Pour comprendre le déroulement d'une phase d'expansion d'IS dans les génomes de *Wolbachia,* nous avons étudié les copies et les groupes d'IS qui ont subi une prolifération récente. L'analyse de la structure des copies d'IS des génomes de *Wolbachia* avait permis de mettre en évidence que 4-37% des copies d'IS des génomes de *w*Mel, *w*Pel, *w*Ri et *w*VulC sont potentiellement fonctionnelles (Figure 3-10 - page 57). Dans chaque génome, les IS potentiellement fonctionnelles sont reparties dans 6 groupes différents, à l'exception du génome de *w*VulC qui ne comprend que 3 groupes (Tableau 4-5).

Il existe une importante diversité d'IS potentiellement fonctionnels dans les génomes de *Wolbachia*. Ces copies sont issues de 16 groupes et 10 familles d'IS différentes. Dans chacun des génomes un nombre limité de groupes (1 à 3) représente plus de la moitié des copies potentiellement fonctionnelles. Cependant, dans le génome de *w*VulC, le nombre de copies potentiellement fonctionnelles est homogène entre les différents groupes et généralement faible (Tableau 4-6 – page suivante). L'opposition entre *w*VulC et les 3 autres génomes possédant des copies potentiellement fonctionnelles nous conforte dans l'idée que l'activité des IS est dans une phase différente de la dynamique dans ce génome. Il est intéressant de noter que ce ne sont pas les mêmes groupes qui ont subi une forte expansion dans les différents génomes.

Ceci participe très probablement à l'hétérogénéité de distribution des groupes d'IS que l'on

Tableau 4-6: Caractéristiques des groupes d'IS utilisés pour construire des réseaux d'haplotypes

Groupe d'IS	Génome de *Wolbachia*	Nombre de copies d'IS	Nombre de copies complètes	Divergence nucléotidique moyenne entre copies complètes
IS3wA	*w*Ri	15	12	9,82%
ISWen2	*w*Ri	23	20	0,13%
ISWpi1	*w*Ri	22	21	0,02%
ISWen3	*w*Ri	44	28	0,04%[a]
IS3wA	*w*Mel	18	14	10,58%
ISWpi1	*w*Mel	14	13	0,02%
ISWpi16	*w*Pel	54	47	0,00%
IS3wB	*w*VulC	11	11	0,19%[b]
ISWpi15	*w*VulC	44	31	7,38%

[a] deux séquences très divergentes ont été exclues

[b] une seul séquence est très divergente, si elle est retiré la divergence moyenne est de 0,052%

observe dans les génomes de *Wolbachia*. La très faible divergence observée entre les IS des groupes ayant subi une forte prolifération dans les génomes de *w*Mel, *w*Ri et *w*Pel suggère que leur expansion a eu lieu durant l'évolution récente de *Wolbachia* et de façon spécifique pour chaque groupe d'IS.

Afin de caractériser l'expansion des groupes d'IS, nous les avons analysé avec une méthode utilisée généralement en génétique des populations mais qui a aussi fait ses preuves pour l'analyse des ET : les réseaux d'haplotypes (Cordaux, et al. 2004a). Nous avons considéré tous les groupes d'IS contenant au moins 10 copies complètes, des 4 génomes de *Wolbachia* ayant des copies potentiellement fonctionnelles. Néanmoins seul le groupe ISWen2 présentait une divergence moyenne inférieure à 1% mais non nul. La divergence moyenne des autres groupes testés était soit quasi nulle, ce qui veut dire que toutes les séquences sont identiques ou presque, soit bien trop élevée, ce qui ne nous aurait pas permis de produire un réseau d'haplotypes interprétable (Tableau 4-5). Les réseaux d'haplotypes dérivés des groupes d'IS ayant une divergence moyenne nulle ne présenterait qu'un seul cercle contenant toutes les copies. A contrario, les réseaux d'haplotypes dérivés de groupes d'IS ayant une divergence supérieure à celle du groupe ISWen2 présentaient de grandes branches, synonymes de la forte divergence entre copies. De plus, les branches se recoupaient fréquemment entre elles formant des structures de type cristallin, ce qui peut être le résultat d'un phénomène de recombinaisons homologues entre copies. Le groupe ISWen2 dans le génome de *w*Ri est donc le seul exemple pour décrire l'expansion d'un groupe d'IS spécifique dans un génome de *Wolbachia*. Des copies du groupe ISWen2 ont été identifiées dans les génomes de *w*VulC, *w*Mel et *w*Ri (Figure 4-6a – page suivante). Nous nous sommes focalisés sur les génomes de *w*Mel et *w*Ri car, bien que 3 génomes possèdent au moins une copie potentiellement fonctionnelle, on peut observer une grande différence d'abondance entre les

génomes de *w*Mel et *w*Ri qui sont pourtant phylogénétiquement très proches. Cette grande différence est

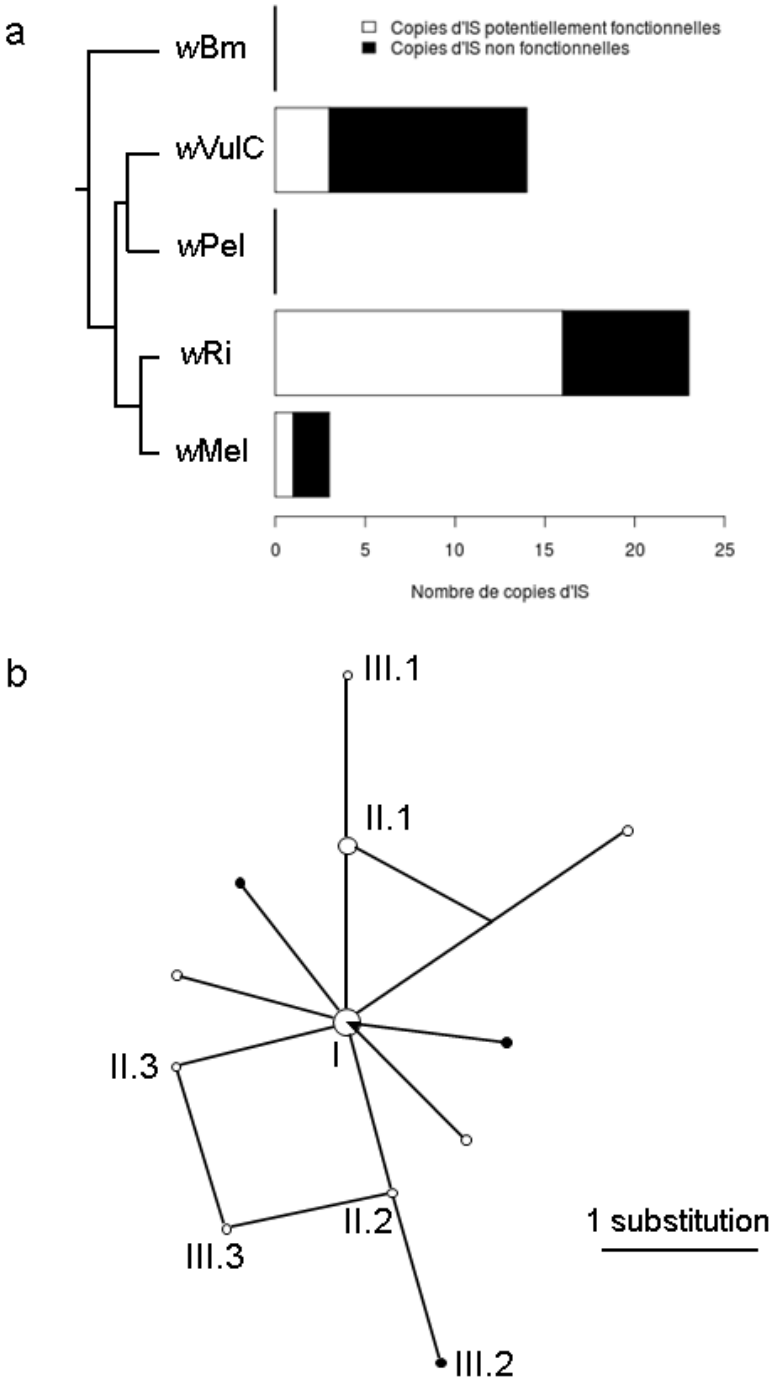

Figure 4-6 : Expansion du groupe ISWen2 dans le génome de *w*Ri.

(a) nombre de copies du groupe ISWen2 détectées dans les cinq génomes de *Wolbachia* étudiés. Les longueurs de branches de l'arbre phylogénétique sont arbitraires.

(b) Réseau d'haplotypes réalisé sur les 20 séquences complètes du groupe ISWen2 détectées dans le génome de *w*Ri. Chaque cercle représente un type de séquence (nœuds) et la taille du cercle est fonction du nombre de séquence ayant le type considéré. Les nœuds qui sont utilisés dans le texte sont nommés I, II 1-3 et III 1-3. Les lignes représentent les substitutions avec l'échelle en bas à droite de la figure. Le code couleur concernant le statut fonctionnel des copies est le même que dans la partie (a) de la figure avec les copies potentiellement fonctionnelles en blanc et les non fonctionnelles en noir.

principalement due à l'augmentation du nombre de copies potentiellement fonctionnelles dans le génome de *w*Ri, ce qui suppose une expansion récente.

Pour analyser cette expansion, nous avons utilisé une approche phylogénétique basée sur les réseaux d'haplotypes via le programme Network ver 4.5.1.6 (Bandelt, et al. 1999). Ce type d'approche a déjà été utilisée pour caractériser l'expansion d'une sous-famille d'éléments *Alu* dans le génome humain (Cordaux, et al. 2004a). Contrairement à la phylogénie classique, l'approche par réseau d'haplotype permet de visualiser les liens qui existent entre les différents types de séquences observées et de comprendre leur évolution. Le réseau réalisé avec les copies du groupe ISWen2 montre une structure en étoile avec 35% des copies dans le nœud central (I dans Figure 4-6b). Certains nœuds périphériques ne sont pas directement connectés au nœud central (III.1, III.2 et III.3 dans Figure 4-6b) ou contiennent plusieurs séquences (II.1 dans Figure 4-6b). Etant donné que l'homoplasie est un phénomène peu courant à cette profondeur phylogénétique et que les IS transposent préférentiellement en *cis* (c'est-à-dire que la transposase vient se fixer sur la copie qui a permis sa production), nous en avons conclu que l'expansion du groupe ISWen2 dans le génome de *w*Ri est supportée par au moins 3 copies (au moins une des nœuds I et II.1 et celle du nœud II.2) et au plus 11 copies (des nœuds I, II.1, II.2 et II.3). Notre analyse suggère que de 15 à 55% des copies peuvent contribuer à l'expansion d'un groupe d'IS. Ceci suggère que seule une partie des copies est transpositionnellement active et participe à l'expansion du groupe. Les copies ne possèderaient-elles pas toutes l'ensemble des structures nécessaires à la transposition ou ne sont-elles pas toutes transcrites ? Cette question sera soulevée dans le chapitre suivant traitant de la transcription des IS.

Tableau 4-7 : Polymorphisme de présence/absence d'insertion de 44 copies d'IS dans les génomes de *w*Mel et *w*Ri.

Famille d'IS	Groupe d'IS	Copies d'IS spécifiques de *w*Mel	Copies d'IS spécifiques de *w*Ri	Copies d'IS partagées par *w*Me et *w*Ri
IS3	IS3			12
IS4	IS4-wB			6
IS5	IS903	1		
	IS1031			1
	ISWpi1	6	4	
IS110	IS1111			1
	ISWen2		4	
	ISWpi12			1
	IS110-wA			1
IS481	ISWpi2		1	
	ISWpi4			1
IS630	ISWpi11	1		
IS982	ISWpi16			1
IS66	ISWen3		2	1
Nombre de copies		8	11	25
Proportion de copies potentiellement fonctionnelles		75%	91%	4%

4.3.4. Recherche de signes d'activité pour d'autres groupes d'IS

Il a précédemment été montré que le groupe ISWpi1, spécifique des génomes de *Wolbachia* et très répandu dans ces derniers, présente des copies qui sont insérées à des sites différents entre souches (Cordaux 2008; Cordaux, et al. 2008). Ce schéma de présence/absence suggère une activité de transposition intense et récente du groupe ISWpi1 dans les génomes de *Wolbachia* (Cordaux 2008; Cordaux, et al. 2008). Pour savoir si ce schéma peut être étendu à un nombre plus important de groupes d'IS, nous avons recherché des polymorphismes d'insertions de copies d'IS dans les génomes des deux souches phylogénétiquement proches *w*Mel et *w*Ri. Nous avons trouvé 19 copies d'IS spécifiquement insérées dans l'un des deux génomes parmi les 44 loci orthologues que nous avons identifié sans ambiguïté entre les génomes de *w*Mel et *w*Ri (Tableau 4-7). Les copies polymorphes représentent 6 groupes d'IS différents, dont ISWpi1, ce qui montre que de multiples groupes d'IS ont transposé dans le passé récent de l'évolution de *Wolbachia*. Le fait que 89% (16/19) des copies polymorphes contre seulement 4% (1/25) des copies partagées entre les souches *w*Mel et *w*Ri soient potentiellement fonctionnelles renforce l'idée que les copies polymorphes sont d'origine récente (Tableau 4-7). De plus, il existe des copies d'IS polymorphe dans les souches de *Wolbachia w*Pel infectant le moustique *Culex pipiens* (Duron, et al. 2005; Sanogo, et al. 2007). Ces résultats suggèrent que la transposition des IS est peut-être toujours en cours dans les génomes des souches de *Wolbachia w*Mel, *w*Ri et *w*Pel.

4.3.5. Quelle origine pour les copies potentiellement fonctionnelles?

La présence d'une grande diversité de groupes d'IS potentiellement actifs dans les génomes de *Wolbachia* pose la question de leur origine. Les simulations de l'évolution de la

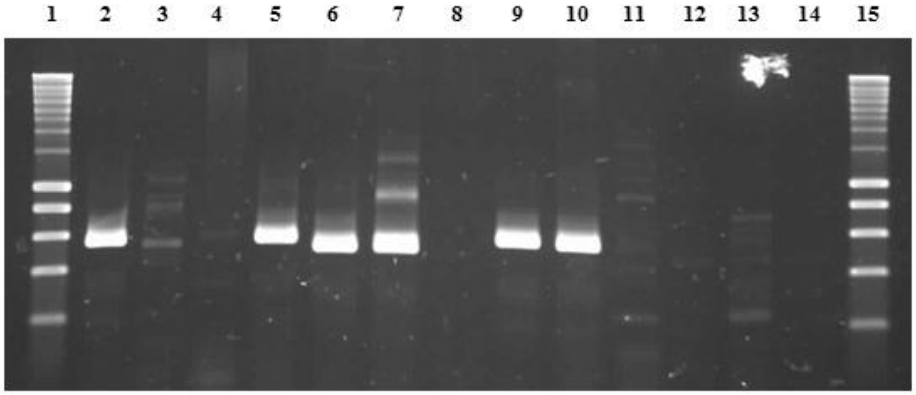

Figure 4-7 : Détection par PCR de la présence du groupe ISWpi11 dans les 11 souches des supergroupes de diversité de *Wolbachia* B et G.

Les amorces utilisées pour la détection du groupe ISWpi11 ont été dessinées à partir des copies présentes dans le génome de *w*Pel. Les pistes 1 et 15 correspondent au marqueur de taille SmartLadder (Eurogentec). La piste 2 est un témoin positif et la piste 14 est un témoin négatif. Les pistes 3 à 13 correspondent aux individus des différentes souches de *Wolbachia* des supergroupes de diversité B et G de notre jeu de données. Les signaux positifs ont été séquencés en séquençage directe pour vérification.

divergence nucléotidique des IS nous ont montré qu'il n'est pas possible que les copies d'IS potentiellement actives qui sont présentes dans les génomes de *Wolbachia* soient issues d'une prolifération ancestrale et se soient maintenues à long terme à un faible niveau de transposition.

L'hypothèse la plus probable concernant l'origine des groupes d'IS est celle des transferts horizontaux. En effet, il a été montré que les transferts horizontaux d'IS tiennent une place cruciale dans la dynamique évolutive des IS, même si d'autres forces influencent plus fortement celle-ci (Wagner 2006). De plus, il a été montré que la dynamique des copies du groupe ISWpi1 est intimement liée à une transmission horizontale fréquente entre souches de *Wolbachia* (Cordaux, et al. 2008). Pour évaluer si ce scénario s'applique également à d'autres groupes d'IS de *Wolbachia*, nous avons testé, par PCR (Figure 4-7), la présence de 17 groupes d'IS dans un échantillon de 22 souches de *Wolbachia* différentes.

Pour tester la fiabilité de notre analyse, nous avons utilisé comme témoins les génomes de *w*Mel, *w*Ri et *w*Pel qui nous ont par ailleurs servi à créer les amorces d'amplification spécifiques de chacun des 17 groupes d'IS. Nous avons comparé la distribution des groupes d'IS obtenue en PCR à celle prédite *in silico*. Nous avons obtenu des résultats cohérents pour 47 des 51 tests PCR réalisés (17 groupes d'IS dans 3 génomes). Pour les 4 résultats incohérents (absence de signal PCR alors que l'analyse *in silico* prédit la présence d'au moins une copie détectable), la divergence observée entre les copies du génome ayant servi à créer les amorces et celles du génome pour lequel nous avons une absence de détection est toujours supérieure à 13%. Etant donné que nous cherchons à détecter des transferts horizontaux récents, donc présentant peu de divergence entre les copies d'IS de différents génomes, nous en avons conclu que notre méthode était suffisamment fiable.

Nous avons comparé la distribution des 17 groupes d'IS à la phylogénie connue des 22 souches de *Wolbachia* en appliquant le principe de parcimonie pour inférer les acquisitions et

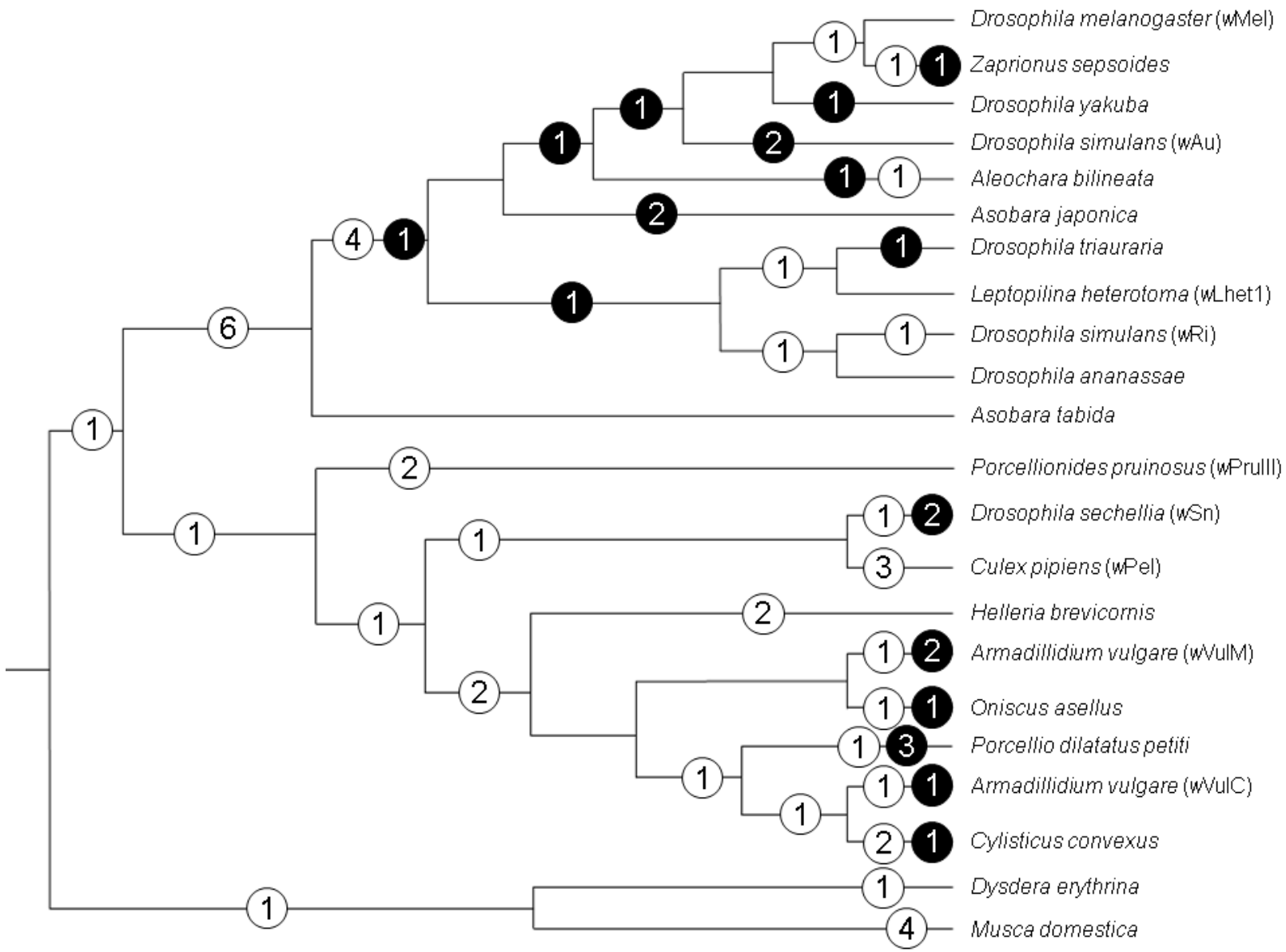
Drosophila melanogaster (wMel)
Zaprionus sepsoides
Drosophila yakuba
Drosophila simulans (wAu)
Aleochara bilineata
Asobara japonica
Drosophila triauraria
Leptopilina heterotoma (wLhet1)
Drosophila simulans (wRi)
Drosophila ananassae
Asobara tabida
Porcellionides pruinosus (wPruIII)
Drosophila sechellia (wSn)
Culex pipiens (wPel)
Helleria brevicornis
Armadillidium vulgare (wVulM)
Oniscus asellus
Porcellio dilatatus petiti
Armadillidium vulgare (wVulC)
Cylisticus convexus
Dysdera erythrina
Musca domestica

Figure 4-8 : Historique des acquisitions et pertes de 17 groupes d'IS dans 22 souches de *Wolbachia*.

La figure présente le nombre d'acquisitions (cercles blancs) et de pertes (cercles noirs) le plus parcimonieux pour expliquer la répartition de 17 groupes d'IS dans 22 souches de *Wolbachia*. Le nombre d'acquisitions et/ou de pertes pour une branche considérée est indiqué dans les cercles. La phylogénie des 22 souches de *Wolbachia* a été réalisée à partir de la littérature (Cordaux, et al. 2001; Cordaux, et al. 2008; Lo, et al. 2007). La taille des branches était arbitraire. Les souches sont nommées par le nom de l'espèce hôte dans laquelle elles ont été isolées.

##########

pertes de groupes d'IS au cours de l'évolution des souches (Figure 4-8). Notre analyse a permis de mettre en évidence que tous les génomes de *Wolbachia* possèdent des copies d'IS d'au moins 3 groupes différents. Ceci confirme que la présence d'IS est bien une caractéristique générale des génomes de *Wolbachia*. De plus, nous avons pu montrer que bien que des IS soient présents dans tous les génomes de *Wolbachia*, leur distribution est hétérogène, ce qui est en accord avec les observations faites chez d'autres bactéries (Qiu, et al. 2010; Sawyer, et al. 1987).

Notre analyse a mis en évidence que la répartition des 17 groupes d'IS dans les 22 souches de *Wolbachia* nécessite au moins 44 évènements de transferts horizontaux (Figure 4-8). Cette estimation est très conservative car (i) nous avons toujours privilégié les scénarios qui minimisent le nombre d'acquisitions lorsque plusieurs scénarios équi-parcimonieux étaient possibles pour un groupe particulier, (ii) nous sommes partis de l'hypothèse que les groupes monophylétiques de souches de *Wolbachia* qui possédaient tous un groupe d'IS donné l'avaient hérité de leur ancêtre commun, or il est possible que les souches l'aient acquis indépendamment comme précédemment décrit pour ISWpi1 (Cordaux, et al. 2008), (iii) il n'est pas improbable qu'un génome ait subi plusieurs évènements d'acquisitions indépendants pour un groupe d'IS considéré.

Les séquences obtenues par séquençage direct des fragments de PCR ne présentent qu'un très faible nombre de site nucléotidique incertain, c'est-à-dire pour lesquels il existe deux séquences différentes avec chacune un acide nucléique différent. Lorsqu'une paire

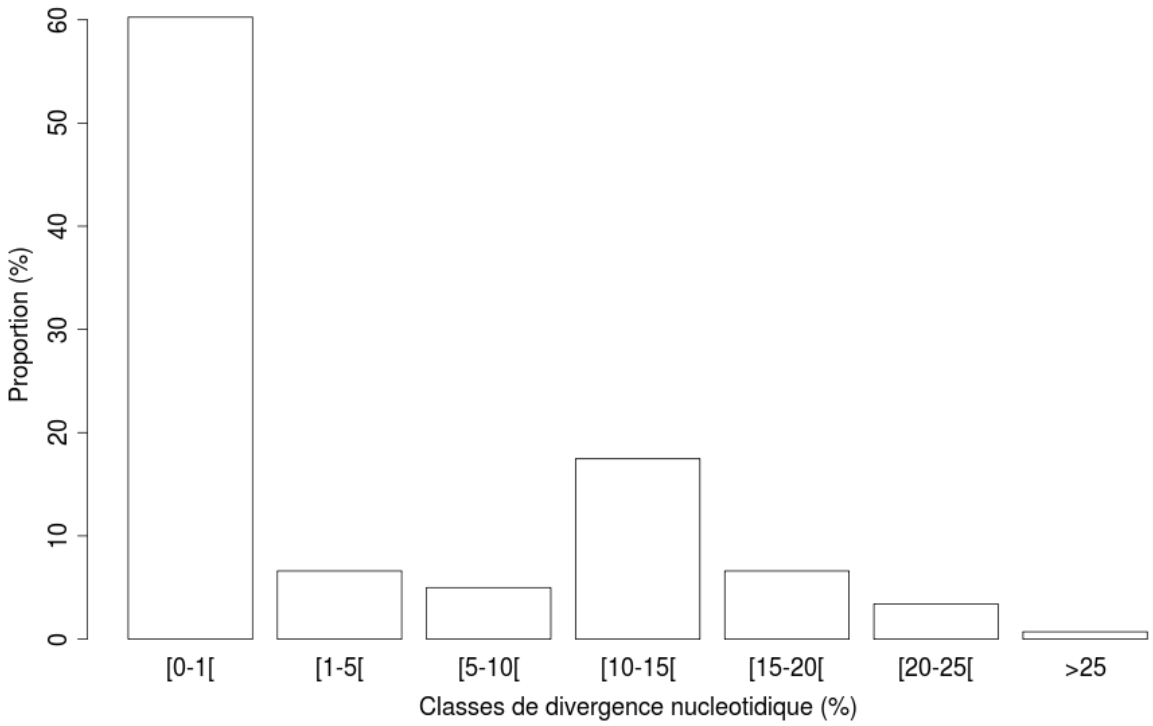

Figure 4-9 : Distribution de la divergence nucléotidique entre paire de séquences obtenues lors de l'analyse de la distribution des 17 groupes d'IS dans les 22 souches de *Wolbachia*.

Les calculs de divergence entre paires de séquences ont été réalisés au sein des 17 groupes d'IS séparément. Les valeurs ont été regroupées en un seul jeu de données.

d'amorces permet l'amplification en PCR de deux types nucléotidiquement différents d'une même séquence, le séquençage direct présente des incertitudes aux niveaux des sites nucléotidiques variant entre les deux types de séquence. Ceci, nous permet de penser que la diversité nucléotidique au sein d'un groupe d'IS est très faible. Les souches possèdent donc soit une copie du groupe considéré soit un grand nombre de copies identiques. De plus, la distribution hétérogène que nous observons suggère des transferts horizontaux fréquents et récents. Cette hypothèse est soutenue par le fait que plus de 60% des comparaisons entre fragments PCR aboutissent à une divergence nucléotidique inférieure à 1% (Figure 4-9). Etant donné que cette divergence est plus faible que la divergence généralement observée entre les souches de *Wolbachia* analysées, les groupes d'IS présents dans différentes souches ne sont vraisemblablement pas directement hérités d'un ancêtre commun. Ainsi, sur la base de la distribution hétérogène et de la faible divergence des groupes d'IS entre souches de *Wolbachia*, on peut penser que les transferts horizontaux sont un facteur déterminant de la distribution des IS dans les souches de *Wolbachia*.

La fréquence des transferts horizontaux pourrait être expliquée par le mode de vie de *Wolbachia*. En effet, Toutes les souches de *Wolbachia* sont intracellulaires strictes. Néanmoins, elles ne sont pas toutes strictement transmises de façon verticale depuis la mère à ses descendants, notamment en ce qui concerne les souches parasites de la reproduction chez les arthropodes. En effet, chez ces dernières, des transferts horizontaux de *Wolbachia* entre hôtes ont pu être identifiés à l'échelle évolutive (Michel-Salzat, et al. 2001; Rigaud and Juchault 1995; Vavre, et al. 1999). Des cas de multi-infections par différentes souches de *Wolbachia* ont été observées (Dedeine, et al. 2004). Ainsi, les cas de transferts horizontaux, de multi-infections d'un hôte et de cohabitations avec d'autres bactéries sont autant de possibilités pour *Wolbachia* d'échanger des portions d'ADN, selon le concept de l'arène

cellulaire (Bordenstein and Wernegreen 2004), et donc d'obtenir de nouveaux IS susceptibles de

Tableau 4-8 : Distribution des copies d'IS insérées dans les prophages de 4 génomes de *Wolbachia*.

Famille d'IS	Groupe d'IS	*w*Mel[a]	*w*Ri[a]	*w*Pel[a]	*w*VulC
IS5	IS1031			1 (0)	
	ISWpi1	1 (1)	3 (3)		
IS66	ISWen3				1 (0)
IS110	ISWen2	1 (1)	2 (2)		
	ISWpi12	1 (0)		1 (1)	
	ISWpi13		1 (1)		1 (0)
	ISWpi14	1 (1)			
IS256	ISWpi15			1 (0)	
IS481	IS481wA				1 (0)
IS630	ISWpi10		1 (0)		
IS982	ISWpi16			1 (1)	
Nombre de copies d'IS		4 (3)	7 (6)	4 (2)	3 (0)
Nombre de prophages		3	4	5	8

[a] Le nombre de copies potentiellement fonctionnelles est indiqué entre parenthèses.

proliférer dans son génome.

Même s'il peut exister une proximité physique entre différentes souches de *Wolbachia* (au moins occasionnellement), la question du transport des IS se pose toujours. Ceci pourrait être facilité par la présence de prophages dans différentes souches de *Wolbachia* qui sont pour certains potentiellement actifs (Bordenstein and Wernegreen 2004; Braquart-Varnier, et al. 2005; Tanaka, et al. 2009) et qui pourraient servir de navettes pour le transport des IS. Nous avons identifié au total 18 copies d'IS, dont 11 potentiellement fonctionnelles, de 11 groupes différents, insérées dans les 20 prophages intégrés aux génomes de *w*Mel, *w*Pel, *w*Ri et *w*VulC (Tableau 4-8, Figure 3-3 page 44). Néanmoins, nous ne pouvons savoir si les bactériophages se sont insérés dans le génome en transportant les copies d'IS ou si les IS se sont insérés après l'arrivée des bactériophages dans le génome. Cependant, une copie d'ISWpi12 potentiellement fonctionnelle a été identifiée dans la séquence du bactériophage libre WOcauB2 de la souche de *Wolbachia w*CauB (Tanaka, et al. 2009), même si les bactériophages contiennent généralement peu d'IS quand ils sont sous forme libre (Leclercq and Cordaux 2011). Il semble que pour *Wolbachia*, la voie de transfert horizontal la plus probable soit les bactériophages car aucun plasmide n'a été caractérisé à ce jour chez ces bactéries.

4.3.6. D'autres sources de transferts?

Pour chercher d'autres sources que *Wolbachia* comme origine des transferts horizontaux d'IS, nous avons comparé par BLASTN les séquences d'IS des 5 génomes que nous avons étudiés à la base de données de NCBI du 1er octobre 2011contenant les séquences non redondantes. La comparaison de la séquence nucléique consensus des 40 groupes d'IS identifiés dans les génomes de *Wolbachia* nous a permis de mettre en évidence que la majorité des groupes présentait peu de similarité avec des séquences déjà connues ou alors une forte

Tableau 4-9 : Recherche de similarité des IS de *Wolbachia* dans la base de données non redondante de NCBI datant du 1er octobre 2011

Famille d'IS	Groupe d'IS	Caractéristique du meilleur résultat		
		Nom de la souche de bactérie	Pourcentage de couverture	Pourcentage d'identité
IS3	IS3wA	*Zunongwangia profunda*	3,92	84,44
	IS3wB	*Escherichia coli* (plasmide)	28,13	67,13
IS4	**ISWosp2**	***Rickettsia akari***	**84,16**	**65,36**
	IS50	/	/	/
	ISWen1	/	/	/
	IS4wA	*Helicobacter hepaticus*	2,49	89,74
	IS4wB	*Methanotorris igneus*	4,58	79,03
	IS4sawA	*Aggregatibacter actinomycetemcomitans*	2,59	96,77
	IS4sawB	*Bacillus thuringiensis*	3,00	91,43
IS5	IS903	*Butyrivibrio fibrisolvens*	3,36	88,89
	IS1031	***Parachlamydia*-related symbiont**	**94,86**	**67,38**
	ISWpi1	*Waddlia chondrophila*	10,26	85,11
	IS427wA	*Legionella pneumophila*	38,42	66,15
	IS427wB	/	/	/
	ISL2wA	***Candidatus Amoebophilus asiaticus***	**69,54**	**68,42**
	ISL2wB	***Orientia tsutsugamushi***	**100,00**	**66,67**
	ISL2wC	*Orientia tsutsugamushi*	40,74	80
IS66	IS6	*Rickettsia peacockii*	12,27	78,22
IS110	IS1111	/	/	/
	ISWen2	*Orientia tsutsugamushi*	2,94	86,36
	ISWpi12	*Rickettsia bellii*	15,53	71,03
	ISWpi13	*Nostoc azollae*	28,24	66,52
	ISWpi14	*Legionella longbeachae*	30,67	68,61
	IS110wA	*Uncultured Desulfobacterium sp;*	7,93	72,83
	IS110wB	/	/	/
	IS110wC	*Mycoplasma mycoides subsp. capri*	5,67	85,19
	IS110wD	*Prochlorococcus marinus*	11,64	79,49
IS200/605	**IS200**	***Thermoanaerobacterium xylanolyticum***	**65,53**	**67,67**
IS256	ISWpi15	*Shewanella putrefaciens 200*	43,30	66,67
IS481	**ISWpi2**	***Rickettsia peacockii***	**85,87**	**67,58**
	ISWpi4	***Nitrosomonas europaea***	**96,44**	**65,23**
	IS481wA	*Rickettsia felis*	8,74	80
IS630	ISWpi11	*Cyanothece sp.*	35,45	70,15
	ISWpi10	*Syntrophobotulus glycolicus*	9,55	73,33
	IS630wA	*Orientia tsutsugamushi*	24,00	70,4
	IS630wB	*Orientia tsutsugamushi*	20,53	72,73
IS982	ISWpi16	*Cyanothece sp.*	11,62	73,68
IS66	ISBst12	*Nostoc punctiforme*	4,21	77,42
	ISWen3	*Brachyspira murdochii*	3,58	81,13
IS1182	IS1182	/	/	/

Les lignes en gras correspondent aux 7 groupes d'IS présentant une forte ressemblance dans des génomes bactéries. Les lignes soulignées correspondent à des souches de bactérie qui sont intracellulaires parmi les génomes présentant une correspondance avec les IS de *Wolbachia*. Le signe "/" signifie qu'aucun match significatif a été obtenu.

similarité mais sur une petite partie de leur séquence (Tableau 4-9). Cependant, 7 groupes d'IS présentaient une similarité de séquence supérieure à 65% sur plus de 50% de leur séquence avec des éléments présents dans d'autres génomes bactériens (Tableau 4-9). Les 7 groupes d'IS de *Wolbachia* présentent des similarités avec des séquences d'IS de 7 bactéries différentes, dont 5 sont des bactéries intracellulaires comme *Wolbachia* (Tableau 4-9).

