PAULA CRISTINA GASPERAZZO TURRINI

**INFLUÊNCIA DE ELEMENTOS GENÉTICOS MÓVEIS EM
LINHAGENS DE *XANTHOMONAS ORYZAE***

Tese apresentada ao Programa de Pós-Graduação
Interunidades em Biotecnologia USP/Instituto Butantan/IPT,
para obtenção do Título de Doutora em Biotecnologia.

São Paulo
2016

PAULA CRISTINA GASPERAZZO TURRINI

**INFLUÊNCIA DE ELEMENTOS GENÉTICOS MÓVEIS EM
LINHAGENS DE *XANTHOMONAS ORYZAE***

Tese apresentada ao Programa de Pós-Graduação
Interunidades em Biotecnologia USP/Instituto Butantan/IPT,
para obtenção do Título de Doutora em Biotecnologia.

Área de Concentração: Biotecnologia

Orientadora: Prof^a Dr^a Marie-Anne Van Sluys

Versão original

São Paulo
2016

CATALOGAÇÃO NA PUBLICAÇÃO (CIP)
Serviço de Biblioteca e informação Biomédica
do Instituto de Ciências Biomédicas da Universidade de São Paulo

Ficha Catalográfica elaborada pelo(a) autor(a)

Turrini, Paula Cristina Gasperazzo
Influência de elementos genéticos móveis em
linhagens de *Xanthomonas oryzae* / Paula Cristina
Gasperazzo Turrini; orientadora Marie-Anne Van
Sluys. -- São Paulo, 2016.
140 p.

Tese (Doutorado) -- Universidade de São Paulo,
Instituto de Ciências Biomédicas.

1. elementos de transposição. 2. *Xanthomonas*
oryzae. 3. plasticidade genômica. 4. elementos IS.
5. MITEs. I. Van Sluys, Marie-Anne, orientador. II.
Título.

Programa de Pós-Graduação Interunidades em Biotecnologia
Universidade de São Paulo, Instituto Butantan, Instituto de Pesquisas Tecnológicas

Candidato(a): Paula Cristina Gasperazzo Turrini

Título da Tese: Influência de elementos genéticos móveis em linhagens de
Xanthomonas oryzae

Orientador(a): Profª Drª Marie-Anne Van Sluys

A Comissão Julgadora dos trabalhos de Defesa da Tese de Doutorado, em sessão
pública realizada a/...../....., considerou

() Aprovada

() Reprovada

Examinador(a): Assinatura:
Nome:
Instituição:

Examinador(a): Assinatura:
Nome:
Instituição:

Examinador(a): Assinatura:
Nome:
Instituição:

Examinador(a): Assinatura:
Nome:
Instituição:

Presidente: Assinatura:
Nome:
Instituição:

AGRADECIMENTOS

Eu agradeço em primeiro lugar aos meus pais, Lina L Gasperazzo Turrini e Ademir Turrini, que me ensinaram o valor do trabalho através do exemplo. Me deram a oportunidade de conhecer novos mundos, os quais eles não tiveram a oportunidade de conhecer. Também me deram a preciosa chance de errar e de recomeçar. À minha irmã, Fernanda Turrini, que sempre está pronta para me ajudar nas minhas dificuldades e para celebrar minhas conquistas. Também são parte de minha família Brutus, Barack e Aspen, que como bons cachorros, me encham de alegria quando se encham de alegria ao me ver.

Agradeço à Daniel Lahr, meu companheiro e minha melhor metade. Obrigada por tudo: pelo todo o amor e carinho que recebo diariamente, pela inspiração na ciência e por acreditar sempre em mim. Eu sou muito feliz de ser sua melhor metade também.

Meus agradecimentos se estendem aos meus sogros, Lurdes Galafasse Lahr e Roberto Lahr, e aos meus cunhados Marcela Lahr e Alex Neves pelo constante apoio e torcida.

À minha orientadora Dr^a Marie-Anne Van Sluys, primeiro pela oportunidade e pela confiança. Sua compreensão e apoio foram imprescindíveis quando eu tive meu trabalho em andamento publicado.

Aos colegas do laboratório GaTE, que são todos ótimas pessoas e profissionais. Um agradecimento especial à técnica do laboratório Tatiana Corrêa, à Nathalia de Setta e colega de Doutorado Gesiele Barros, por dividir as preocupações comuns (reais e inventadas) a este período. Agradeço à Dr^a Jan Leach e aos colegas na Colorado State University pelo acolhimento, ajuda e apoio nos cinco meses de convívio e correria. Um agradecimento especial à Jillian Lang e Alejandra Huerta.

Aos amigos Daniela Soltys e Gabriel Moura, pelas inúmeras discussões sobre Ciência, a Vida, o Universo e Tudo mais. À Rebeca Viana, com quem eu divido tantas reflexões e que rapidamente compreende o que eu quero dizer mesmo quando nem eu sei o que estou dizendo. À Felipe Curcio e Carla Piantoni, que são amigos com quem eu simplesmente sou eu. Aos amigos, Ana Paula, Elizabeth, Karine, Gisele, Carolzinha, Larissa e Pedrão Barral cujas divergências de vida não nos impedem de continuar sendo os mesmos amigos de sempre.

É um imenso prazer trabalhar com pessoas comprometidas com o trabalho. Me sinto muito feliz por essa experiência incrível, que continua me dando a oportunidade de interagir com pessoas inteligentes, em ambientes que me permitem sempre aprender e crescer.

Agradeço à CAPES pela bolsa nos primeiros meses do Doutorado e à Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), que financiou meu projeto através da bolsa de Doutorado Regular 2012/24954-0 e BEPE 2015/24956-1.

" ... I believe we are again experiencing such a revolution. It is altering our concepts of the genome: its component parts, their organization, mobilities and their mode of operation..."

(Barbara McClintock, Nobel Lecture, 1983)

RESUMO

Turrini PCG. Influência de elementos genéticos móveis em linhagens de *Xanthomonas oryzae*. [Tese (Doutorado em Biotecnologia)]. São Paulo: Instituto de Ciências Biomédicas, Universidade de São Paulo; 2016.

Elementos móveis influenciam expressivamente a evolução do grupo Bacteria, uma vez que são integrantes importantes do processo de geração da variabilidade genética e promovem a diversidade da população frente à estocasticidade do ambiente. Nesta tese, investigamos a influência de elementos genéticos móveis em linhagens de *X. oryzae* sob três diferentes aspectos. Primeiramente, investigamos a transferência horizontal de um *operon* de biossíntese do fitormônio GA₄, adquirido exclusivamente pelo patovar *oryzicola* (*Xoc*). A importância da produção do fitormônio GA₄ por *Xoc* é indicada pela alta prevalência do *operon* em diferentes populações, e pela ampla distribuição geográfica em um amplo período de tempo. Também abordamos a influência de elementos genéticos móveis mais amplamente no genoma de *X. oryzae*, descrevendo como elementos de transposição, representados por elementos do tipo IS, são capazes de interferir em um sistema central do metabolismo bacteriano, como o sistema de reparo de DNA. Por fim, descrevemos um novo elemento do tipo MITE encontrado em *X. oryzae* e analisamos sua organização, distribuição, filogenia, seu nível transcricional e o potencial funcional deste elemento na interação de *X. oryzae* com seu hospedeiro *O. sativa*. Nosso conjunto de resultados reitera a importância dos elementos de transposição no aumento de variabilidade genética de populações, e dessa forma, infere sobre seu papel relevante na diversificação em *X. oryzae*. Descrevemos a variabilidade gerada, exploramos as consequências da amplificação de elementos de transposição nesta espécie e evidenciamos a complexidade desta variação em nível populacional.

Palavras-chave: *Xanthomonas oryzae*. Elementos de transposição. Elementos IS. MITE. Reparo de DNA.

ABSTRACT

Turrini PCG. Influence of mobile genetic elements in *Xanthomonas oryzae* lineages. [Ph.D. Thesis (Biotechnology)]. São Paulo: Instituto de Ciências Biomédicas, Universidade de São Paulo; 2016.

Mobile elements are genomic components that trigger genetic variability. They promote the diversity of the population against the stochasticity of the environment, significantly influencing the evolution of the Bacteria group. In this thesis, we investigate the influence of mobile genetic elements on *X. oryzae* lineages under three different aspects. First, we investigated the horizontal transfer event of an operon coding the biosynthesis of the phytohormone GA₄, acquired exclusively by the pathovar *oryzicola* (*Xoc*). The importance of the production of the phytohormone GA₄ by *Xoc* is indicated by its high prevalence in different populations, in a wide geographic distribution and in a large time period. We also address the influence of mobile genetic elements more comprehensively in the *X. oryzae* genome. We describe how transposable elements, represented by IS elements, are capable of interfering in a central system of bacterial metabolism, such as the DNA repair system. Finally, we describe a new MITE element found in *X. oryzae* and analyzed its organization, distribution, phylogeny and transcriptional level. We also speculate its function in the *X. oryzae*-*Oryza sativa* interaction. Our results reiterate the importance of transposable elements enriching genetic variability of populations and infer on their relevant role in the diversification of *X. oryzae* lineage. We describe the variability transposable elements confers, explore the consequences resulting from the amplification of transposition elements in the *X. oryzae* species and we show the complexity of this variation at the population level.

Keywords: *Xanthomonas oryzae*. Transposable elements. IS elements. MITE. DNA repair.

LISTA DE FIGURAS

1 INTRODUÇÃO GERAL

Figura 1.1 - Sintomas de <i>Xanthomonas oryzae</i> em plantas de arroz	14
Figura 1.2 - Previsão da ocorrência e severidade do crestamento foliar causados por Xoo na Ásia	16
Figura 1.3 - Organização de elementos do tipo IS e derivados	17
Figura 1.4 - Mecanismo de inserção de ISs em um novo sítio genômico.	18
Figura 1.5 - Representação de alguns mecanismos de transposição conservativa e replicativa de transposases DDE.....	19

2 DISTRIBUIÇÃO DO OPERON DE GA₄ EM DIFERENTES POPULAÇÕES DE *XANTHOMONAS ORYZAE* PV. *ORYZICOLA*

Figura 2.1 - Estratégia de amplificação do operon de GA ₄ por PCR	27
Figura 2.2 - Perfil de inserção de elementos ISs no contexto genômico no entorno do operon GA ₄ das cepas Xoc atualmente sequenciadas	30
Figura 2.3 - Origem dos isolados de Xoc diagnosticados para a presença do operon biossintético GA ₄	31

3 ANÁLISE DO IMPACTO DE ELEMENTOS SEQUÊNCIAS DE INSERÇÃO NO SISTEMA DE REPARO DE DNA EM *X. ORYZAE*

Figura 3.1 - Proximidade entre genes e elementos IS em diferentes espécies de <i>Xanthomonas</i> - via de reparo por síntese translesão e outras vias acessórias	47
Figura 3.2 - Representação de genes envolvidos nas principais vias de reparo de DNA e suas proximidades a elementos IS	49
Figura 3.3 – Proximidade entre ISs e genes da via de reparo por excisão de bases em cepas de <i>X. oryzae</i>	52
Figura 3.4 – Proximidade entre ISs e genes da via de reparo por excisão nucleotídica em cepas de <i>X. oryzae</i>	53
Figura 3.5 – Proximidade entre ISs e genes da via de reparo de mal emparelhamento de bases em cepas de <i>X. oryzae</i>	54
Figura 3.6 – Proximidade entre ISs e genes da via de reparo por recombinação homóloga em cepas de <i>X. oryzae</i>	55
Figura 3.7 - Proximidade de elementos IS ao gene <i>radA</i> em <i>X. oryzae</i>	57

4 UM NOVO ELEMENTO DO TIPO MITE EM *X. ORYZAE*

Figura 4.1 - Descrição dos elementos MITE-Xo encontrados em <i>X. oryzae</i>	72
Figura 4.2 - Árvore filogenética de MITES-Xo de <i>X. oryzae</i>	74
Figura 4.3 - Grupos sintênicos de MITES-Xo1 em <i>X. oryzae</i>	76
Figura 4.4 - Representação de três grupos sintênicos de MITE-Xo1 em Xoc	77
Figura 4.5 - Árvore filogenética das cópias de MITE-Xo1 em Xoc e Xoo	79
Figura 4.6 - Representação da evolução da sintenia de MITE-Xo2 sob a hipótese de Acctran	81
Figura 4.7 - Representação dos grupos sintênicos de MITE-Xo2 em Xoc	82
Figura 4.8 - Árvore filogenética de MITES-Xo2 de <i>X. oryzae</i>	84
Figura 4.9 - Comparação entre as sequências dos MITES-Xo em Xoc BLS256 e a sonda XOCORF3496	86
Figura 4.10 - Intensidade de fluorescência de MITES-Xo por microarranjos	87
Figura 4.11 - Quantificação de transcritos de MITES-Xo em Xoo PXO99A e Xoo PXO86	89
Figura 4.12 - <i>Motif</i> de 18 nucleotídeos nas cópias de MITES-Xo em Xoo e Xoc	91
Figura 4.13 - Cladograma das espécies de <i>O. sativa</i> e presença do <i>motif</i> de 18 nucleotídeos	95

LISTA DE TABELAS

1 INTRODUÇÃO GERAL

Sem tabelas

2 DISTRIBUIÇÃO DO OPERON DE GA₄ EM DIFERENTES POPULAÇÕES DE *XANTHOMONAS ORYZAE* PV. *ORYZICOLA*

Tabela 2.1 - Iniciadores utilizados na amplificação dos fragmentos do <i>operon</i> GA ₄	27
Tabela 2.2 - Classificação dos genes encontrados no <i>operon</i> de biossíntese de GA ₄	29
Tabela 2.3 - Número de isolados de <i>Xoc</i> diagnosticados para a presença do <i>operon</i> de GA ₄ , por país	31