Parmi les 5 bactéries intracellulaires, *Parachlamydia* et *Candidatus Amoebophilus asiaticus* sont des endosymbiotes d'amibes et les 3 autres (*Rickettsia peacockii*, *R. akari* et *Orientia tsutsugamushi*) sont des Rickettsiales intracellulaires infectant des arthropodes comme *Wolbachia*. Le fait que l'on puisse détecter des IS proches de ceux identifiés chez *Wolbachia* dans le génome de 3 autres bactéries intracellulaires infectant des arthropodes est une parfaite illustration de la théorie de l'arène intracellulaire (Bordenstein and Wernegreen 2004). En effet, il est probable qu'en infectant le même spectre d'hôtes des souches de *Wolbachia* puissent avoir été en contact avec des souches de *Rickettsia* et d'*Orientia*. Les deux autres souches de bactéries intracellulaires (*Candidatus Amoebophilus asiaticus* et *Parachlamydia-related symbiont*) dans lesquelles nous avons détecté des IS proches de ceux de *Wolbachia* sont des symbiotes obligatoires d'amibes. Même dans le cadre de la théorie de l'arène intracellulaire le lien entre symbiotes d'amibe et *Wolbachia* n'est pas évident. Cependant, le séquençage du génome de *Rickettsia. bellii,* une souche de Rickettsiale comme *Wolbachia*, a permis de détecter des gènes très similaires à ceux identifiés chez des symbiotes obligatoires d'amibes (Ogata, et al. 2006). Il a ainsi été proposé que l'ancêtre des Rickettsiales ait utilisé une amibe comme hôte ce qui a pu permettre des échanges de matériel génétique avec les symbiotes d'amibes (Fuxelius, et al. 2007). La présence d'IS dont la séquence est proche de celle d'IS identifiés chez *Wolbachia* dans deux bactéries à vie libre (*Nitrosomonas europaea* et *Thermoanaerobacterium xylanolyticum*) est là encore explicable par l'hypothèse de l'arène intracellulaire mais nécessite un intermédiaire supplémentaire. En effet, il est

possible que *Wolbachia* ait été en contact avec une bactérie intracellulaire facultative avec laquelle elle aurait pu échanger du matériel génétique. Puis cette bactérie intracellulaire facultative pourrait être sortie de ces cellules hôtes et être rentrée en contact avec des bactéries à vie libre avec lesquelles elle aurait échangé du matériel génétique.

4.4. Conclusion

L'importante quantité de copies d'IS dégradées présentes dans les génomes de *Wolbachia* nous a permis de mieux comprendre la dynamique évolutive des IS dans les génomes bactériens. Nous avons étudié la dégradation des IS, qui semble se faire généralement selon un processus d'accumulation concomitante de substitutions et de délétions au cours du temps. La taille des délétions pourrait cependant perturber la relation linéaire qui existe entre l'accumulation des délétions et des substitutions. De plus, l'analyse de la dynamique de transposition des IS nous a permis de mettre en évidence que l'activité de transposition des IS dans les génomes de *Wolbachia* a connue dans le passé des variations avec deux phases importantes d'activité entrecoupées d'une phase de quiescence. La phase d'activité la plus ancienne des IS pourrait avoir commencé au moment de la séparation des supergroupes de diversité A et B de *Wolbachia*, il y a environ entre 20 et 60 millions d'années (Figure 4-4 - page 83) alors que la phase plus récente est peut-être toujours en cours dans les souches *w*Mel, *w*Pel et *w*Ri. Nous avons mis en évidence que les génomes de *w*Mel, *w*Pel et *w*Ri semblaient être en phase dans leur dynamique d'activité des IS alors que le génome de *w*VulC semble être décalé. Sur la base de l'analyse de la divergence entre copies, l'activité des IS du génome de *w*Bm semble être en phase avec celle des génomes de *w*Mel, *w*Ri et *w*Pel. Cependant ce génome ne possède pas de copies potentiellement fonctionnelles et cette apparente synchronicité pourrait être due à l'homogénéisation des séquences provoquée par la

conversion génique. Ainsi, la divergence des copies les plus récentes pourrait être maintenue à un taux faible inférieur à 1%

Dans les génomes de *w*Mel, *w*Pel et *w*Ri, différents groupes d'IS semblent avoir récemment fortement augmenté en nombre de copies. Cette augmentation, qui est indépendante dans les 3 génomes, est apparemment simultanée dans la limite de notre résolution. Cette synchronicité pourrait être l'une des conséquences du grand nombre de transferts horizontaux d'IS dans les génomes de *Wolbachia*. Mais comment expliquer que le génome de *w*VulC ne soit pas synchrone ? Il est possible que la densité déjà très importante d'IS dans le génome de *w*VulC gène la prolifération de nouveaux groupes ou que ce génome soit moins sensible aux transferts horizontaux que les autres génomes.

Quoi qu'il en soit, la réussite évolutive des groupes d'IS est très variable au sein des génomes, ce qui indique que les transferts horizontaux sont nécessaires mais pas suffisants à la prolifération des IS. Cette apparente variation de réussite des groupes d'IS pourrait être liée à différents paramètres intrinsèques des IS ou de leur environnement génomique (Hoffmann and Turelli 1997; Nagy and Chandler 2004).

5. Etude de l'expression des IS du génome de *w*VulC

5.1. Contexte

Dans le chapitre précèdent, nous avons mis en évidence que les IS du génome de *w*VulC ne semblent pas être dans la même phase de dynamique d'activité que ceux des génomes de *w*Mel, *w*Pel et *w*Ri. Nous n'avions cependant pas pu déterminer avec certitude si les IS du génome de *w*VulC se trouvaient au début de la phase d'activité ou à la fin. Afin d'obtenir plus d'informations sur la dynamique d'activité des IS du génome de *w*VulC, nous avons étudié le niveau global de transcription de différents groupes d'IS, ainsi que le niveau de transcription des copies au sein d'un groupe d'IS. Nous avons fait l'hypothèse que si les IS se trouvaient au début d'une phase de forte activité, ils devraient posséder des structures de régulation de l'expression et que leur transcription devrait permettre la production d'ARNm complets. Nous sommes conscients que la transcription ne reflète pas directement l'activité de transposition des IS car d'autres conditions, comme la présence de sites de fixation des ribosomes, doivent être remplies pour permettre la production de transposase. Cependant, la transcription est une condition *sine qua non* à la transposition des IS. De plus, nous ne pouvons cultiver et donc transformer les souches de *Wolbachia*, c'est pourquoi nous n'avons pas pu mettre en place une analyse fonctionnelle classique nécessitant l'utilisation de plasmides pour évaluer le taux de transposition des IS (Feng and Colloms 2007; Tenzen, et al. 1990; Urasaki, et al. 2002).

La présence dans la séquence de nombreuses familles d'ET de structures de régulation de la transcription laisse penser que ces éléments peuvent contrôler leur niveau d'expression. Chez la drosophile, par exemple, une recherche dans les banques d'EST (*expressed sequence tag*) a permis de mettre en évidence que la majorité des familles d'ET sont transcrites, aussi

bien pour les ET de classe 1 que ceux de classe 2 (Deloger, et al. 2009). Il a par ailleurs été démontré, toujours chez la drosophile, que les ET de classe 2 de type *Hobo* étaient exprimés dans les embryons selon un schéma similaire à celui des gènes du développement (Deprá, et al. 2009). Cela suppose que ces ET sont régulés de la même manière que les gènes du développement ou que la propagation des ET de type *Hobo* pourrait avoir contribué à la création d'un nouveau système de régulation de gènes chez la drosophile (Deprá, et al. 2009). La propagation fulgurante de certains ET comme les éléments *P* dans les populations naturelles de drosophiles suppose un niveau de transposition et probablement d'expression très élevé au moment de l'expansion (Kidwell 1983). Ce phénomène de prolifération très rapide d'un ET, souvent peu de temps après son arrivée dans le génome, est nécessaire selon les modèles de dynamique pour expliquer leur maintien dans un génome (Le Rouzic and Capy 2005).

Chez les procaryotes, les modèles d'évolution des ET (Wagner 2006) et les variations très importantes de composition en IS entre des bactéries phylogénétiquement proches laissent penser que les ET sont transpositionnellement actifs au moins à certaines périodes. Cette activité est probablement sous le contrôle des promoteurs internes que les IS possèdent (Nagy and Chandler 2004). Par ailleurs, les promoteurs de certains IS sont inductibles. Par exemple, il a été montré que les spermines, qui sont des facteurs de croissance bactériens, peuvent induire l'expression de deux groupes d'IS de *Francisella tularensis* (Carlson, et al. 2009). Néanmoins, l'activité des IS chez les bactéries est généralement considérée comme maintenue à un niveau très bas. De nombreux mécanismes comme la production d'anti-ARNm (Simons and Kleckner 1988) ou de protéines partielles, via des promoteurs internes ou des décalages de cadre de lecture(Escoubas, et al. 1991; Luthi, et al. 1990), pourraient expliquer le maintien d'une faible activité des IS. De plus, l'impact négatif de l'insertion des IS dans les génomes bactériens laisse penser que l'activité de transposition des ET pourrait être contre sélectionnée

Tableau 5-1 : Liste et caractéristiques des groupes d'IS du génome de *w*VulC choisis pour l'analyse d'expression.

Nom de famille d'IS	Nom de groupe d'IS	Nom de sous-groupe d'IS	Nombre de copies		Divergence	
			Etudiées	Potentiellement fonctionnelles	Intra sous-groupe	Inter sous-groupe
IS3	IS3wB		11	6	0,2%	/
IS4	IS4wA	IS4wA_1	7	5	0%	19,4%
		IS4wA_2	1	1	/	
	ISWosp2	ISWosp2_1	5		0,5%	17,7%
		ISWosp2_2	1		/	
IS5	IS903		4		0	/
	ISL2		3		0,1%	/
IS110	ISWpi13		9		12,51%	/
	ISWen2	ISWen2_1	3	3	0%	15,62%
		ISWen2_2	5		0,1%	
IS256	ISWpi15		10		0,1%	/
IS66	ISWen3		3		0%	/

(Cerveau, et al. 2011a). Etant donné qu'en général plus de 80% des génomes bactériens est composé de séquences codantes (Lawrence, et al. 2001) et que les structures régulatrices d'expression des gènes se trouvent en grande partie dans les régions intergéniques, l'insertion d'une nouvelle copie d'IS a une grande probabilité d'être éliminée par la sélection naturelle. Nous pourrions donc sous-estimer l'activité de transposition des IS et des ET en général.

Ce chapitre présente un travail réalisé sur l'expression de plusieurs groupes d'IS du génome de *w*VulC. Notre objectif était de mieux caractériser la phase d'activité dans laquelle les IS du génome de *w*VulC se situent. Pour cela, nous avons à la fois utilisé une approche bio-informatique pour caractériser les structures de régulation de l'expression des IS, et une approche expérimentale pour étudier l'expression des mêmes IS.

5.2. Méthodes

Nous avons choisi de travailler à deux échelles. D'une part, nous avons fait une analyse globale de la transcription de différents groupes d'IS et d'autre part, nous avons fait une analyse pour chacune des copies du groupe IS4wA_1.

5.2.1. Choix des groupes d'IS

Nous avons sélectionné pour cette étude tous les groupes d'IS qui possèdent au moins une copie potentiellement fonctionnelle ou plus de 3 copies complètes non fonctionnelles, ce qui représente 9 des 35 groupes et 6 des 10 familles d'IS du génome de *w*VulC (Tableau 5-1). La diversité de séquence ou de statut fonctionnel des différentes copies des groupes IS4wA, ISWosp2 et ISWen2 nous ont poussés à séparer les copies de chacun de ces groupes en deux sous-groupes différents pour une meilleure résolution. Les sous-groupes ainsi créés sont composés de copies ayant des séquences homogènes (Tableau 5-1). Néanmoins, pour les

groupes IS3wB et IS4wA_1, nous n'avons pas séparé les copies potentiellement fonctionnelles des non fonctionnelles. En effet, pour le groupe IS3wB, la non fonctionnalité de la transposase de 5 des 11 copies est uniquement due à une insertion de 2 nucléotides. Cette insertion entraîne un décalage de cadre de lecture et donc une perte de fonctionnalité des copies touchées.. Dans le cas du groupe IS4wA_1, les deux copies non fonctionnelles ont une séquence nucléotidique identique aux autres et possèdent une transposase intacte mais elles sont toutes les deux interrompues à leur extrémité 3' par l'insertion d'une copie d'ISWpi15 exactement au même site dans les deux cas. Il est donc possible de faire la différence entre ces deux copies et les autres copies du groupe mais on ne peut pas différencier ces deux copies l'une de l'autre. Le niveau du sous-groupe sera le niveau de référence pour la suite de notre travail dans ce chapitre.

5.2.2. Analyses bio-informatiques

Nous avons cherché deux types de structures, d'une part les promoteurs de type box -35/-10 et d'autre part les terminateurs de transcription Rho indépendant, car elles font parties des structures essentielles dans la régulation de la transcription des gènes.

Afin d'identifier les sites d'initiation de transcription de type box -35/-10, la séquence de chacun des groupes a été analysée avec le logiciel BPROM de la suite SoftBerry (http://linux1.softberry.com/berry.phtml?topic=bprom&group=programs&subgroup=gfindb). BPROM permet d'identifier les séquences promotrices de transcription en amont des ORF ou des opérons avec une spécificité de plus de 80%. De plus, pour le groupe IS4wA_1, groupe choisi pour faire l'analyse de l'activité des copies le composant, 500 pb de séquence flanquant la copie en amont et en aval ont été ajoutées à l'analyse. Ce groupe a été choisi car c'est celui qui possède le plus de copies potentiellement fonctionnelles dans le génome de *w*VulC et le site d'intégration de chaque copie est différent de tous les autres, ce qui permet de les étudier

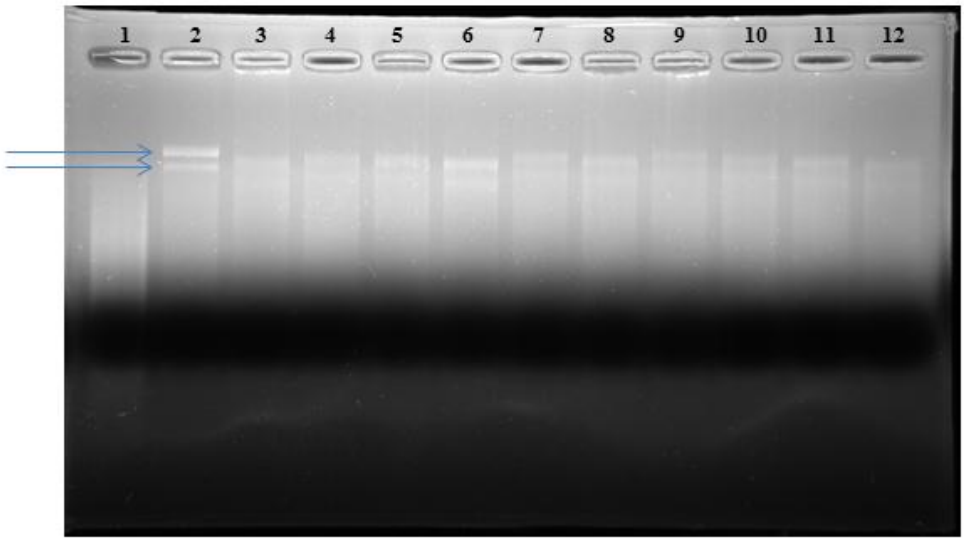

Figure 5-1 : Gel de vérification de la qualité des ARN extraits.

1 µg d'ARN totaux provenant de différentes extractions de pool de 25 ovaires de femelles infectées par *Wolbachia* a été déposé dans chacun des puits de 12 puits. Les deux bandes pointées par les flèches correspondent aux ARN ribosomiques 18S et 28S. La qualité des ARN est évalué visuellement par la présence des bandes d'ARN ribosomiques ou non. Pour la piste ,1 une trainée est très clairement visible et les deux bandes des ARN ribosomiques ne sont pas présentes ce qui signifie que ces ARN sont dégradés. Par la piste 2, les bandes des ARN ribosomiques sont bien visibles et la trainée est très peu visible ce qui signifie que ces ARN sont de bonne qualité. Seuls les ARN de bonne qualité sont conservés pour la suite des expérimentations.

indépendamment.

Afin d'identifier les terminateurs de transcription bactériens les plus courants (type rho) dans les séquences d'IS, une recherche avec le logiciel ARNold a été réalisée (http://rna.igmors.u-psud.fr/toolbox/arnold/index.php). Cet outil fait appel à deux logiciels de recherche de structures que sont ERPIN (Gautheret and Lambert 2001) et RNAmotif (Macke, et al. 2001). Ces deux outils recherchent les boucles caractéristiques des terminateurs de transcription bactériens et ont été éprouvés sur des génomes de référence comme ceux de *Bacillus subtilis* et *E.coli*. Ces deux logiciels ont été utilisés lors d'une précédente étude portant sur 302 génomes bactériens et ont permis de mettre en évidence que les IS contenaient un grand nombre de terminateurs de transcription (Naville and Gautheret 2010).

5.2.3. *Analyse in vivo*

Des ARN déjà extraits par Frédéric Chevalier (doctorant au laboratoire) ont été mis à ma disposition. Les ARN d'ovaires poolés de 25 femelles d'*A.vulgare* infectées par la souche *w*VulC ont été extraits à l'aide du kit RNAeasy Mini kit (QIAGEN) en suivant le protocole « Purification of Total RNA from Animal Tissues ». Les ARN ont été élués dans 50 µL d'eau filtrée à 0,2 µm. La concentration de la solution d'ARN a été mesurée à l'aide d'un Nanodrop (Thermo Scientific) à 260 nm et sa pureté a été estimée grâce au rapport de longueurs d'onde 260/280 nm. La solution d'ARN a ensuite été conservée à −80°C.

L'intégrité des ARN a été vérifiée par électrophorèse sur gel d'agarose 1% (m/v) dans du tampon MOPS 1X (3[N-morpholino]propane sulfonicacide, pH 7,0 / MOPS 5X : MOPS 0,1 M ; acétate de sodium 40 mM ; EDTA 1 mM) auquel sont ajoutés du formaldéhyde (17 %) et 0,57 $\mu g.mL^{-1}$ de bromure d'éthidium (BET) (Figure 5-1). Avant dépôt sur gel, les échantillons sont séchés sous vide (SpeedVac), puis repris dans 10 µL de tampon de charge (formamide 48 % ; formaldéhyde 6,4 % ; bleu de bromophénol 0,25 % ; glycérol 6,6 % dans

Figure 5-2 : Description des amorces utilisées pour réaliser les tests d'expression *in vivo*.

Le rectangle bleu représente la séquence de l'IS qui est encadré par ses TIR représentés en noir. L'IS est lui-même encadré par ses DR en brun et ses séquences flanquantes en vert. Le couple d'amorces rouge a été dessiné afin de détecter l'expression de chacun des groupes d'IS dans sa globalité. Les couples d'amorces bleus et verts ont été dessinés sur chacune des copies du groupe IS4wA_1 pour analyser leur expression.

Tableau 5-2 : Caractéristiques des amorces utilisées pour étudier l'expression des groupes d'IS.

Nom	Séquence de l'amorce	Tm (°C)	Taille du produit (pb)
IS3wB-F	CAGGCTGATTTATTGGGGATT	60	610
IS3wB-R	TTCTCCATAAGCGCTCAACA	60	
IS4wA_1-F	CGGCAAGTGCATTTACTCAA	60	689
IS4wA_1-R	TACTCCCCAGCGCAAGTAAT	60	
IS4wA_2-F	TGGCAGCGTGAAATCTAATG	60	633
IS4wA_2-R	CGATCCATGATGCAGTCTGT	60	
ISWosp2_1-F	GTGGGTTGGGAATTATCTCAAG	60	380
ISWosp2_1-R	TTGTTTATCGTTGGAAAATCTGG	60	
ISWosp2_2-F	CACTGTGTGTGACAGAGAAGCA	60	680
ISWosp2_2-R	GGGCTACATTTGGGCATTTA	60	
IS903-F	ATTGCAATTGCCATAGACAGC	60	390
IS903-R	GCATTCCTTTCAGCCATATAG	60	
ISL2-F	TGGGAGAGAGGTAGAAGAAG	60	440
ISL2-R	CTTCTGGCCTCTGTATTTGC	60	
ISWpi13-F	ACACTCAACGCATGCAAAAG	60	600
ISWpi13-R	CATTTGCAGCACGATTTACG	60	
ISWen2_1-F	ACAGGAACCTTCTCGTTGGA	60	656
ISWen2_1-R	CTTCAATCAACCTCATAGCCC	60	
ISWen2_2-F	CGTTGCTCAGTTGCGTTAAA	60	406
ISWen2_2-R	TGGCTACTATGGCAGGAAAGA	60	
ISWpi15-F	AGGAGGATGGGCATTGTGTA	60	618
ISWpi15-R	CCATGACCCATTTTTGCTCT	60	
ISWen3-F	GGACAAACCAAAAAGCGAAA	60	708
ISWen3-R	TGATAAATGTGCCCAGCAGA	60	

du tampon MOPS 1X). Après migration, les bandes d'ARN sont visualisées par illumination aux UV (Sambrook et al. 1989).

Les rétrotranscriptions produisant des ADN complémentaires (ADNc) (RT+) sont réalisées avec le kit Super-ScriptIII Fist-Strand Synthesis (Invitrogen), à partir de 1 µg d'ARN total à l'aide d'amorces aléatoires (hexamères). Des témoins négatifs (RT−), dans lesquels la rétrotranscriptase est remplacée par de l'eau, sont réalisés dans les mêmes conditions afin de vérifier l'absence de contamination en ADN génomique (ADNg).

Afin d'évaluer le niveau de transcription globale de chacun des groupes, une paire d'amorces internes, localisée dans la mesure du possible dans le gène codant la transposase de l'IS, a été dessinée (Figure 5-2, couple d'amorces rouge ; Tableau 5-2).

Afin d'étudier l'expression de chacune des copies du groupe IS4wA_1, des amorces externes situées dans les régions flanquantes de part et d'autre des différentes copies d'IS ont été dessinées. L'amorce externe dessinée dans la région flanquante en amont ou en aval de la copie d'IS a été utilisée en couple avec une amorce située dans l'IS (Figure 5-2, couples d'amorces bleus et verts ; Tableau 5-3 - page suivante). Deux analyses ont été réalisées. La première analyse permet de savoir si une copie du groupe IS4wA_1 est exprimée au moins dans sa région 3', en utilisant une amorce localisée dans la séquence flanquante en aval de la copie couplée avec une amorce localisée dans la séquence de l'IS. Il n'est pas possible de détecter l'expression spécifique des copies 3 et 7 puisqu'elles sont toutes les deux interrompues dans leur région 3' par une copie d'ISWpi15 et que par conséquent l'amorce externe en aval des copies est identique. La seconde analyse permet de savoir si une copie est transcrite cette fois dans sa région 5'. L'amorce de la région flanquante en amont de l'IS a été couplée avec une amorce localisée dans l'IS. Afin de minimiser la probabilité qu'il y ait des

Tableau 5-3 : Caractéristiques des amorces utilisées pour étudier l'expression des copies du groupe IS4wA_1.

Nom	Séquence de l'amorce	Tm (°C)
IS4wA_e1	ATGAAGGGCAGCTTTCTTTT	58
IS4wA_e4	TGCATTATGGATCGACTATCAC	58
IS4wA_1-copie3-F	AGCAAGAAAGCATCGGCTAA	60
IS4wA_1-copie7-F	AGGTTGGAAAGGGAGCAAAT	60
IS4wA_1-copie3&7-R	TCGCTTTCACCAGACAAGTG	60
IS4wA_1-copie4-F	ATGGCATTTGTCTTGGATGC	60
IS4wA_1-copie4-R	TTGGATCTTTTGAGCCCATC	60
IS4wA_1-copie6-F	AAACGACAGGTTCTCGCAGT	60
IS4wA_1-copie6-R	ATTGGAGAGCAGCAGGAAGA	60
IS4wA_1-copie8-F	GGAAGGAACGGTCAGGAGA	60
IS4wA_1-copie8-R	TCCTTCACAAAACTTTGCTCAT	60
IS4wA_1-copie9-F	AGATAACGGCTTGCAACACC	60
IS4wA_1-copie9-R	TTTTCAGGAATTTTCCACGC	60
IS4wA_1-copie10-F	CGTTGCTCAGTTGCGTTAAA	60
IS4wA_1-copie10-R	CAGCAAATGGTGTCATTCCA	60

Tableau 5-4 : Nombre de promoteurs prédits dans le sens positif pour chaque groupe d'IS étudiés.

Nom du groupe d'IS	Nombre de copies concernées par l'étude	Nombre de promoteurs par copie dans le sens positif
IS3wB*	11	3
IS4wA_1*	7	4
IS4wA_2*	1	4
ISWosp2_1	2	3
ISWosp2_2	1	5
IS903	5	3
ISL2	3	2
ISWpi13	5	4
ISWen2_1*	3	4
ISWen2_2	5	4
ISWpi15	10	4
ISWen3	3	3

* : Groupes d'IS avec au moins une copie potentiellement fonctionnelle.

terminateurs de transcription entre les amorces, elles ont été dessinées après la recherche bio-informatique de ces structures.

Afin de vérifier la spécificité des amorces, chaque paire a été testée sur de l'ADNg de femelle *d'A.vulgare* de la lignée de laboratoire ZN infectée par *Wolbachia*, et qui a été utilisée pour le séquençage du génome de *w*VulC.

Les PCR ont été réalisées selon le protocole décrit dans la section 4-2-5. Les produits obtenus ont ensuite été séquencés directement comme décrit dans la section 4-2-5. Les tests d'expression ont été réalisés sur deux couples de RT+ et RT– issus de deux pools d'ARN différents. Le témoin positif est une extraction d'ADNg de femelle ZN utilisée préalablement pour vérifier la spécificité des couples d'amorces.

5.3. Résultats et discussion

5.3.1. Analyses de l'expression des groupes d'IS

L'analyse de 58 copies d'IS réparties dans 12 groupes a mis en évidence que la séquence de toutes les copies, quel que soit leur statut fonctionnel, contient entre deux et 5 promoteurs potentiels de type Box -35/-10 dans le sens positif (Tableau 5-4). Ces promoteurs sont répartis uniformément tout au long de la séquence. Ceci confirme que les IS sont capables de promouvoir de façon autonome la production de leur ARNm. Néanmoins, parmi les 4 groupes d'IS qui possèdent des copies potentiellement fonctionnelles, seulement un (ISWen2_1) possède un promoteur potentiel localisé en amont de l'ATG initial du gène codant la transposase et qui pourrait permettre de produire un ARNm couvrant completement le gène (Figure 5-3 – page suivante). De plus, le groupe ISWen2_1 possède aussi d'autres promoteurs qui sont localisés plus en aval. Les promoteurs localisés en aval de l'ATG initial,

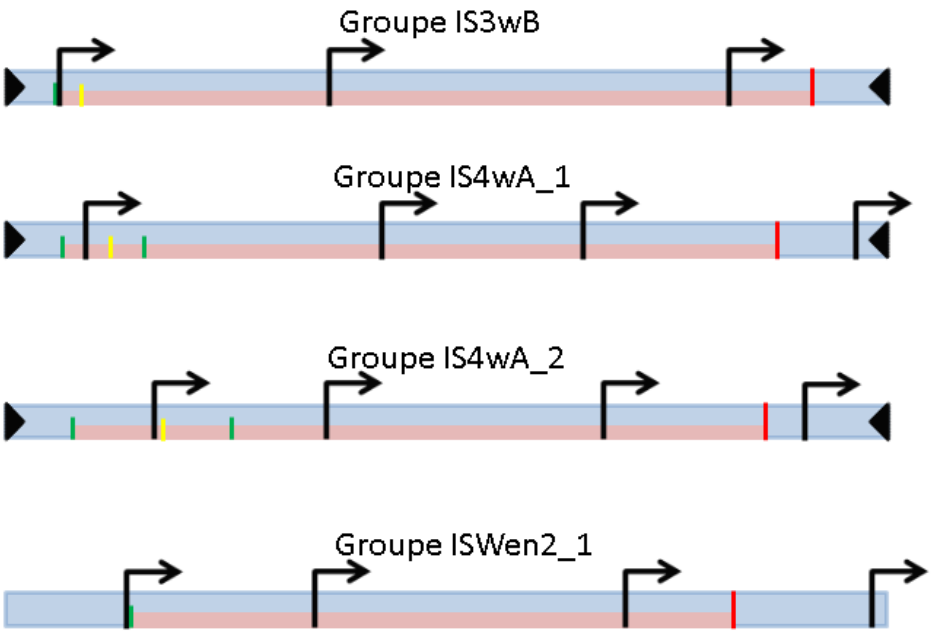

Figure 5-3 : Localisation des promoteurs potentiels des 4 groupes d'IS avec des copies potentiellement fonctionnelles.

Le cadre bleu représente la séquence de l'IS et les triangles noirs aux extrémités symbolisent les TIR lorsqu'ils sont présents. Les traits verts, jaunes et rouges dans la séquence de l'IS représentent respectivement les codons d'initiation ATG, les codons initiation alternatifs et les codons stop. Le cadre rose dans la séquence des IS représente la plus longue protéine qui peut être traduite. Les flèches noires représentent les promoteurs potentiels détectés en sens positif.

s'ils sont actifs, ne pourront induire que la production d'ARNm partiels. Si la transcription se fait dans le même cadre de lecture que que l'ATG initial et qu'ils sont traduits, ces ARNm partiels produiraient des transposases partielles. Par ailleurs, il a été montré que la production de transposases incomplètes dans la région N-terminale peut inhiber l'activité de transposition de la transposase complète (Johnson and Reznikoff 1984). De plus, si les transposases partielles possèdent au moins le site de reconnaissance et de fixation à l'ADN il peut y avoir une compétition entre les transposases complètes et les transposases incomplètes avec pour résultat la limitation de l'activité de transposition de ces groupes d'IS. Les 3 groupes d'IS ne possédant pas d'initiateurs de transcription potentiels avant l'ATG initial peuvent eux aussi produire des ARNm partiels via des promoteurs localisés dans le gène codant la transposase.

L'absence de promoteur en amont du gène codant la transposase de la majorité des copies potentiellement fonctionnelles est surprenante car les IS possèdent généralement leur propre structure de régulation de l'expression (Nagy and Chandler 2004). Cependant, les promoteurs des IS sont faibles et peuvent être difficilement détectable. Cette absence pourrait être due à une dégradation des copies liée à la dérive génétique entraînant l'altération du promoteur. Une telle altération pourrait expliquer la fin des vagues de forte activité de transposition. De plus, nous avons montré que les IS du génome de *w*VulC ne sont pas dans la même phase expansive que ceux des autres génomes de *Wolbachia* infectant des arthropodes. Ils pourraient être en début ou en fin de phase de forte activité. Afin de savoir si l'absence de promoteur identifiable en amont des gènes codant la transposase est une caractéristique spécifique des IS de *w*VulC, nous avons analysé la séquence des copies d'IS potentiellement fonctionnelles présentes dans les autres génomes de *Wolbachia*.

Parmi les 18 groupes d'IS possédant des copies potentiellement fonctionnelles dans les génomes de *w*Mel, *w*Ri et *w*Pel, seuls 6 possèdent un promoteur en amont du gène codant la transposase, permettant donc de produire des ARNm recouvrant la totalité de celui-ci. Ceci

suggère donc que l'absence de promoteur en amont du gène codant la transposase de la majorité des groupes d'IS n'est pas une spécificité du génome de *w*VulC. Cependant, il semble que l'absence de promoteur n'impacte en rien le potentiel de transposition des IS de *Wolbachia*. Nous avons pu montrer que dans les génomes de *w*Pel, *w*Mel et *w*Ri, les groupes d'IS potentiellement fonctionnels ont été probablement actifs très récemment.

Une recherche bio-informatique des terminateurs de transcription de type Rho-indépendant dans les génomes de 302 bactéries a mis en évidence que les IS contenaient très fréquemment ce type de structure et qu'ils pouvaient être considérés comme des atténuateurs de transcription mobiles (Naville and Gautheret 2010). Ces terminateurs de transcription peuvent se trouver soit en aval du gène codant la transposase, et dans ce cas ils n'ont pas d'impact sur la transcription des IS ; soit en amont ou dans le gène codant la transposase, et dans ce cas ils peuvent réduire la transcription et donc l'activité de transposition de l'IS. Lorsqu'un terminateur est localisé dans le gène codant la transposase, il peut y avoir production d'ARNm incomplets avec les mêmes conséquences que décrites précédemment.

Afin d'évaluer l'impact des terminateurs de transcription sur la dynamique des IS du génome de *w*VulC, une recherche bio-informatique des terminateurs de transcription Rho-indépendant a été réalisée à l'aide des logiciels ERPIN et RNAmotif déjà utilisés par (Naville and Gautheret 2010). Aucune des 58 copies d'IS réparties dans les 12 groupes différents ne possède de terminateur de transcription en sens positif. Ces observations suggèrent que si il y a régulation de l'activité de transposition des IS dans le génome de *w*VulC, celle-ci ne se fait pas par des terminateurs de transcription situés dans ou en amont du gène codant la transposase.

Les analyses *in silico* suggèrent que tous les groupes d'IS étudiés peuvent potentiellement être transcrits de façon complète ou partielle. De plus, parmi les groupes ayant des copies potentiellement fonctionnelles, seul un possède un promoteur de

Tableau 5-5: Résultats des RT-PCR (40 cycles) réalisées pour les 12 groupes d'IS

Nom du groupe d'IS	Copie potentiellement fonctionnelle	RT-PCR positive
IS3wB	6	Oui
IS4wA_1	5	Oui
IS4wA_2	1	Non
ISWosp2_1	0	Oui
ISWosp2_2	0	Oui
IS903	0	Non
ISL2	0	Oui
ISWpi13	0	Non
ISWen2_1	3	Oui
ISWen2_2	0	Oui
ISWpi15	0	non
ISWen3	0	Oui

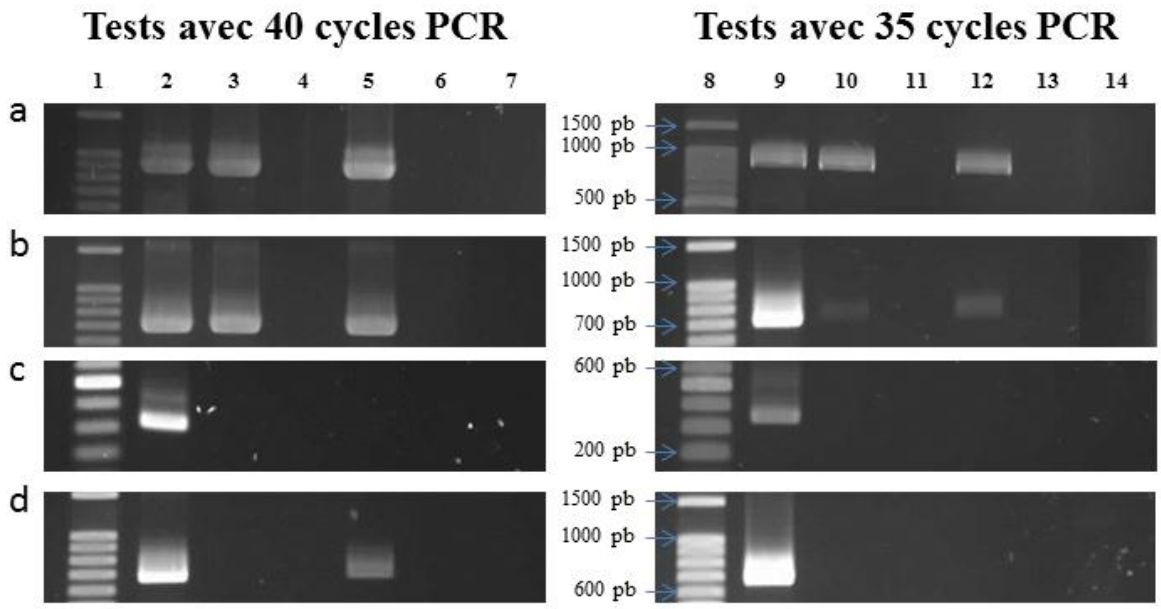

Figure 5-4 : Détection par PCR de l'expression de groupes d'IS présents dans le génome de *w*VulC.

Les amorces utilisées pour la détection de l'expression des différents groupes d'IS ont été dessinées à partir des copies présentes dans le génome de *w*VulC. Les tests PCR ont été réalisés selon deux conditions différentes: 40 et 35 cycles PCR. La ligne (a) correspond au gène *ARNr 16S* de *Wolbachia* qui est notre gène témoin. La ligne (b) correspond au groupe IS4wA_1, la (c) correspond au groupe IS903 et la (d) au groupe ISWen2_1. Les pistes 1 et 8 correspondent au marqueur de taille BenchTop (Promega). Les pistes 2 et 9 correspondent au témoin positif de chaque PCR qui est une extraction d'ADNg de femelle infectée par la souche de *Wolbachia w*VulC. Les pistes 7 et 14 correspondent au témoin négatif. Les pistes 3, 4 et 10, 11 correspondent à la première paire de RT+ et -. Les pistes 5, 6 et 12, 13 correspondent à la seconde paire de RT+ et -.

transcription situé en amont de l'ATG initial du gène codant la transposase. Ceci aurait pu remettre en cause la définition que nous donnons des copies potentiellement fonctionnelles. Néanmoins, l'analyse des copies d'IS des génomes de *w*Pel, *w*Mel et *w*Ri suggère que malgré l'absence pour certains groupes de promoteur interne à l'IS en amont du gène codant la transposase, le potentiel de transposition des groupes n'est pas altéré. Il est possible que le potentiel de transcription soit maintenu par des promoteurs externes. Afin de compléter notre analyse des IS du génome de *w*VulC, nous avons étudié la transcription des différents groupes précédemment utilisés dans l'analyse *in silico*.

La série de tests d'expression réalisée sur les 12 groupes d'IS définis précédemment avec 40 cycles d'amplification PCR nous a permis de détecter l'expression de 8 groupes d'IS (Tableau 5-5). Cependant lorsque le nombre de cycles PCR a été diminué à 35, nous n'avons détecté l'expression que du groupe IS4wA_1 sans que la détection de l'expression du gène témoin (*ARNr 16S* de *Wolbachia*), n'en soit modifiée (Figure 5-4). De plus, pour le groupe ISWen2_1, l'expression n'a été détectée que pour une seule des deux RT+ à 40 cycles de PCR. Les amplifiats obtenus ont été séquencés, ce qui a permis de confirmer la spécificité de tous les signaux positifs. La forte diminution du nombre de groupes d'IS pour lesquels nous avons détecté de l'expression en RT-PCR lors du passage de 40 à 35 cycles, alors même que la détection d'expression de notre gène témoin n'est pas modifiée, laisse penser que les groupes d'IS sont peu ou pas exprimés dans le génome de *w*VulC.

Parmi les 8 groupes d'IS exprimés, 3 possèdent entre 60 et 100% de copies fonctionnelles et 5 uniquement des copies non fonctionnelles (Tableau 5-5). Une étude récente a montré que tous les groupes d'IS possédant au moins 3 copies d'IS intacts présents dans le génome de la bactérie *Amoebophilus asiaticus* étaient exprimés (Schmitz-Esser, et al. 2011). Cependant, l''expression des groupes d'IS ne semble pas être conditionnée par le statut fonctionnel des copies qui les composent. L'absence de signal pour 4 des groupes d'IS étudiés

peut avoir différentes causes : i) un problème de détection de certains terminateurs de transcription, ii) une sensibilité insuffisante de la détection par PCR et iii) une expression maintenue à un niveau très faible voire nul par un répresseur de transcription. Ce type de structure a l'effet inverse du promoteur de transcription : lorsque le facteur de répression est présent il y a inhibition de la transcription soit parce que la polymérase ne peut pas se fixer à l'ADN soit parce qu'elle ne peut pas faire l'élongation de l'ARNm. Ce répresseur peut être une transposase partielle issue de l'IS lui-même (Escoubas, et al. 1991).

En conclusion, nous avons montré que tous les groupes d'IS possèdent des promoteurs de transcription potentiels. De plus, nous avons montré que la majorité des groupes d'IS étaient exprimés mais probablement à un faible niveau. Pour compléter notre analyse, nous avons voulu savoir si, pour les groupes d'IS exprimés, toutes les copies participaient à l'expression.

5.3.2. Analyse de l'expression de chaque copie d'un groupe d'IS

Afin de mener l'analyse concernant l'expression individuelle de différentes copies d'IS au sein d'un groupe, nous nous sommes focalisés sur le groupe IS4wA_1. Ce groupe se compose de 7 copies (identifiant: 3, 4, 6, 7, 8, 9 et 10), dont 5 sont potentiellement fonctionnelles et deux sont non fonctionnelles (identifiant: 3 et 7) car interrompues dans leur partie 3' chacune par l'insertion d'une copie d'ISWpi15. Cependant ces deux copies possèdent un gène codant une transposase de la même taille que celui des copies complètes.

Malgré l'absence de promoteur interne en amont de l'ATG initial pour le groupe IS4wA_1 ayant des copies potentiellement fonctionnelles, des ARNm couvrant la totalité du gène codant la transposase pourraient tout de même être produits via des promoteurs situés

Tableau 5-6: Résultats des RT-PCR réalisées pour les 7 copies du groupe IS4wA_1

Nom de la copie	Partie 5' de l'IS exprimée	Partie 3' de l'IS exprimée
IS4wA_1-copie 4	Oui	Non
IS4wA_1-copie 6	Non	Non
IS4wA_1-copie 8	Oui	Non
IS4wA_1-copie 9	Oui	Non
IS4wA_1-copie 10	Oui	Oui
IS4wA_1-copie 3	Non	Oui[a]
IS4wA_1-copie 7	Oui	

[a] l'expression de la région 3' des copies 3 et 7 ne peut pas être spécifiquement étudiée car l'amorce dessinée dans la région flanquante est située dans un ISWpi15 inséré exactement au même point dans les deux copies.

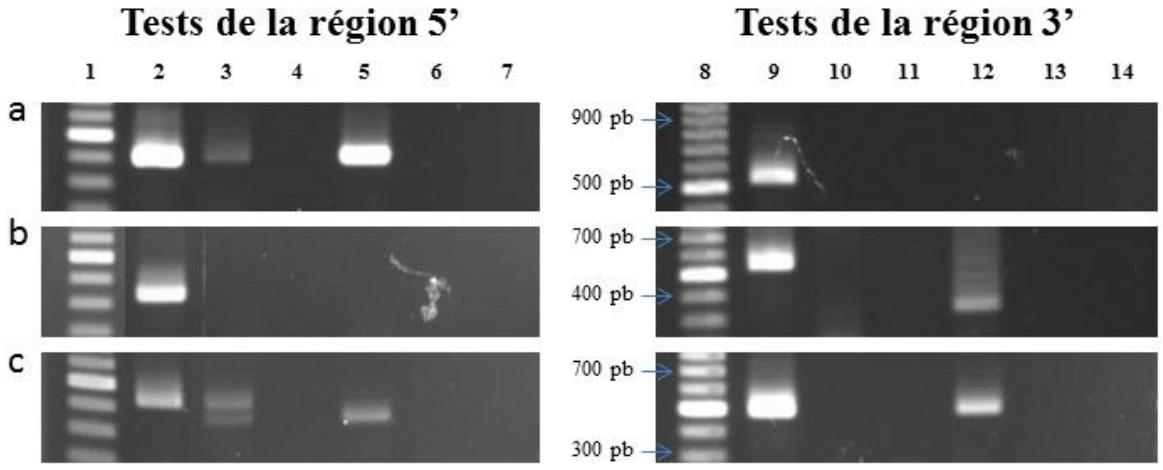

Figure 5-5 : Détection par PCR de l'expression de copies du groupes IS4wA_1 présentes dans le génome de *w*VulC.

Les amorces utilisées pour la détection de l'expression des différents copies du groupe IS4wA_1 ont été dessinées à partir des copies présentes dans le génome de *w*VulC. La ligne (a) correspond à la copie 4, la (c) correspond à la copie 6 et la (d) à la copie 10. Les pistes 1 et 8 correspondent au marqueur de taille BenchTop (Promega). Les pistes 2 et 9 correspondent au témoin positif qui est une extraction d'ADNg de femelle infectée par la souche de *Wolbachia w*VulC. Les pistes 7 et 14 correspondent au témoin négatif. Les pistes 3, 4 et 10, 11 correspondent à la première paire de RT+ et -. Les pistes 5, 6 et 12, 13 correspondent à la seconde paire de RT+ et -.

dans les séquences flanquantes en amont des copies. Pour tester cette hypothèse, une recherche de promoteur a été réalisée dans les 500 pb flanquantes en amont de chaque copie d'IS du groupe IS4wA_1. Pour chaque copie, au moins un promoteur potentiel est prédit entre 36 et 334 pb en amont du début l'IS. La transcription des différentes copies peut donc être initiée en principe par des promoteurs externes.

Nous avons fait une recherche des terminateurs de transcription Rho-indépendants dans les régions flanquantes de chaque copie du groupe IS4wA_1. Aucun terminateur de transcription n'a été détecté dans les 500 pb de séquence flanquante en amont des différentes copies en sens positif. Etant donné que nous n'avons pas détecté de terminateur de transcription Rho indépendant dans la séquence du groupe IS4wA_1, les copies peuvent être transcrites intégralement si l'initiation de la transcription se fait via un promoteur localisé en amont des copies. Afin d'appuyer les observations faites *in silico,* des analyses de transcription ont été réalisées pour chaque copie du groupe IS4wA_1.

Afin d'identifier quelles copies du groupe IS4wA_1 sont transcrites, une paire d'amorces a été dessinée pour chaque copie dans la zone à la jonction entre la partie 5' de l'IS et la région flanquante en amont (Figure 5-1 page 106 couple d'amorces bleu). Un signal d'expression a été détecté pour 5 des 7 copies (Tableau 5-6 et Figure 5-5). De plus, une paire d'amorces a été dessinée dans la zone à la jonction entre la partie 3' de chaque copie d'IS et sa séquence flanquante en aval (Figure 5-1 page 106 couple d'amorces vert). Nous avons pu mettre en évidence que la région 3' de l'IS ainsi que la séquence flanquante en aval de 2 ou 3 des 7 copies étaient transcrites (Tableau 5-6). Il existe une incertitude pour les copies 3 et 7 car le couple d'amorces qu'ils partagent a produit un signal positif en PCR. Toutefois, nous ne savons pas si l'une seulement ou les deux copies sont transcrites. Au total, au maximum 2 copies sur les 7 étudiées sont exprimées à la fois dans leur région 5' et dans leur région 3' (Tableau 5-6). Pour savoir si ces 2 copies étaient exprimées entièrement nous avons fait une

PCR en utilisant les amorces situées en amont et en aval de chacune des copies. Nous n'avons obtenu de signal permettant de dire que les copies sont exprimées entièrement. Cependant nous n'avions pas de contrôle vérifiant que nous pouvons amplifier des ADNc bactériens de la même taille que ceux des IS.

Ces résultats suggèrent que l'expression du groupe IS4wA_1 détectée dans l'analyse par groupes (voir section 5.3.1) est en fait due à l'expression d'un nombre limité de copies de ce groupe. Ces résultats sont cohérents avec le fait que seules 15 à 55% des copies ont apparemment participé à l'expansion récente du groupe ISWen2 présent dans le génome de *w*Ri (voir section 4.3.3). Pour Les 2 ou 3 qui sont transcrites à la fois dans leur région 5' et dans leur région 3', il est donc possible que la transcription soit initiée via un promoteur localisé dans la région flanquante en amont de la copie. Pour les copies transcrites dont la région 5'est transcrite mais pas dans leur région 3', il est possible que la transcription ait été initiée dans l'IS à partir d'un promoteur localisé dans le sens négatif (Annexe 4).

En effet, en plus des promoteurs localisés dans la séquence en sens positif des différents groupes d'IS de *w*VulC, entre 2 et 5 promoteurs ont été détectés en sens négatif (Annexe 4). Ces promoteurs pourraient produire des anti-ARNm, qui pourraient réguler l'activité de transposition des groupes d'IS (Nagy and Chandler 2004). En effet, il a été montré que les anti-ARNm pouvaient interagir avec les ARNm et bloquer la fixation des ribosomes (Simons and Kleckner 1988). De plus, des enzymes comme les RNase III, reconnaissant et détruisant les ARN double brin, ont notamment été identifiées chez *B.subtilis* (Arraiano, et al. 2010). Elles pourraient détruire les complexes d'ARN double brin qui résulteraient de la transcription des IS par les promoteurs situés en sens positif et négatif. Étant donné que nous avons synthétisé les ADNc avec des amorces aléatoires nous ne pouvons pas savoir si l'expression que nous détectons est liée uniquement à des ARNm, à des anti-ARNm ou à un mélange des deux. Pour avoir cette information, il faudrait dessiner une

amorce spécifique des ARNm, une amorce spécifique des anti-ARNm et faire une synthèse d'ADNc avec chacune d'entre elles indépendamment. Néanmoins, on peut supposer que les promoteurs en sens positif et en sens négatif ont la capacité d'activer la synthèse d'ARNm. Il est donc possible que l'on détecte un mélange des deux formes.

5.4. Conclusion

Les analyses *in silico* réalisées ont permis de confirmer que tous les groupes d'IS étudiés possèdent les structures promotrices leur permettant d'être transcrits. Néanmoins, pour 3 des 4 groupes possédant des copies potentiellement fonctionnelles, le premier promoteur potentiel est localisé en aval de l'ATG initial du gène codant la transposase. Cependant, la recherche de promoteurs dans les séquences des copies d'IS potentiellement fonctionnelles des génomes de *w*Mel, *w*Ri et *w*Pel montre là aussi que la majorité des groupes d'IS ne possède pas de promoteur en amont de l'ATG initial du gène codant la transposase. Malgré cela, de nombreux signes d'activité récente des IS ont été relevés dans ces génomes.

De plus, même si un groupe d'IS de *w*VulC possède un promoteur en amont de l'ATG initial du gène codant la transposase, il en possède également plusieurs en aval. Il est donc possible que des ARNm ne couvrant pas la totalité du gène codant la transposase soient produits, comme décrit pour le groupe IS50 (Johnson and Reznikoff 1984). Des promoteurs sont aussi localisés dans le sens négatif des séquences des groupes d'IS étudiés, ce qui peut entraîner la production d'anti-ARN. Ces deux phénomènes, que sont la production d'ARNm tronqués et d'anti-ARN, sont des mécanismes décrits ponctuellement comme pouvant réguler l'activité de transposition des IS (Johnson and Reznikoff 1984; Simons and Kleckner 1988).

Contrairement à ce qui avait été décrit dans un échantillon de 302 génomes bactériens (Naville and Gautheret 2010), les IS de *Wolbachia* ne possèdent pas dans leur séquence de terminateur de transcription Rho-indépendant. Ainsi, la production d'ARNm tronqués par un

terminateur de transcription présent dans le gène codant la transposase n'est probablement pas un mécanisme de régulation de l'activité des IS des génomes de *Wolbachia*.

Les analyses expérimentales ont mis en évidence que, tout comme dans le génome de la drosophile (Deloger, et al. 2009), tous les groupes d'IS de *Wolbachia* ne sont pas transcrits malgré la présence de nombreux promoteurs de transcription dans leurs séquences. La transcription des groupes n'est pas liée au statut fonctionnel des copies qui les composent. En effet, des groupes avec des copies d'IS fonctionnelles ne sont pas transcrits alors que d'autres avec des copies d'IS non fonctionnelles le sont. De plus, les groupes transcrits semblent l'être très faiblement, ce qui confirme que l'activité des promoteurs endogènes des IS est probablement faible (Dalrymple and Arber 1985; Reimmann, et al. 1989). Néanmoins, il est possible que la transcription complète du gène codant la transposase puisse être initiée *via* des promoteurs externes. Il reste à savoir si cette transcription « passive » peut permettre la production d'ARNm complets.

Le faible niveau de transcription des IS que nous avons détecté chez *w*VulC ne semble pas en accord avec les modèles d'invasion rapide des IS dans les génomes bactériens (Wagner 2006). Néanmoins, il est possible que certains groupes soient en phase d'expansion alors que d'autres non, ce qui pourrait expliquer le fait nous n'avons détecté que la transcription du groupe IS4wA_1 à 35 cycles PCR. L'hypothèse que tous les groupes d'IS ne se trouve pas dans la même phase d'activité est confirmée par le fait que le groupe IS4wA_1 est celui qui possède le plus de copies potentiellement fonctionnelles dans le génome de *w*VulC. Le génome de *w*VulC est le génome de *Wolbachia* d'arthropode présentant le moins de copies potentiellement fonctionnelles, ce qui pourrait indiquer qu'il se trouve soit en début d'une phase d'expansion d'IS. Il est possible que le niveau de transcription des IS soit régulé de différentes manières et que, par conséquent, il puisse varier au cours du temps. Nous savons par exemple que certains IS sont capables de synchroniser leur activité de transposition avec

la réplication du génome de l'hôte, probablement dans le but d'augmenter plus efficacement leur nombre de copies (Ton-Hoang, et al. 2010). De plus, il est possible que l'absence de promoteur en amont de l'ATG initial de la plupart des groupes d'IS de *w*VulC ayant des copies potentiellement fonctionnelles altère leurs capacités de transcription. Cependant, l'étude des copies d'IS potentiellement fonctionnelles des 3 autres génomes de *Wolbachia* qui en contiennent ne permet pas de confirmer cette hypothèse.

Au vu des résultats de cette étude sur la transcription des IS du génome de *w*VulC, la phase d'activité dans laquelle se situent les IS de *w*VulC n'est pas clairement identifiable. Cependant, la réponse n'est probablement pas unique si l'on considère les groupes d'IS séparément. Par exemple, les groupes d'IS contenant uniquement des copies non fonctionnelles qui sont transcrites pourraient avoir été actifs dans le passé, les copies ayant été ensuite inactivées par dérive génétique. De plus, il est possible que le groupe IS4wA_1 qui semble présenter un niveau de transcription plus important que tous les autres soit dans une phase d'expansion. Pour confirmer cette hypothèse, il serait intéressant de quantifier l'expression des différents groupes d'IS. Il serait aussi intéressant de comparer le niveau d'expression des IS avec celui de gènes de *Wolbachia* tout en considérant le cycle de vie de l'hôte et des tissus dans lesquels se trouvent *Wolbachia*.

6. Discussion générale

Du fait de leur forte densité en IS (Moran and Plague 2004; Wu, et al. 2004), de leur faible taille de génome (Foster, et al. 2005; Klasson, et al. 2008; Klasson, et al. 2009; Wernegreen 2002; Wu, et al. 2004) et de la diversité des souches, *Wolbachia* apparait comme un outil particulièrement adapté pour étudier la dynamique des ET bactériens. Il a été montré que des IS de *Wolbachia* avaient probablement été actifs dans certaines souches dans un passé récent (Duron, et al. 2005) et qu'il était possible de retracer l'histoire évolutive récente de ces éléments (Cordaux 2008; Cordaux, et al. 2008). Dans ce contexte, les objectifs de ma thèse étaient d'annoter de façon exhaustive les IS des génomes de *Wolbachia* disponibles afin d'étudier plus largement leur dynamique évolutive.

L'annotation et l'analyse des 5 génomes à notre disposition nous a permis de confirmer que les génomes de *Wolbachia* faisaient bien partie des génomes bactériens dans lesquels la densité en IS est la plus importante. De plus, nous avons montré l'importance des MGE et plus particulièrement des IS dans l'évolution de la taille des génomes de *Wolbachia*. Enfin, l'analyse de la structure des copies a mis en évidence que la majorité, et même la totalité chez *w*Bm, des copies d'IS présentes dans les génomes de *Wolbachia* étaient non fonctionnelles (Cordaux 2009). Il est intéressant de noter que *w*Bm, qui est la seule souche mutualiste de *Wolbachia* de notre jeu de données, possède le moins de copies d'IS avec aucune copie potentiellement fonctionnelle. De plus cette souche ne contient aucune copie d'intron de groupe II et aucun prophage. Chez les nématodes, la transmission uniquement verticale de *Wolbachia* n'offre que peu d'opportunités d'acquisition d'ADN par transferts horizontaux. On peut donc supposer que l'évolution génomique des souches de *Wolbachia* à transmission strictement verticale est analogue à celle des endosymbiotes mutualistes nutritionnels tels que *Buchnera aphidicola* chez les pucerons.

Le grand nombre de copies non fonctionnelles et/ou dégradées chez *Wolbachia* va à l'encontre de l'hypothèse conventionnelle d'origine récente des ET dans les génomes bactériens (Wagner 2006; Wagner, et al. 2007). Cette abondance de copies dégradées pourrait être liée au mode de vie intracellulaire de *Wolbachia* qui entraîne une diminution de l'efficacité de sélection et une augmentation de l'effet de la dérive génétique, entraînant la modification des processus d'élimination des copies d'IS. Étant donné que peu de génomes bactériens ont été annotés avec le même effort que celui que nous avons déployé pour *Wolbachia*, nous ne pouvons savoir si ces copies dégradées sont une caractéristique des génomes de *Wolbachia* ou si elles sont présentes dans d'autres génomes bactériens sans avoir été détectées par les méthodes d'annotation classiques. Cependant, le développement de nouveaux outils d'annotation, tels qu'ISsaga, devrait permettre une meilleur détection des copies dégradées.

Quoi qu'il en soit, l'abondance de copies d'IS non fonctionnelles et/ou dégradées nous a donné l'opportunité d'étudier empiriquement la dynamique évolutive des IS de *Wolbachia* et les mécanismes qui entraînent leur dégradation. Nous avons ainsi montré que les copies d'IS étaient dégradées principalement par une lente accumulation de substitutions et de petites délétions au cours du temps. Néanmoins, une variante a été observée et consiste en une réduction rapide de la taille de certains éléments par des délétions de grande taille, probablement via un mécanisme de réparation des cassures de l'ADN.

L'étude de la dynamique passée des IS dans les génomes de *Wolbachia* nous a permis de conclure que l'activité des IS n'a pas été constante au cours de l'histoire évolutive de *Wolbachia*. En effet, nous avons observé une distribution bimodale de la divergence nucléotidique et de la taille relative des copies, ce qui suggère une forte phase d'activité de transposition passée et une plus récente, séparées par une phase de « quiescence ». Globalement, ce schéma a été observé dans tous les génomes de *Wolbachia* avec toutefois une

petite nuance dans le génome de *w*VulC. En effet, les distributions pour *w*VulC sont toujours bimodales mais le schéma global des distributions est différent de celui observé dans les autres génomes. De plus, le schéma d'activité des IS du génome de *w*Bm semble être similaire à celui des génomes de *w*Pel, *w*Mel et *w*Ri. Cependant, nous savons que les IS de *w*Bm n'ont pas eu de phase d'activité récente car aucune copie n'est fonctionnelle. On peut donc penser que la dynamique évolutive des IS chez *w*Bm est influencée par la conversion génique qui tend à homogénéiser les séquences nucléotidiques des éléments répétés (Cordaux 2009).

L'analyse de la distribution des groupes en fonction de la taille relative des copies et la recherche de copies d'IS insérées dans des sites orthologues dans différents génomes ont permis de proposer un scénario concernant les phases d'activités de transposition au cours de l'évolution des souches de *Wolbachia*. Il est ainsi possible que la phase d'activité ancienne des IS ait débuté à une période proche de la séparation des souches des supergroupes A (*w*Mel/*w*Ri) et B (*w*Pel/*w*VulC) et qu'elle ait continuée indépendamment dans ces deux lignées. Nous avons ainsi suggéré que la séparation entre les souches *w*VulC et *w*Pel (supergroupe B) s'est passée durant cette phase d'activité ancienne. De plus, nous avons montré que la phase d'activité récente s'est déroulée de façon indépendante dans les génomes de *Wolbachia*. Elle aurait commencé peu avant la divergence entre les souches *w*Mel et *w*Ri. Nous avons donc pu observer que les deux phases d'activité des IS pourraient être concomitantes avec la séparation de différentes lignées et souches de *Wolbachia*. Les phases d'activité d'IS et la diversification des souches de *Wolbachia* sont peut être survenus indépendamment l'une de l'autre. Cependant, l'augmentation du nombre de copies par transposition peut entraîner une augmentation de la variabilité intra souche. L'augmentation de la variabilité entre les souches finirait par provoquer un évènement de diversification. Chez la drosophile la dysgénésie des hybride qui est un système d'altération de la reproduction dû à des ET pourrait aboutir à la mise en place d'un isolement reproducteur (Bingham, et al. 1982).

En effet, dans les deux systèmes connus I-R et P-M, il existe des lignées P et I qui contiennent des ET du même nom et des lignées dites sensibles (R et M) qui permettent l'explosion de l'activité de transposition des ET (Bucheton, et al. 1984). Lorsqu'il y a croisement entre un individu de la lignée sensible et un individu de la lignée contenant les ET, il y a une forte augmentation du nombre de copies d'ET car le système de régulation des ET des lignées (R et M) n'est pas efficace. Ceci entraîne une diminution de la fertilité, l'augmentation des mutations et des aberrations chromosomiques à cause de l'activité de transposition des ET, ce qui peut mettre en place une barrière de reproduction et donc un phénomène de spéciation (Bingham, et al. 1982)

Cependant, il est aussi possible que la phase de diversification de souche permet l'augmentation du nombre de copies d'IS, notamment à cause du goulot d'étranglement au moment de la diversification qui entraîne une diminution de l'effet de la sélection et une augmentation de l'effet de la dérive génétique. Les deux phases de forte activité des IS pourraient ainsi être concomitantes de la divergence entre les supergroupes A et B ainsi que de la radiation récente des souches de *Wolbachia* au sein du supergroupe A (Werren, et al. 1995). Il est admis que les souches de *Wolbachia* du supergroupe A sont moins divergentes les unes des autres que celles de supergroupe B. Ceci suppose une diversification plus récente des souches du supergroupe A, ce qui est parfaitement en accord avec l'intense activité du groupe ISWpi1 observée dans les souches du supergroupe A (Cordaux, et al. 2008). Concernant la souche *w*Pel, elle a subi une diversification récente (Atyame, et al. 2011), ce qui pourrait avoir permis une augmentation de l'activité de transposition. La coïncidence entre vagues d'activité de transposition et évènements de spéciation du génome hôte a par ailleurs été déjà observée chez les salmonidés (De Boer, et al. 2007). Il a été proposé que l'effet de fondation induisant des goulots d'étranglements et la réduction de la taille de population pouvait entraîner une augmentation de l'activité des ET (Hedges et al., 2004 ; Cordaux and

Batzer. 2006), ce qui est plutôt en faveur de l'hypothèse d'une expansion d'ET facilitée par la diversification de souches.

Afin de tester la relation entre diversification des souches de *Wolbachia* et dynamique des IS, il serait intéressant d'influencer expérimentalement l'évolution des souches de *Wolbachia* pour en obtenir de nouvelles souches et d'observer la modification d'abondance en IS. Une autre possibilité pour tester cette relation serait de séquencer les génomes de différentes souches de *Wolbachia* issues d'un même supergroupe mais ayant des degrés de divergence variables. Nous pourrions ainsi caractériser les populations d'IS des différents génomes et comparer leur dynamique évolutive afin de voir si elle est cohérente avec l'hypothèse du lien entre diversification et phase d'activité d'IS.

La mise en évidence d'une succession de phases d'activité de transposition d'IS dans les génomes de *Wolbachia* pose la question de leur initiation et de leur arrêt. Concernant l'initiation des phases d'activité, il est fort probable qu'elles soient dues à l'arrivée de nouvelles copies d'IS par des transferts horizontaux. En effet, il a été montré que les transferts horizontaux sont nécessaires au maintien des IS dans les génomes bactériens (Wagner 2006; Wagner, et al. 2007). Nous avons aussi pu montrer expérimentalement que les génomes de *Wolbachia* avaient subi de nombreux transferts horizontaux récents d'IS. Ces transferts horizontaux récents pourraient être à l'origine de la phase d'activité récente dans les génomes de *Wolbachia*. Ainsi, le décalage de phase d'activité de transposition d'IS suspecté dans le génome de *w*VulC pourrait être lié à une absence de transferts horizontaux récents ou à une fréquence plus faible que dans les autres souches de *Wolbachia* séquencées.

Contrairement à ce qui a été observé chez des insectes (Vavre et al., 1999 ; (Narita, et al. 2007)), aucun cas de multi-infections par *Wolbachia* au sein d'un même individu hôte n'a été décrit dans la littérature chez les isopodes terrestres, même si on sait que plusieurs souches peuvent coexister dans une même espèce (Bouchon, et al. 2008; Cordaux, et al. 2004b;

Michel-Salzat, et al. 2001). De plus, des simulations de dynamique des populations prédisent que les multi-infections intra-individuelles de *Wolbachia* sont très instables dans le cas de souches féminisantes telles que *w*VulC (Yves Caubet, communication personnelle). Ces deux arguments suggèrent donc que les co-infections par des souches de *Wolbachia* féminisantes sont au plus transitoires dans la nature (Verne, et al. 2007). Enfin, des transferts horizontaux expérimentaux de souches de *Wolbachia* entre espèces d'isopodes terrestres, qui sont les principaux hôtes des souches féminisantes, ont montré que plus les espèces hôtes sont éloignées phylogénétiquement, plus la probabilité de réussite d'un transfert de *Wolbachia* est faible (Bouchon, et al. 2008). Cela suggère une forte spécificité des relations *Wolbachia*-hôte isopode, réduisant encore les possibilités de multi-infections stables dans le temps. Ainsi, l'une des clés de la présence des IS chez *Wolbachia* et de leur dynamique évolutive semble être l'abondance de transferts horizontaux. De plus, ce sont encore les transferts horizontaux qui pourraient expliquer le décalage de phase d'activité des IS observé dans le génome de *w*VulC. En effet, l'effet féminisant de la souche *w*VulC pourrait expliquer une diminution de sa capacité à cohabiter avec d'autres souches de *Wolbachia* au sein d'un même hôte isopode diminuant les possibilités d'échanger de matériel génétique.

L'analyse de l'expression des IS de *w*VulC ne nous a pas permis de clairement identifier la phase d'activité dans laquelle ils se trouvent. Nous avons confirmé que tous les groupes d'IS possédaient dans leur séquence des promoteurs de transcription. Cependant, les promoteurs de transcription de la plupart des groupes ayant des copies potentiellement fonctionnelles sont prédits en aval du codon initial la transposase. L'absence de promoteurs en amont du codon initial du gène codant la transposase pourrait faire partie du mécanisme d'inactivation des IS entraînant la fin d'une période d'activité pour le groupe considéré. Néanmoins, notre étude des groupes d'IS contenant des copies potentiellement fonctionnelles des génomes de *w*Mel, *w*Pel et *w*Ri a montré que, pour la majorité des groupes, le promoteur

se trouve là aussi en aval du codon initiation initial. Ceci n'a semble-t-il pas affecté l'activité des groupes d'IS, qui présentent de nombreux signes d'activité récente. Nous pouvons donc supposer que la transcription des copies d'IS est dépendante, au moins en partie, de l'environnement génomique des copies. L'observation de la co-transcription des régions flanquantes en aval et en amont de copies d'IS permet de conforter cette hypothèse. La dépendance de l'expression des IS vis-à-vis de leur environnement génomique pourrait expliquer les variations d'abondance observées entre souches bactériennes phylogénétiquement proches. De plus, nous avons montré que seule une partie des copies d'un groupe d'IS participent à la production de nouvelles copies, ce qui renforce l'idée de l'importance de l'environnement génomique. Ainsi, la réussite évolutive d'une famille ou d'un groupe pourrait être fonction du site d'insertion de la première copie d'IS dans le génome de l'hôte. Afin de tester cette hypothèse, il faudrait réaliser expérimentalement le transfert horizontal d'une copie d'IS dans une souche de bactérie naïve, c'est-à-dire ne contenant pas de copies d'IS du même groupe que celle transférée, et suivre l'évolution du nombre de copies au cours des générations en fonction du site d'insertion de la première copie.

En conclusion, nos travaux sur les IS des génomes de *Wolbachia* ont permis de mettre en évidence l'exceptionnelle abondance de copies d'IS dégradées. La présence de ces copies dégradées nous a permis d'apporter des confirmations empiriques au modèle de dynamique évolutive des ET procaryotes. Nous avons par ailleurs confirmé l'importance des transferts horizontaux d'IS entre souches de bactéries dans la dynamique évolutive de ces derniers. En outre, nous avons identifié différents modes de dégradation des IS, qui participent à l'évolution des IS dans les génomes de *Wolbachia*. Enfin, nous avons suggéré que malgré leur capacité à promouvoir leur propre transcription, les IS pourraient être dépendants de

l’environnement génomique dans lequel ils s’insèrent, ce qui a une influence sur leur dynamique évolutive.

Bibliographie

Aravin AA, Hannon GJ, Brennecke J 2007. The Piwi-piRNA pathway provides an adaptive defense in the transposon arms race. Science 318: 761-764.