3 ANÁLISE DO IMPACTO DE ELEMENTOS SEQUÊNCIAS DE INSERÇÃO NO SISTEMA DE REPARO DE DNA EM *XANTHOMONAS ORYZAE*

Tabela 3.1 - Predição do número de elementos IS em Xanthomonadaceae	43
Tabela 3.2 - Perfil de distribuição de elementos IS em cepas de <i>X. oryzae</i>	45

4 UM NOVO ELEMENTO DO TIPO MITE EM *X. ORYZAE*

Tabela 4.1 - Identificação das cepas de <i>X. oryzae</i> utilizadas e o número total de cópias de MITEs-Xo completas e parciais	73
Tabela 4.2 - Local cromossômico e anotações de genes potencialmente alvos do <i>motif</i> de 18 nucleotídeos de MITE-Xo1	94

ABREVIATÖES

Xoo	<i>Xanthomonas oryzae</i> pv. <i>oryzae</i>
Xoc	<i>Xanthomonas oryzae</i> pv. <i>oryzae</i>
IS	elementos do tipo sequência de inserção
MITE	<i>miniature inverted-repeat transposable element</i>
BER	via de reparo por excisão de bases
NER	via de reparo por excisão nucleotídica
HR	via de reparo por recombinação homóloga
MMR	via de reparo de bases mal emparelhadas
ORF	<i>open reading frame</i>
mRNA	RNA mensageiro
sRNA	pequeno RNA

SUMÁRIO

1 INTRODUÇÃO GERAL	13
1.1 Características de <i>Xanthomonas oryzae</i>	13
1.2 Importância agro-econômica de <i>X. oryzae</i>	15
1.3 Elementos Sequência de Inserção	16
1.4 Influência de IS na plasticidade genômica	20
1.5 Objetivo	22
REFERÊNCIAS	23
 2 DISTRIBUIÇÃO DO OPERON DE GA₄ EM DIFERENTES POPULAÇÕES DE XANTHOMONAS ORYZAE PV. ORYZICOLA	25
2.1 Introdução	25
2.2 Materiais e Métodos	26
2.2.1 Análise bioinformática	26
2.2.2 Isolados de <i>X. oryzae</i>	26
2.2.3 Iniciadores de PCR	27
2.2.4 Detecção dos fragmentos	27
2.3 Resultados	28
2.3.1 Avaliação da adaptação do uso de códon no operon GA ₄	28
2.3.2 Distribuição biogeográfica do operon de biossíntese de GA ₄	30
2.4 Discussão	32
REFERÊNCIAS	36
 3 ANÁLISE DO IMPACTO DE ELEMENTOS DO TIPO IS NO SISTEMA DE REPARO DE DNA EM XANTHOMONAS ORYZAE	38
3.1 Introdução	38
3.2 Materiais e Métodos	40
3.2.1 Predição do número de ISs	40
3.2.2 Análise de proximidade a partir de anotações genômicas	41
3.2.3 Análise de proximidade gene a gene	41
3.3 Resultados	42
3.3.1 Amplificação de elementos IS em <i>X. oryzae</i> no grupo Xanthomonadaceae	42
3.3.2 Distribuição das famílias de IS em <i>X. oryzae</i>	44
3.3.3 Repertório de genes de reparo em <i>X. oryzae</i>	46
3.3.4 <i>X. oryzae</i> apresenta ISs próximos a genes de reparo	48
3.3.5 Avaliação gene a gene dos eventos de proximidade IS-gene de reparo	50
3.4 Discussão	57
REFERÊNCIAS	65
 4 UM NOVO ELEMENTO DO TIPO MITE EM XANTHOMONAS ORYZAE	69
4.1 Introdução	69
4.2 Materiais e Métodos	70

4.3 Resultados	71
4.3.1 Descrição do elemento MITE -Xo	71
4.3.2 Estudo de sintonia entre as cópias de MITE-Xo em 16 genomas de <i>X. oryzae</i>	75
4.3.3 Expressão de MITE-Xo em linhagens de <i>Xanthomonas oryzae</i>	83
4.3.4 Análise de expressão de MITE-Xo por RNA seq.....	88
4.3.5 Potencial caráter funcional dos MITEs-Xo.....	89
4.4 Discussão.....	96
REFERÊNCIAS	101
5 CONCLUSÃO.....	105
APÊNDICES	
A - Submissão do artigo referente aos dados de biogeografia do <i>operon</i> de biossíntese de GA ₄	108
B - Géis de eletroforese das amplificações dos fragmentos que compõem o <i>operon</i> GA ₄ em isolados de <i>X. oryzae</i>	109
C - Isolados de <i>X. oryzae</i> utilizados para verificação da presença do <i>operon</i> GA ₄	113
D - Lista de genes de reparo utilizados	116
E - Proximidade entre genes e elementos IS em diferentes espécies de <i>Xanthomonas</i> - via de reparo de simples fita	118
F - Proximidade entre genes e elementos IS em diferentes espécies de <i>Xanthomonas</i> - via de reparo por recombinação homóloga	119
G - Proximidade de ISs e genes de reparo de vias acessórias do sistema de reparo de DNA em cepas de <i>X. oryzae</i>	120
H - Variação nucleotídica encontrada nas cópias de MITEs-Xo na cepa <i>Xoc</i> CFBP2286	121
I - Lista completa das cópias de MITEs-Xo encontradas em <i>X. oryzae</i>	122
J - Principais grupos sintênicos de MITE-Xo1 conservados em <i>Xoc</i>	130
K - Quantificação de transcritos de MITEs e genes de <i>Xoo</i> PXO99A	135
L - Quantificação de transcritos de MITEs e genes de <i>Xoo</i> PXO86	137
M - Identificação das montagens dos genomas de arroz	139
N - Predição da estrutura secundária de MITE-Xo1 e MITE-Xo2	140

1 INTRODUÇÃO GERAL

Microorganismos ocupam os mais diversos habitats e exibem uma enorme diversidade metabólica, em estrutura e composição celulares, estilos de vida e nichos ecológicos (Arber, 2000). A expressiva capacidade adaptativa que garante o sucesso evolutivo deste grupo depende de processos de geração da variabilidade genética que ocorrem em nível celular individual.

Elementos móveis influenciam a evolução do grupo Bacteria. São integrantes importantes do processo de geração da variabilidade genética e promovem a diversidade da população frente à estocasticidade do ambiente.

Neste projeto de Doutorado intitulado '**Influência de elementos genéticos móveis em linhagens de *Xanthomonas oryzae***' investigamos como elementos de transposição podem ter influenciado a história evolutiva *Xanthomonas oryzae* e contribuído para o processo de diversificação desta linhagem.

1.1 Características de *Xanthomonas oryzae*

O gênero *Xanthomonas* compõe a família Xanthomonadaceae, dentro do grupo Gamaproteobacteria. Membros deste gênero são majoritariamente fitopatógenos que podem apresentar um elevado grau de especificidade quanto ao hospedeiro vegetal e tecido de colonização. Aproximadamente 400 espécies vegetais de mono- ou dicotiledônias servem como hospedeiros para as 27 espécies até hoje conhecidas de *Xanthomonas* (Lu et al., 2008).

Como exemplo, podemos citar *Xanthomonas campestris* que infectam diferentes brassicáceas e solanáceas, *X. citri* pv. *citri* que causa a doença do cancro em espécies de citros, *X. albilineans*, causadora da escaldadura-das-folhas em cana-de-açúcar, e *Xanthomonas oryzae*, que infectam especificamente culturas de arroz.

Espécies de *Xanthomonas* podem ser compostas por múltiplas variantes patogênicas (patovares, pv.). Dentro uma mesma espécie, diferentes patovares são assim classificados porque colonizam diferentes hospedeiros (como *X. campestris* pv. *campestris* e *X. campestris* pv. *carotae*, que infectam brássicas e cenoura, respectivamente) ou porque apresentam diferentes sintomatologias em uma mesma espécie vegetal. A espécie *Xanthomonas oryzae* é composta pelos patovares *X. oryzae* pv. *oryzae* e *X. oryzae* pv. *oryzicola*, daqui em diante referidos como Xoo e Xoc por toda a tese. Causam, respectivamente, as doenças *crestamento foliar bacteriana* e *estriamento foliar bacteriano* (do inglês, *bacterial blight* e *bacterial leaf streak*) em culturas de arroz (*Oryza sativa* L.). Além de *O. sativa*, somente *Leersia oryzoides*, *L. oryzoides* var. *japonica* e *Zizania latifolia* foram reportadas como hospedeiros de Xoo (Mew et al., 2014).

Embora *Xoo* e *Xoc* sejam intimamente relacionados filogeneticamente e possuam alta especificidade com o hospedeiro, o processo de infecção é marcadamente distinto. O patovar *oryzae* (*Xoo*) invade o hospedeiro por hidatódios das pontas e margens da folha. São capazes de colonizar os espaços intercelulares e atingem o xilema, sistema vascular pelo qual *Xoo* veicula por toda a planta, comprometendo sistemicamente o hospedeiro (Figura 1.1A).

O patovar *oryzicola* (*Xoc*) penetra na folha principalmente pelos estômatos. As bactérias se proliferam na cavidade sub-estomatal e colonizam os espaços intercelulares do tecido mesofilico. Diferentemente de *Xoo*, *Xoc* é incapaz de invadir o sistema vascular (xilema), caracterizando-se como uma linhagem de menor agressividade (Niño-Liu et al., 2006; Ryan et al., 2011). Ferimentos no tecido vegetal também representam uma importante porta de entrada bacteriana para ambos patovares.

Figura 1.1 - Sintomas de *Xanthomonas oryzae* em plantas de arroz. A) Crestamento foliar causado por *Xoo* em folhas de arroz. As lesões são feixes úmidos que se iniciam nas margens e pontas e progridem até a base foliar. B) Lesões lineares, localizadas em tecidos entre os vasos foliares causados por *Xoc*, como estrias nas folhas. As lesões são translúcidas quando expostas contra luz. C) em condições de umidade, gotículas amarelas colonizadas por *Xoo* podem ser observadas na superfície foliar (cima) e quando secam, formam pequenas placas amarelas dentro da folha (baixo). Imagens do website Rice Knowledge Bank desenvolvido pelo International Rice Research Institute (IRRI), Filipinas.

Os sintomas iniciais observados na planta também podem ser utilizados para a distinção entre *Xoo* e *Xoc*. Em estágios iniciais da infecção, *Xoo* apresenta lesões umedificadas nas pontas e margens da folha e à medida que progride as lesões se tornam cloróticas (amareladas) e necróticas que se estendem da ponta da folha ao longo das nervuras. As primeiras lesões causadas por *Xoc* são pequenas e úmidas e, uma vez que a nervura foliar serve como barreira física, colonizam o tecido no sentido longitudinal da folha, formando “estrias” translúcidas e amareladas (Figura 1.1B). À medida que a infecção progride, as lesões se tornam acinzentadas levando à morte do tecido vegetal (Niño-Liu et al., 2006). Sob condições de alta umidade, ou pela manhã, lesões jovens podem gerar exsudatos, gotículas que contêm grandes quantidades do patógeno (Figura 1.1C e D).

1.2 Importância agro-econômica de *X. oryzae*

O patovar *Xoo* é um patógeno de grande importância agrônômica principalmente devido a sua alta agressividade no desenvolvimento da doença. Está entre os cinco primeiros fitopatógenos de maior importância mundial (Mansfield et al., 2012) e é considerada o patógeno mais importante no cultivo de arroz na Ásia, região responsável por 90% da produção mundial do grão.

Epidemias na década de 50 chegaram a provocar perdas de 30 a 40% em províncias chinesas. O desenvolvimento e o amplo uso de variedades híbridas resistentes tornaram o crestamento foliar uma ameaça endêmica e esporádica, porém não menos preocupante. Apesar do extenso controle, surtos da doença chegaram a causar perdas de aproximadamente 20% em 2005, na China (Mew et al., 2014).

A predição dos danos causados por *Xoo*, considerando as atuais condições climáticas são alarmantes num cenário sem controle da doença (Figura 1.2). E estima-se que as perdas hoje observadas se tornem mais severas e frequentes com climas extremos, escassez de água e chuvas irregulares decorrente das mudanças climáticas.

Na África, o crestamento e o estriamento foliar bacteriano respondem respectivamente por 70% e 20% da perda na produção de arroz irrigado no Oeste Africano, ameaçando a segurança alimentar desta região. São considerados doenças emergentes também no Leste da África e Madagascar (Wonni, 2016).

O Brasil é responsável por 65% da área produtora de arroz na América Latina e o 9º maior produtor mundial do grão (11.3 milhões de toneladas/ano). É o único país não-asiático dentro dos 10 maiores produtores mundiais, liderado por China, Índia e Indonésia (dados referentes ao período de 2006-2010, FAOSTAT). Em 2010, 61.1% da produção estava concentrada no Rio Grande do Sul, 9.2% em Santa Catarina, 6.1% em Mato Grosso e 5.2% em Maranhão. O aumento da produção na safra de 2003/2004 permitiu o país se tornar

autossuficiente e hoje exporta cerca de 5% da produção nacional. Além de exportar, o Brasil importa arroz principalmente do Paraguai, Uruguai e Argentina.

Figura 1.2 - Previsão da ocorrência e severidade (tons de azul) do crestamento foliar causados por *Xoo* na Ásia. A previsão foi calculada a partir de dados do período de 2001 a 2010, considerando a ausência de medidas de controle da doença. O mapa foi baseado nas condições climáticas ocorridas no período de plantio determinado pelo laboratório IRRI GIS. Adaptado de Global Rice Science Partnership (GRiSP) (2013).