Arraiano CM, Andrade JM, Domingues S, Guinote IB, Malecki M, Matos RG, Moreira RN, Pobre V, Reis FP, Saramago M, Silva IJ, Viegas SC 2010. The critical role of RNA processing and degradation in the control of gene expression. FEMS Microbiol Rev 34: 883-923.

Atyame CM, Delsuc F, Pasteur N, Weill M, Duron O 2011. Diversification of Wolbachia Endosymbiont in the Culex pipiens Mosquito. Mol Biol Evol 28: 2761-2772.

Aziz RK, Breitbart M, Edwards RA 2010. Transposases are the most abundant, most ubiquitous genes in nature. Nucleic Acids Res 38: 4207-4217.

Bandelt HJ, Forster P, Rohl A 1999. Median-joining networks for inferring intraspecific phylogenies. Mol Biol Evol 16: 37-48.

Bandi C, Anderson TJ, Genchi C, Blaxter ML 1998. Phylogeny of Wolbachia in filarial nematodes. Proc Biol Sci 265: 2407-2413.

Baranov PV, Fayet O, Hendrix RW, Atkins JF 2006. Recoding in bacteriophages and bacterial IS elements. Trends Genet 22: 174-181.

Barker CS, Pruss BM, Matsumura P 2004. Increased motility of Escherichia coli by insertion sequence element integration into the regulatory region of the flhD operon. J Bacteriol 186: 7529-7537.

Batzer MA, Deininger PL 2002. Alu repeats and human genomic diversity. Nat Rev Genet 3: 370-379.

Belancio VP, Hedges DJ, Deininger P 2008. Mammalian non-LTR retrotransposons: for better or worse, in sickness and in health. Genome Res 18: 343-358.

Bennetzen JL 2000. Transposable element contributions to plant gene and genome evolution. Plant Mol Biol 42: 251-269.

Bennetzen JL, Ma J, Devos KM 2005. Mechanisms of recent genome size variation in flowering plants. Ann Bot 95: 127-132.

Beuzon CR, Marques S, Casadesus J 1999. Repression of IS200 transposase synthesis by RNA secondary structures. Nucleic Acids Res 27: 3690-3695.

Bezier A, Annaheim M, Herbiniere J, Wetterwald C, Gyapay G, Bernard-Samain S, Wincker P, Roditi I, Heller M, Belghazi M, Pfister-Wilhem R, Periquet G, Dupuy C, Huguet E, Volkoff AN, Lanzrein B, Drezen JM 2009. Polydnaviruses of braconid wasps derive from an ancestral nudivirus. Science 323: 926-930.

Bichsel M, Barbour AD, Wagner A 2010. The early phase of a bacterial insertion sequence infection. Theor Popul Biol 78: 278-288.

Bingham PM, Kidwell MG, Rubin GM 1982. The molecular basis of P-M hybrid dysgenesis: the role of the P element, a P-strain-specific transposon family. Cell 29: 995-1004.

Bordenstein SR, Wernegreen JJ 2004. Bacteriophage flux in endosymbionts (Wolbachia): infection frequency, lateral transfer, and recombination rates. Mol Biol Evol 21: 1981-1991.

Bouchon D, Cordaux R, Grève P. 2008. Feminizing Wolbachia and the evolution of sex determination in isopods. In: Bourtzis K, Miller T, editors. Insect Symbiosis: CRC Press. p. 273-294.

Braquart-Varnier C, Greve P, Felix C, Martin G 2005. Bacteriophage WO in Wolbachia infecting terrestrial isopods. Biochem Biophys Res Commun 337: 580-585.

Brennecke J, Aravin AA, Stark A, Dus M, Kellis M, Sachidanandam R, Hannon GJ 2007. Discrete small RNA-generating loci as master regulators of transposon activity in Drosophila. Cell 128: 1089-1103.

Brookfield JF 2005. The ecology of the genome - mobile DNA elements and their hosts. Nat Rev Genet 6: 128-136.

Brugger K, Torarinsson E, Redder P, Chen L, Garrett RA 2004. Shuffling of Sulfolobus genomes by autonomous and non-autonomous mobile elements. Biochem Soc Trans 32: 179-183.

Bucheton A, Paro R, Sang HM, Pelisson A, Finnegan DJ 1984. The molecular basis of I-R hybrid dysgenesis in Drosophila melanogaster: identification, cloning, and properties of the I factor. Cell 38: 153-163.

Bureau TE, Wessler SR 1992. Tourist: a large family of small inverted repeat elements frequently associated with maize genes. Plant Cell 4: 1283-1294.

Carlson PE, Jr., Horzempa J, O'Dee DM, Robinson CM, Neophytou P, Labrinidis A, Nau GJ 2009. Global transcriptional response to spermine, a component of the intramacrophage environment, reveals regulation of Francisella gene expression through insertion sequence elements. J Bacteriol 191: 6855-6864.

Cerveau N, Leclercq S, Bouchon D, Cordaux R. 2011a. Evolutionary Dynamics and Genomic Impact of Prokaryote Transposable Elements. In: Pontarotti P, editor. Evolutionary Biology – Concepts, Biodiversity, Macroevolution and Genome Evolution: Springer Berlin Heidelberg. p. 291-312.

Cerveau N, Leclercq S, Leroy E, Bouchon D, Cordaux R 2011b. Short and long-term evolutionary dynamics of bacterial insertion sequences: insights from Wolbachia endosymbionts. Genome Biol Evol 3: 1175-1186.

Chandler M, Mahillon J. 2002. Insertion Sequences Revisited. In: Craig NL, Craigie R, Gellert M, Lambowitz AM, editors. Mobile DNA II. Washington, D.C.: ASM Press. p. 305-366.

Chapman EJ, Carrington JC 2007. Specialization and evolution of endogenous small RNA pathways. Nat Rev Genet 8: 884-896.

Charlier D, Piette J, Glansdorff N 1982. IS3 can function as a mobile promoter in E. coli. Nucleic Acids Research 10: 5935-5948.

Chevalier F, Herbiniere-Gaboreau J, Bertaux J, Raimond M, Morel F, Bouchon D, Greve P, Braquart-Varnier C 2011. The Immune Cellular Effectors of Terrestrial Isopod Armadillidium vulgare: Meeting with Their Invaders, Wolbachia. PLoS One 6: e18531.

Chillon I, Martinez-Abarca F, Toro N 2010. Splicing of the Sinorhizobium meliloti RmInt1 group II intron provides evidence of retroelement behavior. Nucleic Acids Res 39: 1095-1104.

Cho NH, Kim HR, Lee JH, Kim SY, Kim J, Cha S, Darby AC, Fuxelius HH, Yin J, Kim JH, Lee SJ, Koh YS, Jang WJ, Park KH, Andersson SG, Choi MS, Kim IS 2007. The

Orientia tsutsugamushi genome reveals massive proliferation of conjugative type IV secretion system and host-cell interaction genes. Proc Natl Acad Sci U S A 104: 7981-7986.

Cordaux R 2009. Gene conversion maintains nonfunctional transposable elements in an obligate mutualistic endosymbiont. Mol Biol Evol 26: 1679-1682.

Cordaux R 2008. ISWpi1 from Wolbachia pipientis defines a novel group of insertion sequences within the IS5 family. Gene 409: 20-27.

Cordaux R, Batzer MA 2009. The impact of retrotransposons on human genome evolution. Nat Rev Genet 10: 691-703.

Cordaux R, Bouchon D, Greve P 2011. The impact of endosymbionts on the evolution of host sex-determination mechanisms. Trends Genet 27: 332-341.

Cordaux R, Hedges DJ, Batzer MA 2004a. Retrotransposition of Alu elements: how many sources? Trends Genet 20: 464-467.

Cordaux R, Michel-Salzat A, Bouchon D 2001. Wolbachia infection in crustaceans: novel hosts and potential routes for horizontal transmission. Journal of Evolutionary Biology 14: 237-243.

Cordaux R, Michel-Salzat A, Frelon-Raimond M, Rigaud T, Bouchon D 2004b. Evidence for a new feminizing Wolbachia strain in the isopod Armadillidium vulgare: evolutionary implications. Heredity 93: 78-84.

Cordaux R, Pichon S, Ling A, Perez P, Delaunay C, Vavre F, Bouchon D, Greve P 2008. Intense transpositional activity of insertion sequences in an ancient obligate endosymbiont. Mol Biol Evol 25: 1889-1896.

Curcio MJ, Derbyshire KM 2003. The outs and ins of transposition: from mu to kangaroo. Nat Rev Mol Cell Biol 4: 865-877.

Czech B, Malone CD, Zhou R, Stark A, Schlingeheyde C, Dus M, Perrimon N, Kellis M, Wohlschlegel JA, Sachidanandam R, Hannon GJ, Brennecke J 2008. An endogenous small interfering RNA pathway in Drosophila. Nature 453: 798-802.

Daboussi MJ, Capy P 2003. Transposable elements in filamentous fungi. Annu Rev Microbiol 57: 275-299.

Dale C, Moran NA 2006. Molecular interactions between bacterial symbionts and their hosts. Cell 126: 453-465.

Dalrymple B, Arber W 1986. The characterization of terminators of RNA transcription on IS30 and an analysis of their role in IS element-mediated polarity. Gene 44: 1-10.

Dalrymple B, Arber W 1985. Promotion of RNA transcription on the insertion element IS30 of E. coli K12. EMBO J 4: 2687-2693.

Darby AC, Cho NH, Fuxelius HH, Westberg J, Andersson SG 2007. Intracellular pathogens go extreme: genome evolution in the Rickettsiales. Trends Genet 23: 511-520.

De Boer JG, Yazawa R, Davidson WS, Koop BF 2007. Bursts and horizontal evolution of DNA transposons in the speciation of pseudotetraploid salmonids. BMC Genomics 8: 422.

De Setta N, Van Sluys MA, Capy P, Carareto CM 2011. Copia retrotransposon in the Zaprionus genus: another case of transposable element sharing with the Drosophila melanogaster subgroup. J Mol Evol 72: 326-338.

Dedeine F, Vavre F, Shoemaker DD, Bouletreau M 2004. Intra-individual coexistence of a Wolbachia strain required for host oogenesis with two strains inducing cytoplasmic incompatibility in the wasp Asobara tabida. Evolution 58: 2167-2174.

Deloger M, Cavalli FM, Lerat E, Biemont C, Sagot MF, Vieira C 2009. Identification of expressed transposable element insertions in the sequenced genome of Drosophila melanogaster. Gene 439: 55-62.

Deprá M, Valente VL, Margis R, Loreto EL 2009. The hobo transposon and hobo-related elements are expressed as developmental genes in Drosophila. Gene 448: 57-63.

Diao Y, Qi Y, Ma Y, Xia A, Sharakhov I, Chen X, Biedler J, Ling E, Tu ZJ 2011. Next-generation sequencing reveals recent horizontal transfer of a DNA transposon between divergent mosquitoes. PLoS One 6: e16743.

Duron O, Bouchon D, Boutin S, Bellamy L, Zhou L, Engelstadter J, Hurst GD 2008. The diversity of reproductive parasites among arthropods: Wolbachia do not walk alone. BMC Biol 6: 27.

Duron O, Lagnel J, Raymond M, Bourtzis K, Fort P, Weill M 2005. Transposable element polymorphism of Wolbachia in the mosquito Culex pipiens: evidence of genetic diversity, superinfection and recombination. Mol Ecol 14: 1561-1573.

Eickbush TH, Eickbush DG. 2005. Transposable Elements: Evolution: John Wiley & Sons, Ltd.

Eickbush TH, Jamburuthugoda VK 2008. The diversity of retrotransposons and the properties of their reverse transcriptases. Virus Res 134: 221-234.

Eickbush TH, Malik HS. 2002. Origins and Evolution of Retrotransposons. In: Craig NL, Craigie R, Gellert M, Lambowitz AM, editors. Mobile DNA II. Washington, D.C.: ASM Press. p. 305-366.

Escoubas JM, Prere MF, Fayet O, Salvignol I, Galas D, Zerbib D, Chandler M 1991. Translational control of transposition activity of the bacterial insertion sequence IS1. EMBO J 10: 705-712.

Esnault C, Maestre J, Heidmann T 2000. Human LINE retrotransposons generate processed pseudogenes. Nat Genet 24: 363-367.

Felsheim RF, Kurtti TJ, Munderloh UG 2009. Genome sequence of the endosymbiont Rickettsia peacockii and comparison with virulent Rickettsia rickettsii: identification of virulence factors. PLoS One 4: e8361.

Feng Q, Moran JV, Kazazian HH, Jr., Boeke JD 1996. Human L1 retrotransposon encodes a conserved endonuclease required for retrotransposition. Cell 87: 905-916.

Feng X, Colloms SD 2007. In vitro transposition of ISY100, a bacterial insertion sequence belonging to the Tc1/mariner family. Mol Microbiol 65: 1432-1443.

Feschotte C, Jiang N, Wessler SR 2002. Plant transposable elements: where genetics meets genomics. Nat Rev Genet 3: 329-341.

Feschotte C, Pritham EJ. 2007a. Computational analysis of paleogenomics of interspersed repeats in eukaryotes. In: Stojanovic N, editor. Computational Genomics: Current Methods: Horizon Scientific Press. p. 31-53.

Feschotte C, Pritham EJ 2007b. DNA transposons and the evolution of eukaryotic genomes. Annu Rev Genet 41: 331-368.

Filee J, Siguier P, Chandler M 2007. Insertion sequence diversity in archaea. Microbiol Mol Biol Rev 71: 121-157.

Finnegan DJ 1989. Eukaryotic transposable elements and genome evolution. Trends Genet 5: 103-107.

Foster J, Ganatra M, Kamal I, Ware J, Makarova K, Ivanova N, Bhattacharyya A, Kapatral V, Kumar S, Posfai J, Vincze T, Ingram J, Moran L, Lapidus A, Omelchenko M, Kyrpides N, Ghedin E, Wang S, Goltsman E, Joukov V, Ostrovskaya O, Tsukerman K, Mazur M, Comb D, Koonin E, Slatko B 2005. The Wolbachia genome of Brugia malayi: endosymbiont evolution within a human pathogenic nematode. PLoS Biol 3: e121.

Fuxelius HH, Darby A, Min CK, Cho NH, Andersson SG 2007. The genomic and metabolic diversity of Rickettsia. Res Microbiol 158: 745-753.

Garcillán-Barcia PM, Bernales I, Mendiola MV, De la Cruz F. 2002. IS91 Rolling-Circle Transposition. In: Craig NL, Craigie R, Gellert M, Lambowitz AM, editors. Mobile DNA II. Washington, D.C.: ASM Press. p. 305-366.

Gautheret D, Lambert A 2001. Direct RNA motif definition and identification from multiple sequence alignments using secondary structure profiles. J Mol Biol 313: 1003-1011.

Gilbert C, Schaack S, Pace JK, 2nd, Brindley PJ, Feschotte C 2010. A role for host-parasite interactions in the horizontal transfer of transposons across phyla. Nature 464: 1347-1350.

Gomez-Valero L, Latorre A, Gil R, Gadau J, Feldhaar H, Silva FJ 2008. Patterns and rates of nucleotide substitution, insertion and deletion in the endosymbiont of ants Blochmannia floridanus. Mol Ecol 17: 4382-4392.

Gottesman S 2005. Micros for microbes: non-coding regulatory RNAs in bacteria. Trends Genet 21: 399-404.

Haegeman A, Vanholme B, Jacob J, Vandekerckhove TT, Claeys M, Borgonie G, Gheysen G 2009. An endosymbiotic bacterium in a plant-parasitic nematode: member of a new Wolbachia supergroup. Int J Parasitol 39: 1045-1054.

Hall TA 1999. BioEdit: a user-friendly biological sequence alignement editor and analysis program for Windows 95/98/NT. Nucl. Acids. Symp. Ser. 41: 95-98.

Haniford DB, Benjamin HW, Kleckner N 1991. Kinetic and structural analysis of a cleaved donor intermediate and a strand transfer intermediate in Tn10 transposition. Cell 64: 171-179.

Hertig M, Wolbach SB 1924. Studies on Rickettsia-Like Micro-Organisms in Insects. J Med Res 44: 329-374 327.

Hilgenboecker K, Hammerstein P, Schlattmann P, Telschow A, Werren JH 2008. How many species are infected with Wolbachia?--A statistical analysis of current data. FEMS Microbiol Lett 281: 215-220.

Hoffmann AA, Turelli M. 1997. Cytoplasmic incompatibility in insects. In: O'Neill SL, Hoffmann AA, Werren JH, editors. Influential Passengers: Oxford University Press. p. 42-80.

Horvath P, Barrangou R 2010. CRISPR/Cas, the immune system of bacteria and archaea. Science 327: 167-170.

Hosokawa T, Koga R, Kikuchi Y, Meng XY, Fukatsu T 2010. Wolbachia as a bacteriocyte-associated nutritional mutualist. Proc Natl Acad Sci U S A 107: 769-774.

Houck CM, Rinehart FP, Schmid CW 1979. A ubiquitous family of repeated DNA sequences in the human genome. J Mol Biol 132: 289-306.

Hua-Van A, Le Rouzic A, Boutin TS, Filee J, Capy P 2011. The struggle for life of the genome's selfish architects. Biol Direct 6: 19.

Hua-Van A, Le Rouzic A, Maisonhaute C, Capy P 2005. Abundance, distribution and dynamics of retrotransposable elements and transposons: similarities and differences. Cytogenet Genome Res 110: 426-440.

Jansen R, Embden JD, Gaastra W, Schouls LM 2002. Identification of genes that are associated with DNA repeats in prokaryotes. Mol Microbiol 43: 1565-1575.

Jaurin B, Normark S 1983. Insertion of IS2 creates a novel ampC promoter in Escherichia coli. Cell 32: 809-816.

Jeyaprakash A, Hoy MA 2000. Long PCR improves Wolbachia DNA amplification: wsp sequences found in 76% of sixty-three arthropod species. Insect Mol Biol 9: 393-405.

Johnson RC, Reznikoff WS 1984. Role of the IS50 R proteins in the promotion and control of Tn5 transposition. J Mol Biol 177: 645-661.

Kapitonov VV, Jurka J 2003. Molecular paleontology of transposable elements in the Drosophila melanogaster genome. Proc Natl Acad Sci U S A 100: 6569-6574.

Kent BN, Bordenstein SR 2010. Phage WO of Wolbachia: lambda of the endosymbiont world. Trends Microbiol 18: 173-181.

Kent BN, Funkhouser LJ, Setia S, Bordenstein SR 2011. Evolutionary Genomics of a Temperate Bacteriophage in an Obligate Intracellular Bacteria (Wolbachia). PLoS One 6: e24984.

Khan H, Smit A, Boissinot S 2006. Molecular evolution and tempo of amplification of human LINE-1 retrotransposons since the origin of primates. Genome Res 16: 78-87.

Kichenaradja P, Siguier P, Perochon J, Chandler M 2010. ISbrowser: an extension of ISfinder for visualizing insertion sequences in prokaryotic genomes. Nucleic Acids Res 38: D62-68.

Kidwell MG 1983. Evolution of hybrid dysgenesis determinants in Drosophila melanogaster. Proc Natl Acad Sci U S A 80: 1655-1659.

Kidwell MG 2002. Transposable elements and the evolution of genome size in eukaryotes. Genetica 115: 49-63.

Klasson L, Walker T, Sebaihia M, Sanders MJ, Quail MA, Lord A, Sanders S, Earl J, O'Neill SL, Thomson N, Sinkins SP, Parkhill J 2008. Genome evolution of Wolbachia strain wPip from the Culex pipiens group. Mol Biol Evol 25: 1877-1887.

Klasson L, Westberg J, Sapountzis P, Naslund K, Lutnaes Y, Darby AC, Veneti Z, Chen L, Braig HR, Garrett R, Bourtzis K, Andersson SG 2009. The mosaic genome structure of the Wolbachia wRi strain infecting Drosophila simulans. Proc Natl Acad Sci U S A 106: 5725-5730.

Koga A, Iida A, Hori H, Shimada A, Shima A 2006. Vertebrate DNA transposon as a natural mutator: the medaka fish Tol2 element contributes to genetic variation without recognizable traces. Mol Biol Evol 23: 1414-1419.

Kumar S, Nei M, Dudley J, Tamura K 2008. MEGA: a biologist-centric software for evolutionary analysis of DNA and protein sequences. Brief Bioinform 9: 299-306.

Kuo CH, Ochman H 2009. Inferring clocks when lacking rocks: the variable rates of molecular evolution in bacteria. Biol Direct 4: 35.

Lambowitz AM, Zimmerly S 2010a. Group II Introns: Mobile Ribozymes that Invade DNA. Cold Spring Harb Perspect Biol.

Lambowitz AM, Zimmerly S 2010b. Group II Introns: Mobile Ribozymes that Invade DNA. Annu Rev Genet 38: 1-35.

Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W, Funke R, Gage D, Harris K, Heaford A, Howland J, Kann L, Lehoczky J, LeVine R, McEwan P, McKernan K, Meldrim J, Mesirov JP, Miranda C, Morris W, Naylor J, Raymond C, Rosetti M, Santos R, Sheridan A, Sougnez C, Stange-Thomann N, Stojanovic N, Subramanian A, Wyman D, Rogers J, Sulston J, Ainscough R, Beck S, Bentley D, Burton J, Clee C, Carter N, Coulson A, Deadman R, Deloukas P, Dunham A, Dunham I, Durbin R, French L, Grafham D, Gregory S, Hubbard T, Humphray S, Hunt A, Jones M, Lloyd C, McMurray A, Matthews L, Mercer S, Milne S, Mullikin JC, Mungall A, Plumb R, Ross M, Shownkeen R, Sims S, Waterston RH, Wilson RK, Hillier LW, McPherson JD, Marra MA, Mardis ER, Fulton LA, Chinwalla AT, Pepin KH, Gish WR, Chissoe SL, Wendl MC, Delehaunty KD, Miner TL, Delehaunty A, Kramer JB, Cook LL, Fulton RS, Johnson DL, Minx PJ, Clifton SW, Hawkins T, Branscomb E, Predki P, Richardson P, Wenning S, Slezak T, Doggett N, Cheng JF, Olsen A, Lucas S, Elkin C, Uberbacher E, Frazier M, Gibbs RA, Muzny DM, Scherer SE, Bouck JB, Sodergren EJ, Worley KC, Rives CM, Gorrell JH, Metzker ML, Naylor SL, Kucherlapati RS, Nelson DL, Weinstock GM, Sakaki Y, Fujiyama A, Hattori M, Yada T, Toyoda A, Itoh T, Kawagoe C, Watanabe H, Totoki Y, Taylor T, Weissenbach J, Heilig R, Saurin W, Artiguenave F, Brottier P, Bruls T, Pelletier E, Robert C, Wincker P, Smith DR, Doucette-Stamm L, Rubenfield M, Weinstock K, Lee HM, Dubois J, Rosenthal A, Platzer M, Nyakatura G, Taudien S, Rump A, Yang H, Yu J, Wang J, Huang G, Gu J, Hood L, Rowen L, Madan A, Qin S, Davis RW, Federspiel NA, Abola AP, Proctor MJ, Myers RM, Schmutz J, Dickson M, Grimwood J, Cox DR, Olson MV, Kaul R, Shimizu N, Kawasaki K, Minoshima S, Evans GA, Athanasiou M, Schultz R, Roe BA, Chen F, Pan H, Ramser J, Lehrach H, Reinhardt R, McCombie WR, de la Bastide M, Dedhia N, Blocker H, Hornischer K, Nordsiek G, Agarwala R, Aravind L, Bailey JA, Bateman A, Batzoglou S, Birney E, Bork P, Brown DG, Burge CB, Cerutti L, Chen HC, Church D, Clamp M, Copley RR, Doerks T, Eddy SR, Eichler EE, Furey TS, Galagan J, Gilbert JG, Harmon C, Hayashizaki Y, Haussler D, Hermjakob H, Hokamp K, Jang W, Johnson LS, Jones TA, Kasif S, Kaspryzk A, Kennedy S, Kent WJ, Kitts P, Koonin EV, Korf I, Kulp D, Lancet D, Lowe TM, McLysaght A, Mikkelsen T, Moran JV, Mulder N, Pollara VJ, Ponting CP, Schuler G, Schultz J, Slater G, Smit AF, Stupka E, Szustakowski J, Thierry-Mieg D, Thierry-Mieg J, Wagner L, Wallis J, Wheeler R, Williams A, Wolf YI, Wolfe KH, Yang SP, Yeh RF, Collins F, Guyer MS, Peterson J, Felsenfeld A, Wetterstrand KA, Patrinos A, Morgan MJ, de Jong P, Catanese JJ, Osoegawa K, Shizuya H, Choi S, Chen YJ 2001. Initial sequencing and analysis of the human genome. Nature 409: 860-921.

Lawrence JG, Hendrix RW, Casjens S 2001. Where are the pseudogenes in bacterial genomes? Trends Microbiol 9: 535-540.

Lawrence JG, Ochman H, Hartl DL 1992. The evolution of insertion sequences within enteric bacteria. Genetics 131: 9-20.

Le Rouzic A, Boutin TS, Capy P 2007a. Long-term evolution of transposable elements. Proc Natl Acad Sci U S A 104: 19375-19380.

Le Rouzic A, Capy P 2005. The first steps of transposable elements invasion: parasitic strategy vs. genetic drift. Genetics 169: 1033-1043.

Le Rouzic A, Dupas S, Capy P 2007b. Genome ecosystem and transposable elements species. Gene 390: 214-220.

Leclercq S, Cordaux R 2011. DO PHAGES EFFICIENTLY SHUTTLE TRANSPOSABLE ELEMENTS AMONG PROKARYOTES? Evolution: no-no.

Leclercq S, Giraud I, Cordaux R 2011. Remarkable abundance and evolution of mobile group II introns in Wolbachia bacterial endosymbionts. Mol Biol Evol 28: 685-697.

Lerat E 2010. Identifying repeats and transposable elements in sequenced genomes: how to find your way through the dense forest of programs. Heredity 104: 520-533.

Lo N, Paraskevopoulos C, Bourtzis K, O'Neill SL, Werren JH, Bordenstein SR, Bandi C 2007. Taxonomic status of the intracellular bacterium Wolbachia pipientis. Int J Syst Evol Microbiol 57: 654-657.

Luan DD, Korman MH, Jakubczak JL, Eickbush TH 1993. Reverse transcription of R2Bm RNA is primed by a nick at the chromosomal target site: a mechanism for non-LTR retrotransposition. Cell 72: 595-605.

Luthi K, Moser M, Ryser J, Weber H 1990. Evidence for a role of translational frameshifting in the expression of transposition activity of the bacterial insertion element IS1. Gene 88: 15-20.

Ma C, Simons RW 1990. The IS10 antisense RNA blocks ribosome binding at the transposase translation initiation site. EMBO J 9: 1267-1274.

Macke TJ, Ecker DJ, Gutell RR, Gautheret D, Case DA, Sampath R 2001. RNAMotif, an RNA secondary structure definition and search algorithm. Nucleic Acids Res 29: 4724-4735.

Makarova KS, Wolf YI, van der Oost J, Koonin EV 2009. Prokaryotic homologs of Argonaute proteins are predicted to function as key components of a novel system of defense against mobile genetic elements. Biol Direct 4: 29.

McClintock B 1950. The origin and behavior of mutable loci in maize. Proc Natl Acad Sci U S A 36: 344-355.

McClintock B 1961. Some Parallels Between Gene Control Systems in Maize and in Bacteria. The American Naturalist 95: 265-277.

Mendiola MV, de la Cruz F 1992. IS91 transposase is related to the rolling-circle-type replication proteins of the pUB110 family of plasmids. Nucleic Acids Res 20: 3521.

Michel-Salzat A, Cordaux R, Bouchon D 2001. Wolbachia diversity in the Porcellionides pruinosus complex of species (Crustacea: Oniscidea): evidence for host-dependent patterns of infection. Heredity 87: 428-434.

Min KT, Benzer S 1997. Wolbachia, normally a symbiont of Drosophila, can be virulent, causing degeneration and early death. Proc Natl Acad Sci U S A 94: 10792-10796.

Mira A, Ochman H, Moran NA 2001. Deletional bias and the evolution of bacterial genomes. Trends Genet 17: 589-596.

Mizuuchi K 1992. Transpositional recombination: mechanistic insights from studies of mu and other elements. Annu Rev Biochem 61: 1011-1051.

Mohr G, Ghanem E, Lambowitz AM 2010. Mechanisms used for genomic proliferation by thermophilic group II introns. PLoS Biol 8: e1000391.

Moon S, Byun Y, Kim HJ, Jeong S, Han K 2004. Predicting genes expressed via -1 and +1 frameshifts. Nucleic Acids Res 32: 4884-4892.

Moran NA 2003. Tracing the evolution of gene loss in obligate bacterial symbionts. Curr Opin Microbiol 6: 512-518.

Moran NA, McCutcheon JP, Nakabachi A 2008. Genomics and evolution of heritable bacterial symbionts. Annu Rev Genet 42: 165-190.

Moran NA, McLaughlin HJ, Sorek R 2009. The dynamics and time scale of ongoing genomic erosion in symbiotic bacteria. Science 323: 379-382.

Moran NA, Plague GR 2004. Genomic changes following host restriction in bacteria. Curr Opin Genet Dev 14: 627-633.

Mount SM, Rubin GM 1985. Complete nucleotide sequence of the Drosophila transposable element copia: homology between copia and retroviral proteins. Mol Cell Biol 5: 1630-1638.

Moya A, Pereto J, Gil R, Latorre A 2008. Learning how to live together: genomic insights into prokaryote-animal symbioses. Nat Rev Genet 9: 218-229.

Nagy Z, Chandler M 2004. Regulation of transposition in bacteria. Res Microbiol 155: 387-398.

Nakayama K, Yamashita A, Kurokawa K, Morimoto T, Ogawa M, Fukuhara M, Urakami H, Ohnishi M, Uchiyama I, Ogura Y, Ooka T, Oshima K, Tamura A, Hattori M, Hayashi T 2008. The Whole-genome sequencing of the obligate intracellular bacterium Orientia tsutsugamushi revealed massive gene amplification during reductive genome evolution. DNA Res 15: 185-199.

Narita S, Nomura M, Kageyama D 2007. Naturally occurring single and double infection with Wolbachia strains in the butterfly Eurema hecabe: transmission efficiencies and population density dynamics of each Wolbachia strain. FEMS Microbiol Ecol 61: 235-245.

Naville M, Gautheret D 2010. Premature terminator analysis sheds light on a hidden world of bacterial transcriptional attenuation. Genome Biol 11: R97.

Ogata H, La Scola B, Audic S, Renesto P, Blanc G, Robert C, Fournier PE, Claverie JM, Raoult D 2006. Genome sequence of Rickettsia bellii illuminates the role of amoebae in gene exchanges between intracellular pathogens. PLoS Genet 2: e76.

Oliver KM, Degnan PH, Hunter MS, Moran NA 2009. Bacteriophages encode factors required for protection in a symbiotic mutualism. Science 325: 992-994.

Orgel LE, Crick FH 1980. Selfish DNA: the ultimate parasite. Nature 284: 604-607.

Pace JK, 2nd, Feschotte C 2007. The evolutionary history of human DNA transposons: evidence for intense activity in the primate lineage. Genome Res 17: 422-432.

Paques F, Haber JE 1999. Multiple pathways of recombination induced by double-strand breaks in Saccharomyces cerevisiae. Microbiol Mol Biol Rev 63: 349-404.

Parkhill J, Sebaihia M, Preston A, Murphy LD, Thomson N, Harris DE, Holden MT, Churcher CM, Bentley SD, Mungall KL, Cerdeno-Tarraga AM, Temple L, James K, Harris B, Quail MA, Achtman M, Atkin R, Baker S, Basham D, Bason N, Cherevach I, Chillingworth T, Collins M, Cronin A, Davis P, Doggett J, Feltwell T, Goble A, Hamlin N, Hauser H, Holroyd S, Jagels K, Leather S, Moule S, Norberczak H, O'Neil S, Ormond D, Price C, Rabbinowitsch E, Rutter S, Sanders M, Saunders D, Seeger K, Sharp S, Simmonds M, Skelton J, Squares R, Squares S, Stevens K, Unwin L, Whitehead S, Barrell BG, Maskell DJ 2003. Comparative analysis of the genome sequences of Bordetella pertussis, Bordetella parapertussis and Bordetella bronchiseptica. Nat Genet 35: 32-40.

Pitcher RS, Brissett NC, Doherty AJ 2007. Nonhomologous end-joining in bacteria: a microbial perspective. Annu Rev Microbiol 61: 259-282.

Price AL, Jones NC, Pevzner PA 2005. De novo identification of repeat families in large genomes. Bioinformatics 21 Suppl 1: i351-358.

Qiu N, He J, Wang Y, Cheng G, Li M, Sun M, Yu Z 2010. Prevalence and diversity of insertion sequences in the genome of Bacillus thuringiensis YBT-1520 and comparison with other Bacillus cereus group members. FEMS Microbiol Lett 310: 9-16.

Reimmann C, Moore R, Little S, Savioz A, Willetts NS, Haas D 1989. Genetic structure, function and regulation of the transposable element IS21. Mol Gen Genet 215: 416-424.

Rigaud T, Juchault P 1995. Success and failure of horizontal transfers of feminizing Wolbachia endosymbionts in woodlice. Journal of Evolutionary Biology 8: 249-255.

Robart AR, Seo W, Zimmerly S 2007. Insertion of group II intron retroelements after intrinsic transcriptional terminators. Proc Natl Acad Sci U S A 104: 6620-6625.

Rocha EP 2008. Evolutionary patterns in prokaryotic genomes. Curr Opin Microbiol 11: 454-460.

Rodriguez H, Snow ET, Bhat U, Loechler EL 1992. An Escherichia coli plasmid-based, mutational system in which supF mutants are selectable: insertion elements dominate the spontaneous spectra. Mutat Res 270: 219-231.

Rozen S, Skaletsky H 2000. Primer3 on the WWW for general users and for biologist programmers. Methods Mol Biol 132: 365-386.

Saedler H, Besemer J, Kemper B, Rosenwirth B, Starlinger P 1972. Insertion mutations in the control region of the Gal operon of E. coli. I. Biological characterization of the mutations. Mol Gen Genet 115: 258-265.

Saedler H, Reif HJ, Hu S, Davidson N 1974. IS2, a genetic element for turn-off and turn-on of gene activity in E. coli. Molecular and General Genetics MGG 132: 265-289.

Sanogo YO, Dobson SL, Bordenstein SR, Novak RJ 2007. Disruption of the Wolbachia surface protein gene wspB by a transposable element in mosquitoes of the Culex pipiens complex (Diptera, Culicidae). Insect Mol Biol 16: 143-154.

Sawyer SA, Dykhuizen DE, DuBose RF, Green L, Mutangadura-Mhlanga T, Wolczyk DF, Hartl DL 1987. Distribution and abundance of insertion sequences among natural isolates of Escherichia coli. Genetics 115: 51-63.

Schaack S, Gilbert C, Feschotte C 2010. Promiscuous DNA: horizontal transfer of transposable elements and why it matters for eukaryotic evolution. Trends Ecol Evol 25: 537-546.

Schmid S, Seitz T, Haas D 1998. Cointegrase, a naturally occurring, truncated form of IS21 transposase, catalyzes replicon fusion rather than simple insertion of IS21. J Mol Biol 282: 571-583.

Schmitz-Esser S, Penz T, Spang A, Horn M 2011. A bacterial genome in transition - an exceptional enrichment of IS elements but lack of evidence for recent transposition in the symbiont Amoebophilus asiaticus. BMC Evol Biol 11: 270.

Schnable PS, Ware D, Fulton RS, Stein JC, Wei F, Pasternak S, Liang C, Zhang J, Fulton L, Graves TA, Minx P, Reily AD, Courtney L, Kruchowski SS, Tomlinson C, Strong C, Delehaunty K, Fronick C, Courtney B, Rock SM, Belter E, Du F, Kim K, Abbott RM, Cotton M, Levy A, Marchetto P, Ochoa K, Jackson SM, Gillam B, Chen W, Yan L, Higginbotham J, Cardenas M, Waligorski J, Applebaum E, Phelps L, Falcone J, Kanchi K, Thane T, Scimone A, Thane N, Henke J, Wang T, Ruppert J, Shah N, Rotter K, Hodges J, Ingenthron E, Cordes M, Kohlberg S, Sgro J, Delgado B, Mead K, Chinwalla A, Leonard S, Crouse K, Collura K, Kudrna D, Currie J, He R, Angelova A, Rajasekar S, Mueller T, Lomeli R, Scara G, Ko A, Delaney K, Wissotski M, Lopez G, Campos D, Braidotti M, Ashley E, Golser W, Kim H, Lee S, Lin J, Dujmic Z, Kim W, Talag J, Zuccolo A, Fan C, Sebastian A, Kramer M, Spiegel L, Nascimento L, Zutavern T, Miller B, Ambroise C, Muller S, Spooner W, Narechania A, Ren L, Wei S, Kumari S, Faga B, Levy MJ, McMahan L, Van Buren P, Vaughn MW, Ying K, Yeh CT, Emrich SJ, Jia Y, Kalyanaraman A, Hsia AP, Barbazuk WB, Baucom RS, Brutnell TP, Carpita NC, Chaparro C, Chia JM, Deragon JM, Estill JC, Fu Y, Jeddeloh JA, Han Y, Lee H, Li P, Lisch DR, Liu S, Liu Z, Nagel DH, McCann MC, SanMiguel P, Myers AM, Nettleton D, Nguyen J, Penning BW, Ponnala L, Schneider KL, Schwartz DC, Sharma A, Soderlund C, Springer NM, Sun Q, Wang H, Waterman M, Westerman R, Wolfgruber TK, Yang L, Yu Y, Zhang L, Zhou S, Zhu Q, Bennetzen JL, Dawe RK, Jiang J, Jiang N, Presting GG, Wessler SR, Aluru S, Martienssen RA, Clifton SW, McCombie WR, Wing RA, Wilson RK 2009. The B73 maize genome: complexity, diversity, and dynamics. Science 326: 1112-1115.

Shigenobu S, Watanabe H, Hattori M, Sakaki Y, Ishikawa H 2000. Genome sequence of the endocellular bacterial symbiont of aphids Buchnera sp. APS. Nature 407: 81-86.

Siguier P, Filee J, Chandler M 2006a. Insertion sequences in prokaryotic genomes. Curr Opin Microbiol 9: 526-531.

Siguier P, Perochon J, Lestrade L, Mahillon J, Chandler M 2006b. ISfinder: the reference centre for bacterial insertion sequences. Nucleic Acids Res 34: D32-36.

Silva FJ, Latorre A, Moya A 2001. Genome size reduction through multiple events of gene disintegration in Buchnera APS. Trends Genet 17: 615-618.

Simons RW, Kleckner N 1988. Biological regulation by antisense RNA in prokaryotes. Annu Rev Genet 22: 567-600.

Strauch E, Beutin L 2006. Imprecise excision of insertion element IS5 from the fliC gene contributes to flagellar diversity in Escherichia coli. FEMS Microbiol Lett 256: 195-202.

Tanaka K, Furukawa S, Nikoh N, Sasaki T, Fukatsu T 2009. Complete WO phage sequences reveal their dynamic evolutionary trajectories and putative functional elements required for integration into the Wolbachia genome. Appl Environ Microbiol 75: 5676-5686.

Tenzen T, Matsutani S, Ohtsubo E 1990. Site-specific transposition of insertion sequence IS630. J Bacteriol 172: 3830-3836.

Thomas J, Sorourian M, Ray D, Baker RJ, Pritham EJ 2011. The limited distribution of Helitrons to vesper bats supports horizontal transfer. Gene 474: 52-58.

Ton-Hoang B, Pasternak C, Siguier P, Guynet C, Hickman AB, Dyda F, Sommer S, Chandler M 2010. Single-stranded DNA transposition is coupled to host replication. Cell 142: 398-408.

Toro N, Jimenez-Zurdo JI, Garcia-Rodriguez FM 2007. Bacterial group II introns: not just splicing. FEMS Microbiol Rev 31: 342-358.

Touchon M, Rocha EP 2007. Causes of insertion sequences abundance in prokaryotic genomes. Mol Biol Evol 24: 969-981.

Toussaint A, Merlin C 2002. Mobile elements as a combination of functional modules. Plasmid 47: 26-35.

Turcotte K, Srinivasan S, Bureau T 2001. Survey of transposable elements from rice genomic sequences. Plant J 25: 169-179.

Urasaki A, Sekine Y, Ohtsubo E 2002. Transposition of cyanobacterium insertion element ISY100 in Escherichia coli. J Bacteriol 184: 5104-5112.

Varani AM, Siguier P, Gourbeyre E, Charneau V, Chandler M 2011. ISsaga is an ensemble of web-based methods for high throughput identification and semi-automatic annotation of insertion sequences in prokaryotic genomes. Genome Biol 12: R30.

Vavre F, Fleury F, Lepetit D, Fouillet P, Bouletreau M 1999. Phylogenetic evidence for horizontal transmission of Wolbachia in host-parasitoid associations. Mol Biol Evol 16: 1711-1723.

Venner S, Feschotte C, Biemont C 2009. Dynamics of transposable elements: towards a community ecology of the genome. Trends Genet 25: 317-323.

Verne S, Johnson M, Bouchon D, Grandjean F 2007. Evidence for recombination between feminizing Wolbachia in the isopod genus Armadillidium. Gene 397: 58-66.

Vogele K, Schwartz E, Welz C, Schiltz E, Rak B 1991. High-level ribosomal frameshifting directs the synthesis of IS150 gene products. Nucleic Acids Res 19: 4377-4385.

Wagner A 2006. Periodic extinctions of transposable elements in bacterial lineages: evidence from intragenomic variation in multiple genomes. Mol Biol Evol 23: 723-733.

Wagner A, Lewis C, Bichsel M 2007. A survey of bacterial insertion sequences using IScan. Nucleic Acids Res 35: 5284-5293.

Wernegreen JJ 2005. For better or worse: genomic consequences of intracellular mutualism and parasitism. Curr Opin Genet Dev 15: 572-583.

Wernegreen JJ 2002. Genome evolution in bacterial endosymbionts of insects. Nat Rev Genet 3: 850-861.

Werren JH, O'Neill SL. 1997. The evolution of heritable symbionts. In: O'Neill SL, Hoffmann AA, Werren JH, editors. Influential Passengers: Oxford University Press. p. 1-41.

Werren JH, Zhang W, Guo LR 1995. Evolution and phylogeny of Wolbachia: reproductive parasites of arthropods. Proc Biol Sci 261: 55-63.

Wery J, Hidayat B, Kieboom J, de Bont JA 2001. An insertion sequence prepares Pseudomonas putida S12 for severe solvent stress. J Biol Chem 276: 5700-5706.

Wicker T, Sabot F, Hua-Van A, Bennetzen JL, Capy P, Chalhoub B, Flavell A, Leroy P, Morgante M, Panaud O, Paux E, SanMiguel P, Schulman AH 2007. A unified classification system for eukaryotic transposable elements. Nat Rev Genet 8: 973-982.

Wu M, Sun LV, Vamathevan J, Riegler M, Deboy R, Brownlie JC, McGraw EA, Martin W, Esser C, Ahmadinejad N, Wiegand C, Madupu R, Beanan MJ, Brinkac LM, Daugherty SC, Durkin AS, Kolonay JF, Nelson WC, Mohamoud Y, Lee P, Berry K, Young MB, Utterback T, Weidman J, Nierman WC, Paulsen IT, Nelson KE, Tettelin H, O'Neill SL, Eisen JA 2004. Phylogenomics of the reproductive parasite Wolbachia pipientis wMel: a streamlined genome overrun by mobile genetic elements. PLoS Biol 2: E69.

Xing J, Hedges DJ, Han K, Wang H, Cordaux R, Batzer MA 2004. Alu element mutation spectra: molecular clocks and the effect of DNA methylation. J Mol Biol 344: 675-682.

Xing J, Wang H, Belancio VP, Cordaux R, Deininger PL, Batzer MA 2006. Emergence of primate genes by retrotransposon-mediated sequence transduction. Proc Natl Acad Sci U S A 103: 17608-17613.

Yang F, Yang J, Zhang X, Chen L, Jiang Y, Yan Y, Tang X, Wang J, Xiong Z, Dong J, Xue Y, Zhu Y, Xu X, Sun L, Chen S, Nie H, Peng J, Xu J, Wang Y, Yuan Z, Wen Y, Yao Z, Shen Y, Qiang B, Hou Y, Yu J, Jin Q 2005. Genome dynamics and diversity of Shigella species, the etiologic agents of bacillary dysentery. Nucleic Acids Res 33: 6445-6458.

Zhang Z, Saier MH, Jr. 2009. A novel mechanism of transposon-mediated gene activation. PLoS Genet 5: e1000689.

Annexe 1 : article de synthèse publié dans l'ouvrage « Evolutionary Biology – Concepts, Biodiversity, Macroevolution and Genome Evolution »

Chapter 17
Evolutionary Dynamics and Genomic Impact of Prokaryote Transposable Elements

Nicolas Cerveau, Sébastien Leclercq, Didier Bouchon, and Richard Cordaux

Abstract Transposable elements (TEs) are one of the major forces that drive prokaryote genome evolution. Analyses of TE evolutionary dynamics revealed extensive variability in TE density between prokaryote genomes, even closely related ones. To explain this variability, a model of recurrent invasion/proliferation/extinction cycles has been proposed. In this chapter, we examine different parameters that influence these cycles in two of the simplest TE classes: insertion sequences and group II introns. In particular, we discuss TE transposition efficiency (mechanisms and regulation), ability to transfer horizontally (through plasmids and phages), and impact on genome evolution (gene activation/inactivation and structural variation). Finally, we describe TE dynamics in bacterial endosymbionts, especially in *Wolbachia*, to illustrate the importance of host population size in prokaryote TE evolution.

17.1 Introduction

Mobile genetic elements are one of the major forces that drive genome evolution in all living organisms. Among mobile genetic elements, transposable elements (TEs) can be defined as elements able to move from one genomic location to another. While TEs sometimes represent a large fraction of eukaryote genomes (up to 80% in maize), they generally do not account for more than a few percent of prokaryote genomes (Siguier et al. 2006). In this chapter, we focus on two of the simplest prokaryotic TEs, namely, insertion sequences (IS) and group II introns.

IS elements range in size from 0.8 to 2.5 kb and encode a transposase (Tpase) protein allowing mobility (Chandler and Mahillon 2002) (Fig. 17.1). IS are

N. Cerveau and S. Leclercq are co-first authors of the chapter.

N. Cerveau • S. Leclercq • D. Bouchon • R. Cordaux
Université de Poitiers, UMR CNRS 6556 Ecologie Evolution Symbiose, 40 Avenue du Recteur Pineau, 86022 Poitiers, France
e-mail: richard.cordaux@univ-poitiers.fr

P. Pontarotti (ed.), *Evolutionary Biology – Concepts, Biodiversity, Macroevolution and Genome Evolution*, DOI 10.1007/978-3-642-20763-1_17,

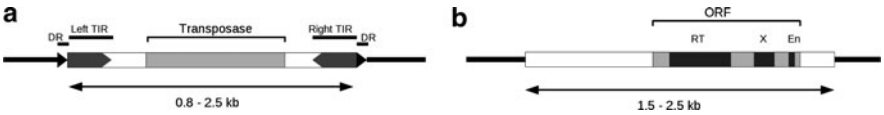

Fig. 17.1 Schematic representation of an insertion sequence (**a**) and a group II intron (**b**). *DR* direct repeat, *TIR* terminal inverted repeat, *ORF* open reading frame, *RT* reverse transcriptase, *X* maturase, *EN* endonuclease (lacking in several group II introns). Flanking sequences are shown in *black*. Drawings are not to scale

typically bounded by terminal inverted repeats (TIRs) ranging in size from 10 to 40 bp, which are recognized and bound by the Tpase during the transposition process. Most IS elements create 2–8 bp long direct repeats when inserting in a new genomic location. They are divided in around 20 families (Chandler and Mahillon 2002). Bacterial group II introns are 1.5–2.5 kb long elements, which generally encode a multi-domain protein promoting self-splicing of the element and reintegration into another genomic location (Fig. 17.1). They are distributed in nine major classes (Lambowitz and Zimmerly 2010). Contrary to IS elements which use a DNA intermediate during their transposition, group II introns use an RNA intermediate with a typical 6-domain secondary structure.

The recent sequencing of hundreds of genomes revealed that IS and group II introns are not uniformly distributed among prokaryotes (Touchon and Rocha 2007; Leclercq et al. 2011). This trend holds true even at the strain-level scale (Sawyer et al. 1987; Tourasse and Kolsto 2008; Leclercq et al. 2011; Qiu et al. 2010). The basis of this variability depends on the underlying TE evolutionary dynamics. Here, we summarize several aspects of TE dynamics, including mechanisms of transpositional activity and ability to horizontally transfer, which are essential for TE invasion and propagation. Next, we focus on TE genomic impact and on the selective pressures, which determine TE maintenance or loss in prokaryotic genomes.

17.2 Mobility of Prokaryote Transposable Elements

17.2.1 IS Transposition

17.2.1.1 Transposition Mechanisms

Excellent reviews on IS transposition mechanisms have already been published (Chandler and Mahillon 2002; Curcio and Derbyshire 2003). Briefly, IS transposition typically starts with the transcription of the Tpase gene. Most IS elements are autonomous as they carry their own promoter. After translation, the resulting Tpase binds to the TIRs of the target IS element, and cleaves both DNA strands at the 5′ and 3′ ends of the element to physically excise the IS element from the donor

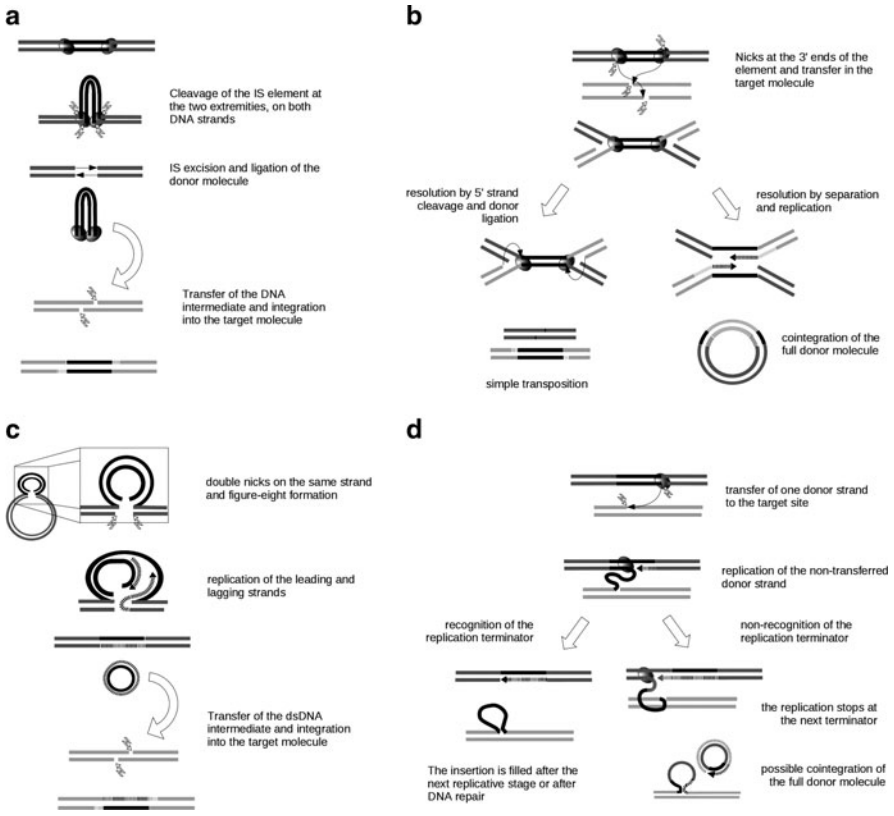

Fig. 17.2 Major IS transposition mechanisms: classical excision-based transposition (**a**), cointegrative transposition (**b**), figure-eight transposition (**c**), and rolling-circle transposition (**d**). IS DNA is represented in *black*, the donor molecule in *dark gray*, and the target molecule in *light gray*. Newly synthetized DNA is hatched. *Shaded ovals* represent Tpase proteins, and *sparks* represent cleavage events

location (Fig. 17.2a). The linear excised IS element forms a synaptic complex with the Tpase, which targets a new genomic location and promotes element integration. This pathway normally does not increase IS copy number in the genome, and is referred to as non-replicative transposition. However, if transposition occurs during replication, the IS element can excise from one newly replicated molecule and insert in the other (or at a position not yet replicated), thus leading to duplication of the element in the target molecule.

Other IS elements, such as those of the IS1 and IS6 families, use an alternative pathway called cointegrative transposition, in which a single DNA strand at the 5′ and 3′ ends of the element is cleaved. The element is thus not excised when it is integrated into the new genomic location, resulting in a covalent association between the donor and the target locations (Fig. 17.2b). The second strand at the 5′ and 3′ ends can then be cleaved, which leads to the simple transfer of the element

from the donor position to the target position. The covalent association can also be resolved by IS strand separation and replication. In this case, the IS element is duplicated at the new genomic location and not excised from the donor location. This replicative process also leads to the integration of the whole donor molecule within the target molecule in a case of an intermolecular transfer, and to genomic inversion in the case of an intramolecular transfer.

Another replicative pathway, termed "figure-eight" transposition, was identified for IS911 elements and uses a circularized DNA intermediate (Duval-Valentin et al. 2004). Tpase binding creates specific molecular figure-eight structures, which are resolved by replication of the IS element using the host replication machinery (Fig. 17.2c). The circularized replicated IS DNA sequence can be inserted into a new genomic location. This transposition pathway is probably used by members of the IS3 family, such as IS2 and IS3, which produce circularized DNA intermediates (Lewis and Grindley 1997; Ohtsubo et al. 2004). Contrary to cointegrative transposition, figure-eight transposition just duplicates the IS element without integrating the donor molecule in the target molecule.

Rolling-circle transposition is another mechanism used by elements such as IS91, in which a single-stranded IS molecule is transferred to the target molecule, and concurrently replicated in the donor molecule (Garcillan-Barcia et al. 2002) (Fig. 17.2d). Replication stops at a replication terminator located at the IS termini, leading to the duplication of the IS element. However, the whole cointegration of the donor molecule sometimes happens when the host replication system does not recognize the replication terminator and performs several rounds of replication.

The four major transposition mechanisms discussed above are the most documented pathways, but other more atypical pathways have been described or probably remain to be discovered.

17.2.1.2 Control of IS Transposition

IS elements use several strategies to regulate their transposition, presumably to limit their negative effect on the host genome, as detailed in Nagy and Chandler (2004). For example, the production of a fully active Tpase may be conditional on a ribosome frameshift during translation (Escoubas et al. 1991; Vögele et al. 1991; Lewis and Grindley 1997). This reduces transposition activity by half in the IS3 family (Vögele et al. 1991) and by more than 99% for IS1 and IS2 elements (Escoubas et al. 1991; Lewis and Grindley 1997). In other cases, transposition may be regulated through impinging transcription, i.e., sequestering of the Tpase translation initiation site in a secondary structure of the mRNA, generally induced by inverted repeat sequences (Beuzon et al. 1999). This hinders the ribosomal complex to initiate Tpase translation.

Other elements, such as IS10 and IS50, carry Dam sites that may be methylated, leading to transcriptional inactivation (Roberts et al. 1985; Tomcsanyi and Berg 1989). During genome replication, methylated sites become hemimethylated, leading to the reactivation of the IS element. This suggests that such IS elements

increase their transposition rate during the replication phase of their host genome (Roberts et al. 1985; Dodson and Berg 1989). Moreover, as explained above, these IS elements use a non-replicative transposition pathway and they may benefit from DNA replication to promote their duplication. Thus, replication-induced activity could be an efficient evolutionary strategy for proliferation of these elements. Another example of intrinsic regulation is given by the recently described IS608 elements. They are excised only as single-stranded DNA. Thus, they can transpose only when the DNA molecule is opened, i.e., at replication forks (Ton-Hoang et al. 2010) or during repair after DNA fragmentation (Pasternak et al. 2010).

Transposition may also be limited by the availability of free insertion sites. Although most IS families are able to insert at nonspecific positions, some families show nonrandom insertion patterns. Insertion generally targets 2–5 bp DNA sequences (Chandler and Mahillon 2002). Target insertion sites are sometimes much more restrictive, as for elements of the IS30 family, which precisely insert in a ~25-bp long palindromic sequence that resembles their own TIRs (Olasz et al. 1998; Kiss et al. 2007). Lack of this specific motif in a genome is a strong limitation for insertion, as exemplified by *Salmonella typhimurium*, which naturally lacks the IS30 target site and shows a very low IS30 integration rate (Casadesus et al. 1999). When the relevant insertion site was experimentally added, IS30 transposition greatly increased. This demonstrates that the site specificity of IS insertions is a key factor for IS invasion and proliferation in bacterial genomes.

17.2.2 Group II Intron Mobility

Contrary to IS elements, group II introns transpose via an RNA-intermediate, which necessarily leads to element duplication. They do not carry transcription promoters, so they must be inserted in a transcribed region to be active. Mobility starts with the transcription of the region containing the intron. The intron-encoded protein (IEP) produced from the intron mRNA binds to the intron ribozyme to form a ribonucleoprotein (RNP) complex and catalyzes intron self-splicing. The remaining mRNA is religated and can then be normally translated, while the RNP complex targets a new genomic location to insert the intron mRNA via reverse-splicing/reverse-transcription mechanisms, as reviewed in Toro et al. (2007); Lambowitz and Zimmerly (2010). Many group II introns insert directly within double-stranded DNA using the endonuclease activity of their IEP to open one DNA strand. However, some group II introns lack the endonuclease domain and are thought to insert in single-stranded DNA at replication forks.

Group II intron mobility is often called retrohoming, as introns were primarily observed to target intronless alleles of orthologous genes, defined as homing sites. Intron insertions are highly site-specific and require a conserved region of approximatively 30 bp. This very stringent insertion capacity presumably relies on the need of the intron ribozyme to bind to specific DNA motifs to be correctly spliced. As intron survival depends on its splicing ability (when inserted within

genes and to promote mobility), it is more relevant to target genomic locations that may promote intron transcription and activity (Mohr et al. 2010). Insertions at ectopic sites, i.e., at genomic locations with limited similarity with the typical insertion site motif were also observed (Cousineau et al. 2000; Martinez-Abarca and Toro 2000). These events occur with a much lower frequency but they are believed to contribute to group II intron proliferation, by ultimately diversifying insertion sites (Leclercq et al. 2011; Mohr et al. 2010). Finally, some group II intron families preferentially target structural regions rather than nucleotide motifs, such as class C introns, which insert downstream of transcriptional terminators (Robart et al. 2007) and the Avi.GroEL group II intron and relatives, which insert at or near initiation/stop codons (Michel et al. 2007).

Regulation of group II intron mobility is poorly documented, except site specificity and the need to be inserted in a transcribed region to be active. It was recently observed, though, that environmental conditions, such as temperature, may affect mobility of natural group II introns (Mohr et al. 2010).

17.2.3 Horizontal Transfers

TE survival and evolutionary success in bacteria is intimately linked to their ability to spread through horizontal transfers (HT). HT can be unambiguously detected when two divergent bacteria share identical or almost identical TEs. Evidence for HT within bacterial genera has been reported for both IS elements (Lawrence et al. 1992; Bisercic and Ochman 1993; Wagner and de la Chaux 2008) and group II introns (Dai and Zimmerly 2002; Fernandez-Lopez et al. 2005; Tourasse and Kolsto 2008; Leclercq et al. 2011). A typical example is provided by ISWpi1 elements in *Wolbachia*. Average nucleotide divergence between copies in 22 different strains is only of 0.22%, while the average divergence between highly conserved housekeeping genes for the same strains is ~3.7% (Cordaux et al. 2008).

More ancient HT can also be detected by comparing presence/absence patterns of IS families or intron classes in a set of prokaryote genomes and the phylogenetic relationships of these prokaryotes. Using this method, it was inferred that no more than 30 detectable HT are needed to explain the distribution of the 20 most abundant IS elements in 450 fully sequenced bacterial genomes (Wagner and de la Chaux 2008). HT events were unequally distributed among nine IS families, with seven HT inferred for the IS1 family, while other IS families displayed only one HT.

IS elements are commonly viewed as vectors for genetic exchanges between bacterial strains because they can form composite transposons that may carry virulence, resistance, or metabolic genes (Toussaint and Merlin 2002). However, IS elements and composite transposons, just like group II introns, lack the genetic material enabling HT between bacterial cells, and consequently, they are unable to perform HT by themselves (Toussaint and Merlin 2002). They need to shuttle via larger mobile elements, such as plasmids and bacteriophages, to be horizontally transferred (Frost et al. 2005).

17.2.3.1 Plasmid-Mediated Transfers

TE shuttling via plasmids has long been proposed, as IS elements and group II introns are recurrently detected in plasmid sequences of prokaryote species (Hall et al. 1989; Ng et al. 1998; Sundin 2007), in which they sometimes represent more than half of the detected open reading frames (ORF), as in *Shigella* plasmids (Venkatesan et al. 2001).

The first step for a plasmid-mediated HT is the transposition of the TE from a chromosomal location to a plasmid location. This was demonstrated in vitro and in vivo with mobility assays for IS (Schwartz et al. 1988; Wilde et al. 2003) and group II introns (Martinez-Abarca and Toro 2000; Ichiyanagi et al. 2003). Chromosome-to-plasmid transposition also occurs in natural populations, as exemplified in *Bacillus subtilis*. Many strains of *B. subtilis* carry multiple ISBsu2 copies in their chromosome and some strains also carry two plasmids only differentiated by an ISBu2 insertion (Poluektova et al. 2002).

After HT of a TE-containing plasmid from one bacterial cell to another by conjugation (Frost et al. 2005), the TE must move from the plasmid to the bacterial chromosome. This transfer can be achieved through several ways. Direct transfer through transposition is the most obvious possibility, and it is used in IS and group II intron mobility assays in vitro and in vivo (Vögele et al. 1991; Olasz et al. 1998; Cousineau et al. 2000). TEs can also be transferred to the host chromosome indirectly via the integration of genomic islands, also known as conjugative transposons, which are DNA regions containing virulence or adaptive genes and occasionally TEs (Mullany et al. 1996; Burrus et al. 2002). Finally, integrative plasmids are able to fully integrate into host chromosomes (Burrus et al. 2002), leading to concomitant integration of plasmid-borne TEs.

17.2.3.2 Phage-Mediated Transfers

HT through bacteriophages is also commonly assumed for TEs but far less documented than HT through plasmids. IS elements and group II introns are frequently found in prophage sequences, i.e., silent phages integrated in the host genome, but infrequently in active bacteriophage sequences. For example, several IS and group II intron copies are found in WO prophages of *Wolbachia* genomes (Leclercq et al. 2011; unpublished results). However, no intron and only one IS element is inserted in the genome of the active WO phage sequenced by Tanaka et al. (2009). When present, IS in active phages are found in only one or two copies, and they often are defective (Lobocka et al. 2004; Creuzburg et al. 2005). One exception is the C neurotoxin-converting phage of *Clostridium botulinum*, which contains 12 IS elements belonging to 7 families (Sakaguchi et al. 2005). To our knowledge, no intact group II intron has been reported to date in an active bacteriophage sequence.

17.3 Genomic Impact of Prokaryote Transposable Elements

IS elements and group II introns may impact genomic instability and variation in various ways, which can be classified in two major categories: (1) insertional mutagenesis, which refers to genomic consequences that directly follow insertions at novel genomic sites, and (2) structural variation, which refers to genomic consequences that take place at a post-insertional stage and are coupled with cell mechanisms like recombination. These two major types of genomic impact mediated by TEs are well illustrated by studies of experimental evolution. For example, analyses of two *Escherichia coli* populations after 10,000 generations of growth on glucose minimal medium led to the identification of several IS-associated mutations (Schneider et al. 2000), such as four IS150 copies disrupting genes and two additional IS150 copies, which had recombined, resulting in the inversion of the intervening sequence.

17.3.1 Insertional Mutagenesis

17.3.1.1 Coding Sequence Disruption

Gene inactivation by TE insertion, particularly IS, is very frequent in bacterial genomes. A study performed on *E. coli* demonstrated that the mutational spectrum of a reporter gene is largely linked to IS insertions (60% of the mutants) that disrupt the ORF (Rodriguez et al. 1992). Beyond studies using experimental systems, many cases of specific genes interrupted by IS have been reported to naturally occur. For example, the study of an unusual case of nonmobile and nonpathogenic strain of *Rickettsia peacockii* led to the discovery that two genes are disrupted by IS insertions (Simser et al. 2005). One disrupted gene is involved in actin-tail polymerization and the second gene is suspected to be involved in cell adhesion and bacterial virulence. These observations provided new insight into pathogenesis mechanisms in these bacteria.

While TE insertion in a coding gene is often thought to inactivate the gene through introduction of a premature stop codon resulting in truncated, nonfunctional proteins, this is not always the case. For example, IS insertion in the gene coding the ribosomal protein S1 of *E. coli* did not preclude protein production, although it lacked the last of six imperfect repeats (Skorski et al. 2007). Growth of the *E. coli* strain expressing the smaller S1 protein was lower compared to the wild type. However, growth delay is not due to the absence of the final repeat. Indeed, experimental introduction of an early stop codon that suppressed the final repeat in the wild-type S1 gene had no impact on bacterial growth. Instead, it was the presence of the IS in the S1 gene that probably created an unnatural 3′ end, which favored exonuclease degradation and induced growth delay.

The increasing availability of bacterial genomes provides the opportunity to more systematically track IS-mediated gene disruptions. For example, analysis of *Anabaena* sp. strain PCC 7120 genome identified 145 IS divided into several families (Wolk et al. 2010). More than 20% of these IS are inserted in putative protein-coding genes. By contrast, a systematic analysis of the pseudogenes in the *Sodalis glossinidius* genome revealed that only 18 out of 1,051 pseudogenes were associated with IS insertions, suggesting that IS elements are not a significant source of gene disruption in this species (Belda et al. 2010).

In comparison with IS elements, there are far fewer examples of gene inactivation mediated by group II introns. This is at least partly attributable to their splicing ability, which restores normal ORFs at the mRNA level. Furthermore, group II introns are less prevalent than IS elements in bacterial genomes, so less likely to disrupt genes (Touchon and Rocha 2007; Leclercq et al. 2011). One case of gene disrupted by a group IIC intron has been reported in *Geobacillus stearothermophilus* (Moretz and Lampson 2010). Group IIC introns generally target transcription terminators and thus avoid the disruption of host genes. Consistently, all but one of the 20 intron copies identified in *G. stearothermophilus* strain 10 are inserted in transcription terminators. The remaining copy is inserted in the rRNA methylase gene. Experiments were performed to detect splicing of the intron copy without success (Moretz and Lampson 2010). Thus, the authors suspected that this intron is able to splice in vivo to avoid blocking of methylase synthesis, which is an essential gene for the bacterial host.

Wolbachia is one of the few bacterial organisms for which information on the mutagenic potential of both IS and group II introns is currently available. Three of the four completely sequenced *Wolbachia* genomes harbor genes interrupted by IS and intron insertions: up to 45 genes are disrupted by IS in the *Wolbachia* *w*Mel, *w*Ri, and *w*Pel genomes (Wu et al. 2004; Klasson et al. 2008; Klasson et al. 2009), and 12 of 18 introns detected in these *Wolbachia* genomes are inserted in conserved genes (Leclercq et al. 2011). By contrast, the genome devoid of IS- and intron-disrupted genes is also the only one that lacks introns and potentially functional IS copies (Foster et al. 2005; Cordaux 2009; Leclercq et al. 2011).

TE insertions in genes being mainly deleterious, they are widely used as experimental tools to identify gene function or biosynthesis pathways (Reznikoff 2008). For example, the last unknown gene of the histidine biosynthesis pathway of *Corynebacterium glutamicum* was identified by random IS6100 insertional mutagenesis (Mormann et al. 2006). IS6100 had initially been described in a *C. glutamicum* plasmid and experimentally shown to be potentially active in this species using a transposition assay (Tauch et al. 2002). IS6100 was subsequently used to create a transposon mutant library of *C. glutamicum*. One mutant exhibited a histidine-auxotrophic phenotype. Analysis of the IS6100-disrupted gene in the mutant revealed that it encoded an L-histidinol-phosphate phosphatase. This example nicely illustrates the usefulness of IS elements as experimental tools.

17.3.1.2 Impact on Gene Expression

In the previous section, we discussed consequences of IS and group II intron insertions in the coding sequences of bacterial genomes. In this section, we focus on TE insertions in non-coding regulatory regions.

Some IS elements carry transcriptional promoters (Chandler and Mahillon 2002; Nagy and Chandler 2004). Thus, IS insertions can radically change expression levels of neighboring genes. For example, the internal promoter of IS3 has been shown to activate *argE* transcription in *E. coli* (Charlier et al. 1982). More recently, it was shown that glycerol use is modified by an IS insertion in *E. coli* (Zhang and Saier 2009). Expression of essential proteins for glycerol use (encoded by the *glpFK* operon) is normally activated by the cyclic AMP receptor protein encoded by the *crp* gene. However, expression of the *glpFK* operon was found to be activated by an IS5 insertion in mutant strains lacking *crp* (Zhang and Saier 2009). IS5 partial truncation experiments further demonstrated that only a short sequence is fully responsible for the activation of the *glpFK* operon. Thus, IS insertions followed by degradation can lead to the creation of new bacterial promoters.

IS insertions cannot only activate gene and operon expression, but they can also increase expression levels of already expressed genes. For example, the virulence level of group B *Streptococcus*, which mainly causes neonatal sepsis and meningitis, is linked to an IS1548 insertion in the *scpB-lmb* intergenic region (Al Safadi et al. 2010). This leads to overexpression of the *lmb* gene, which encodes laminin, a surface protein that probably plays a crucial role in binding and invasion of different host surfaces. Consequently, laminin-binding ability is increased and its density on cell surface rises, which results in the induction of neonatal meningitis. In another example, *ampC* gene transcription and β-lactamase protein production were increased by 20-fold following an IS2 insertion in the *ampC* promoter of *E. coli* (Jaurin and Normark 1983). In this case, the increase was not due to the internal IS2 promoter, but rather to a cryptic -35 box-like sequence, which became activated following IS2 insertion in a configuration restoring an optimal distance between this cryptic -35 box and the endogenous -10 box of the *ampC* gene.