X. oryzae é um patógeno quarentenário A1 (definição: ausente no país, com alto risco de dano econômico) no Brasil. Por este motivo, sementes importadas pelo país passam por diagnósticos fitossanitários. Interessantemente, 21 isolados bacterianos de *Xanthomonas* sp. provenientes de sementes providas do Uruguai e Argentina com resultados positivos para *Xoo* pelo teste de ELISA, não são identificados como *X. oryzae* por testes moleculares (Júnior et al., 2015), evidenciando a necessidade de uma melhor caracterização da diversidade e ocorrência de *X. oryzae* na América Latina.

1.3 Elementos Sequência de Inserção

Elementos Sequências de Inserção (ISs, *insertion sequences*) é um tipo de elemento de transposição, pertencente a Classe II (transposon de DNA), amplamente distribuído no domínio Bacteria. São segmentos compactos de DNA, que variam entre 0.7 a 3Kb, que codificam enzimas responsáveis pela sua própria mobilidade. A transposase pode ser organizada em uma ou duas ORFs (do inglês, *open reading frame*) – e o elemento geralmente apresenta sequências terminais invertidas (TIRs, do inglês, *terminal inverted*

repeats), essenciais para o processo de transposição (Mahillon, Chandler, 1998; Siguier et al., 2015; Varani et al., 2015) (Figura 1.3A).

Figura 1.3 - Organização de elementos do tipo IS e elementos derivados. ISs codificam a transposase, enzima que cataliza sua mobilização (setas internas preta), geralmente apresentam repetições terminais invertidas (TIRs, setas rosa), repetições diretas (DR, setas azul-claro). Representam derivações autônomas de IS: IS transportador ou tIS (B) e transposon completo (C). Representam derivações não-autônomas: MITEs (D) e MICs (E) representam algumas derivações deste tipo de elemento de transposição. Genes passageiros (blocos roxos); p, região promotora da transposase; (E) esquerda; (D) direita.

O mecanismo de mobilidade é um dos principais critérios de classificação de elementos IS, e portanto, depende do tipo de transposase codificada. Atualmente, mais de 4000 ISs distintos estão classificados dentro de 25 famílias e, em sua maioria, utilizam a DDE-transposase para mobilização (Curcio, Derbyshire, 2003; Siguier et al., 2015).

DDE-transposons são assim denominados porque codificam transposases que compartilham o motivo DDE – três aminoácidos polares ácidos Asp(D), Asp(D) e Glu(E) presentes em todas as enzimas desta família. O resíduo conservado DDE, posicionado de maneira não-contígua na sequência de aminoácidos da enzima - os aminoácidos DD separados por aproximadamente 35 resíduos do aminoácido E -, é essencial para o processo de transposição uma vez que facilita a catálise pela coordenação de íons metal no sítio catalítico da transposase (Davies et al., 2000; Lovell et al., 2002).

O mecanismo de transposição de muitas famílias de IS começa com a hidrólise da cadeia fosfodiéster do DNA do sítio *doador* de transposon, que gera uma par de 3'-OH terminais (Figura 1.4). Em seguida, forma-se um complexo sináptico (transpososoma) com as pontas 3'-OH terminais expostas do transposon, a enzima transposase e o sítio *alvo* da transposição, que clivam as fitas de DNA do sítio *alvo* e juntam as pontas 3'-OH livres do elemento ao sítio de inserção por uma reação de trans-esterificação.

A quebra da fita do DNA *alvo* é feita através de um ataque nucleofílico pelos dois grupos 3'-OH e é geralmente deslocada em poucos nucleotídeos – geralmente de 2 a 8 nucleotídeos – e resulta na duplicação do sítio alvo de inserção ao final da transposição do

elemento (DR, do inglês *direct repeats*; Figura 1.4) (Curcio, Derbyshire, 2003; Mizuuchi, 1992).

Figura 1.4 - Mecanismo de inserção de ISs em um novo sítio genômico. Catalizado pela transposase, as duas pontas de DNA com os grupos 3'-OH expostos realizam um ataque nucleofílico ao DNA alvo. Repetições diretas (DR) são produzidas ao ser realizado o reparo do genoma após a transposição. Traduzido de Hickman, Dyda (2015).

A transposição de elementos IS pode simplesmente gerar a mudança de um elemento de um lugar a outro no genoma (transposição conservativa) ou gerar o aumento do número de cópias no genoma do organismo que reside (transposição replicativa) (Figura 1.5).

Na transposição conservativa, o elemento deve ser extraído do sítio *doador* através de quebras duplas no DNA. A principal diferença entre os mecanismos de diversas famílias de elementos é o processamento (clivagem) da segunda fita de DNA para liberação completa do transposon do sítio *doador*. Ocorre um primeiro ataque nucleofílico gerando quebras de fita simples nas pontas do transposon (gerando o grupo 3'-OH, como descrito anteriormente). Um segundo subsequente evento de hidrólise pode gerar a quebra da segunda fita, como os elementos Tc1/*mariner* (Figura 1.5B).

Nos grupos IS10, IS50, Tn5 e Tn10 (família IS4), o grupo 3'-OH formado pelo primeiro evento de hidrólise atua como um nucleófilo e ataca as pontas 5' do transposon, liberando o elemento, mas resultando na formação de um *hairpin* em suas pontas (Figura 1.5A). As pontas do *hairpin* precisam ser novamente clivadas para reconstituir o grupo 3'-OH e realizar a transferência de fita ao novo sítio.

Na transposição replicativa ocorre o aumento do número de cópias e a replicação do DNA é geralmente empregada no processo. O primeiro evento de hidrólise ocorre em somente uma das pontas do transposon, que por sua vez atua como nucleófilo para atacar a mesma fita de DNA, na outra extremidade do elemento (Figura 1.5D). As duas pontas do elemento são então conectadas por uma ponte de simples fita e o mecanismo de replicação da célula hospedeira converte esta estrutura a um transposon circular. O transposon circular precisa ser novamente clivado para gerar os grupos 3'-OH e realizar a

transferência de fita. Este mecanismo é utilizado nas famílias IS3, IS30, IS110, IS256, IS21, ISL3 e IS911.

Figura 1.5 - Representação de alguns mecanismos de transposição conservativa e replicativa de transposases DDE. Na transposição conservativa é necessário a quebra de fita dupla de DNA. O processamento da segunda fita pode ocorrer pela formação de um *hairpin* nas pontas do transposon, que então é posteriormente clivado (A); alternativamente, a segunda fita pode ser processada simplesmente por um segunda reação de hidrólise, como ocorrido na clivagem da primeira fita (B). A transposição replicativa conta com o mecanismo de replicação da célula hospedeira. A transferência de fita por ser através da formação de um co-integrado entre os dois replicons envolvidos (C) ou pela formação de um transposon circular. Tn, transposon; DR, repetições diretas. Adaptado de Curcio, Derbyshire (2003) e Hickman, Dyda (2015).

Comparativamente, a transposição dos elementos Mu, Tn3 e IS6 envolve dois replicons (e.g. um cromossomo e um plasmídeo), que representam um doador e o outro receptor do elemento. Em um processo sequencial, cada ponta clivada do transposon (com a formação do grupo 3'-OH), ataca e se une ao sítio *alvo* de inserção, resultando na fusão dos replicons envolvidos (Figura 1.5C). A replicação a partir das novas pontas 3'-OH no sítio

do DNA alvo resulta em uma molécula co-integrada, na qual os replicons doador e receptor são unidos pelas duas cópias do transposon (Curcio, Derbyshire, 2003).

Comum a todos os mecanismos, o final do processo de transposição exige o reparo de *gaps* nas fitas de DNA resultante das quebras provocadas na excisão do elemento do sítio *doador* e na integração do elemento no novo sítio. Detalhes deste processo são escassos ambos *in vivo* e *in vitro*. Assume-se que o reparo do genoma é feito por enzimas de replicação do hospedeiro ou por recombinação homóloga, quando há a presença de uma cromátide irmã (após replicação do DNA e antes da divisão celular) (Hickman, Dyda, 2015). Novos estudos apontam o auxílio de enzimas do sistema de reparo do hospedeiro, como foi demonstrado no transposon bacteriófago Mu, que utiliza a via de reparo de quebra dupla do DNA (do inglês, DSB, *double strand repair*) (Jang et al., 2012).

Existem outros tipos de elementos, semelhantes a ISs, porém com variações importantes quanto à organização. Entre eles, o IS transportador (*tlS*, *transporter IS*) codifica, além da transposase, genes não relacionados com sua mobilização (Figura 1.3B); transposons compostos (*composite transposons*) são formados por mais de um ISs flanqueando um conjunto de genes entre eles (Figura 1.3C). Os MITEs (discutidos no Capítulo 4, Figura 1.3D) representam unidades não autônomas de elementos, porque não codificam transposases e dependem da atividade de um outro elemento para se mobilizarem, assim como MICs (*mobile insertion cassettes*), que carregam somente gene(s) passageiro(s) por entre repetições terminais invertidas (Figura 1.3E) (Siguier et al., 2015; Varani et al., 2015).

1.4 Influência de IS na plasticidade genômica

Em uma escala de tempo evolutiva, genomas bacterianos equilibram sua natureza clonal - herança vertical do material genético pela reprodução assexuada - com mecanismos que alteram sua composição e organização.

Uma importante evidência da fluidez de genomas bacterianos é a grande variação do conteúdo gênico encontrado em espécies proximamente relacionadas, em linhagens dentro de uma mesma espécie, ou ainda, entre diferentes isolados de uma mesma população bacteriana (Cordero, Polz, 2014; Dobrindt, Hacker, 2001; Ochman et al., 2000; Polz et al., 2013; Vos et al., 2015).

A diversidade genética entre linhagens pode ser enriquecida pelo acúmulo de mutações nucleotídicas pontuais em sequências codificantes e regiões regulatórias existentes. De uma maneira mais rápida, o surgimento de novos genes pode ocorrer por duplicação gênica ou transferência lateral. Enquanto a diversificação do conteúdo gênico por duplicação gênica é observada com maior frequência em eucariotos, transferência horizontal

é considerado o principal meio de aquisição de novos genes em Bactéria (Gogarten, Townsend, 2005). Rearranjo cromossômico também é considerado um processo de aumento de variabilidade, uma vez que reorganiza o conteúdo gênico e estabelece novas associações regulatórias do mesmo conjunto gênico. Eventos de recombinação e perda gênica contrabalançam as tendências de expansão e divergência genética (Dobrindt, Hacker, 2001).

Elementos IS compõem uma fração importante de elementos repetitivos de genomas bacterianos (Treangen et al., 2009) e são mediadores de processos que introduzem variação genômica no grupo Bactéria.

A natureza de elementos móveis, que incluem os elementos IS, é traduzida em oportunidades e desafios nos genomas que os contém. Por promoverem sua própria mobilidade, são capazes de se propagar pelo genoma indiscriminadamente. Isto se torna particularmente desafiador no compacto genoma bacteriano, cujas sequências codificantes (de proteínas ou moléculas de RNA centrais ao funcionamento) totalizam 85-90% do genomas. Ainda, parte do restante corresponde a regiões regulatórias. Assim, a fração genômica que está de fato livre de pressões seletivas é limitada e a transposição inerentemente tem uma alta probabilidade de interromper genes importantes à célula ou interferir em regiões controladoras de genes.

A mesma capacidade disruptiva de elementos de transposição pode ser interpretada como oportunidade de geração de heterogeneidade genética, trazendo novos substratos para evolução, que abrange a aquisição de capacidade de colonização e adaptação a novos nichos.

Por codificarem elementos regulatórios, ISs influenciam localmente genomas de seus hospedeiros bacterianos alterando a expressão de genes adjacentes ao local de inserção, por silenciamento ou amplificação da transcrição (Bohn, Bouloc, 1998; Darmon, Leach, 2014; Gregory, Dahlberg, 2008; Reynolds et al., 1981; Saedler et al., 1974; SaiSree et al., 2001; Zhang, Saier, 2009).

A presença de múltiplas cópias de elementos de repetição aumenta a probabilidade de eventos de rearranjo. Assim, ISs podem promover rearranjos intragenômicos de grande magnitude como deleções, inversões ou duplicações genômicas, utilizando a maquinaria celular nativa de recombinação. Ainda, elementos IS podem ser permutados entre diferentes genomas por mecanismos de transferência lateral de genes (Chain et al., 2004; Parkhill et al., 2003).

Esta tese de Doutorado investiga a influência de elementos móveis em linhagens de *X. oryzae*. O estudo comparativo de espécies filogeneticamente próximas ainda que patogenicamente distintas, como *Xoo* e *Xoc*, tem como objetivo revelar complexidades na

evolução de patógenos e o papel de interações sutis entre patógeno e hospedeiro. Os dois patovares de *X. oryzae* apresentam um conjunto de características que os colocam como modelo para o estudo proposto: são filogeneticamente muito próximas; fenotipicamente, apresentam o mesmo organismo hospedeiro porém possuem modo de infecção marcadamente distinto, que determina sua agressividade e; genotipicamente, possuem perfis de distribuição de elementos IS significativamente distintos entre si e comparado às outras espécies presentes no grupo *Xanthomonas* (abordado com mais detalhe no Capítulo 3).

Abordamos a influência de elementos móveis em *X. oryzae* nos três próximos capítulos desta tese. O Capítulo 2 se refere a um evento de transferência horizontal da ilha genômica denominada XODB602, encontrada em *Xoc* porém ausente em *Xoo*. Avaliamos o caráter adaptativo da aquisição da ilha genômica por *Xoc* através de uma investigação biogeográfica, em populações de *Xoc* de diferentes regiões geográficas.