IS insertions can also interact with gene expression by inactivating their repression. For example, the expression of *SrpABC*, a gene encoding multidrug efflux pump, was derepressed by IS insertions in *Pseudomonas putida*. It was shown that the vast majority of *P. putida* strains are able to resist to 1% toluene shock, resulting from IS*S12* insertion-mediated inactivation of *SrpS*, which is a *SrpABC* repressor (Wery et al. 2001). In another example, extended incubation of a nonmotile *E. coli* strain led to discrimination of two motile subpopulations harboring an IS5 insertion at one of two different sites in the *flhD* operon promoter region, which is the master operon of the flagellar regulon (Barker et al. 2004). The IS5 insertions did not alter the transcriptional start site of the operon. In addition, they cannot activate *flhd* operon transcription because they are inserted in opposite orientation relative to the operon. Thus, the authors suggested a disturbance of transcriptional repression due to IS5 insertions (Barker et al. 2004).

Other effects on gene expression involve transcriptional attenuation. Transcriptional terminators were recently identified in IS elements (Naville and Gautheret 2010). Two types of terminators were identified: (1) terminators located upstream of the Tpase gene, which could limit IS proliferation, and (2) terminators located in IS-borne sequences and immediately upstream of cellular genes. Many IS-related terminators are conserved, suggesting that they may have an important impact on genome evolution.

17.3.2 Impact on Genomic Structural Variation

17.3.2.1 Genome Size Variation

IS elements generally represent less than ~3% of prokaryote genomes (Siguier et al. 2006). However, IS elements sometimes cover more than 10% of the genome, as in *Sulfolobus solfataricus*, *Orientia tsutsugamushi*, and *Wolbachia* (Brügger et al. 2004; Nakayama et al. 2008; unpublished results). Overall, genome size is positively correlated with IS number in prokaryote genomes (Touchon and Rocha 2007). This correlation is also suggested in *Wolbachia* genomes (Fig. 17.3). This may not be so surprising because these bacterial endosymbionts have reduced genomes in the Mb size range and high IS densities (48–118 copies/Mb, unpublished results), as compared to other bacterial genomes in which IS density is usually closer to ~3.5 IS copies/Mb on average (Touchon and Rocha 2007).

17.3.2.2 Genomic Rearrangements

The presence of multiple TE copies in a genome may induce a variety of changes on chromosomal structure. TEs are generally of recent origin in bacterial genomes

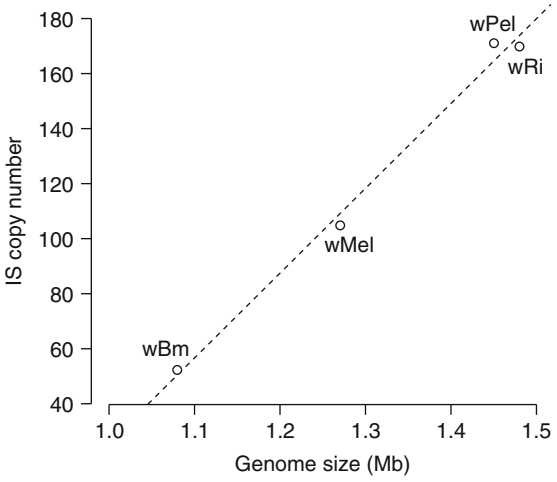

Fig. 17.3 Scatter plot of IS copy number according to genome size in the four completely sequenced *Wolbachia* genomes (*empty circles*). The dotted line indicates the correlation between genome size and IS copy number

(Wagner 2006), and their high sequence similarity makes them ideal substrates for ectopic (nonhomologous) recombination events. In addition, sequence homogeneity between copies within genomes may be maintained by gene conversion, as described for IS elements in *Wolbachia w*Bm (Cordaux 2009).

Sequence inversions are one of the potential genome rearrangements mediated by recombination between TE sequences. Inversions may result from recombination between the two TIRs of a single IS copy (Ling and Cordaux 2010), but they are more easily detected as recombination events between IS copies. For example, the *sgaA* mutant in *E. coli* has been shown to result from a rearrangement between two IS5 copies located ~60 kb apart from each other (Zinser et al. 2003). The IS5-mediated genomic inversion in the *sgaA* mutant relocated the *ybeJ* operon under the control of the promoter of another gene. As a result, the *sgaA* mutant was able to grow with aspartate and grow faster with glutamate, asparagine, and proline as carbon sources compared to the wild-type strain.

Deletions can also be mediated by IS elements. Growth of *E. coli* on glucose minimal medium during 2,000 generations resulted in rapid and parallel losses of D-ribose catabolism function (Cooper et al. 2001). The mutation frequency between the wild type and the mutant was ~5.1×10^{-5} per cell generation. PCR investigation of the *rbs* operon in 11 strains demonstrated that a fragment of variable size of the operon was deleted in all strains. Interestingly, an IS150 insertion constituted a boundary shared by all deletions. The authors suggested that the deletion process involved a second IS150 insertion at various sites in all strains, followed by recombination events between the IS150 pairs.

In the last decade, the number of available genome sequences dramatically increased. Comparative analyses have allowed quantifying IS-mediated rearrangements. For example, the comparison of *Rickettsia peacockii* and *R. rickettsii* genomes highlighted a lack of synteny, which is associated with the recent presence of 42 ISRpe1 copies in the *R. peacockii* genome (while ISRpe1 is lacking in *R. rickettsii*) (Felsheim et al. 2009). It turns out that 84% of syntenic block rearrangements are associated with ISRpe1 and 71% of genomic deletions are flanked by at least one ISRpe1 copy. These IS-mediated rearrangements might have deeply impacted *R. peacockii* lifestyle, as *R. rickettsii* is virulent whereas *R. peacockii* is nonpathogenic (Felsheim et al. 2009). Similar patterns of chromosomal rearrangements linked to IS elements have been reported in other genomes, such as *Bordetella* (Parkhill et al. 2003) and *Shigella* (Yang et al. 2005). *Wolbachia* genomes do not infringe the rule. Comparison between the two closely related strains *w*Mel and *w*Ri identified 35 gene-order breakpoints, half of which are flanked by IS elements (Klasson et al. 2009). Comparison of *Wolbachia* genomes also provided evidence for the implication of group II introns in genomic rearrangements, notably inversions (Leclercq et al. 2011). This is in contrast with the traditional view that group II introns are rarely involved in recombination events (Tourasse and Kolsto 2008). Thus, TEs may have an important impact on prokaryote chromosomal architecture.

17.4 Transposable Element Evolutionary Dynamics in Prokaryote Genomes

In the previous sections, we described mechanisms of TE mobility and impact on bacterial genome integrity. However, these properties cannot fully explain TE abundance in bacteria without considering population effects. Theoretical models of TE dynamics have been developed, involving positive or negative selection, horizontal transfers, and/or genetic drift (Sawyer et al. 1987; Basten and Moody 1991; Bichsel et al. 2010). How do these parameters influence TE spread and maintenance in bacteria?

17.4.1 Positive Selection

The most obvious way for a TE to be maintained in a genome is to be beneficial enough to the host to be positively selected. Several cases of gene activation following IS insertions have been shown to increase host fitness. For instance, the *fucAO* operon in *E. coli* became constitutively expressed after an IS5 insertion, allowing growth on propanediol substrates (Chen et al. 1989). Also, IS-mediated disruption of gene repressors can increase gene expression in a beneficial way, as in the case of the multidrug efflux pump gene repressor disrupted by IS*Ppu21* in *Pseudomonas putida*, which resulted in a 17,000-fold higher resistance of the mutant to a solvent stress compared to the wild type (Sun and Dennis 2009). Positive selection of IS elements was also observed in studies of experimental evolution of *E. coli* populations, in which IS insertions interrupted host genes and conferred selective advantages in minimal-glucose environment or under conditions of freezing/thawing cycles (Cooper et al. 2001; Sleight et al. 2008).

In a more general context, TEs are often considered as promoters of adaptability, via the various genomic rearrangements they may induce (Naas et al. 1994; Yang et al. 2005; Felsheim et al. 2009). A nice example is an inversion of a genomic fragment located between two IS5 copies in *E. coli* that induced a change in an operon expression pattern (Zinser et al. 2003) (see Sect. 17.3.2.2 above). This led to faster growth on diverse substrates, but reduced fitness during starvation. TE proliferation might thus be positively selected in bacteria living in highly unstable environments. However, the exact beneficial effect of TE-mediated recombination and resulting genomic variability on adaptability is still a matter of great debate (Dale and Moran 2006; Rocha 2008). Indeed, most bacteria for which high TE-mediated recombination levels were reported have small population sizes, thus allowing the fixation of slightly deleterious recombination events because of genetic drift and not necessarily because they are beneficial (see Sect. 17.4.4 below).

17.4.2 Negative Selection

Despite several cases of positively selected IS insertions, TE effect on bacterial genomes is thought to be mainly deleterious. Indeed, our vision of selection underestimates TE deleteriousness due to several biases. First, one of the major drawbacks of experimental evolution studies is that they proceed in controlled environments. Consequently, IS-interrupted genes may not be considered as essential although they could be in natural environments. Observed fitness increases may thus reflect a deregulation only sustainable in the experimental environment. A second and more important bias is that we cannot detect lethal insertions in genes. Similarly, sublethal transposition events are quickly discarded by natural selection and they are difficult to observe in natural prokaryote genomes. As protein-coding genes generally cover at least 80% of bacterial genomes (Lawrence et al. 2001), deleterious insertions in these genomic regions are expected to occur frequently. Consistently, cases of TE deleterious activity are regularly reported. For example, IS carry transcriptional terminators that negatively influence downstream gene expression (Naville and Gautheret 2010), and they can also induce mRNA destabilization (Skorski et al. 2007). Interestingly, some IS elements and group II introns have been found to specifically insert in other TEs, which may be a way to attenuate their deleteriousness in other genomic regions (Copertino and Hallick 1991; Olasz et al. 1998; Fernandez-Lopez et al. 2005). Overall, negative selection is an important evolutionary force in TE dynamics.

17.4.3 The Influence of Genetic Drift

In addition to TE intrinsic properties and deleteriousness, host effective population size is another critical factor in TE dynamics because it determines the efficiency of natural selection. Most bacterial species are assumed to have huge population sizes, at least large enough to rapidly discard most deleterious mutations. In this case, the spread of TE copies within genomes is strongly limited by negative selection, leading to the relatively reduced TE abundance generally observed in free-living bacteria (Moran and Plague 2004; Touchon and Rocha 2007). However, a reduction in effective population size leads to decreased efficiency of selection on slightly deleterious mutations, as well as increased fixation probability because of enhanced genetic drift. This can ultimately lead to an increase in TE copies. There is ample empirical evidence for this population genetics prediction, e.g., *Pyrococcus*, a thermophilic archaeon with a fragmented habitat (Escobar-Paramo et al. 2005), and recent pathogenic bacteria of human populations, cultured plants, and domesticated animals (Mira et al. 2006).

17.4.4 Cycles of Invasions and Extinctions

Deleteriousness raises the question of TE persistence in prokaryote genomes. Analyses of IS elements among related strains of given species usually show very high similarity between copies within strains, and striking differences in copy number between closely related strains (Sawyer et al. 1987; Lawrence et al. 1992; Parkhill et al. 2003). This suggests that IS elements are not stably maintained in bacterial genomes, but they rather undergo periodic phases of invasion, spread, and extinction (Wagner 2006). Interestingly, our data on *Wolbachia* IS elements provide direct evidence for such long-term dynamics (unpublished results).

The extinction–recolonization hypothesis is based on a balance between HT rate, transposition rate, and strength of natural selection. HT is a major driving force of bacterial genome evolution (Gogarten and Townsend 2005). Indeed, it has been estimated that up to 17% of bacterial genome sequences were acquired by recent HT (Ochman et al. 2000). Specifically, the HT rate depends on TE ability to invade vectors such as plasmids or phages (see Sect. 17.2.3 above). However, genes acquired by HT, and particularly TEs, are eliminated more rapidly than core genes (Fuxelius et al. 2008). Transposition rates are directly linked to transposition mechanisms (replicative vs. non-replicative), transposition regulation, and insertion site availability (see Sect. 17.2.1 above). Thus, the issue of selection is critical in this model. Hence, some authors consider that at least some TE copies must be beneficial to the host for elements to be maintained (Blot 1994; Schneider and Lenski 2004). Yet, recent models suggest that TE invasion and persistence in bacterial genomes does not necessarily require positive selection (Bichsel et al. 2010).

Interestingly, the extinction–recolonization hypothesis seems to also apply to group II intron dynamics (Tourasse and Kolsto 2008; Leclercq et al. 2011), despite their splicing ability that should minimize their deleteriousness. In fact, the deleterious effect of group II introns could arise from less-than-100% splicing efficiency (Chillon et al. 2011), potential loss of splicing activity after inactivating mutations, or because of their recombinational power (Leclercq et al. 2011).

17.4.5 Transposable Elements in Bacterial Endosymbionts

Bacterial endosymbionts nicely illustrate the different forces that play on TE dynamics. Indeed, the first step of endosymbiosis, i.e., the shift from a free-living lifestyle to intracellularity, is generally linked to a sharp increase in TE density (Moran and Plague 2004). Such TE proliferations reflect drastic reductions in effective population size and ensuing relaxed selection and enhanced genetic drift. Effective population size reduction is initially caused by the founding effect during which a bacterial population becomes intracellular, and subsequently accentuated by the recurrent bottlenecks undergone at each cellular transmission

(Mira and Moran 2002; Moran and Plague 2004; Wernegreen 2005). Subsequently, TEs are slowly degraded, but recolonization is prevented by the cellular confinement, which limits HT. Finally, long-term endosymbionts are completely depleted in TEs, as illustrated by the genomes of mutualistic endosymbionts such as *Buchnera aphidicola* (Shigenobu et al. 2000). TE loss in long-term endosymbionts can be attributed to genomic reduction, which corresponds to a sharp decrease in genome size, by loosing all nonessential genes, including mobile genetic elements (Wernegreen 2002).

Interestingly, *Wolbachia* endosymbionts harbor many characteristics of long-term endosymbionts (i.e., small genome, host dependence). Yet, their genomes are littered with IS elements, group II introns and phages, a good part of which is of recent origin (Cordaux 2008; Cordaux et al. 2008; Kent and Bordenstein 2010; Leclercq et al. 2011; unpublished results). Such TE dynamics is analogous to that observed in recently host-associated bacteria. It is probably linked to *Wolbachia* ability to switch hosts, leading to coinfections and breaking intracellular confinement (Cordaux et al. 2001). This may facilitate DNA exchange and import of new TEs between strains (Bordenstein and Reznikoff 2005).

17.5 Conclusion

A complex interplay between activity and HT, genomic impact and various selective pressures shapes TE dynamics in prokaryote genomes. These observations suggest a model in which TEs are recurrently acquired and rapidly lost. By contrast, eukaryote genomes often contain more ancient TE sequences, thereby offering a perspective on TE dynamics at broader evolutionary timescales (Kapitonov and Jurka 2003; Han et al. 2005; Cordaux and Batzer 2009). The fast-growing number of available bacterial genomes will certainly provide the opportunity to further our understanding of prokaryote TE dynamics at broader evolutionary timescales.

References

Al Safadi R, Amor S, Hery-Arnaud G, Spellerberg B, Lanotte P, Mereghetti L, Gannier F, Quentin R, Rosenau A (2010) Enhanced expression of lmb gene encoding laminin-binding protein in *Streptococcus agalactiae* strains harboring IS1548 in scpB-lmb intergenic region. PLoS ONE 5:e10794

Barker CS, Pruss BM, Matsumura P (2004) Increased motility of *Escherichia coli* by insertion sequence element integration into the regulatory region of the flhD operon. J Bacteriol 186:7529–7537

Basten CJ, Moody ME (1991) A branching-process model for the evolution of transposable elements incorporating selection. J Math Biol 29:743–761

Belda E, As M, Bentley S, Silva FJ (2010) Mobile genetic element proliferation and gene inactivation impact over the genome structure and metabolic capabilities of *Sodalis glossinidius*, the secondary endosymbiont of tsetse flies. BMC Genomics 11:449

Beuzon CR, Marquas S, Casadesus J (1999) Repression of IS200 transposase synthesis by RNA secondary structures. Nucleic Acids Res 27:3690–3695
Bichsel M, Barbour AD, Wagner A (2010) The early phase of a bacterial insertion sequence infection. Theor Popul Biol 78:278–288
Bisercic M, Ochman H (1993) The ancestry of insertion sequences common to *Escherichia coli* and *Salmonella typhimurium*. J Bacteriol 175:7863–7868
Blot M (1994) Transposable elements and adaptation of host bacteria. Genetica 93:5–12
Bordenstein SR, Reznikoff WS (2005) Mobile DNA in obligate intracellular bacteria. Nat Rev Microbiol 3:688–699
Brügger K, Torarinsson E, Redder P, Chen L, Garrett RA (2004) Shuffling of *Sulfolobus* genomes by autonomous and non-autonomous mobile elements. Biochem Soc Trans 32:179–183
Burrus V, Pavlovic G, Decaris B, Guedon G (2002) Conjugative transposons: the tip of the iceberg. Mol Microbiol 46:601–610
Casadesus J, Naas T, Garzon A, Arini A, Torreblanca J, Arber W (1999) Lack of hotspot targets: a constraint for IS30 transposition in *Salmonella*. Gene 238:231–239
Chandler M, Mahillon J (2002) Insertion sequences revisited. In: Craig NL et al (eds) Mobile DNA II. ASM Press, Washington, DC, pp 305–366
Charlier D, Piette J, Glansdorff N (1982) IS3 can function as a mobile promoter in *E. coli*. Nucleic Acids Res 10:5935–5948
Chen YM, Lu Z, Lin EC (1989) Constitutive activation of the fucAO operon and silencing of the divergently transcribed fucPIK operon by an IS5 element in *Escherichia coli* mutants selected for growth on L-1,2-propanediol. J Bacteriol 171:6097–6105
Chillon I, Martinez-Abarca F, Toro N (2011) Splicing of the *Sinorhizobium meliloti* RmInt1 group II intron provides evidence of retroelement behavior. Nucleic Acids Res 39:1095–1104
Cooper VS, Schneider D, Blot M, Lenski RE (2001) Mechanisms causing rapid and parallel losses of ribose catabolism in evolving populations of *Escherichia coli* B. J Bacteriol 183:2834–2841
Copertino DW, Hallick RB (1991) Group II twintron: an intron within an intron in a chloroplast cytochrome b-559 gene. EMBO J 10:433–442
Cordaux R (2008) ISWpi1 from *Wolbachia pipientis* defines a novel group of insertion sequences within the IS5 family. Gene 409:20–27
Cordaux R (2009) Gene conversion maintains nonfunctional transposable elements in an obligate mutualistic endosymbiont. Mol Biol Evol 26:1679–1682
Cordaux R, Batzer MA (2009) The impact of retrotransposons on human genome evolution. Nat Rev Genet 10:691–703
Cordaux R, Michel-Salzat A, Bouchon D (2001) *Wolbachia* infections in crustaceans: novel hosts and potential routes for horizontal transmission. J Evol Biol 14:237–243
Cordaux R, Pichon S, Ling A, Perez P, Delaunay C, Vavre F, Bouchon D, Greve P (2008) Intense transpositional activity of insertion sequences in an ancient obligate endosymbiont. Mol Biol Evol 25:1889–1896
Cousineau B, Lawrence S, Smith D, Belfort M (2000) Retrotransposition of a bacterial group II intron. Nature 404:1018–1021
Creuzburg K, Jr R, Kuhle V, Herold S, Hensel M, Schmidt H (2005) The Shiga toxin 1-converting bacteriophage BP-4795 encodes an NleA-like type III effector protein. J Bacteriol 187:8494–8498
Curcio MJ, Derbyshire KM (2003) The outs and ins of transposition: from mu to kangaroo. Nat Rev Mol Cell Biol 4:865–877
Dai L, Zimmerly S (2002) The dispersal of five group II introns among natural populations of *Escherichia coli*. RNA 8:1294–1307
Dale C, Moran NA (2006) Molecular interactions between bacterial symbionts and their hosts. Cell 126:453–465
Dodson KW, Berg DE (1989) Factors affecting transposition activity of IS50 and Tn5 ends. Gene 76:207–213

Duval-Valentin G, Marty-Cointin B, Chandler M (2004) Requirement of IS911 replication before integration defines a new bacterial transposition pathway. EMBO J 23:3897–3906

Escobar-Paramo P, Ghosh S, DiRuggiero J (2005) Evidence for genetic drift in the diversification of a geographically isolated population of the hyperthermophilic archaeon *Pyrococcus*. Mol Biol Evol 22:2297–2303

Escoubas JM, Prere MF, Fayet O, Salvignol I, Galas D, Zerbib D, Chandler M (1991) Translational control of transposition activity of the bacterial insertion sequence IS1. EMBO J 10:705–712

Felsheim RF, Kurtti TJ, Munderloh UG (2009) Genome sequence of the endosymbiont *Rickettsia peacockii* and comparison with virulent *Rickettsia rickettsii*: identification of virulence factors. PLoS ONE 4:e8361

Fernandez-Lopez M, Munoz-Adelantado E, Gillis M, Willems A, Toro N (2005) Dispersal and evolution of the *Sinorhizobium meliloti* group II RmInt1 intron in bacteria that interact with plants. Mol Biol Evol 22:1518–1528

Foster J, Ganatra M, Kamal I, Ware J, Makarova K, Ivanova N, Bhattacharyya A, Kapatral V, Kumar S, Posfai J, Vincze T, Ingram J, Moran L, Lapidus A, Omelchenko M, Kyrpides N, Ghedin E, Wang S, Goltsman E, Joukov V, Ostrovskaya O, Tsukerman K, Mazur M, Comb D, Koonin E, Slatko B (2005) The *Wolbachia* genome of *Brugia malayi*: endosymbiont evolution within a human pathogenic nematode. PLoS Biol 3:e121

Frost LS, Leplae R, Summers AO, Toussaint A (2005) Mobile genetic elements: the agents of open source evolution. Nat Rev Microbiol 3:722–732

Fuxelius H-H, Darby AC, Cho N-H, Andersson SGE (2008) Visualization of pseudogenes in intracellular bacteria reveals the different tracks to gene destruction. Genome Biol 9:R42

Garcillan-Barcia MP, Bernales I, Mendiola MV, De la Cruz F (2002) IS91 Rolling-circle transposition. In: Craig NL et al (eds) Mobile DNA II. ASM Press, Washington, DC, pp 891–904

Gogarten JP, Townsend JP (2005) Horizontal gene transfer, genome innovation and evolution. Nat Rev Microbiol 3:679–687

Hall BG, Parker LL, Betts PW, DuBose RF, Sawyer SA, Hartl DL (1989) IS103, a new insertion element in *Escherichia coli*: characterization and distribution in natural populations. Genetics 121:423–431

Han K, Xing J, Wang H, Hedges DJ, Garber RK, Cordaux R, Batzer MA (2005) Under the genomic radar: the stealth model of Alu amplification. Genome Res 15:655–664

Ichiyanagi K, Beauregard A, Belfort M (2003) A bacterial group II intron favors retrotransposition into plasmid targets. Proc Natl Acad Sci USA 100:15742–15747

Jaurin B, Normark S (1983) Insertion of IS2 creates a novel ampC promoter in *Escherichia coli*. Cell 32:809–816

Kapitonov VV, Jurka J (2003) Molecular paleontology of transposable elements in the *Drosophila melanogaster* genome. Proc Natl Acad Sci USA 100:6569–6574

Kent BN, Bordenstein SR (2010) Phage WO of *Wolbachia*: lambda of the endosymbiont world. Trends Microbiol 18:173–181

Kiss J, Nagy Z, Toth G, Kiss G, Jakab J, Chandler M, Olasz F (2007) Transposition and target specificity of the typical IS30 family element IS1655 from *Neisseria meningitidis*. Mol Microbiol 63:1731–1747

Klasson L, Walker T, Sebaihia M, Sanders MJ, Quail MA, Lord A, Sanders S, Earl J, O'Neill SL, Thomson N, Sinkins SP, Parkhill J (2008) Genome evolution of *Wolbachia* strain wPip from the *Culex pipiens* group. Mol Biol Evol 25:1877–1887

Klasson L, Westberg J, Sapountzis P, Naslund K, Lutnaes Y, Darby AC, Veneti Z, Chen L, Braig HR, Garrett R, Bourtzis K, Andersson SGE (2009) The mosaic genome structure of the *Wolbachia* wRi strain infecting *Drosophila simulans*. Proc Natl Acad Sci USA 106:5725–5730

Lambowitz AM, Zimmerly S (2010) Group II introns: mobile ribozymes that invade DNA. Cold Spring Harb Perspect Biol. doi:10.1101/cshperspect.a003616

Lawrence JG, Ochman H, Hartl DL (1992) The evolution of insertion sequences within enteric bacteria. Genetics 131:9–20
Lawrence JG, Hendrix RW, Casjens S (2001) Where are the pseudogenes in bacterial genomes? Trends Microbiol 9:535–540
Leclercq S, Giraud I, Cordaux R (2011) Remarkable abundance and evolution of mobile group II introns in *Wolbachia* bacterial endosymbionts. Mol Biol Evol 28:685–697
Lewis LA, Grindley ND (1997) Two abundant intramolecular transposition products, resulting from reactions initiated at a single end, suggest that IS2 transposes by an unconventional pathway. Mol Microbiol 25:517–529
Ling A, Cordaux R (2010) Insertion sequence inversions mediated by ectopic recombination between terminal inverted repeats. PLoS ONE 5:e15654
Lobocka MB, Rose DJ, Plunkett G, Rusin M, Samojedny A, Lehnherr H, Yarmolinsky MB, Blattner FR (2004) Genome of bacteriophage P1. J Bacteriol 186:7032–7068
Martinez-Abarca F, Toro N (2000) RecA-independent ectopic transposition in vivo of a bacterial group II intron. Nucleic Acids Res 28:4397–4402
Michel F, Costa M, Doucet AJ, Ferat J-L (2007) Specialized lineages of bacterial group II introns. Biochimie 89:542–553
Mira A, Moran NA (2002) Estimating population size and transmission bottlenecks in maternally transmitted endosymbiotic bacteria. Microb Ecol 44:137–143
Mira A, Pushker R, Rodriguez-Valera F (2006) The neolithic revolution of bacterial genomes. Trends Microbiol 14:200–206
Mohr G, Ghanem E, Lambowitz AM (2010) Mechanisms used for genomic proliferation by thermophilic group II introns. PLoS Biol 8:e1000391
Moran NA, Plague GR (2004) Genomic changes following host restriction in bacteria. Curr Opin Genet Dev 14:627–633
Moretz SE, Lampson BC (2010) A group IIC-type intron interrupts the rRNA methylase gene of *Geobacillus stearothermophilus* strain 10. J Bacteriol 192:5245–5248
Mormann S, Lomker A, Ruckert C, Gaigalat L, Tauch A, Puhler A, Kalinowski J (2006) Random mutagenesis in *Corynebacterium glutamicum* ATCC 13032 using an IS6100-based transposon vector identified the last unknown gene in the histidine biosynthesis pathway. BMC Genomics 7:205
Mullany P, Pallen M, Wilks M, Stephen JR, Tabaqchali S (1996) A group II intron in a conjugative transposon from the gram-positive bacterium, *Clostridium difficile*. Gene 174:145–150
Naas T, Blot M, Fitch WM, Arber W (1994) Insertion sequence-related genetic variation in resting *Escherichia coli* K-12. Genetics 136:721–730
Nagy Z, Chandler M (2004) Regulation of transposition in bacteria. Res Microbiol 155:387–398
Nakayama K, Yamashita A, Kurokawa K, Morimoto T, Ogawa M, Fukuhara M, Urakami H, Ohnishi M, Uchiyama I, Ogura Y, Ooka T, Oshima K, Tamura A, Hattori M, Hayashi T (2008) The Whole-genome sequencing of the obligate intracellular bacterium *Orientia tsutsugamushi* revealed massive gene amplification during reductive genome evolution. DNA Res 15:185–199
Naville M, Gautheret D (2010) Premature terminator analysis sheds light on a hidden world of bacterial transcriptional attenuation. Genome Biol 11:R97
Ng WV, Ciufo SA, Smith TM, Bumgarner RE, Baskin D, Faust J, Hall B, Loretz C, Seto J, Slagel J, Hood L, DasSarma S (1998) Snapshot of a large dynamic replicon in a halophilic archaeon: megaplasmid or minichromosome? Genome Res 8:1131–1141
Ochman H, Lawrence JG, Groisman EA (2000) Lateral gene transfer and the nature of bacterial innovation. Nature 405:299–304
Ohtsubo E, Minematsu H, Tsuchida K, Ohtsubo H, Sekine Y (2004) Intermediate molecules generated by transposase in the pathways of transposition of bacterial insertion element IS3. Adv Biophys 38:125–139
Olasz F, Kiss J, Konig P, Buzas Z, Stalder R, Arber W (1998) Target specificity of insertion element IS30. Mol Microbiol 28:691–704

Parkhill J, Sebaihia M, Preston A, Murphy LD, Thomson N, Harris DE, Holden MTG, Churcher CM, Bentley SD, Mungall KL, Cerdeno-Tarraga AM, Temple L, James K, Harris B, Quail MA, Achtman M, Atkin R, Baker S, Basham D, Bason N, Cherevach I, Chillingworth T, Collins M, Cronin A, Davis P, Doggett J, Feltwell T, Goble A, Hamlin N, Hauser H, Holroyd S, Jagels K, Leather S, Moule S, Norberczak H, O'Neil S, Ormond D, Price C, Rabbinowitsch E, Rutter S, Sanders M, Saunders D, Seeger K, Sharp S, Simmonds M, Skelton J, Squares R, Squares S, Stevens K, Unwin L, Whitehead S, Barrell BG, Maskell DJ (2003) Comparative analysis of the genome sequences of *Bordetella pertussis*, *Bordetella parapertussis* and *Bordetella bronchiseptica*. Nat Genet 35:32–40

Pasternak C, Ton-Hoang B, Coste G, Bailone A, Chandler M, Sommer S (2010) Irradiation-induced *Deinococcus radiodurans* genome fragmentation triggers transposition of a single resident insertion sequence. PLoS Genet 6:e1000799

Poluektova EU, Holsappel S, Gagarina EI, Bron S, Prozorov AA (2002) The ISBsu2 mobile element is present in a plasmid of a soil strain and in the chromosomes of several other strains of *Bacillus subtilis*. Genetika 38:1719–1722

Qiu N, He J, Wang Y, Cheng G, Li M, Sun M, Yu Z (2010) Prevalence and diversity of insertion sequences in the genome of *Bacillus thuringiensis* YBT-1520 and comparison with other *Bacillus cereus* group members. FEMS Microbiol Lett 310:9–16

Reznikoff WS (2008) Transposon Tn5. Annu Rev Genet 42:269–286

Robart AR, Seo W, Zimmerly S (2007) Insertion of group II intron retroelements after intrinsic transcriptional terminators. Proc Natl Acad Sci USA 104:6620–6625

Roberts D, Hoopes BC, McClure WR, Kleckner N (1985) IS10 transposition is regulated by DNA adenine methylation. Cell 43:117–130

Rocha EPC (2008) The organization of the bacterial genome. Annu Rev Genet 42:211–233

Rodriguez H, Snow ET, Bhat U, Loechler EL (1992) An *Escherichia coli* plasmid-based, mutational system in which supF mutants are selectable: insertion elements dominate the spontaneous spectra. Mutat Res 270:219–231

Sakaguchi Y, Hayashi T, Kurokawa K, Nakayama K, Oshima K, Fujinaga Y, Ohnishi M, Ohtsubo E, Hattori M, Oguma K (2005) The genome sequence of *Clostridium botulinum* type C neurotoxin-converting phage and the molecular mechanisms of unstable lysogeny. Proc Natl Acad Sci USA 102:17472–17477

Sawyer SA, Dykhuizen DE, DuBose RF, Green L, Mutangadura-Mhlanga T, Wolczyk DF, Hartl DL (1987) Distribution and abundance of insertion sequences among natural isolates of *Escherichia coli*. Genetics 115:51–63

Schneider D, Lenski RE (2004) Dynamics of insertion sequence elements during experimental evolution of bacteria. Res Microbiol 155:319–327

Schneider D, Duperchy E, Coursange E, Lenski RE, Blot M (2000) Long-term experimental evolution in *Escherichia coli*. IX. Characterization of insertion sequence-mediated mutations and rearrangements. Genetics 156:477–488

Schwartz E, Herberger C, Rak B (1988) Second-element turn-on of gene expression in an IS1 insertion mutant. Mol Gen Genet 211:282–289

Shigenobu S, Watanabe H, Hattori M, Sakaki Y, Ishikawa H (2000) Genome sequence of the endocellular bacterial symbiont of aphids *Buchnera* sp. APS. Nature 407:81–86

Siguier P, Filee J, Chandler M (2006) Insertion sequences in prokaryotic genomes. Curr Opin Microbiol 9:526–531

Simser JA, Rahman MS, Dreher-Lesnick SM, Azad AF (2005) A novel and naturally occurring transposon, ISRpe1 in the *Rickettsia peacockii* genome disrupting the rickA gene involved in actin-based motility. Mol Microbiol 58:71–79

Skorski P, Proux F, Cheraiti C, Dreyfus M, Hermann-Le Denmat S (2007) The deleterious effect of an insertion sequence removing the last twenty percent of the essential *Escherichia coli* rpsA gene is due to mRNA destabilization, not protein truncation. J Bacteriol 189:6205–6212

Sleight SC, Orlic C, Schneider D, Lenski RE (2008) Genetic basis of evolutionary adaptation by *Escherichia coli* to stressful cycles of freezing, thawing and growth. Genetics 180: 431–443

Sun X, Dennis JJ (2009) A novel insertion sequence derepresses efflux pump expression and preadapts *Pseudomonas putida* S12 for extreme solvent stress. J Bacteriol 191:6773–6777

Sundin GW (2007) Genomic insights into the contribution of phytopathogenic bacterial plasmids to the evolutionary history of their hosts. Annu Rev Phytopathol 45:129–151

Tanaka K, Furukawa S, Nikoh N, Sasaki T, Fukatsu T (2009) Complete WO phage sequences reveal their dynamic evolutionary trajectories and putative functional elements required for integration into the *Wolbachia* genome. Appl Environ Microbiol 75:5676–5686

Tauch A, Gotker S, Puhler A, Jr K, Thierbach G (2002) The 27.8-kb R-plasmid pTET3 from *Corynebacterium glutamicum* encodes the aminoglycoside adenyltransferase gene cassette aadA9 and the regulated tetracycline efflux system Tet 33 flanked by active copies of the widespread insertion sequence IS6100. Plasmid 48:117–129

Tomcsanyi T, Berg DE (1989) Transposition effect of adenine (Dam) methylation on activity of O end mutants of IS50. J Mol Biol 209:191–193

Ton-Hoang B, Pasternak C, Siguier P, Guynet C, Hickman AB, Dyda F, Sommer S, Chandler M (2010) Single-stranded DNA transposition is coupled to host replication. Cell 142:398–408

Toro N, Jimenez-Zurdo J, Garcia-Rodriguez FM (2007) Bacterial group II introns: not just splicing. FEMS Microbiol Rev 31:342–358

Touchon M, Rocha EPC (2007) Causes of insertion sequences abundance in prokaryotic genomes. Mol Biol Evol 24:969–981

Tourasse NJ, Kolsto A-B (2008) Survey of group I and group II introns in 29 sequenced genomes of the *Bacillus cereus* group: insights into their spread and evolution. Nucleic Acids Res 36:4529–4548

Toussaint A, Merlin C (2002) Mobile elements as a combination of functional modules. Plasmid 47:26–35

Venkatesan MM, Goldberg MB, Rose DJ, Grotbeck EJ, Burland V, Blattner FR (2001) Complete DNA sequence and analysis of the large virulence plasmid of *Shigella flexneri*. Infect Immun 69:3271–3285

Vögele K, Schwartz E, Welz C, Schiltz E, Rak B (1991) High-level ribosomal frameshifting directs the synthesis of IS150 gene products. Nucleic Acids Res 19:4377–4385

Wagner A (2006) Periodic extinctions of transposable elements in bacterial lineages: evidence from intragenomic variation in multiple genomes. Mol Biol Evol 23:723–733

Wagner A, de la Chaux N (2008) Distant horizontal gene transfer is rare for multiple families of prokaryotic insertion sequences. Mol Genet Genomics 280:397–408

Wernegreen JJ (2002) Genome evolution in bacterial endosymbionts of insects. Nat Rev Genet 3:850–861

Wernegreen JJ (2005) For better or worse: genomic consequences of intracellular mutualism and parasitism. Curr Opin Genet Dev 15:572–583

Wery J, Hidayat B, Kieboom J, de Bont JA (2001) An insertion sequence prepares *Pseudomonas putida* S12 for severe solvent stress. J Biol Chem 276:5700–5706

Wilde C, Escartin F, Kokeguchi S, Latour-Lambert P, Lectard A, Clement J-M (2003) Transposases are responsible for the target specificity of IS1397 and ISKpn1 for two different types of palindromic units (PUs). Nucleic Acids Res 31:4345–4353

Wolk CP, Lechno-Yossef S, Jager KM (2010) The insertion sequences of *Anabaena* sp. strain PCC 7120 and their effects on its open reading frames. J Bacteriol 192:5289–5303

Wu M, Sun LV, Vamathevan J, Riegler M, Deboy R, Brownlie JC, McGraw EA, Martin W, Esser C, Ahmadinejad N, Wiegand C, Madupu R, Beanan MJ, Brinkac LM, Daugherty SC, Durkin AS, Kolonay JF, Nelson WC, Mohamoud Y, Lee P, Berry K, Young MB, Utterback T, Weidman J, Nierman WC, Paulsen IT, Nelson KE, Tettelin H, O'Neill SL, Eisen JA (2004) Phylogenomics of the reproductive parasite *Wolbachia pipientis* wMel: a streamlined genome overrun by mobile genetic elements. PLoS Biol 2:E69

Yang F, Yang J, Zhang X, Chen L, Jiang Y, Yan Y, Tang X, Wang J, Xiong Z, Dong J, Xue Y, Zhu Y, Xu X, Sun L, Chen S, Nie H, Peng J, Xu J, Wang Y, Yuan Z, Wen Y, Yao Z, Shen Y, Qiang B, Hou Y, Yu J, Jin Q (2005) Genome dynamics and diversity of *Shigella* species, the etiologic agents of bacillary dysentery. Nucleic Acids Res 33:6445–6458

Zhang Z, Saier MH Jr (2009) A novel mechanism of transposon-mediated gene activation. PLoS Genet 5:e1000689

Zinser ER, Schneider D, Blot M, Kolter R (2003) Bacterial evolution through the selective loss of beneficial genes. Trade-offs in expression involving two loci. Genetics 164:1271–1277

Annexe 2: épreuves non corrigées de l'article de recherche accepté pour publication en septembre 2011 dans la revue Genome Biology and Evolution

DOI : 10.1093/gbe/evr096

http://gbe.oxfordjournals.org/content/early/2011/09/21/gbe.evr096.short?rss=1

GENOME BIOLOGY AND EVOLUTION

SMBE

GBE

Short- and Long-term Evolutionary Dynamics of Bacterial Insertion Sequences: Insights from *Wolbachia* Endosymbionts

Nicolas Cerveau, Sébastien Leclercq, Elodie Leroy, Didier Bouchon, and Richard Cordaux*

UMR CNRS 6556, Ecologie, Evolution, Symbiose, Université de Poitiers, France

*Corresponding author: E-mail: richard.cordaux@univ-poitiers.fr.