No Capítulo 3 mostramos que a espécie *X. oryzae* amplificou o número de elementos IS em seus genomas, comparado a outras espécies de *Xanthomonas*. Mostramos que genes pertencentes ao sistema de reparo de DNA podem estar sob influência de elementos e avaliamos *in silico* o possível envolvimento do sistema nativo de reparo na proliferação genômica de transposons do tipo IS em *X. oryzae*.

E finalmente o Capítulo 4, descrevemos um novo elemento do tipo MITE, atualmente restrito à espécie *X. oryzae*. Além de sua organização, descrevemos a frequência deste elemento nos genomas disponíveis de *X. oryzae*, apresentamos evidências de sua expressão celular e avaliamos uma possível funcionalidade no processo de interação entre *X. oryzae* e seu hospedeiro *O. sativa*.

1.5 Objetivo

O objetivo geral deste trabalho é avaliar o impacto de elementos genéticos móveis presentes nas linhagens de *Xanthomonas oryzae* e inclui 1) identificar o contexto evolutivo da ilha genômica XODB602 presente em *Xanthomonas oryzae* pv. *oryzicola*, seu papel na inovação evolutiva do genoma e na garantia do sucesso evolutivo do organismo hospedeiro e 2) identificar o envolvimento do sistema de reparo na proliferação de elementos de inserção nas linhagens de *X. oryzae*.

REFERÊNCIAS*

- Arber W. Genetic variation: molecular mechanisms and impact on microbial evolution. *FEMS Microbiol Rev.* 2000;24(1):1–7.
- Cordero OX, Polz MF. Explaining microbial genomic diversity in light of evolutionary ecology. *Nat Rev Microbiol.* 2014;12(4):263–73.
- Chain PSG, Carniel E, Larimer FW, Lamerdin J, Stoutland PO, Regala WM, et al. Insights into the evolution of *Yersinia pestis* through whole-genome comparison with *Yersinia pseudotuberculosis*. *Proc Natl Acad Sci U S A.* 2004;101(38):13826–31.
- Curcio MJ, Derbyshire KM. The outs and ins of transposition: from Mu to Kangaroo. *Nat Rev Mol Cell Biol.* 2003;4(11):865–77.
- Davies DR, Goryshin IY, Reznikoff WS, Rayment I. Three-Dimensional Structure of the Tn5 Synaptic Complex Transposition Intermediate. *Science.* 2000;289(5476):77–85.
- Dobrindt U, Hacker J. Whole genome plasticity in pathogenic bacteria. *Curr Opin Microbiol.* 2001;4(5):550–7.
- Global Rice Science Partnership (GRiSP). *Rice Almanac*. 4th Edition. International Rice Research Institute; 2013.
- Gogarten JP, Townsend JP. Horizontal gene transfer, genome innovation and evolution. *Nat Rev Microbiol.* 2005;3(9):679–87.
- Hickman AB, Dyda F. Mechanisms of DNA Transposition. *Microbiol Spectr [Internet]*. 2015;3(2):e1-22
- Jang S, Sandler SJ, Harshey RM. Mu Insertions Are Repaired by the Double-Strand Break Repair Pathway of *Escherichia coli*. *PLOS Genet.* 2012;8(4):e1002642.
- Júnior S, De IT, Santos M de O, Oliveira AMR de, Moraes MG de, Duarte V, et al., Isolates characterization of *Xanthomonas* sp. Associated with rice seed and differentiation of *Xanthomonas oryzae* strains. *Ciênc Rural.* 2015;45(12):2099–105.
- Lovell S, Goryshin IY, Reznikoff WR, Rayment I. Two-metal active site binding of a Tn5 transposase synaptic complex. *Nat Struct Mol Biol.* 2002 Apr;9(4):278–81.
- Lu H, Patil P, Sluys M-AV, White FF, Ryan RP, Dow JM, et al., Acquisition and Evolution of Plant Pathogenesis–Associated Gene Clusters and Candidate Determinants of Tissue-Specificity in *Xanthomonas*. *PLOS ONE.* 2008;3(11):e3828.
- Mahillon J, Chandler M. Insertion Sequences. *Microbiol Mol Biol Rev.* 1998;62(3):725–74.
- Mansfield J, Genin S, Magori S, Citovsky V, Sriariyanum M, Ronald P, et al., Top 10 plant pathogenic bacteria in molecular plant pathology. *Mol Plant Pathol.* 2012;13(6):614–29.
- Mew TW, Hibino H, Savary S, Vera Cruz CM. Rice diseases: their biology and selected management practices [Internet]. International Rice Research Institute (IRRI), Africa Rice Center (AfricaRice), and the International Center for Tropical Agriculture (CIAT); 2014.

*De acordo com:

International Committee of Medical Journal Editors. [Internet]. Uniform requirements for manuscripts submitted to Biomedical Journal: sample references. [updated 2011 Jul 15]. Available from: <http://www.icmje.org>

Mizuuchi K. Transpositional Recombination: Mechanistic Insights from Studies of Mu and Other Elements. *Annu Rev Biochem.* 1992;61(1):1011–51.

Niño-Liu DO, Ronald PC, Bogdanove AJ. *Xanthomonas oryzae* pathovars: model pathogens of a model crop. *Mol Plant Pathol.* 2006;7(5):303–24.

Ochman H, Lawrence JG, Groisman EA. Lateral gene transfer and the nature of bacterial innovation. *Nature.* 2000;405(6784):299–304.

Parkhill J, Sebaihia M, Preston A, Murphy LD, Thomson N, Harris DE, et al. Comparative analysis of the genome sequences of *Bordetella pertussis*, *Bordetella parapertussis* and *Bordetella bronchiseptica*. *Nat Genet.* 2003;35(1):32–40.

Polz MF, Alm EJ, Hanage WP. Horizontal gene transfer and the evolution of bacterial and archaeal population structure. *Trends Genet.* 2013;29(3):170–5.

Ryan RP, Vorhölter F-J, Potnis N, Jones JB, Van Sluys M-A, Bogdanove AJ, et al., Pathogenomics of *Xanthomonas*: understanding bacterium–plant interactions. *Nat Rev Microbiol.* 2011;9(5):344–55.

Siguier P, Gourbeyre E, Varani A, Ton-Hoang B, Chandler M. Everyman's Guide to Bacterial Insertion Sequences. *Microbiol Spectr.* 2015;3(2):e1-35

Treangen TJ, Abraham A-L, Touchon M, Rocha EPC. Genesis, effects and fates of repeats in prokaryotic genomes. *FEMS Microbiol Rev.* 2009;33(3):539–71.

Varani A, Carvalho L, Zerillo MM, Monteiro-Vitorello CB. Elementos de Transposição: Classificação e Mecanismos de Mobilização. In: *Elementos de Transposição Diversidade, evolução aplicações e impacto nos genomas dos seres vivos.* FIOCRUZ; 2015. p. 11–43.

Vos M, Hesselman MC, te Beek TA, van Passel MWJ, Eyre-Walker A. Rates of Lateral Gene Transfer in Prokaryotes: High but Why? *Trends Microbiol.* 2015;23(10):598–605.

Wonni I. Evaluation of Elite Rice Varieties Unmasks New Sources of Bacterial Blight and Leaf Streak Resistance for Africa. *Rice Res Open Access.* 2016; 4(1):e1-8

2 DISTRIBUIÇÃO DO OPERON DE GA₄ EM DIFERENTES POPULAÇÕES DE *XANTHOMONAS ORYZAE* PV. *ORYZICOLA*

2.1 Introdução

As duas variedades patogênicas em estudo, *Xoo* e *Xoc*, são evolutivamente muito próximas, no entanto, importantes diferenças fenotípicas e genéticas distinguem os dois patovares. O *operon* XODB602 é um dos exemplos que caracteriza a singularidade de cada patovar. Ausente no patovar *Xoo* e presente em *Xoc*, este *operon* também é encontrado em ilhas simbióticas de organismos do grupo Rhizobiales, como *Mesorhizobium loti* e *Bradyrhizobium japonicum*, o plasmídeo p42d de *Rhizobium etli*, cepas de *Sinorhizobium*.

Em Rhizobia, o *operon* XODB602 é constituído por 9 genes. As enzimas GGPP sintase (IDI/GGPS), ent-copalil difostato (CPS) e ent-kaurene sintase (KS) são responsáveis pela produção de ent-kaurene (Morrone et al., 2009; Hershey et al., 2014). Os três genes da família citocromo P450 (CYP112, CYP114 e CYP117), juntamente com ferredoxina (Fd_{GA}) e desidrogenase (SDR_{GA}) convertem ent-kaurene em giberelina A₉ (GA₉) (Tatsukami and Ueda 2016) (Figura 2.1).

Em *Xoc*, esta via metabólica contém um gene adicional em relação ao encontrado em Rhizobiales. A enzima CYP115, também da classe citocromo P450, cataliza a β-hidroxilação de GA₉, que é então convertido no hormônio bioativo giberelina A₄ (GA₄) (Nett et al., 2016). Em função desta caracterização renomeamos o *operon* XODB602 de *operon* de síntese do hormônio GA₄.

Assumimos que o patovar *Xoc* adquiriu o *operon* de síntese de GA₄ por transferência lateral. A indicação mais robusta deste fenômeno é a distribuição filogenética irregular do *operon* GA₄, que está presente em linhagens filogeneticamente distantes (como *Xoc*, do grupo Gamaproteobacteria e em Rhizobiales, do grupo Alphaproteobacteria) porém ausentes em linhagens muito próximas (presente em *Xoc* e ausente em *Xoo* e outras espécies do gênero *Xanthomonas*). Ainda, elementos ISs flanqueiam todo o *operon* de síntese de GA₄ e denotam mobilidade a este conjunto gênico (Figura 2.2).

A incorporação de novos genes (como os transferidos horizontalmente) serve como matéria prima aos processos de seleção: genes que contribuem positivamente para o *fitness* do organismo, ou facilitam a adaptação a novos nichos ecológicos, possuem uma maior probabilidade de se fixarem na população. Sob a hipótese nula, o *operon* se fixou por processos evolutivos estocásticos na população de *Xoc* e não confere vantagem adaptativa ao organismo. Alternativamente, a manutenção deste *operon* no novo contexto genômico depende da incorporação da nova característica (*i.e.* produção do fitormônio GA₄) ao contexto celular e modo de vida de *Xoc*. Dessa forma, entender a frequência de um caracter em nível populacional é um componente importante para a investigação de seu papel

adaptativo e de colonização de novos nichos. Por ser um fitopatógeno extracelular não obrigatório, a contribuição do operon para o *fitness* de *oryzicola* pode atuar em sua fase patogênica ou em sua fase de vida livre, enquanto persiste no solo, no ar ou como epífita sobre folhas.

Neste capítulo, descrevemos primeiramente o potencial transcricional pela adaptação de códons dos genes do operon de biossíntese de GA₄ em *Xoc* através da análise de codon. A fim de entender a frequência genotípica (e potencialmente fenotípica) do operon no patovar *oryzicola*, investigamos a presença do operon em diferentes populações de *Xoc* isoladas de diferentes regiões geográficas e em diferentes momentos.

Este trabalho foi feito em colaboração com Dra. Jan Leach (Colorado State University, EUA), responsável por uma coleção representativa de linhagens vivas e DNA de *X. oryzae*. Adicionalmente, o resultado deste trabalho foi submetido e aceito para publicação, em colaboração com Dr. Reuben Peters (Iowa State University, EUA) (Apêndice A).

2.2 Materiais e Métodos

2.2.1 Análise bioinformática

A procura bioinformática por cópias de operon de biossíntese de GA₄ em outros organismos foram feitas através do algoritmo BLAST em espécies disponíveis no banco de dados NCBI. Elementos sequências de inserção (ISs) na vizinhança de operon GA4 foram identificados pela ferramenta ISfinder (Siguier et al., 2006). O índice CAI (Codon Adaptation Index) foi calculado pelo servidor CAIcal (<http://genomes.urv.cat/CAIcal/>) (Puigbò et al., 2008).

Também calculamos o uso de códon nativo, descrito por Davis, Olsen (2011). Neste método, identifica-se primeiramente o uso de códon que caracteriza a maior parte dos genes de uma genoma (fazendo uma moda das frequências de códons de todos os genes do genoma - uso modal). Da mesma forma, uma segunda frequência modal é feita a partir de genes altamente expressos, por possuírem um uso atípico, altamente enviesado de uso de códon. Assim, o *códon nativo* é constituído por um eixo de frequência de códons que começa no uso modal e continua até o uso de códons de alta expressão. Após a construção deste eixo, é calculado para cada gene do genoma se a frequência dos códons nele encontrada é estatisticamente semelhante ou estatisticamente diferente ao codon nativo, e denominado "modal" ou "não semelhante", respectivamente.

2.2.2 Isolados de *X. oryzae*

Foram utilizados isolados de *X. oryzae* previamente descritos, provenientes do Oeste Africano (Wonni et al., 2013) e Ásia (Raymundo et al., 1999), obtidos da coleção mantida na

Colorado State University (Fort Collins, EUA), sob responsabilidade da Dr. Jan E. Leach. A quantidade de DNA utilizado como substrato para reação de PCR variou entre 10 a 50ng por reação.

2.2.3 Iniciadores de PCR

Iniciadores foram desenhados para anelar dentro de genes do *operon* de GA₄, excluindo os elementos de transposição adjacentes. Fragmentos foram amplificados usando a polimerase Phusion High-Fidelity (Thermo Scientific), de acordo com as instruções do fabricante, em reações otimizadas para DNAs ricos em GC, e adição de 3% de DMSO. Para isolados que não amplificaram o *operon*, a integridade do DNA foi testada através de reações de PCR não-específicas. Linhagens que não amplificaram o *operon* e também não amplificaram produtos não-específicos foram descartadas de nossa análise.