Accepted: 13 September 2011

Abstract

Transposable elements (TE) are one of the major driving forces of genome evolution, raising the question of the long-term dynamics underlying their evolutionary success. Long-term TE evolution can readily be reconstructed in eukaryotes, thanks to many degraded copies constituting genomic fossil records of past TE proliferations. By contrast, bacterial genomes usually experience high sequence turnover and short TE retention times, thereby obscuring ancient TE evolutionary patterns. We found that *Wolbachia* bacterial genomes contain 52–171 insertion sequence (IS) TEs. IS account for 11% of *Wolbachia w*Ri, which is one of the highest IS genomic coverage reported in prokaryotes to date. We show that many IS groups are currently expanding in various *Wolbachia* genomes and that IS horizontal transfers are frequent among strains, which can explain the apparent synchronicity of these IS proliferations. Remarkably, >70% of *Wolbachia* IS are nonfunctional. They constitute an unusual bacterial IS genomic fossil record providing direct empirical evidence for a long-term IS evolutionary dynamics following successive periods of intense transpositional activity. Our results show that comprehensive IS annotations have the potential to provide new insights into prokaryote TE evolution and, more generally, prokaryote genome evolution. Indeed, the identification of an important IS genomic fossil record in *Wolbachia* demonstrates that IS elements are not always of recent origin, contrary to the conventional view of TE evolution in prokaryote genomes. Our results also raise the question whether the abundance of IS fossils is specific to *Wolbachia* or it may be a general, albeit overlooked, feature of prokaryote genomes.

Key words: insertion sequence, transposable element, evolutionary dynamics, prokaryote, *Wolbachia*, molecular palaeontology.

Introduction

Transposable elements (TE) are discrete pieces of DNA that can move within (and sometimes, between) genomes. They are widely distributed in eukaryotes and prokaryotes, and they sometimes represent substantial fractions of genomes. For example, TEs encompass about half of the human genome (Lander et al. 2001) and nearly 85% of the maize genome (Schnable et al. 2009). Because of their mobility and accumulation, TEs are major drivers of genome evolution, with effects ranging from generating insertion mutations and genomic instability to altering gene expression and contributing to genetic innovation (Feschotte and Pritham 2007; Cordaux and Batzer 2009; Cerveau et al. 2011). Given their tremendous genomic impact, abundance, and widespread taxonomic distribution, the question arises as to what long-term dynamics have made TEs so prolific and evolutionary successful during the evolution of life.

In eukaryotes, long-term TE dynamics can readily be investigated because genomes often carry highly mutated and degraded TE relics constituting a genomic fossil record of past TE proliferations and evolution at various time depths (Lander et al. 2001; Kapitonov and Jurka 2003). For example, analyses of the human genomic fossil record have revealed that DNA transposons became extinct ~40 million years ago in the primate lineage, after having experienced intense activity during the mammalian radiation and early primate evolution, 60–150 million years ago (Lander et al. 2001; Pace and Feschotte 2007). By contrast, Alu and L1 retrotransposons have proliferated throughout primate evolution, although their activity has declined within the past ~20 million years (Lander et al. 2001; Xing et al. 2004; Khan et al. 2006). The relevance of genomic fossil records is also well illustrated with the emerging field of paleovirology, consisting in the study of ancient extinct viruses unearthed

from genome sequences, which are witnesses of ancient viral infections, and the effects these agents have had on their host evolution (Emerman and Malik 2010; Gilbert and Feschotte 2010).

In sharp contrast with eukaryotes, the gene repertoires of prokaryotes change quickly by lateral (or horizontal) gene transfer and gene deletion (Rocha 2008). This high turnover is well illustrated by pseudogenes and TEs in which retention times appear to be particularly short (Wagner 2006; Touchon and Rocha 2007; Wagner et al. 2007; Kuo and Ochman 2010; Cerveau et al. 2011). As a consequence, recent TE insertions are overrepresented in bacterial genomes, and our ability to infer ancient evolutionary patterns vanishes with the erosion of the past TE fossil record. For example, insertion sequences (IS), which are simple transposase-encoding TEs frequently found in prokaryotic genomes (Chandler and Mahillon 2002; Siguier, Filee, et al. 2006) are generally considered to be of recent origin. This is reflected in the very low nucleotide divergence generally observed between IS sequences within genomes. This result has been reported in early IS studies, as exemplified by three IS families of *Escherichia coli* in which copies are >99.7% similar to their family consensus sequences (Lawrence et al. 1992). Broader-scale studies on several hundreds of bacterial genomes and up to 20 IS families confirmed this trend as more than two thirds of transposase genes are identical within genomes (Wagner 2006; Wagner et al. 2007). IS recent origin is further supported by their usually patchy distribution among closely related strains (Sawyer et al. 1987; Parkhill et al. 2003; Yang et al. 2005; Cordaux et al. 2008; Qiu et al. 2010).

The analysis of recently integrated IS elements in bacterial genomes has suggested that IS may undergo extinction–reinfection cycles on the long term (Wagner 2006). Under this scenario, periodic IS reintroductions in genomes mediated by lateral transfers are crucial for their long-term survival (Wagner 2006; Bichsel et al. 2010). However, IS copy number is not directly correlated to the rate of lateral gene transfer in bacteria, suggesting that horizontal transfer may not be a major determinant of IS abundance in genomes (Touchon and Rocha 2007). Overall, our current understanding of long-term TE dynamics in prokaryotes lags far behind that of eukaryotic TE evolutionary dynamics, in part because no IS genomic fossil record has been reported and analyzed in prokaryotes.

In this study, we report an analysis of IS elements in the genomes of *Wolbachia* bacterial endosymbionts. These ancient obligate intracellular microorganisms have been associated with arthropod and nematode hosts for >100 million years, and they are considered one of the most abundant endosymbionts on Earth (Werren et al. 1995; Cordaux, Michel-Salzat, et al. 2004; Bouchon et al. 2008; Saridaki and Bourtzis 2010; Cordaux et al. 2011). Despite their reduced sizes, *Wolbachia* genomes show an unusually high proportion of repetitive and mobile DNA, including IS elements (Moran and Plague 2004; Wu et al. 2004; Bordenstein and Reznikoff 2005; Foster et al. 2005; Klasson et al. 2008, 2009; Cordaux 2009; Leclercq et al. 2011). Strikingly, we found that the vast majority of *Wolbachia* IS copies are more or less severely degraded as a result of the accumulation of nucleotide substitutions and deletions across time. Thus, they constitute an uncommon genomic fossil record for these bacterial TEs. This rich genomic archive gave us an opportunity to directly investigate the long-term dynamics of IS elements (and provide the first empirical test of Wagner's hypothesis that IS elements experience extinction–reinfection cycles on the long term) and the microevolutionary processes governing IS expansions in bacterial genomes.

Materials and Methods

Identification and Classification of IS Elements

The complete genome sequences and annotations of *Wolbachia* strains *w*Mel (Wu et al. 2004), *w*Bm (Foster et al. 2005), *w*Pel (Klasson et al. 2008), and *w*Ri (Klasson et al. 2009) were consulted and downloaded from the National Center for Biotechnology Information (NCBI) website (http://www.ncbi.nlm.nih.gov/genomes/MICROBES/microbial_taxtree.html).

We used three different strategies to identify IS elements. We first queried the original genome annotations available in GenBank with the keywords "transposase" and "transposon." Next, we performed similarity searches against the IS reference database ISFinder (Siguier, Perochon, et al. 2006) using ISsaga (Varani et al. 2011). Finally, we performed a de novo repeat detection for each genome with Repeatscout software, using l-mers size of 15 bp (Price et al. 2005). These approaches were complementary because each method alone has its own advantages and drawbacks. For example, IS identification with Repeatscout requires at least three copies in the genome, but it does not require any a priori knowledge of IS sequences. By contrast, ISsaga can detect single copy IS elements, but it requires a library of IS sequences for querying genomes.

All originally annotated IS elements were recovered by ISsaga. Repeatscout results were used as queries for BlastN searches against GenBank to identify non-IS repeats. All repeats with significant identity to known non-IS repeats (e.g., phages, group II introns, duplicated genes, etc.) were discarded. The remaining repeats were subjected to TBlastX searches against ISFinder to identify known IS elements. For the few Repeatscout repeats with no homology to known transposases remaining at this stage, we manually aligned copies and searched for IS hallmarks, such as terminal inverted repeats and target site duplications (Chandler and Mahillon 2002). None of these repeats exhibited hallmarks of IS elements, and they were therefore discarded. In sum, all IS elements identified with Repeatscout were also recovered by ISsaga.

To refine our IS annotations (i.e., to identify fragments and highly divergent copies that may have been missed

before), we generated a library of *Wolbachia* IS sequences based on the IS elements detected as described above. Next, BlastN searches against the four *Wolbachia* genomes were performed using as queries the aforementioned *Wolbachia* IS library. BlastN searches were performed with default parameters without low-complexity region filter, using a minimal subject size of 40 bp, minimal similarity of 75%, maximal *e* value of 0.05, reward of 2, and penalty of 3. To eliminate potentially redundant or overlapping IS matches, the positions of each IS copy were picked up and compared with all others for each genome. This procedure allowed us to identify 18 cases of IS copies split in two parts by nested IS insertions; each of the disrupted IS copies was counted as a single insertion event (3 in *w*Mel, 6 in *w*Ri, and 9 in *w*Pel). Overall, this analysis yielded a total of 511 candidate IS copies from the four *Wolbachia* genomes. We discarded 13 candidates as false positives (e.g., *DnaA* mistakenly assigned to IS21 and *XerC/D* recombinases mistakenly assigned to IS91 in all four genomes). Thus, the final data set consisted of 498 validated IS copies.

Each IS copy was assigned to an IS family by TBlastX searches against ISFinder (Siguier, Perochon, et al. 2006). The sequences of IS copies assigned to the same IS family were aligned using ClustalW as implemented in the software Bioedit ver 7.0 (Hall 1999), followed by manual adjustments. Due to high sequence divergence, some IS sequences could not be aligned to each other within some IS families. Therefore, we defined groups within IS families as IS nucleotide sequences that can reliably be aligned to each other within groups but cannot be aligned with sequences from other groups. As a quality control, BlastN searches were performed using all IS sequences as queries against all four *Wolbachia* genomes. For all queries, the returned matches exclusively comprised copies from the query IS group, thereby confirming the validity of the defined IS groups. *Wolbachia* IS group names were assigned based on ISFinder best IS group matches in TBlastX searches. *Wolbachia* IS groups with no functional representative (which thus could not be deposited in ISFinder) were named as follows: IS family name, followed by "-w" (for *Wolbachia*) and a specific upper-case letter. For example, IS110-wA represents a *Wolbachia* IS group from the IS110 family, which has no known representative in ISFinder.

Structure and Nucleotide Divergence of IS Elements

To investigate IS structure, for each IS group, IS sequences along with 500 bp of 5′ and 3′ flanking genomic sequences were aligned. These alignments were used to identify IS boundaries and to determine terminal inverted repeats and direct repeats generated upon insertion, whenever present. Transposase genes were identified through open reading frame detection using the NCBI online tool ORFFinder (http://www.ncbi.nlm.nih.gov/) and the FSFinder software (Moon et al. 2004). For each IS group with at least two alignable copies in a given genome, we calculated pairwise nucleotide divergence between copies with the MEGA ver 4.0 software based on observed nucleotide substitutions (Kumar et al. 2008).

Orthology Analyses

To identify IS copies inserted at orthologous genomic sites between genomes, we performed BlastN searches (minimal subject size of 40 bp, minimal similarity of 75%, maximal *e* value of 0.05, reward of 2, and penalty of 3) using as queries 300 bp of upstream and downstream genomic regions flanking each IS copy. For each IS locus, orthologous flanking regions from queried genomes were aligned and compared to identify orthologous IS insertions. However, for many IS loci, BlastN searches yielded no, partial, or multiple matches in queried genomes, thereby preventing reliable identification of orthologous IS insertion sites. Therefore, the final data set used for orthology analyses exclusively consisted of IS loci with unambiguous matches for both flanking sequences in queried genomes.

Simulations of IS Evolutionary Dynamics

Scenarios of IS dynamics are based on the simulation of a set of evolving IS sequences in a haploid genome. Simulations were designed to allow qualitative comparisons between scenarios and thus only roughly represent the biological complexity of IS evolution. Four major processes of IS evolution were considered: acquisition by horizontal transfer, copy number expansion from a resident copy, degradation through random substitutions, and loss through deletion. IS elements were represented by sequences of 300 numbers ranging in value from 0 to 63. Each number represented a hypothetical codon, with three codons representing stop codons (as in the bacterial genetic code). Each initial IS sequence did not carry any stop codon and was considered as functional.

All simulations counted a fixed number of generations. At each generation, IS sequences were allowed to mutate or be deleted at fixed constant rates. We arbitrarily chose 10,000 generations and a mutation rate of 15×10^{-6} mutation per codon to produce an average pairwise divergence of 30% between the oldest copies at the end of the simulations and keep our simulations manageable in terms of computational time. Mutations changed the value of the mutated codon to another random value. If the new value corresponded to a stop codon, the IS sequence irreversibly moved from functional to nonfunctional status. Four deletion rates were tested: equal to the mutation rate, 10 or 100 times slower than the mutation rate, and equal to 0 (no deletion).

Two types of copy number expansions were implemented: 1) instantaneous expansion (or burst), in which a horizontally transferred (and functional) copy was duplicated several times in the genome at once and 2) slow expansion, in which

a resident (and functional) copy was regularly selected at random and duplicated once in the genome. Horizontal transfers were simulated by adding a random functional copy from a source genome to the target genome. The source genome was a reservoir of 10 copies of the initial functional IS sequence evolving in parallel with the target genome. IS copies in the source genome underwent frequent random bursts, thus providing an unlimited reservoir of functional elements for future horizontal transfers.

Five different scenarios (1–5) were implemented, all designed to provide *n* IS copies:

Scenario 1 (single ancient burst): A single horizontal transfer at generation 1 immediately followed by an instantaneous expansion to *n* copies at generation 2.

Scenario 2 (single recent burst): A single horizontal transfer at generation 9,900 immediately followed by an instantaneous expansion to *n* copies at generation 9,901.

Scenario 3: (slow expansion): A single horizontal transfer at generation 1 followed by one copy duplication every 10,000/*n* generations.

Scenario 4 (two recent bursts): Two independent horizontal transfers at generation 9,900, each immediately followed by instantaneous expansions of *n*/2 copies at generation 9,901.

Scenario 5 (ancient and recent bursts): Two independent horizontal transfers at generations 1 and 9,900, each immediately followed by instantaneous expansions of *n*/2 copies at generations 2 and 9,901, respectively.

After the simulation ended, pairwise codon divergence between all functional and nonfunctional copies was calculated on the whole sequence length. Each distribution in figure 1 represents the pooled pairwise divergence distribution from 23 simulations, each of which has an expected final number of elements equal to the size of 1 of the 23 *Wolbachia* IS groups comprising at least two alignable IS copies (supplementary table S1, Supplementary Material online).

IS Survey across *Wolbachia* Strains

We assessed the presence or absence of 17 IS groups in a panel of 22 diverse *Wolbachia* strains from the A, B, and G supergroups, available from a previous study (Cordaux et al. 2008). Supplementary table S2, Supplementary Material online, provides details on the *Wolbachia* strains. The single *Wolbachia* infection status of each of the 22 samples was confirmed by polymerase chain reaction (PCR) amplification and sequencing of two to three chromosomal markers (*wsp*, 16S rRNA, and *GroE*) (Cordaux et al. 2008). The 17 IS groups were selected from the three sequenced genomes from the A and B supergroups (i.e., *w*Mel, *w*Ri, and *w*Pel) based on the occurrence of at least one potentially functional IS copy and/or several full-length copies. For each IS group, within-IS specific oligonucleotide primer pairs were designed to amplify 499- to 706-bp long fragments, using the program Primer3 (Rozen and Skaletsky 2000). PCR amplification, separation, and visualization were performed using a standard protocol (Cordaux et al. 2006, 2008). The D.mel, D.sim, and Slab DNA samples corresponding to the *w*Mel, *w*Ri, and *w*Pel *Wolbachia* strains, respectively, were used as positive controls (supplementary table S2, Supplementary Material online). Water controls were used in all PCR assays. PCR conditions for each IS group, including primer sequences and expected PCR product sizes, are shown in supplementary table S3, Supplementary Material online. To confirm the results, all PCR amplifications were performed twice independently, and PCR fragments were sequenced, as previously described (Cordaux et al. 2001). For each IS group, sequences obtained from PCR fragments were aligned with Bioedit, and pairwise nucleotide divergence between each pair of sequences was calculated with MEGA ver 4.0 based on observed nucleotide substitutions (sequence alignments are available upon request).

To infer the number of IS group acquisitions and losses, the distribution of each of the 17 IS groups was mapped onto a phylogeny of the 22 tested *Wolbachia* strains (Cordaux et al. 2001, 2008; Lo et al. 2007). For each IS group, we favored the most parsimonious scenario, that is, the scenario requiring the smallest number of acquisitions and losses to explain the distribution of the IS group according to *Wolbachia* strain phylogenetic relationships. For IS groups with two or more equiparsimonious scenarios, we conservatively favored the scenario minimizing the number of acquisitions.

Results and Discussion

Abundance and Distribution of IS Elements in *Wolbachia*

We used three independent and complementary strategies to identify IS elements in the four completely sequenced *Wolbachia* genomes from the *w*Mel, *w*Ri, *w*Pel, and *w*Bm strains (see "Materials and Methods"). This analysis revealed an IS copy number (at least 40 bp in length) per genome ranging from 52 in *w*Bm to 171 in *w*Ri (table 1). This is considerable given that prokaryotic genomes generally carry relatively few IS copies as illustrated by a survey of 262 genomes that identified a median number of 12 IS copies (range 0–342) per genome (Touchon and Rocha 2007). Overall, IS copies account for up to 11% (~160 kb) of *Wolbachia* genomes (table 1). Such IS genomic coverage exceeds that described in most other prokaryotic genomes, which is generally below 3% (Siguier, Filee, et al. 2006), except in a few rare cases such as *Shigella dysenteriae*, *Orientia tsutsugamushi*, or *Sulfobolus solfataricus* where it can reach >10% (Brugger et al. 2004; Yang et al. 2005; Cho et al. 2007; Filee et al. 2007; Nakayama et al. 2008).

Wolbachia IS elements encompass a total of 11 IS families (table 1), out of the ~20 major recognized IS families (Chandler and Mahillon 2002; Siguier, Filee, et al. 2006;

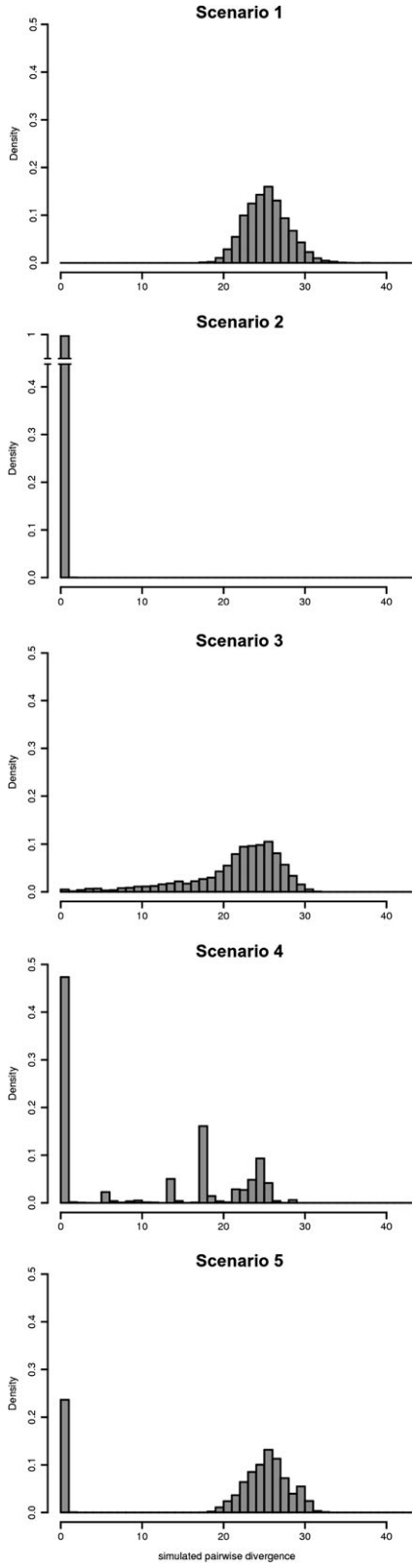

Fig. 1.—Frequency distribution of simulated pairwise IS divergence in a haploid genome under five models of IS dynamics. Scenario 1: single ancient burst; scenario 2: single recent burst; scenario 3: slow expansion; scenario 4: two independent recent bursts; and scenario 5: ancient and recent bursts. Each distribution represents the pooled pairwise divergence distribution from 23 simulations, each of which has an expected final number of elements equal to the size of 1 of the 23 *Wolbachia* IS groups comprising at least two alignable IS copies.

Siguier, Perochon, et al. 2006). Except for the IS6 and IS200/605 families that are specific to *w*Pel, all other IS families (9/11 or 82%) are shared by at least three of the four *Wolbachia* genomes (table 1). However, there are significant differences in the distribution of IS families among the various *Wolbachia* genomes (chi-square test, $P < 10^{-16}$) (table 1). Because IS families contain large numbers of heterogeneous IS types (Chandler and Mahillon 2002; Siguier, Perochon, et al. 2006), we refined our analysis by classifying all *Wolbachia* IS copies into 1 of 33 IS groups (see "Materials and Methods"). The group-level analysis confirmed the family-level analysis, in that most IS groups (27/33 or 82%) are shared by multiple genomes. However, the most frequent IS groups per genome are largely specific to each genome, and globally, the distribution of IS groups is significantly different among the various *Wolbachia* genomes (chi-square test, $P < 10^{-16}$) (supplementary table S1, Supplementary Material online).

Overall, these results demonstrate that *Wolbachia* genomes qualitatively carry IS elements from the same families and groups, but they substantially differ in the families and groups that mostly contributed to their IS genomic landscapes. This is particularly striking for the *w*Mel and *w*Ri genomes, which are phylogenetically closely related (Wu et al. 2004; Klasson et al. 2009). But the two genomes display very different IS profiles in terms of copy number, genomic coverage, and most frequent families and groups, despite the fact that *w*Mel and *w*Ri virtually possess IS elements from the same families and groups (table 1; supplementary table S1, Supplementary Material online).

Diversity of Potentially Functional IS Copies in *Wolbachia*

To investigate the causes of differential IS abundance and distribution among *Wolbachia* genomes, we searched for potentially functional IS copies, defined as full-length copies with intact transposase genes. This analysis revealed that the *w*Mel, *w*Ri, and *w*Pel genomes possess 20–64 potentially functional IS copies (table 2). Each of these *Wolbachia* genomes possesses at least one potentially functional copy from six different IS groups. Overall, there is an important diversity of potentially functional IS copies in *Wolbachia*, belonging to 14 different IS groups from 9 different IS families. Not surprisingly, IS groups exhibiting the highest numbers of potentially functional copies within individual *Wolbachia* genomes are also the IS groups with the highest overall copy numbers. Altogether, these results suggest that multiple IS copy number expansions have taken place during recent

Table 1
Distribution of IS Elements in Four Completely Sequenced *Wolbachia* Genomes

IS Family	*w*Bm[a]	*w*Mel[a]	*w*Ri[a]	*w*Pel[a]
IS3		18 (17)	15 (9)	2 (1)
IS4	3 (6)	13 (12)	12 (7)	5 (3)
IS5	7 (13)	26 (25)	30 (17)	26 (15)
IS6				11 (6)
IS110	21 (40)	18 (17)	39 (23)	6 (4)
IS200/605				5 (3)
IS256	1 (2)	3 (3)	2 (1)	32 (19)
IS481	3 (6)	4 (4)	12 (7)	5 (3)
IS630	3 (6)	7 (7)	12 (7)	15 (9)
IS982	1 (2)	2 (2)	3 (2)	54 (32)
IS66	13 (25)	14 (13)	46 (27)	9 (5)
Total IS copy number	52	105	171	170
Total IS family number	8	9	9	11
IS density (copies/Mb)	48	83	118	115
IS genomic coverage (bp)	28,188	76,886	159,767	123,823
IS genomic proportion	2.6	6.1	11.0	8.4

[a]The first number relates to observed copy number. The number in brackets indicates the proportion of the IS family among all IS copies in the genome.

Wolbachia evolution as a result of IS group-specific and genome-specific expansions.

The ISWen2 group in *w*Ri provides an excellent example of a group-specific expansion in a specific *Wolbachia* genome (fig. 2*a*). ISWen2 copies were identified only in the closely related *w*Mel and *w*Ri genomes. Despite the fact that both genomes carry at least one potentially functional ISWen2 copy, *w*Ri carries as many as 23 ISWen2 copies, whereas *w*Mel carries only 3 copies. To investigate the recent amplification dynamics of the ISWen2 group in *w*Ri, we performed a phylogenetic analysis using the median-joining network approach, as implemented in the Network ver 4.5.1.6 software (Bandelt et al. 1999; Cordaux, Hedges, et al. 2004). The ISWen2 group network displays a star-like structure in which 35% of the copies fall in the central, most likely ancestral, node (I in fig. 2*b*). Interestingly, several peripheral nodes in the network are not directly connected to the central node (III.1, III.2, and III.3 in fig. 2*b*) or encompass

Table 2
Distribution of Potentially Functional IS Copies Inserted in Four *Wolbachia* Genomes

IS Family	IS Group	*w*Bm	*w*Mel	*w*Ri	*w*Pel
IS4	ISWen1		3		
IS5	ISWpi1		13	20	
IS110	ISWen2		1	16	
	ISWpi12				3
	ISWpi13			1	
	ISWpi14		1		
IS200/605	ISW1				2
IS256	ISWpi15		1		6
IS481	ISWpi2			5	
	ISWpi4		1	1	
IS630	ISWpi11				4
	ISWpi10				1
IS982	ISWpi16				44
IS66	ISWen3			21	
Number of potentially functional IS copies		0	20	64	60
Proportion (%)		0	19	37	35
Number of nonfunctional IS copies		52	85	107	110
Proportion (%)		100	81	63	65

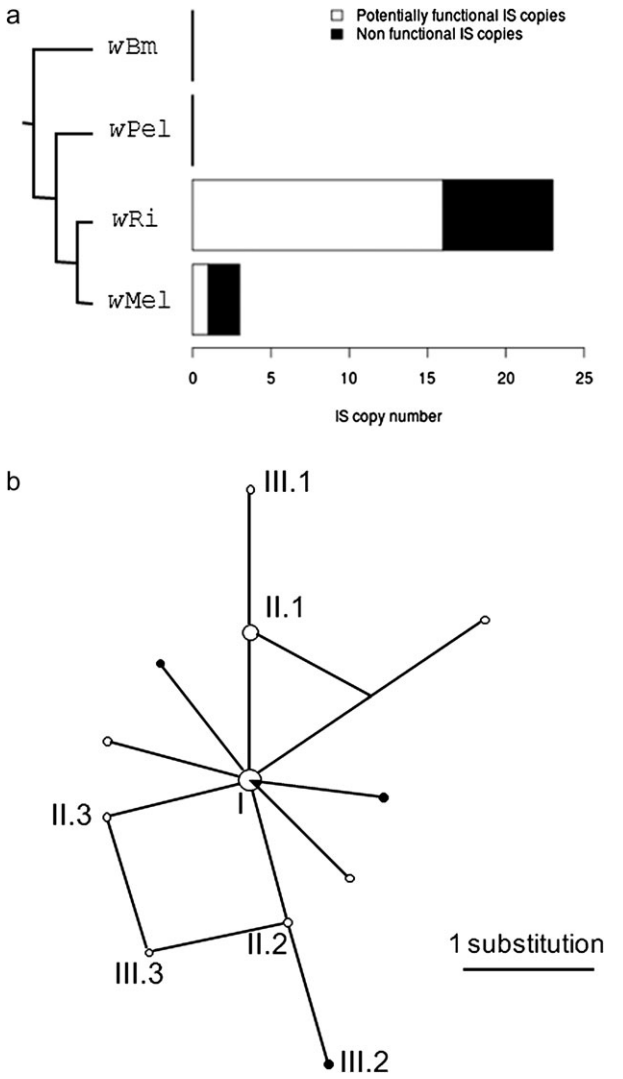

Fig. 2.—Expansion of the ISWen2 group in the *w*Ri genome. (*a*) Copy number of the ISWen2 group in four completely sequenced *Wolbachia* genomes. Branch lengths of the phylogenetic tree are arbitrary. (*b*) Median-joining network of the 20 full-length ISWen2 copies from the *w*Ri genome. Circles denote IS sequence types (nodes). Nodes discussed in the main text were labeled I, II.1–3, and III.1–3. Node size is proportional to IS copy number: $n = 1$ for all nodes except nodes I ($n = 7$) and II.1 ($n = 3$). Lines denote substitution steps, with a one-step distance being indicated in the lower right corner. Potentially functional and nonfunctional copies are shown in white and black, respectively.

Table 3
Insertion Presence/Absence Polymorphism Patterns of 44 IS Copies in the *w*Mel and *w*Ri *Wolbachia* Genomes

IS Family	IS Group	IS Copies Specific to *w*Mel	IS Copies Specific to *w*Ri	IS Copies Shared by *w*Mel and *w*Ri
IS3	IS3			12
IS4	IS4-wB			6
IS5	IS903	1		
	IS1031			1
	ISWpi1	6	4	
IS110	IS1111			1
	ISWen2		4	
	ISWpi12			1
	IS110-wA			1
IS481	ISWpi2		1	
	ISWpi4			1
IS630	ISWpi11	1		
IS982	ISWpi16			1
IS66	ISWen3		2	1
Number of IS copies		8	11	25
Proportion of IS copies that are potentially functional (%)		75	91	4

several ISWen2 sequences (II.1 in fig. 2*b*). Given that bacterial IS elements generally exhibit strong cis preference for transpositional activity (i.e., preferential interaction of a transposase with the element from which it is expressed) (Chandler and Mahillon 2002; Nagy and Chandler 2004) and assuming that homoplasy is negligible at this phylogenetic depth, we conclude that the ISWen2 expansion in *w*Ri may have been mediated by at least 3 IS copies (at least one copy from nodes I and II.1 and the copy at the node II.2) and at most 11 copies (from nodes I, II.1, II.2, and II.3). These results suggest that 15–55% of the copies may have contributed to ISWen2 expansion during recent *w*Ri evolution. It is generally thought that multiple copies may contribute to the expansion of DNA transposon families (Deininger and Batzer 1993; Robertson 2002). To our knowledge, our analysis provides the first quantitative estimate of the proportion of "source" copies that may have contributed to the expansion of a DNA transposon family.