Figura 2.1 - Estratégia de amplificação do *operon* de GA₄ por PCR. Genes do *operon* estão indicados pelos blocos pretos; barras cinzas representam os produtos de amplificação; setas serrilhadas indicam a posição dos iniciadores utilizados. Identificação dos primers pode ser encontrada em Tabela 2.1.

2.2.4 Detecção dos fragmentos

Fragmentos amplificados foram detectados via gel de eletroforese, usando géis preparados com buffer TAE e 0.8% de agarose. Géis foram corados pós-corrida com GelRed (Biotium Inc.) (Apêndice B).

Tabela 2.1 - Iniciadores utilizados na amplificação dos fragmentos do *operon* GA₄

ID	Sequência do iniciador (5' para 3')	Observação	Tamanho do produto
P1	TCCGCGAGGTAATACCTAATCCAGTG	Amplificação do fragmento 1	4929bp
P2	CCGCTGCTGCATTCTC		
P3	CTGGCTCGAACACATTGATA	Amplificação do fragmento 2	5045bp
P4	ACCGTACTGGGAGGATTGA		
P5	GCGGGGCCGAAGTCC	Amplificação do fragmento 3	4146bp
P6	CGCCGCTGCCTTAGATGT		
P7	GCGTCGGCGGCTCCTT	Amplificação do fragmento 4	3001bp
P8	GGTCGGGGCTAATCAGGTTCT		

2.3 Resultados

O rápido aumento do número de genomas bacterianos sequenciados nos últimos quatro anos permitiu a identificação do operon GA₄ em outras linhagens bacterianas. Além da presença no grupo Rhizobiales (Alphaproteobacteria) e na variedade patogênica *X. oryzae* pv. *oryzicola*, algumas cepas de *Xanthomonas translucens* codificam o operon. Apesar da identificação destas novas cepas, a distribuição filogenética do operon GA₄ continua altamente irregular dentro do domínio Bacteria e representa uma forte evidência que a aquisição deste *cluster* gênico ocorreu de forma lateral pelos organismos que o contém.

A composição nucleotídica de genes recentemente adquiridos lateralmente reflete a composição nucleotídica do genoma de origem. Após integração em um novo contexto genômico, as sequências transferidas passam a estar sujeitas aos processos mutacionais inerentes ao novo organismo e, gradativamente, a composição nucleotídica do novo fragmento se torna semelhante ao restante dos genes (Lawrence, Ochman, 1997).

Através da análise de uso de códon, avaliamos a semelhança da composição nucleotídica entre os genes do *cluster* biossintético de GA₄ e o restante do conjunto gênico em *Xoc* a fim de verificar o potencial do operon estar transcricionalmente adaptado a este genoma.

2.3.1 Avaliação da adaptação do uso de códon no operon GA₄

O código genético especifica a incorporação de aminoácidos a cada tripla de nucleotídeos (códon) no processo de tradução de genes que codificam proteínas. A degeneração do código genético permite que diferentes códons possam levar à incorporação de um mesmo aminoácido. No entanto, códons sinônimos não são utilizados de maneira uniforme em diferentes genes, e resulta em um fenômeno denominado viés do “uso de codon” (ou somente viés de codon). Segundo a “hipótese genômica” (tradução livre do inglês *Genome Hypothesis*, Grantham et al., 1980), genomas de diferentes taxons diferem significativamente quanto ao viés de codon, porém, genes pertencentes a um mesmo genoma adotam uma mesma estratégia de uso de códons, ou seja, genes utilizam de uma maneira semelhante o catálogo de codons e empregam escolhas semelhantes dos codons sinônimos.

Calculamos dois tipos de índice de uso de códon. O índice de “uso de códon” mais utilizado para análise de adaptação do codon é o CAI (do inglês, *Codon Adaptation Index*). Este índice compara o uso de codons dos genes de interesse a um conjunto de genes de referência, que por serem altamente transcritos, assume-se que possuem uma ótima utilização de codons. O viés em códons apresentado por estes genes de referência resulta em uma alta eficiência de tradução e portanto, alta disponibilidade das proteínas. Assim, o índice CAI – que varia de 0 a 1 - representa o quanto os códons do gene de interesse é

semelhante aos códons dos genes de referência, de alta expressão. Assim, um alto CAI indica que o gene é altamente adaptado, com alta semelhança aos genes de referência. No entanto, um CAI baixo indica que o uso de códon é diferente da referência e não é possível distinguir entre genes nativos que não utilizam códons altamente enviesados e genes de origem estrangeira. Além do CAI, foi também utilizado o método de *códon nativo* (Davis, Olsen, 2011), que caracteriza o uso modal de todos os genes do genoma e analisa se o gene de interesse é estatisticamente semelhante ao conjunto total (ver Materiais e Métodos).

A análise do genoma de *Xanthomonas oryzae* pv. *oryzicola* pelo método de *códon nativo* indicou que 30% dos 4474 genes deste genoma foram adquiridos horizontalmente, de acordo análises previamente reportadas (Lima et al., 2005). O uso de códons dos genes do operon GA₄ foram todos classificados como estatisticamente semelhantes ao uso de códons do genoma ("modal") (Tabela 2.2). A título de comparação com método de códon nativo, o índice CAI também foi calculado e todos os genes apresentam valores de acima de 0,6.

Assim, os dois métodos de análise de códon utilizados indicam que os genes do operon de GA₄ estão adaptados ao genoma hospedeiro, do ponto de vista de integração dos genes na rede transcricional de *Xoc*.

Interessantemente, a proteína hipotética identificada como xoc86, foi inicialmente designada como parte do operon de GA₄ e classificada pela análise de uso de codon como 'não semelhante'. Esta ORF foi, de fato, predita e anotada erroneamente e trata-se de um novo elemento do tipo MITE (descrito com mais detalhes no Capítulo 4).

Tabela 2.2 - Classificação dos genes encontrados no operon de biossíntese de GA₄

Locus/ Gene	Descrição do gene*	Uso nativo de códon	CAI
xoc73	ISXo1 transposase, IS5 family	MODAL	0.684
xoc74	ISXo8 transposase, IS4 family	NÃO SEMELHANTE	0.589
xoc75	ISXo1 transposase, IS5 family	MODAL	0.689
IDI	isopentenyl-diphosphate delta-isomerase	MODAL	0.655
KS	hypothetical protein XOC_0077	MODAL	0.694
CPS	hypothetical protein XOC_0078	MODAL	0.671
IDS	polyprenyl synthetase	MODAL	0.703
CYP117	cytochromeP450 BJ-4	MODAL	0.739
SDR _{GA}	7-alpha-hydroxysteroid dehydrogenase	MODAL	0.674
Fd _{GA}	cytochromeP450, CpxP3	MODAL	0.707
CYP114	cytochromeP450 BJ-3	MODAL	0.744
CYP112	cytochromeP450 BJ-1	MODAL	0.736
CYP115	cytochromeP450 107B1	MODAL	0.644
xoc86	hypothetical protein XOC_0086	NÃO SEMELHANTE	0.489
xoc87	ISXo8 transposase, IS4 family	NÃO SEMELHANTE	0.589

* descrição dos genes de acordo com publicado no banco de dados NCBI.

116 isolados testados (91.3%), e ausente somente em 10 isolados. Em nível regional, o *operon* está ausente em um único isolado nas Filipinas (Ásia), e esta ausência é observada de forma mais frequente nos isolados do Oeste Africano (9 de 67; 13.4%)(triângulos amarelos em Figura 2.3).

Tabela 2.3 - Número de isolados de *Xoc* diagnosticados para a presença do *operon* de GA₄, por país

País	Número de isolados		
	Total	GA ₄ +	GA ₄ -
Mali	35	31	4
Burkina Faso	32	27	5
Filipinas	41	40	1
Malásia	3	3	0
China	4	4	0
Índia	1	1	0
Total	116	106	10

Também foram testadas 18 cepas de *Xoo* (5 isolados asiáticos e 13 sequenciadas), um isolado de *X. oryzae* (sem designação de variedade patogênica, US-X8-1A) dos Estados Unidos e um isolado de *X. campestris* pv. *leersiae* (BAI23) (Apêndice C). Nenhum destes isolados apresentam o *operon* GA₄ em seus genomas.

Figura 2.3 - Origem dos isolados de *Xoc* diagnosticados para a presença do *operon* biossintético GA₄. A) Países onde *Xoc* foram isolados. Área azuis indicam Mali e Burkina Faso (Oeste Africano), Índia, Malásia, China e Filipinas (Ásia). Detalhes da área de coleta dos isolados no Oeste Africano (B) e Filipinas (C). Áreas vermelhas indicam províncias ou cidades (quando disponíveis); triângulos amarelos indicam regiões provenientes de isolados que não possuem o *operon* de GA₄.

2.4 Discussão

Transferência horizontal de genes é um processo chave na diversificação e evolução do Domínio Bacteria. Há um crescente acúmulo de evidências que demonstram que este processo evolutivo atua como força predominante na expansão do repertório gênico em genomas microbianos (Kuo, Ochman, 2009).

Assumimos a origem lateral do operon GA₄ em *Xoc*, embora este fenômeno não pôde ser detectado por ferramentas computacionais comumente utilizadas para detecção de transferência horizontal, como IslandViewer (Dhillon et al., 2015). A inferência sobre a aquisição do operon GA₄ vem da observação que este *cluster* está confinado a linhagem *X. oryzae* pv. *oryzicola*, e ausentes em grande parte dos membros filogeneticamente próximos.

O *operon* não ser detectado computacionalmente como um fragmento de origem lateral pode ser explicado pelo fato dos genes que o compõem exibirem conteúdo GC e uso de códons semelhantes ao resto do genoma de *Xoc*. Esse resultado pode indicar que *Xoc* adquiriu o operon de GA₄ de um organismo com uso de códon semelhante, ou ainda, um evento de aquisição antigo com subsequente adaptação do códon ao longo do tempo (Brown, 2003; Lawrence, Ochman, 1997). Notadamente, o operon GA₄ é sintênico nas cepas de *Xoc* - encontrado em um mesmo contexto nestes genomas - sugerindo que o *operon* foi adquirido em um único evento ancestral de transferência.

O objetivo da análise de códon realizada foi identificar se os genes do operon de GA₄ estão inseridos na rede de transcrição de *Xoc*. O índice CAI aponta uma tendência de viés de códon nestes genes, que pode indicar a otimização do códon para maior expressão dos genes ou um forte viés mutacional inerente a *Xoc*. Ambos índices de códon utilizados apontam que os genes do *operon* de biossíntese de GA₄ apresentam um uso de códon comum ao contexto genômico de *Xoc*. Adicionalmente, o *operon* de GA₄ representa mais um exemplo de que conjuntos gênicos apresentam uma grande vantagem sobre genes solitários quando transferidos a um novo contexto genômico. Por serem inseridos *en bloc*, como uma unidade funcional, podem carregar os elementos regulatórios para expressão, já fazem parte de uma rede proteica e assim, são incorporados rapidamente pelo genoma recipiente, capazes de proporcionar imediatas mudanças fenotípicas ao organismo hospedeiro (Groisman, Ochman, 1996).

A avaliação da presença do *operon* em diferentes populações de *Xoc* é fundamental para contextualizar sua história evolutiva e entender sua importância para o *fitness* celular. A importância da produção do fitormônio GA₄ por *Xoc* é indicada pela alta prevalência do *operon* em diferentes populações, e pela ampla distribuição geográfica assim como, em um amplo período de tempo.

Embora sutilmente menos prevalente nos isolados do Oeste africano, o *operon* de GA₄ ainda está presente em alta frequência nesta região (>85%). Interessantemente, o fator

de virulência AvrRxo1 (juntamente com a proteína chaperona Arc1), similarmente flanqueado por elementos IS (Liu et al., 2014; Zhao et al., 2004), é mais prevalente nas linhagens asiáticas do que africanas (Triplett et al., 2016). Elementos móveis do tipo MITEs-Xo descritos nesta tese também apresentam uma dinâmica de mobilidade mais frequente nas linhagens africanas. Apesar desta tendência, os parâmetros de localidade ou data de coleta não permitem a visualização de um padrão e a razão pela qual a prevalência do *operon* de biossíntese de GA₄ é menor nesta região permanece sob investigação.

Até o momento, a síntese de GA₄ é restrita ao patovar *oryzicola* em *X. oryzae* e nosso resultado põe em evidência o papel importante do *operon* de GA₄ na história evolutiva de *Xoc*. Anteriormente à caracterização de GA₄ como produto final do operon, a molécula tinha sido classificada como um composto derivado de *ent*-kaurene (Lu et al., 2015). Neste estudo, mutantes de *Xoc* incapazes de sintetizar GA₄ pela deleção de genes iniciais da via biossintética, estão expostos a uma resposta mais acentuada de defesa por ácido jasmônico (JA) por parte do hospedeiro (indicada pelo aumento da expressão de genes marcadores OsLOX e OsAOS2). Dessa forma, foi especulado que *Xoc* modula a sinalização hormonal de arroz suprimindo a resposta de defesa mediada por JA no hospedeiro, através da produção de GA₄.

De fato, a presença do hormônio giberelina (GA) interfere na susceptibilidade de *O. sativa* a patógenos hemibiotróficos como *X. oryzae* (De Vleeschauwer et al., 2013). Diversos estudos demonstram que o acúmulo de giberelina gera maior suscetibilidade de *O. sativa* a *Xoo*, seja por aplicação ectópica de GA₃ anterior à inoculação, pela deleção de genes que degradam GAs bioativos (mutantes *eui*) (Yang et al., 2008), ou ainda pela superexpressão de genes da via biossintética de GA (como o gene OsGA20ox3 em arroz). Reciprocamente, a menor disponibilização de GA aumenta resistência de *O. sativa* a *Xoo* (Qin et al., 2012).

Classicamente investigada pelo seu papel no crescimento e desenvolvimento vegetal, a giberelina interfere na resposta de defesa de plantas porque também integra uma rede regulatória complexa de fitormônios na resposta de defesa a patógenos, que inclui JA, ácido salicílico (SA), etileno (Et), ácido absísico (ABA), auxina e citocina (Robert-Seilanianz et al., 2011). O componente compartilhado entre essas diversas vias de transdução hormonais, capaz de realizar o ajuste fino destes fitormônios são as proteínas DELLA.

As giberelinas regulam a degradação de proteínas DELLA. E estas últimas são responsáveis pela inibição de fatores de transcrição responsivos a giberelina. Brevemente, giberelinas se ligam primeiramente a um receptor solúvel (GID1), causando uma mudança de conformação que facilita a formação do complexo GA-GID1-DELLA. Ocorre a subsequente ubiquitinação da proteína DELLA e sua destruição pelo proteassomo 26S. Ou seja, o aumento da concentração de GA leva à degradação de proteínas DELLA (mediadas

por GID1) (Hedden, Thomas, 2012). A modulação da estabilidade de proteínas DELLA é o mecanismo pelo qual a giberelina participa da rede de respostas fitormonal no sistema de defesa da planta (Grant, Jones, 2009).

Interessantemente, a estabilidade da proteína DELLA do arroz (denominada SLR1, encontrada em cópia única no genoma) aumenta a resistência do arroz a *Xoo* e ativa ambas respostas de JA e SA de defesa, que atuam de forma integrada na resposta a patógenos biotróficos (Vleesschauwer et al., 2016). Se GA₄ produzido por *Xoc* é capaz de modular DELLA, diminuindo sua disponibilidade e consequentemente interferindo na resposta SA e JA de *O. sativa* ao patógeno *Xoc* permanece sob investigação.

Algumas espécies bacterianas simbióticas também são capazes de produzir giberelinas. O cluster de biossíntese de giberelina em simbioses se encontra geralmente imerso em ilhas e plasmídios simbióticos, sugerindo a atuação da giberelina no processo de formação de nódulos de fixação. Embora não haja definição precisa, GAs parecem ser importantes em ambas fase inicial de infecção e estabelecimento do nódulo primordial e fase de maturação do nódulo (Hayashi et al., 2014).

Comparado com o operon encontrado nos organismos simbioses, o operon de *Xoc* possui um gene a mais, o gene CYP115, responsável pela conversão de GA₉ ao hormônio GA₄ (biologicamente ativo). Em *B. japonicum*, GA₉ é sintetizado no bacteróide, provavelmente transportado a célula vegetal para então ser convertido a uma giberelina bioativa (Méndez et al., 2014). Dado que a atividade de hormônios vegetais são altamente reguladas, essa distinção entre os produtos finais do operon nos dois organismos pode indicar que a participação da giberelina no processo de infecção pelo rizóbios está, em algum nível, sob regulação da célula hospedeira. Enquanto em *Xoc*, a produção do composto bioativo pode atuar prontamente como um composto de interferência da transdução de sinalização, desencadeando uma resposta rápida.

A presença do operon de síntese GA₄ no domínio Bactéria, em organismos simbioses e fitopatógeno *Xanthomonas oryzae* pv. *oryzicola* está, por um lado, de acordo com a percepção que ambos estilos de vida compartilham estratégias que permitam entrada, replicação, mecanismos de evasão das defesas do sistema imune do hospedeiro vegetal, persistência, replicação e finalmente a saída para infecção de novos hospedeiros (Ochman, Moran, 2001). Patógenos e simbioses ainda compartilham mecanismos moleculares semelhantes no estabelecimento de interações com hospedeiros, como a utilização de *quorum-sensing*, sistema regulador de dois componentes para sincronização de expressão para se estabelecer a associação e produtos gênicos específicos para estabelecimento da interação (Hentschel, Steinert, Hacker, 2000).

No entanto, é importante considerar, neste caso do operon de GA₄, que existem particularidades fundamentais nos dois tipos de sistemas, que influenciam a resposta de

defesa do hospedeiro. Por exemplo, rizóbios estabelecem associações com diferentes dicotilédneas, enquanto *Xoc* infesta a monocotiledônia *O. sativa*. A sinalização hormonal da resposta de defesa também pode variar entre tecidos distintos, como raízes, invadida pelos simbioss, e folhas, invadidas por *Xoc*. E finalmente, *Xoc* ainda pode estabelecer o processo de infestação através de lesões foliares, além da colonização da câmara sub-estomática após entrada pelos estômatos.

Neste momento, não é possível definir se a diferença entre os modos de infecção de *Xoc* (propagação pelo mesófilo) e *Xoo* (invasão do xilema) pode ser atribuída à síntese de GA₄. É necessário entender de forma mais precisa qual etapa molecular do processo de colonização por *Xoc* está sendo influenciado pela produção de GA₄ e avaliar o efeito da aquisição desta característica no modo de vida de *Xoc*.

REFERÊNCIAS*

- Brown JR. Ancient horizontal gene transfer. *Nat Rev Genet.* 2003;4(2):121–32.
- Davis JJ, Olsen GJ. Characterizing the Native Codon Usages of a Genome: An Axis Projection Approach. *Mol Biol Evol.* 2011;28(1):211–21.
- De Vleeschauwer D, Gheysen G, Höfte M. Hormone defense networking in rice: tales from a different world. *Trends Plant Sci.* 2013;18(10):555–65.
- Dhillon BK, Laird MR, Shay JA, Winsor GL, Lo R, Nizam F, et al., IslandViewer 3: more flexible, interactive genomic island discovery, visualization and analysis. *Nucleic Acids Res.* 2015;43(W1):W104–8.
- Grantham R, Gautier C, Gouy M, Mercier R, Pavé A. Codon catalog usage and the genome hypothesis. *Nucleic Acids Res.* 1980;8(1):r49–62.
- Hayashi S, Gresshoff PM, Ferguson BJ. Mechanistic action of gibberellins in legume nodulation. *J Integr Plant Biol.* 2014;56(10):971–8.
- Hedden P, Thomas SG. Gibberellin biosynthesis and its regulation. *Biochem J.* 2012;444(1):11–25.
- Hershey DM, Lu X, Zi J, Peters RJ. Functional Conservation of the Capacity for ent-Kaurene Biosynthesis and an Associated Operon in Certain Rhizobia. *J Bacteriol.* 2014;196(1):100–6.
- Lawrence JG, Ochman H. Amelioration of bacterial genomes: rates of change and exchange. *J Mol Evol.* 1997;44(4):383–97.
- Lima WC, Sluys M-AV, Menck CFM. Non-Gamma-Proteobacteria Gene Islands Contribute to the *Xanthomonas* Genome. *OMICS J Integr Biol.* 2005;9(2):160–72.
- Liu H, Chang Q, Feng W, Zhang B, Wu T, Li N, et al., Domain Dissection of AvrRxo1 for Suppressor, Avirulence and Cytotoxicity Functions. *PLOS ONE.* 2014 Dec 1;9(12):e113875.
- Lu X, Hershey DM, Wang L, Bogdanove AJ, Peters RJ. An ent-kaurene-derived diterpenoid virulence factor from *Xanthomonas oryzae* pv. *oryzicola*. *New Phytol.* 2015;206(1):295–302.
- Méndez C, Baginsky C, Hedden P, Gong F, Carú M, Rojas MC. Gibberellin oxidase activities in *Bradyrhizobium japonicum* bacteroids. *Phytochemistry.* 2014;98:101–9.
- Morrone D, Chambers J, Lowry L, Kim G, Anterola A, Bender K, et al., Gibberellin biosynthesis in bacteria: Separate ent-copalyl diphosphate and ent-kaurene synthases in *Bradyrhizobium japonicum*. *FEBS Lett.* 2009;583(2):475–80.
- Nett RS, Montanares M, Marcassa A, Lu X, Nagel R, Charles TC, et al. Elucidation of gibberellin biosynthesis in bacteria reveals convergent evolution. *Nat Chem Biol.* 2016; advance online publication.
- Puigbò P, Bravo IG, Garcia-Vallve S. CALcal: A combined set of tools to assess codon usage adaptation. *Biol Direct.* 2008;3:38.

*De acordo com:

International Committee of Medical Journal Editors. [Internet]. Uniform requirements for manuscripts submitted to Biomedical Journal: sample references. [updated 2011 Jul 15]. Available from: <http://www.icmje.org>

Qin X, Liu JH, Zhao WS, Chen XJ, Guo ZJ, Peng YL. Gibberellin 20-Oxidase Gene OsGA20ox3 Regulates Plant Stature and Disease Development in Rice. *Mol Plant Microbe Interact.* 2012;26(2):227–39.

Raymundo AK, Briones AM, Ardales EY, Perez MT, Fernandez LC, Leach JE, et al., Analysis of DNA Polymorphism and Virulence in Philippine Strains of *Xanthomonas oryzae* pv. *oryzicola*. *Plant Dis.* 1999;83(5):434–40.

Robert-Seilanianz A, Grant M, Jones JDG. Hormone Crosstalk in Plant Disease and Defense: More Than Just Jasmonate-Salicylate Antagonism. *Annu Rev Phytopathol.* 2011;49(1):317–43.

Tatsukami Y, Ueda M. Rhizobial gibberellin negatively regulates host nodule number. *Sci Rep.* 2016;6: 27998.

Triplett LR, Shidore T, Long J, Miao J, Wu S, Han Q, et al., AvrRxo1 Is a Bifunctional Type III Secreted Effector and Toxin-Antitoxin System Component with Homologs in Diverse Environmental Contexts. *PLOS ONE.* 2016;11(7):e0158856.

Vleesschauwer DD, Seifi HS, Filipe O, Haeck A, Huu SN, Demeestere K, et al., The DELLA Protein SLR1 Integrates and Amplifies Salicylic Acid- and Jasmonic Acid-Dependent Innate Immunity in Rice. *Plant Physiol.* 2016;170(3):1831–47.

Wilkins KE, Booher NJ, Wang L, Bogdanove AJ. TAL effectors and activation of predicted host targets distinguish Asian from African strains of the rice pathogen *Xanthomonas oryzae* pv. *oryzicola* while strict conservation suggests universal importance of five TAL effectors. *Plant Biot Interact.* 2015;536.

Wonni I, Cottyn B, Detemmerman L, Dao S, Ouedraogo L, Sarra S, et al. Analysis of *Xanthomonas oryzae* pv. *oryzicola* Population in Mali and Burkina Faso Reveals a High Level of Genetic and Pathogenic Diversity. *Phytopathology.* 2013;104(5):520–31.

Yang D-L, Li Q, Deng Y-W, Lou Y-G, Wang M-Y, Zhou G-X, et al., Altered Disease Development in the *eui* Mutants and *Eui* Overexpressors Indicates that Gibberellins Negatively Regulate Rice Basal Disease Resistance. *Mol Plant.* 2008;1(3):528–37.

Zhao B, Ardales EY, Raymundo A, Bai J, Trick HN, Leach JE, et al., The *avrRxo1* Gene from the Rice Pathogen *Xanthomonas oryzae* pv. *oryzicola* Confers a Nonhost Defense Reaction on Maize with Resistance Gene *Rxo1*. *Mol Plant Microbe Interact.* 2004;17(7):771–9.

3 ANÁLISE DO IMPACTO DE ELEMENTOS DO TIPO IS NO SISTEMA DE REPARO DE DNA EM *XANTHOMONAS ORYZAE*

3.1 Introdução

A evolução de genomas bacterianos resulta do equilíbrio entre estabilidade e flexibilidade do genoma. A *estabilidade* funciona como um mecanismo de manutenção das condições operantes celulares e previnem o organismo de estocasticamente diminuir seu *fitness* em seu nicho. Os mecanismos de reparo de DNA têm papel fundamental para a estabelecimento da estabilidade estrutural da molécula e manutenção e restauração da sequência nucleotídica presente, garantindo a integridade do genoma.

A flexibilidade genômica provê a variabilidade genética que garante a sobrevivência do organismo mediante mudanças ambientais. Em contrapartida às mutações pontuais que geram variação nucleotídica lentamente, rearranjos genômicos como deleções, duplicações, ampliações, inserções e translocações são processos mutacionais de rápido impacto. São responsáveis pela geração de novos conjuntos de informações, seja por redução ou aumento do conjunto gênico, ou ainda, pela reconfiguração das informações já existentes (Darmon, Leach, 2014).

Juntamente com outros tipos de elementos genéticos móveis, elementos IS são mediadores importantes de processos de variabilidade e portanto, na plasticidade genômica do grupo Bacteria. Fornecem base estrutural para duplicações, deleções, rearranjos e incorporação de DNA exógeno, que podem acarretar em mudanças fenotípicas de relevância evolutiva.

Apesar de contribuírem de forma distintas na evolução de genomas bacterianos, o sistema de reparo de DNA e a transposição de elementos são processos associados celularmente. Ao mesmo tempo que o sistema de reparo tem uma participação fundamental no restabelecimento da integridade genômica após o processo de transposição, a atividade transpositiva de algumas famílias de IS são estimuladas por danos ao DNA (e que naturalmente acionam o sistema de reparo).