Intense and Global IS Transpositional Activity

We have previously shown that one IS group (ISWpi1) is widespread among *Wolbachia* strains, but individual ISWpi1 copies are inserted at given genomic loci in a single or very few closely related *Wolbachia* strains (Cordaux 2008; Cordaux et al. 2008). Such insertion presence/absence polymorphism patterns demonstrate intense ISWpi1 transpositional activity during recent *Wolbachia* evolution (Cordaux 2008; Cordaux et al. 2008). To investigate whether this evolutionary trend can be extended to other IS groups, we searched for IS insertion presence/absence polymorphisms at orthologous sites for all full-length IS elements inserted in the *w*Mel and *w*Ri genomes. We found 19 IS copies specifically inserted in *w*Mel or *w*Ri out of 44 unambiguously orthologous loci identified between *w*Mel and *w*Ri (table 3). The 19 polymorphic insertions encompass six different IS groups (including ISWpi1) from five different IS families, indicating that multiple IS groups have been transpositionaly active during recent *Wolbachia* evolution. The fact that 84% (16/19) of polymorphic insertions versus only 4% (1/25) of shared insertions by *w*Mel and *w*Ri are potentially functional IS copies (table 3) further corroborates the recent origin of the polymorphic IS copies (i.e., they have not resided in the genomes long time enough as to accumulate inactivating mutations). This result also suggests that IS transpositional activity may be ongoing in these *Wolbachia* genomes.

Exceptional Amount of Nonfunctional Copies in *Wolbachia* Genomes

The analysis of IS copy structure revealed that the *w*Mel, *w*Ri, and *w*Pel *Wolbachia* genomes contain several tens of potentially functional IS copies (table 2). However, these copies only account for a small fraction of all IS copies inserted in *Wolbachia* genomes. In fact, 71% of all IS elements inserted in *Wolbachia* genomes are nonfunctional (table 2). Nonfunctional IS copies are defined here as full-length copies with pseudogenized transposase genes or non–full-length copies (i.e., truncated copies and fragments). In the most extreme case, all 52 IS copies inserted in the *w*Bm genome are nonfunctional (Cordaux 2009). Given that IS elements are usually considered to be of recent origin in bacterial genomes and subject to rapid turnover (Wagner 2006; Wagner et al. 2007; Rocha 2008), the occurrence of so many disrupted and degraded IS elements in *Wolbachia* genomes is all the more surprising.

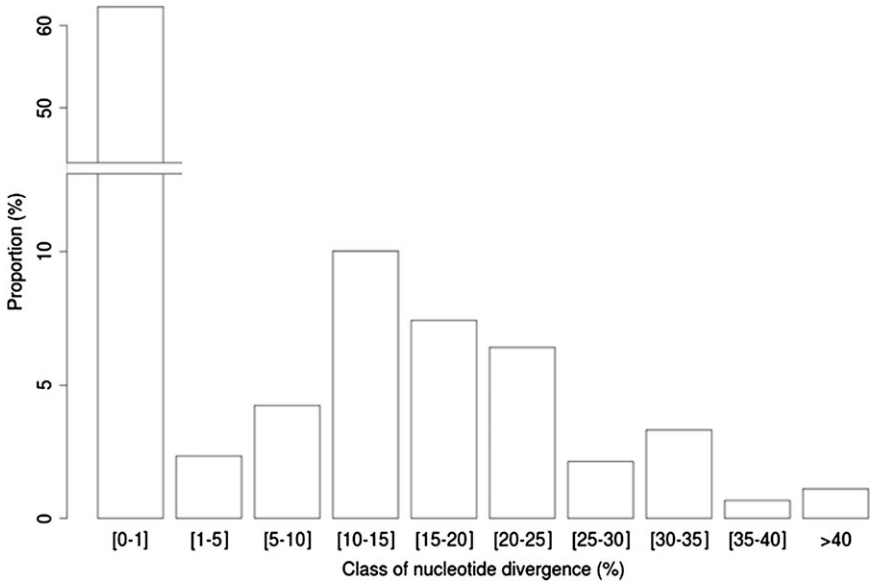

Fig. 3.—Frequency distribution of pairwise IS nucleotide divergence for four *Wolbachia* genomes. IS copies from 23 IS groups comprising at least two alignable IS copies are considered ($n = 454$). The distribution is based on a total of 4,312 pairwise comparisons (169 for *w*Bm, 379 for *w*Mel, 2152 for *w*Pel, and 1612 for *w*Ri).

Investigating Long-term Evolutionary Dynamics of IS Elements

The many degraded IS copies in *Wolbachia* genomes offer a unique opportunity to directly investigate the long-term evolutionary dynamics of bacterial TEs. We analyzed 454 copies from the 23 IS groups comprising at least two copies with alignable sequences from the four *Wolbachia* genomes. For each IS group, we calculated intragenomic nucleotide divergence between pairs of IS copies. The majority of pairwise comparisons exhibited <1% nucleotide divergence (fig. 3). This is consistent with a recent origin of these IS copies and the high copy numbers of recently expanded IS groups, which generates many pairwise comparisons with no or very low divergence. In addition, we found that in the four *Wolbachia* genomes, 22–47% of the pairwise comparisons displayed at least 10% nucleotide divergence (fig. 3). This indicates that *Wolbachia* genomes contain an important amount of ancient IS copies that are witnesses of past IS expansions during *Wolbachia* evolution.

Interestingly, the distribution of IS copies is bimodal, with a first peak corresponding to identical or nearly identical IS copies (<1% divergence) and a second peak at 10–15% divergence (fig. 3). Importantly, this pattern holds when *Wolbachia* genomes are analyzed separately (supplementary fig. S1, Supplementary Material online). This demonstrates that the global bimodal pattern cannot be ascribed to an artifact due to pooling data from multiple genomes that would exhibit different individual distribution patterns. To explore the evolutionary causes of this bimodal distribution, we simulated the evolution of an IS population in a haploid (bacterial) genome under different scenarios. Four major processes of IS evolution were considered: acquisition by horizontal transfer, copy number expansion from a resident copy, degradation through random substitutions, and loss (see "Materials and Methods"). These processes were combined to test five different evolutionary scenarios differing in the tempo of IS acquisitions and bursts.

The first two scenarios simulated a single IS acquisition immediately followed by a sudden burst, at the start (scenario 1) or near the end (scenario 2) of the simulation. We observed a single peak in both simulations, with high divergence and variance for the scenario of ancient IS acquisition and burst (scenario 1) and low divergence and variance for the scenario of recent IS acquisition and burst (scenario 2) (fig. 1). Next, scenario 3 simulated a constant but low transpositional activity for the duration of the simulation, following an initial IS copy acquisition. The resulting distribution showed a single peak at high divergence skewed toward lower divergence (fig. 1). Finally, scenarios 4 and 5 each simulated two independent IS acquisitions and bursts: Both were recent in scenario 4 (i.e., corresponding to scenario 2 repeated twice), and one was ancient and one recent in scenario 5 (i.e., combining scenarios 1 and 2). Scenario 5 is analogous to the model of recurrent horizontal transfers and bursts proposed in the literature (Wagner 2006). The distribution pattern resulting from two recent IS acquisitions bursts (scenario 4) showed multiple peaks with low variance at irregular divergence levels (fig. 1). Interestingly, scenario 5 was the only one that displayed a clear bimodal distribution with a flat peak at high divergence (corresponding to the ancient IS acquisition and burst) and

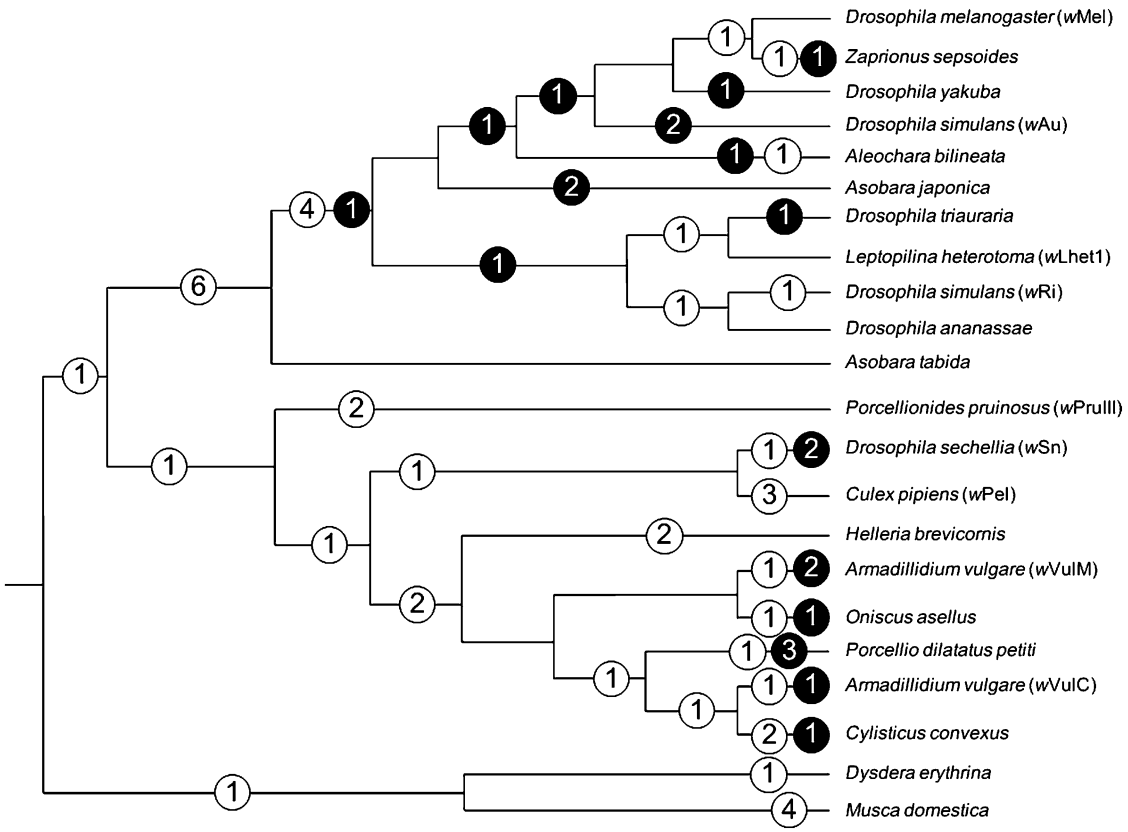

Fig. 4.—History of IS group acquisitions and losses in 22 *Wolbachia* strains. The most parsimonious distribution of acquisitions (white circles) and losses (black circles) of 17 IS groups according to the phylogenetic relationships of 22 *Wolbachia* strains is shown. Numbers of acquisitions and losses are indicated in the circles. Phylogenetic relationships between *Wolbachia* strains are adapted from Cordaux et al. (2001, 2008) and Lo et al. (2007); branch lengths of the phylogenetic tree are arbitrary. *Wolbachia* strains are named after the host species from which they were isolated.

a straight peak at low divergence (corresponding to the recent IS acquisition and burst) (fig. 1), as observed for *Wolbachia* IS elements (fig. 3). Note that higher deletion rates resulted in lower amounts of ancient copies in scenario 5, but the second peak corresponding to ancient copies was apparent whatever the deletion rate (supplementary fig. S2, Supplementary Material online).

Overall, the bimodal distribution indicates that IS transposition activity in *Wolbachia* genomes was not constant over time. Instead, most *Wolbachia* IS elements may have been generated in at least two major periods of intense transpositional activity, including an ancient expansion embodied by ~15% divergent IS copies and a very recent (and perhaps ongoing) expansion corresponding to identical or nearly identical IS copies. Our simulations also emphasize that multiple horizontal IS acquisitions are required to explain the observed distribution pattern.

Frequent IS Horizontal Transfers among *Wolbachia* Strains

To test whether *Wolbachia* IS dynamics is intimately linked with frequent horizontal transmission among strains, as suggested by our simulations above and by evidence from the ISWpi1 group (Cordaux et al. 2008), we screened a panel of 22 diverse *Wolbachia* strains (supplementary table S2, Supplementary Material online) for the presence of 17 IS groups using group-specific PCR detection assays and verification by sequencing (supplementary table S4, Supplementary Material online). To evaluate the reliability of our PCR assays, we compared PCR amplification results in our *w*Mel, *w*Ri, and *w*Pel DNA samples with genome sequence predictions. We obtained results in agreement with expectations in 47 of 51 combinations tested (i.e., 17 IS groups in three reference strains). For the four cases in which conflicting results were recorded (i.e., no PCR result, although an amplification is predicted based on presence of at least one copy of the tested IS group in the genome sequence), average sequence divergence between IS copies among genomes was high in all cases (>13%). Therefore, we conclude that our PCR detection assays are generally highly reliable, at least for copies with low to moderate divergence, hence enabling us to confidently detect recent events of IS horizontal transfers that may have occurred between *Wolbachia* strains.

By mapping IS group distribution onto a phylogeny of the *Wolbachia* strains, we inferred the most parsimonious scenario of IS group acquisitions and losses during the

evolutionary history of the 22 *Wolbachia* strains under investigation (fig. 4). This analysis revealed that all investigated *Wolbachia* genomes (encompassing three different supergroups) possess IS elements from at least three different IS groups. Our large-scale screening thus demonstrates that IS elements are broadly albeit patchily distributed among *Wolbachia* strains. Thus, IS represents a general feature of *Wolbachia* strains, not merely a characteristic of the few sequenced genomes.

Our results also indicate that the presence of the 17 IS groups in the 22 *Wolbachia* strains requires at least 44 independent acquisitions at this level of resolution (fig. 4). This is most likely a very conservative estimate because 1) we favored the scenario minimizing the number of acquisitions when several equiparsimonious scenarios were possible for particular IS groups; 2) we assumed that individual *Wolbachia* strains or monophyletic groups of *Wolbachia* strains possessing a given IS group resulted from a single ancestral acquisition; however, multiple independent acquisitions could also explain such distribution patterns, as previously shown for ISWpi1 (Cordaux et al. 2008); and 3) a larger screening of *Wolbachia* strains for IS group presence might uncover additional acquisition events.

Such a patchy distribution strongly suggests that horizontal transfers of IS copies occur frequently in *Wolbachia*. This is further substantiated by the fact that >60% of the pairwise comparisons of IS sequences obtained by sequencing of PCR fragments displayed nucleotide divergence <1% (supplementary fig. S3, Supplementary Material online). As this is lower than the divergence between most of the analyzed *Wolbachia* strains, such IS groups are unlikely to have been vertically inherited from a common ancestor. Based on patchy distribution of IS groups and generally high similarity of IS sequences between strains, we conclude that horizontal transmission is a major determinant of the current IS distribution in *Wolbachia* strains.

The frequent horizontal transmission of IS elements across *Wolbachia* strains poses the question of the underlying mechanisms of these transfers. The intracellular confinement of bacterial endosymbionts is generally thought to limit exchange of genetic material with other bacterial populations or species (Wernegreen 2002; Moran et al. 2008; Moya et al. 2008). However, a distinguishing feature of *Wolbachia* endosymbionts is their propensity to switch between arthropod hosts (Vavre et al. 1999; Cordaux et al. 2001). Such dynamics favors the occasional co-occurrence of divergent *Wolbachia* strains within the same host cells, either stably or transiently (Vavre et al. 1999; Bordenstein and Wernegreen 2004; Verne et al. 2007). In addition to physical proximity, exchange of genetic material between *Wolbachia* strains might be facilitated by the presence of bacteriophages in many *Wolbachia* endosymbionts (Bordenstein and Wernegreen 2004; Braquart-Varnier et al. 2005; Tanaka et al. 2009) that might serve as shuttles for transferring IS elements among strains. We identified a total of 15 IS elements from 9 different IS groups, including 11 potentially functional copies, inserted in the 12 prophages integrated in the *w*Mel, *w*Ri, and *w*Pel genomes (table 4). However, it is unclear whether these IS copies inserted in prophage genomes following phage integration into *Wolbachia* genomes or the IS elements were already present in bacteriophage genomes and were imported in *Wolbachia* genomes during bacteriophage genome integration. Nevertheless, a potentially functional ISWpi12 copy is inserted in the genome of the active bacteriophage WOcauB2 of the *w*CauB *Wolbachia* strain (Tanaka et al. 2009), whereas bacteriophage genomes most generally lack IS elements (Leclercq and Cordaux 2011). This is consistent with the notion that bacteriophages might be able to shuttle IS elements between *Wolbachia* strains.

Table 4
Distribution of IS Copies Inserted in Prophage Regions of Three *Wolbachia* Genomes

IS Family	IS Group	*w*Mel[a]	*w*Ri[a]	*w*Pel[a]
IS5	IS1031			1 (0)
	ISWpi1	1 (1)	3 (3)	
IS110	ISWen2	1 (1)	2 (2)	
	ISWpi12	1 (0)		1 (1)
	ISWpi13		1 (1)	
	ISWpi14	1 (1)		
IS256	ISWpi15			1 (0)
IS630	ISWpi10		1 (0)	
IS982	ISWpi16			1 (1)
Number of IS copies		4 (3)	7 (6)	4 (2)
Number of prophage regions		3	4	5

[a]Number of potentially functional IS copies shown in brackets.

Conclusions

Our analyses highlighted the patchy distribution of IS groups in *Wolbachia* genomes. The identification of multiple IS groups experiencing independent copy number expansions in different *Wolbachia* genomes is notable because it suggests that IS expansions may occur simultaneously in different genomes (i.e., *w*Mel, *w*Ri, and *w*Pel) through a global activation of transposition. This synchronicity may be linked to the high rate of recent IS horizontal transfers we identified in *Wolbachia* strains. Nevertheless, the evolutionary success of IS families and groups within genomes is highly variable, indicating that horizontal transfer is a necessary but not sufficient condition to IS proliferation. The apparently stochastic loss or success of individual IS families or groups within bacterial strains following import by horizontal transfer may be the result of a complex interplay between various parameters, such as IS intrinsic transpositional efficiency, cellular factors involved in transpositional control, and genomic environment (Chandler and Mahillon 2002; Nagy and

Chandler 2004; Cerveau et al. 2011). However, such targeted effects can hardly explain a global activation of transposition simultaneously involving multiple IS families and groups. This suggests that population-level effects may also play a role in the evolutionary dynamics of bacterial IS elements.

Remarkably, our results show that *Wolbachia* genomes contain an important archive of past IS evolution, as the vast majority of *Wolbachia* IS copies actually are more or less severely degraded. The rich IS fossil record buried in *Wolbachia* genomes provides direct empirical evidence for a long-term evolutionary dynamics of IS elements following a scenario of cyclic bursts of transposition separated by periods of relative transpositional quiescence as previously suggested based on the analysis of exclusively recent IS copies (Wagner 2006; Wagner et al. 2007). This raises the question whether the abundance of IS fossils is specific to *Wolbachia* genomes or it may be a general, albeit overlooked, feature of prokaryote genomes. Unfortunately, IS annotation is rarely optimal in completely sequenced prokaryotic genomes, and currently, it is often limited at best to identification of potentially functional transposase genes (Varani et al. 2011). Therefore, it is possible that many other prokaryote genomes carry an abundance of IS relics, but they cannot be detected using standard annotation procedures. In any event, our detailed analysis of IS elements in *Wolbachia* bacteria shows that comprehensive TE annotations have the potential to uncover unexpected patterns of prokaryote genome evolution. Indeed, the identification of an IS fossil genomic record in *Wolbachia* demonstrates that IS elements are not always of recent origin, contrary to the conventional view of TE evolution in prokaryote genomes.

Supplementary Material

Supplementary tables S1–S4 and supplementary figures S1–S3 are available at *Genome Biology and Evolution* online (http://www.gbe.oxfordjournals.org/).

Acknowledgments

We are grateful to Mylène Weill, Fabrice Vavre, Hervé Merçot, and Denis Poinsot for providing samples. We thank Mick Chandler and Patricia Siguier for access to ISsaga prior to publication and discussions and Daniel Guyonnet for technical assistance. This research was funded by a Young Investigator ATIP Award from the Centre National de la Recherche Scientifique (CNRS) to R.C., a European Research Council Starting Grant (FP7/2007-2013 grant 260729 EndoSexDet) to R.C., and the Research Interdisciplinary Programme "Infectious Diseases and Environment" from the CNRS. N.C. was supported by a PhD fellowship from the French Ministère de l'Education Nationale, de l'Enseignement Supérieur et de la Recherche. S.L. was supported by a postdoctoral fellowship from the CNRS.

Literature Cited

Bandelt HJ, Forster P, Rohl A. 1999. Median-joining networks for inferring intraspecific phylogenies. Mol Biol Evol. 16(1):37–48.

Bichsel M, Barbour AD, Wagner A. 2010. The early phase of a bacterial insertion sequence infection. Theor Popul Biol. 78(4):278–288.

Bordenstein SR, Reznikoff WS. 2005. Mobile DNA in obligate intracellular bacteria. Nat Rev Microbiol. 3(9):688–699.

Bordenstein SR, Wernegreen JJ. 2004. Bacteriophage flux in end (Wolbachia): infection frequency, lateral transfer, and recombination rates. Mol Biol Evol. 21(10):1981–1991.

Bouchon D, Cordaux R, Grève P. 2008. Feminizing Wolbachia and the evolution of sex determination in isopods. In: Bourtzis K, Miller T, editors. Insect Symbiosis. Insect Symbiosis, Vol. 3. Boca Raton (FL): Taylor and Francis Group LLC. p. 273–294.

Braquart-Varnier C, Greve P, Felix C, Martin G. 2005. Bacteriophage WO in Wolbachia infecting terrestrial isopods. Biochem Biophys Res Commun. 337:580–585.

Brugger K, Torarinsson E, Redder P, Chen L, Garrett RA. 2004. Shuffling of Sulfolobus genomes by autonomous and non-autonomous mobile elements. Biochem Soc Trans. 32(Pt 2):179–183.

Cerveau N, Leclercq S, Bouchon D, Cordaux R. 2011. Evolutionary dynamics and genomic impact of prokaryote transposable elements. In: Pontarotti P, editor. Evolutionary biology: concepts, biodiversity, macroevolution and genome evolution. Berlin (Germany): Springer. p. 291–312.

Chandler M, Mahillon J. 2002. Insertion sequences revisited. In: Craig NL, Craig R, Gellert M, Lambowitz AM, editors. Mobile DNA II. Washington (DC): ASM Press. p. 305–366.

Cho NH, et al. 2007. The Orientia tsutsugamushi genome reveals massive proliferation of conjugative type IV secretion system and host-cell interaction genes. Proc Natl Acad Sci U S A. 104(19):7981–7986.

Cordaux R. 2008. ISWpi1 from Wolbachia pipientis defines a novel group of insertion sequences within the IS5 family. Gene 409(1–2):20–27.

Cordaux R. 2009. Gene conversion maintains nonfunctional transposable elements in an obligate mutualistic endosymbiont. Mol Biol Evol. 26(8):1679–1682.

Cordaux R, Batzer MA. 2009. The impact of retrotransposons on human genome evolution. Nat Rev Genet. 10(10):691–703.

Cordaux R, Bouchon D, Greve P. 2011. The impact of endosymbionts on the evolution of host sex-determination mechanisms. Trends Genet. 27:332–341.

Cordaux R, Hedges DJ, Batzer MA. 2004. Retrotransposition of Alu elements: how many sources? Trends Genet. 20(10):464–467.

Cordaux R, Lee J, Dinoso L, Batzer MA. 2006. Recently integrated Alu retrotransposons are essentially neutral residents of the human genome. Gene 373:138–144.

Cordaux R, Michel-Salzat A, Bouchon D. 2001. Wolbachia infection in crustaceans: novel hosts and potential routes for horizontal transmission. J Evol Biol. 14(2):237–243.

Cordaux R, Michel-Salzat A, Frelon-Raimond M, Rigaud T, Bouchon D. 2004. Evidence for a new feminizing Wolbachia strain in the isopod Armadillidium vulgare: evolutionary implications. Heredity 93(1):78–84.

Cordaux R, et al. 2008. Intense transpositional activity of insertion sequences in an ancient obligate endosymbiont. Mol Biol Evol. 25(9):1889–1896.

Deininger PL, Batzer MA. 1993. Evolution of retroposons. Evol Biol. 27:157–196.

Emerman M, Malik HS. 2010. Paleovirology—modern consequences of ancient viruses. PLoS Biol. 8(2):e1000301.

Feschotte C, Pritham EJ. 2007. DNA transposons and the evolution of eukaryotic genomes. Annu Rev Genet. 41:331–368.
Filee J, Siguier P, Chandler M. 2007. Insertion sequence diversity in archaea. Microbiol Mol Biol Rev. 71(1):121–157.
Foster J, et al. 2005. The Wolbachia genome of Brugia malayi: endosymbiont evolution within a human pathogenic nematode. PLoS Biol. 3(4):e121.
Gilbert C, Feschotte C. 2010. Genomic fossils calibrate the long-term evolution of hepadnaviruses. PLoS Biol. 8(9):e1000495.
Hall TA. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symp Ser. 41:95–98.
Kapitonov VV, Jurka J. 2003. Molecular paleontology of transposable elements in the Drosophila melanogaster genome. Proc Natl Acad Sci U S A. 100(11):6569–6574.
Khan H, Smit A, Boissinot S. 2006. Molecular evolution and tempo of amplification of human LINE-1 retrotransposons since the origin of primates. Genome Res. 16(1):78–87.
Klasson L, et al. 2008. Genome evolution of Wolbachia strain wPip from the Culex pipiens group. Mol Biol Evol. 25(9):1877–1887.
Klasson L, et al. 2009. The mosaic genome structure of the Wolbachia wRi strain infecting Drosophila simulans. Proc Natl Acad Sci U S A. 106(14):5725–5730.
Kumar S, Nei M, Dudley J, Tamura K. 2008. MEGA: a biologist-centric software for evolutionary analysis of DNA and protein sequences. Brief Bioinform. 9(4):299–306.
Kuo CH, Ochman H. 2010. The extinction dynamics of bacterial pseudogenes. PLoS Genet. 6(8):e1001050.
Lander ES, et al. 2001. Initial sequencing and analysis of the human genome. Nature 409(6822):860–921.
Lawrence JG, Ochman H, Hartl DL. 1992. The evolution of insertion sequences within enteric bacteria. Genetics 131(1):9–20.
Leclercq S, Cordaux R. 2011. Do phages efficiently shuttle transposable elements among prokaryotes? Evolution. Advance Access published September 13, 2011, doi:10.1111/j.1558-5646.2011.01395.x.
Leclercq S, Giraud I, Cordaux R. 2011. Remarkable abundance and evolution of mobile group II introns in Wolbachia bacterial endosymbionts. Mol Biol Evol. 28(1):685–697.
Lo N, et al. 2007. Taxonomic status of the intracellular bacterium Wolbachia pipientis. Int J Syst Evol Microbiol. 57(Pt 3):654–657.
Moon S, Byun Y, Kim HJ, Jeong S, Han K. 2004. Predicting genes expressed via −1 and +1 frameshifts. Nucleic Acids Res. 32(16):4884–4892.
Moran NA, McCutcheon JP, Nakabachi A. 2008. Genomics and evolution of heritable bacterial symbionts. Annu Rev Genet. 42:165–190.
Moran NA, Plague GR. 2004. Genomic changes following host restriction in bacteria. Curr Opin Genet Dev. 14(6):627–633.
Moya A, Pereto J, Gil R, Latorre A. 2008. Learning how to live together: genomic insights into prokaryote-animal symbioses. Nat Rev Genet. 9(3):218–229.
Nagy Z, Chandler M. 2004. Regulation of transposition in bacteria. Res Microbiol. 155(5):387–398.
Nakayama K, et al. 2008. The whole-genome sequencing of the obligate intracellular bacterium Orientia tsutsugamushi revealed massive gene amplification during reductive genome evolution. DNA Res. 15(4):185–199.
Pace JK 2nd, Feschotte C. 2007. The evolutionary history of human DNA transposons: evidence for intense activity in the primate lineage. Genome Res. 17(4):422–432.
Parkhill J, et al. 2003. Comparative analysis of the genome sequences of Bordetella pertussis, Bordetella parapertussis and Bordetella bronchiseptica. Nat Genet. 35(1):32–40.
Price AL, Jones NC, Pevzner PA. 2005. De novo identification of repeat families in large genomes. Bioinformatics 21(Suppl 1):i351–i358.
Qiu N, et al. 2010. Prevalence and diversity of insertion sequences in the genome of Bacillus thuringiensis YBT-1520 and comparison with other Bacillus cereus group members. FEMS Microbiol Lett. 310(1):9–16.
Robertson HM. 2002. Evolution of DNA transposons in eukaryotes. In: Craig NL, Craig R, Gellert M, Lambowitz AM, editors. Mobile DNA II. Washington (DC): ASM Press. p.1093–1110.
Rocha EP. 2008. Evolutionary patterns in prokaryotic genomes. Curr Opin Microbiol. 11(5):454–460.
Rozen S, Skaletsky H. 2000. Primer3 on the WWW for general users and for biologist programmers. Methods Mol Biol. 132:365–386.
Saridaki A, Bourtzis K. 2010. Wolbachia: more than just a bug in insects genitals. Curr Opin Microbiol. 13(1):67–72.
Sawyer SA, et al. 1987. Distribution and abundance of insertion sequences among natural isolates of Escherichia coli. Genetics 115(1):51–63.
Schnable PS, et al. 2009. The B73 maize genome: complexity, diversity, and dynamics. Science 326(5956):1112–1115.
Siguier P, Filee J, Chandler M. 2006. Insertion sequences in prokaryotic genomes. Curr Opin Microbiol. 9(5):526–531.
Siguier P, Perochon J, Lestrade L, Mahillon J, Chandler M. 2006. ISfinder: the reference centre for bacterial insertion sequences. Nucleic Acids Res. 34(Database issue):D32–D36.
Tanaka K, Furukawa S, Nikoh N, Sasaki T, Fukatsu T. 2009. Complete WO phage sequences reveal their dynamic evolutionary trajectories and putative functional elements required for integration into the Wolbachia genome. Appl Environ Microbiol. 75(17):5676–5686.
Touchon M, Rocha EP. 2007. Causes of insertion sequences abundance in prokaryotic genomes. Mol Biol Evol. 24(4):969–981.
Varani AM, Siguier P, Gourbeyre E, Charneau V, Chandler M. 2011. ISsaga is an ensemble of web-based methods for high throughput identification and semi-automatic annotation of insertion sequences in prokaryotic genomes. Genome Biol. 12(3):R30.
Vavre F, Fleury F, Lepetit D, Fouillet P, Bouletreau M. 1999. Phylogenetic evidence for horizontal transmission of Wolbachia in host-parasitoid associations. Mol Biol Evol. 16(12):1711–1723.
Verne S, Johnson M, Bouchon D, Grandjean F. 2007. Evidence for recombination between feminizing Wolbachia in the isopod genus Armadillidium. Gene 397(1–2):58–66.
Wagner A. 2006. Periodic extinctions of transposable elements in bacterial lineages: evidence from intragenomic variation in multiple genomes. Mol Biol Evol. 23(4):723–733.
Wagner A, Lewis C, Bichsel M. 2007. A survey of bacterial insertion sequences using IScan. Nucleic Acids Res. 35(16):5284–5293.
Wernegreen JJ. 2002. Genome evolution in bacterial endosymbionts of insects. Nat Rev Genet. 3(11):850–861.
Werren JH, Zhang W, Guo LR. 1995. Evolution and phylogeny of Wolbachia: reproductive parasites of arthropods. Proc R Soc Lond Ser B Biol Sci. 261(1360):55–63.
Wu M, et al. 2004. Phylogenomics of the reproductive parasite Wolbachia pipientis wMel: a streamlined genome overrun by mobile genetic elements. PLoS Biol. 2(3):E69.
Xing J, et al. 2004. Alu element mutation spectra: molecular clocks and the effect of DNA methylation. J Mol Biol. 344(3):675–682.
Yang F, et al. 2005. Genome dynamics and diversity of Shigella species, the etiologic agents of bacillary dysentery. Nucleic Acids Res. 33(19):6445–6458.

Associate editor: Emmanuelle Lerat

Résumé

Les éléments transposables (ET) sont l'une des principales forces guidant l'évolution des génomes. La présence de copies dégradées a permis de bien caractérisée la dynamique des ET eucaryotes. A contrario, chez les procaryotes, les ET sont considérés comme récents, ce qui complique l'étude de leur dynamique. Les séquences d'insertion (IS) sont les ET procaryotes les plus abondants. Les modèles prédisent que les IS, arrivés par transferts horizontaux, subissent une forte augmentation de leur nombre, puis sont éliminés. De plus, les modèles prédisent que les génomes des bactéries intracellulaires devraient avoir peu ou pas d'IS.

Le séquençage des génomes de bactéries intracellulaires obligatoires, comme *Wolbachia*, a remis en cause les modèles, car certains ont une grande quantité d'IS. Notre travail portait sur l'étude des génomes de cinq souches de *Wolbachia* ayant des caractéristiques diverses. Nous avons réalisé une annotation détaillée des IS pour chaque génome et testé nos hypothèses basés issues de l'analyse *in silico* sur un panel de souches.

Nous avons confirmé que les génomes de *Wolbachia* ont une forte abondance d'IS. La majorité des copies d'IS étaient dégradées, ce qui a permis d'étudier leur dynamique. L'activité passée des IS de *Wolbachia* n'a pas été constante au cours du temps avec des alternances de phases de forte activité et de quiescence. Les phases de forte activité doivent être précédées de transferts horizontaux, qui ont été expérimentalement détectés en abondance. Enfin, des analyses d'expression suggèrent que l'activité des IS n'est pas uniquement contrôlée par les éléments eux-mêmes, mais dépend également de l'environnement génomique des copies.

Mots clés : éléments transposables, séquences d'insertion, dynamique évolutive, procaryotes, *Wolbachia*

Abstract

Transposable elements (TE) are one of the major driving forces of genome evolution. Eukaryotic TE evolution can easily be reconstructed thanks to many degraded copies. By contrast, in prokaryotes, TE are considered to be recently acquired, which complicates the study of their dynamics. Insertion sequences (IS) are the most abundant TE in prokaryotes. Models predict a high turn-over that starts by TE acquisition by horizontal transfers, followed by copy number increase and subsequent elimination. In addition, models predict that genomes of intracellular bacteria should have few or no IS.

Genome sequencing of obligate intracellular bacteria, as *Wolbachia*, have questioned models, because some have a considerable abundance of IS. Our work was based on the study of five sequenced *Wolbachia* genomes. We have realized a detailed IS annotation for each genome. In addition, experimental analyses were performed to test the hypothesis based on *in silico* analyses.

We confirmed that *Wolbachia* genomes contained a considerable abundance of IS. Surprisingly, the majority of IS copies were degraded, which allowed us to study their evolutionary dynamics. Past IS activity in *Wolbachia* genomes was not constant during time. We identified phases of high activity alternating with phases of relative quiescence. High activity phases needed to be preceded by horizontal transfers of IS that we experimentally detected in abundance in *Wolbachia* genomes. Finally, expression analyses suggested that IS activity is not exclusively controlled by IS themselves, but it also depends on the genomic environment of the IS copies.

Key words: transposable elements, insertion sequences, evolutionary dynamics, prokaryotes, *Wolbachia*