O processo de transposição de elementos ocorre pela ação de transposases, enzimas versáteis que catalisam a clivagem das extremidades do elemento e sua transferência para um novo sítio de inserção. O processo de transposição inerentemente gera quebras nas moléculas de DNA e determina que a mobilização do elemento seja acoplada ao sistema de reparo de DNA da célula hospedeira. O tipo de quebra de DNA depende do tipo de transposase, podendo clivar somente uma ou ambas fitas de DNA. Atualmente, transposons de DNA das famílias Tc1-Mariner, hAT e elementos P conhecidamente realizam corte na fita dupla de DNA, enquanto outras famílias, como

Helitrons e *Mavericks*, realizam sua inserção através de cortes de fita simples no sítio de inserção (Varani et al., 2015).

Como descrito no capítulo 1, elementos IS, compostos em sua maioria por transposases do tipo DDE, são capazes de se transporem de maneira conservativa ou replicativa. Na transposição conservativa, os elementos são excisados por completo do sítio doador e integrados a um novo sítio genômico, e sistema de reparo de DNA é evocado para o sanar os danos consequentes das clivagens às fitas de DNA envolvidas neste processo. O reparo de DNA também é evocado em transposições replicativas, que acarreta no aumento do número de cópias do elemento. Na transposição replicativa, o sítio doador e receptor do transposon podem se fundir, resultando em uma estrutura semelhante a uma forquilha de replicação. Além das clivagens realizadas no momento de excisão do elemento, é necessário o reconhecimento e a resolução estrutural da molécula de DNA pela replicação.

O mecanismo de transposição não é conhecido para muitas das 25 diferentes famílias de IS e, consequentemente, o papel de diferentes componentes do sistema de reparo no mecanismo de transposição ainda é incipiente. No entanto, existem estudos que associam genes de reparo à modulação do processo de transposição.

Os genes *priA* e *ssb* exemplificam como o repertório de genes do sistema de reparo da célula pode modular a atividade transpositiva de elementos e representar um meio de regulação de transposição por parte da célula hospedeira (Nagy, Chandler, 2004). A ausência da helicase *priA*, essencial para via de recomeço da forquilha de replicação, aumenta a taxa de transposição do elemento Tn7, provavelmente pela existência mais prolongada de estruturas resultantes do sistema de reparo menos eficiente. E em membros da família IS220/IS605, a proteína Ssb resulta em uma baixa eficiência de integração do elemento *in vivo* (Lavatine et al., 2016). Ainda, a transposase presente em membros da família IS1634 interagem fisicamente com o *clamp* β de DNA em *Acidiphilium*, *Leptospirillum* e *E. coli*. Os *clamp* β de DNA são essenciais para replicação e reparo do DNA, e associação destes fatores às diversas proteínas aumentam eficiência enzimática deste processo celular (Bloom, 2009). O aumento dessa força de interação entre a transposase e o *clamp* β aumenta a taxa de transposição do elemento *in vivo* (Díaz-Maldonado et al., 2015).

Interessantemente, a atividade de transposição pode observada como uma consequência celular após stresses que resultam em dano ao DNA. Foi demonstrado que a transposição do elemento IS10 é induzida por luz ultravioleta, dependente do sistema de resposta SOS (Eichenbaum, Livneh, 1998); a transposição de Tn7 é estimulada por quebra de fita dupla de DNA através do reconhecimento da replicação de DNA durante o reparo deste tipo de dano (pela enzima TnsE, codificada em Tn7) (Shi et al., 2008); e atividade constitutiva de *recA*, que inicia a resposta SOS a dano, aumenta a excisão de transposição de Tn5 (Kuan et al., 1991).

Ainda, o aumento do número de cópias de *ISDra2* (da família *IS200/IS605*) é fortemente induzida por irradiação celular de *Deinococcus radiodurans*. Neste caso, presume-se que o mecanismo de ativação da transposição se deve ao aumento da disponibilidade de fita simpels de DNA após irradiação, uma vez que esta classe de elementos tanto depende de um intermediário de de fita simples de DNA como é inserido nas moléculas de DNA nesta conformação (Pasternak et al., 2010). O processo de transposição induzido por stress é geralmente entendido como uma forma de gerar diversidade em condições ambientais adversas e promover adaptação.

Assumimos que o equilíbrio entre o processo de transposição e o sistema de reparo de DNA deve refletir o equilíbrio entre a *flexibilidade* e *estabilidade* necessários para sobrevivência de uma bactéria. Neste capítulo, revisamos o repertório de genes de reparo em *X. oryzae* e investigamos o potencial dos elementos ISs modularem o processo celular central de reparo de DNA em *X. oryzae*, que apresenta uma amplificação de elementos IS em relação às outras espécies de *Xanthomonas*.

Assim, a amplificação do número de elementos do tipo IS observada em *X. oryzae* pode ser fenotipicamente silenciosa e tem caráter evolutivamente neutro. Alternativamente, a amplificação dos elementos do tipo IS nas linhagens de *X. oryzae* é resultante de um sistema de reparo de DNA ineficiente da célula, constituindo um contexto que favoreça a atividade transpositiva. Um dos cenários especulados, por exemplo, é de um “efeito esparadrapo” adotado pelo genoma hospedeiro, no qual o processo de reparo utilizado no momento pós-transposição sirva para recrutar elementos de reparo para um dano que está prolongadamente sem restauração.

Inicialmente, comparamos a quantidade de ISs preditos para genomas do grupo Xanthomonadaceae. Em seguida, observamos a presença de elementos IS próximos a genes de reparo em *X. oryzae* comparado a outras espécies de *Xanthomonas*, e finalmente investigamos com maior detalhe os eventos de proximidade entre um gene de reparo e um IS em diferentes cepas de *X. oryzae*.

3.2 Materiais e Métodos

3.2.1 Predição do número de ISs

A predição do número de ISs em genomas de Xanthomonadaceae foi realizada pela ferramenta ISSaga (Varani et al., 2011). Os 69 genomas analisados estão disponíveis no banco de dado NCBI. Todos os genomas utilizados apresentam genomas completos (Tabela 3.1). Para as cepas de *X. oryzae*, além do número total de ISs preditos, tabulamos os dados referentes a distribuição de cada família de IS (Tabela 3.2)

3.2.2 *Análise de proximidade a partir de anotações genômicas*

Analizamos a proximidade entre genes de reparo e elementos IS em 11 genomas escolhidos como representantes do gênero *Xanthomonas*. As distâncias entre genes de reparo e ISs em número de genes foram calculadas a partir de anotações genômicas disponíveis nos arquivos .gff de cada um dos 11 genomas, baixadas do servidor ftp do banco de dados do NCBI (<ftp://ftp.ncbi.nlm.nih.gov/>). Localizados os genes de reparo, foi então calculado quantos genes existiam até o próximo gene com anotação referente a elementos de transposição (e.g. 'TE', 'IS', 'transposase', etc).

As 11 linhagens escolhidas como representantes das espécies de *Xanthomonas* são *X. oryzae* pv. *oryzicola* BLS256, *X. oryzae* pv. *oryzae* MAFF311018, *X. albilineans* GPE PC73, *X. axonopodis* Xac29-1, *X. axonopodis* pv. *citrumelo* F1, *X. axonopodis* pv. *citri* 306, *X. campestris* pv. *raphani* 756C, *X. campestris* pv. *campestris* ATCC 33193, *X. campestris* pv. *campestris* 8004, *X. campestris* pv. *vesicatoria* 85-10 e *X. citri* pv. *citri* Aw12879.

A distância entre os genes de reparo e IS foram medidos em número de ORFs e classificados por cor. Magenta indica que um IS é imediatamente vizinho ao gene de reparo; vermelho, que um elemento IS está de 1 a 5 genes de distância do gene de reparo; laranja, 6 a 10 genes de distância; amarelo, 11 a 15 genes de distância; azul, mais de 15 genes de distância.

3.2.3 *Análise de proximidade gene a gene*

Na análise gene a gene, investigamos em maior detalhe a proximidade entre um gene de reparo e elementos IS nas diferentes cepas de *X. oryzae*. Primeiramente a sequência nucleotídica dos genes de reparo utilizadas como referência foram extraídas do banco de dados KEGG (Apêndice D). Em geral, o organismo do qual o gene de reparo referência foi extraído pertence ao gênero *Xanthomonas*. Se nenhuma espécie do gênero *Xanthomonas* apresentava o gene de referência, a sequência foi extraída de um organismo dentro do grupo Xanthomonadaceae.

Os genes do sistema de reparo de DNA analisados foram identificados a partir do banco de dados KEGG (Kyoto Encyclopedia of Genes and Genomes) como parte da rede gênica 'Reparo e replicação', dentro de 'Processamento da Informação Gênica'. A tabela completa de genes de reparo analisados e os organismos de onde foram extraídas as sequências referência encontra-se em Apêndice D.

Usando a sequência referência do gene de reparo, foi feita análise de semelhança local (blastn) contra um banco de dados composto pelo conjunto de genomas descritos em Tabela 3.1. Dessa forma, extraímos as coordenadas de cada gene de referência em cada genoma analisado. A partir destas coordenadas do gene de reparo, ampliamos 2500bp a montante e a jusante, extraindo a sequência nucleotídica de toda a da região genômica (para cada gene, em cada genoma). *In-house perl scripts* foram utilizados para organização dos arquivos de saída.

As regiões genômicas ampliadas foram então submetidas à procura semi-automática de ISs na plataforma ISSaga (Varani et al., 2011). Para a regiões que acusaram a presença de ISs, foi feita a análise manual do elemento, identificando o tipo de IS, a orientação e a distância do gene de reparo. Elementos sequências de inserção (ISs) foram identificados pelo ISfinder (servidor <http://www-is.biotoul.fr/>) (Siguier et al., 2006b). A distância entre genes de reparo e os elementos IS identificada foi calculada pela procura de BLAST imbutida nesta plataforma ISFinder.

As 10 cepas atualmente sequenciadas de *Xoc* são provenientes da Ásia e Oeste Africano: Filipinas (BLS256, BLS279), Malásia (CFBP2286), China (RS105, B812, L8), Índia (BXORI), Burkina Faso (CFBP7341, CFBP7342) e Mali (CFBP7331). Das seis cepas de Xoo, cinco são asiáticas: Coreia (KACC), Japão (MAFF), Filipinas (PXO83, PXO86 e PXO99A). A única cepas de Xoo africana é proveniente de Camarões (AXO1947). A Tabela 4.1, no Capítulo 4, descreve as identificações destas cepas no banco NCBI, origem e data de coleta.

3.3 Resultados

3.3.1 Amplificação de elementos IS em *X. oryzae* no grupo Xanthomonadaceae

A presença de IS em genomas bacterianos é abundante, diversa, amplamente distribuída, e muitas vezes independe da proximidade filogenética entre eles (Chain et al., 2004; Parkhill et al., 2003; Siguier et al., 2006a). Esta característica também está presente no grupo Xanthomadaceae. As três espécies do gênero *Stenotrophomonas* apresentam de 8 a 52 IS preditos, sendo que as cinco cepas pertencentes à mesma espécie de *S. maltophilia* variam entre 8 e 38 ISs preditos (Tabela 3.1).

As outras espécies de Xanthomonadaceae (*Pseudoxanthomonas*, *Lysobacter*, *Rhodanobacter* e *Dokdonella*) também não ultrapassam de 40 ISs preditos em seus genomas. Somente *Xylella fastidiosa* apresentam quantidade iguais de elementos nas três cepas analisadas, cada uma com 15 ISs preditos.

O gênero *Xanthomonas* apresenta uma tendência de aumento do número de ISs nos genomas comparado às outras espécies do grupo Xanthomoandaceae. A ampla distribuição

Tabela 3.1 - Predição do número de elementos IS em diferente espécies em Xanthomonadaceae

Espécie	NCBI ID	Tamanho do genoma (Gb)	Total de IS	Total de ORFs	ORFs		
					Completo	Parcial	NC
Xanthomonas							
X. oryzae pv. oryzicola BLS256	CP003057.2	4.83	349	384	190	188	6
X. oryzae pv. oryzicola CFBP2286	CP011962.1	4.97	356	391	190	194	6
X. oryzae pv. oryzicola RS105	CP011961.1	4.78	345	388	178	206	4
X. oryzae pv. oryzicola B8-12	CP011955.1	4.79	337	380	178	198	4
X. oryzae pv. oryzicola BLS279	CP011956.1	4.79	328	371	177	190	4
X. oryzae pv. oryzicola L8	CP011960.1	4.80	333	376	178	194	4
X. oryzae pv. oryzicola BXOR1	CP011957.1	4.69	278	317	151	160	6
X. oryzae pv. oryzicola CFBP7341	CP011959.1	5.02	425	491	275	209	7
X. oryzae pv. oryzicola CFBP7331	CP011958.1	5.01	440	514	288	220	6
X. oryzae pv. oryzicola CFBP7342	CP007221.1	5.08	428	528	317	202	9
X. oryzae pv. oryzae AXO1947	NZ_CP013666.1	4.67	406	468	247	215	5
X. oryzae pv. oryzae KACC10331	NC_006834.1	4.94	832	920	285	622	12
X. oryzae pv. oryzae PXO86	NZ_CP007166.1	5.02	818	914	362	541	8
X. oryzae pv. oryzae MAFF311018	NC_007705.1	4.94	785	881	322	550	7
X. oryzae pv. oryzae PXO83	NZ_CP012947.1	5.03	820	914	366	535	10
X. oryzae pv. oryzae PXO99A	NC_010717.2	5.24	826	926	367	547	10
X. axonopodis Xac29-1	NC_020800.1	5.15	46	72	48	21	3
X. axonopodis pv. dieffenbachiae LMG 695	NZ_CP014347.1	5.04	26	40	20	14	6
X. axonopodis pv. citrumelo F1	NC_016010.1	4.97	32	51	23	27	1
X. axonopodis pv. citri 306	NC_003919.1	5.18	51	84	71	10	3
X. citri pv. citri jx-6	NZ_CP011827.2	5.12	45	75	65	8	2
X. citri subsp. citri gd2	NZ_CP009019.1	5.13	45	75	65	8	2
X. citri subsp. citri gd3	NZ_CP009016.1	5.13	45	75	65	8	2
X. citri subsp. citri jx4	NZ_CP009013.1	5.12	45	75	65	8	2
X. citri subsp. citri jx5	NZ_CP009010.1	5.12	45	75	65	8	2
X. citri subsp. citri MN10	NZ_CP009004.1	5.12	45	75	65	8	2
X. citri subsp. citri MN12	NZ_CP008998.1	5.12	45	75	65	8	2
X. citri subsp. citri UI7	NZ_CP008989.1	5.12	45	75	65	8	2
X. citri subsp. citri UI6	NZ_CP008992.1	5.12	45	75	65	8	2
X. citri subsp. citri MN11	NZ_CP009001.1	5.16	49	80	69	9	2
X. citri subsp. citri A306	NZ_CP006857.1	5.18	51	84	71	10	3
X. citri subsp. citri BL18	NZ_CP009025.1	5.18	51	84	71	10	3
X. citri subsp. citri FB19	NZ_CP009022.1	5.18	51	84	71	10	3
X. citri subsp. citri mf20	NZ_CP009007.1	5.18	51	84	71	10	3
X. citri subsp. citri NT17	NZ_CP008995.1	5.18	51	84	71	10	3
X. citri subsp. citri Aw12879	NC_020815.1	5.32	67	110	87	21	2
X. citri subsp. citri AW13	NZ_CP009031.1	5.32	67	110	87	21	2
X. citri subsp. citri AW16	NZ_CP009040.1	5.32	67	110	84	24	2
X. citri subsp. citri AW15	NZ_CP009037.1	5.32	68	111	86	23	2
X. citri subsp. citri AW14	NZ_CP009034.1	5.32	70	112	84	26	2
X. campestris pv. campestris 17	NZ_CP011946.1	4.99	90	113	60	46	7
X. campestris pv. campestris B100	NC_010688.1	5.08	100	132	75	51	6
X. campestris pv. vesicatoria 8510	NC_007508.1	5.18	85	132	97	29	6
X. campestris 17	NZ_CP011256.1	5.25	51	62	29	23	10
X. campestris pv. raphani 756C	NC_017271.1	4.94	61	78	48	26	4
X. campestris pv. campestris 4013	NZ_CP012146.1	4.91	108	137	76	54	7
X. campestris pv. campestris 8004	NC_007086.1	5.15	111	139	79	53	7
X. campestris pv. campestris ATCC 33913	NC_003902.1	5.08	107	139	81	51	7
X. campestris pv. campestris 32	NZ_CP012145.1	4.91	104	139	68	60	11
X. arboricola pv. juglandis Xaj 417	NZ_CP012251.1	5.22	107	147	113	22	12
X. fuscans ssp. fuscans 4834-R	NC_022541.1	4.98	127	133	88	43	2
X. translucens pv. undulosa Xtu4699	NZ_CP008714.1	4.56	101	136	89	37	8
X. sacchari R1	NZ_CP010409.1	4.50	34	39	27	8	4
X. albilineans GPE PC73	NC_013722.1	3.77	65	95	25	63	6
Xylella							
Xylella fastidiosa 9a5c	AE003849.1	2.68	15	18	1	6	11
Xylella fastidiosa M23	CP001011.1	2.54	15	17	0	7	10
Xylella fastidiosa Temecula1	AE009442.1	2.52	15	17	0	7	10
Stenotrophomonas							
Stenotrophomonas acidaminiphila	NZ_CP012900.1	4.14	52	59	26	25	8
Stenotrophomonas rhizophila DSM14405	CP007597.1	4.65	19	29	15	11	3
Stenotrophomonas maltophilia D457	HE798556.1	4.77	25	37	29	8	0
Stenotrophomonas maltophilia 13637	CP008838.1	4.99	34	41	24	14	3
Stenotrophomonas maltophilia JV3	CP002986.1	4.54	8	12	8	0	4
Stenotrophomonas maltophilia K279a	AM743169.1	4.85	38	53	45	8	0
Stenotrophomonas maltophilia R551-3	CP001111.1	4.57	19	22	20	1	1
Outras Xanthomonadaceae							
Pseudoxanthomonas spadix BD-a59	CP003093.2	3.45	26	29	6	14	9
Lysobacter capsici 55	NZ_CP011130.1	6.39	22	35	30	2	3
Lysobacter enzymogenes C3	NZ_CP013140.1	6.16	27	27	16	5	6
Rhodanobacter denitrificans 2APBS1	CP003470.1	4.23	40	44	22	5	16
Dokdonella koreensis DS-123	CP015249.1	4.45	16	20	11	8	1

NC - não classificado

de ISs preditos pode ser observada entre *X. oryzae* pv. *oryzae* KACC10331, que apresenta 32 vezes mais elementos preditos (832 ISs) do que *Xanthomonas axonopodis* pv. *dieffenbachiae*, patógeno de antúrio e espécie da família Araceae, que apresenta 26 ISs preditos, menor número de elementos dentro do gênero.

Diversos genomas de *Xanthomonas citri* pv. *citri* foram sequenciados devido sua grande importância agrônoma, causadora de cancro em diversas espécies de citros, e o número de ISs preditos varia entre 45 e 70 inserções. Linhagens de *X. campestris* apresentam um número de ISs preditos maior em relação a *X. citri*, e a quantidade total de elementos encontra-se em torno de 100 ISs no genoma.

A restrição de habitat ao xilema de plantas de cana-de-açúcar é uma das hipóteses aventadas acerca da redução do genoma de *X. albilineans* (Pieretti et al., 2009). Embora esse processo evolutivo leve à consequente redução de elementos presentes no genoma, *X. albilineans* apresenta 65 ISs preditos, quantidade maior do que outras espécies de fora do gênero *Xanthomonas* analisadas.

Elementos ISs são considerados elementos determinantes na diversificação de linhagens de *X. oryzae* (Ochiai et al., 2005; Triplett et al., 2011). Dentro do gênero *Xanthomonas*, a espécie *X. oryzae* é caracterizada pela amplificação do número de elementos IS. Nos genomas das 16 cepas disponíveis, o número de elementos varia entre 278 e 832 elementos.

3.3.2 Distribuição das famílias de IS em *X. oryzae*

Em concordância com resultados anteriormente reportados (Monteiro-Vitorello et al., 2005; Ryan et al., 2011), o conjunto de elementos ISs na espécie *X. oryzae* é majoritariamente distribuído em 9 famílias distintas – IS1595, IS256, IS3, IS30, IS4, IS5, IS630, IS701 e ISL3. As famílias IS481, IS110, IS200/IS605 e ISNCY estão presentes no genoma porém com número de cópias muito reduzido; e as famílias IS21 e Tn3 estão presentes somente em algumas das cepas analisadas (Tabela 3.2).

Os dois patovares de *X. oryzae*, *Xoo* e *Xoc*, apresentam diferenças importantes na quantidade e padrão de distribuição dos elementos. Cepas de *Xoo* geralmente apresentam um maior número de elementos IS, aproximadamente o dobro do que *Xoc*, devido à amplificação das famílias ISL3, IS701, IS5ssgrlIS427 e IS1595ssgrlIS1595. A família IS5ssgrlIS1031 também está amplificada em *Xoo*, no entanto, a maioria dos elementos estão parciais.

Em relação ao patovar *oryzicola* (*Xoc*), a disponibilidade de 10 cepas sequenciadas oferece a oportunidade de avaliar a variação de elementos intra-populacionalmente. Cepas de *Xoc* apresentam um perfil muito semelhante, principalmente entre as cepas provenientes da Ásia (BLS256, BLS279, CFBP2286, RS105, B8-12, L8 e BXORI).

Tabela 3.2 - Perfil de distribuição de elementos IS em cepas de *X. oryzae*.

Cepa de do genoma	Tamanho (Gb)	Famílias de elementos IS (em números de ORF)																										
		IS1595		IS256		IS3		IS30		IS4		IS5		IS630		IS701		IS481		IS200/IS605		ISL3		ISNCY		IS21		
		ISs	sgrs	ISs	sgrs	ISs	sgrs	ISs	sgrs	ISs	sgrs	ISs	sgrs	ISs	sgrs	ISs	sgrs	ISs	sgrs	ISs	sgrs	ISs	sgrs	ISs	sgrs	ISs	sgrs	
<i>X. oryzae</i>		384	29 (23/6)	5 (0/5)	1	3 (0/1)	71 (36/35)	28 (8/20)	38 (31/7)	7 (4/3)	1 (1/0)	16 (2/14)	3 (2/1)	142 (64/77)	1 (0/1)	15 (6/9)	15 (10/5)	3 (0/3)	1 (1/0)	1 (1/0)	2	1 (1/0)	1 (0/1)	2	1 (1/0)	1 (0/1)	-	-
BLS256	4.83	391	29 (23/6)	5 (0/5)	1	3 (0/1)	70 (32/38)	27 (8/19)	38 (31/7)	7 (4/3)	1 (1/0)	16 (2/14)	3 (2/1)	149 (67/81)	1 (0/1)	15 (6/9)	16 (11/5)	3 (0/3)	1 (1/0)	1 (1/0)	2	1 (1/0)	1 (0/1)	2	1 (1/0)	1 (0/1)	-	-
CFBP2286	4.97	388	29 (19/10)	5 (0/5)	1	1 (0/1)	86 (40/46)	29 (9/20)	47 (32/15)	6 (4/2)	1 (1/0)	15 (3/12)	3 (2/1)	127 (52/74)	1 (0/1)	14 (4/10)	14 (9/5)	3 (0/3)	1 (1/0)	1 (1/0)	2	1 (1/0)	1 (0/1)	2	1 (1/0)	1 (0/1)	-	-
RS105	4.78	380	29 (19/10)	5 (0/5)	1	1 (0/1)	86 (40/46)	29 (9/20)	39 (32/7)	6 (4/2)	1 (1/0)	15 (3/12)	3 (2/1)	127 (52/74)	1 (0/1)	14 (4/10)	14 (9/5)	3 (0/3)	1 (1/0)	1 (1/0)	2	1 (1/0)	1 (0/1)	2	1 (1/0)	1 (0/1)	-	-
B8-12	4.79	380	29 (19/10)	5 (0/5)	1	1 (0/1)	86 (40/46)	29 (9/20)	39 (32/7)	6 (4/2)	1 (1/0)	15 (3/12)	3 (2/1)	127 (52/74)	1 (0/1)	14 (4/10)	14 (9/5)	3 (0/3)	1 (1/0)	1 (1/0)	2	1 (1/0)	1 (0/1)	2	1 (1/0)	1 (0/1)	-	-
BLS279	4.79	371	24 (19/5)	5 (0/5)	1	1 (0/1)	84 (42/42)	29 (9/20)	37 (29/8)	5 (3/2)	1 (1/0)	15 (3/12)	3 (2/1)	126 (51/74)	1 (0/1)	14 (4/10)	16 (11/5)	3 (0/3)	1 (1/0)	1 (1/0)	2	1 (1/0)	1 (0/1)	2	1 (1/0)	1 (0/1)	-	-
L8	4.8	376	29 (19/10)	5 (0/5)	1	1 (0/1)	86 (40/46)	29 (9/20)	39 (33/6)	5 (3/2)	1 (1/0)	15 (3/12)	3 (2/1)	124 (52/71)	1 (0/1)	14 (4/10)	14 (9/5)	3 (0/3)	1 (1/0)	1 (1/0)	2	1 (1/0)	1 (0/1)	2	1 (1/0)	1 (0/1)	-	-
BXORI	4.69	317	29 (20/9)	4 (0/4)	1	1 (0/1)	72 (45/27)	25 (10/14)	32 (24/8)	3 (1/2)	1 (1/0)	17 (2/15)	9 (3/5)	87 (27/59)	1 (0/1)	12 (6/6)	13 (9/4)	4 (0/4)	1 (1/0)	1 (1/0)	2	1 (1/0)	1 (0/1)	2	1 (1/0)	1 (0/1)	-	-
CFBP7331	5.02	514	29 (19/10)	5 (0/5)	1	-	132 (96/36)	28 (12/15)	62 (51/11)	4 (1/3)	1 (1/0)	30 (1/29)	22 (8/14)	136 (56/79)	1 (0/1)	34 (28/6)	17 (11/6)	3 (0/3)	1 (1/0)	1 (1/0)	2	2 (1/0)	1 (0/1)	2	2 (1/0)	1 (0/1)	1 (1/0)	1 (0/1)
CFBP7341	5.01	491	30 (20/10)	6 (0/6)	1	-	113 (84/29)	28 (12/15)	70 (58/12)	4 (1/3)	1 (1/0)	26 (1/24)	21 (8/13)	128 (49/78)	1 (0/1)	34 (28/6)	16 (9/7)	3 (0/3)	1 (1/0)	1 (1/0)	2	2 (1/0)	1 (0/1)	2	2 (1/0)	1 (0/1)	1 (1/0)	1 (0/1)
CFBP7342	5.08	528	28 (15/13)	4 (0/4)	1	-	192 (151/41)	27 (8/17)	59 (38/20)	2 (0/2)	1 (1/0)	18 (3/15)	15 (7/7)	125 (59/65)	1 (0/1)	34 (28/6)	12 (6/6)	3 (0/2)	1 (1/0)	1 (1/0)	2	1 (1/0)	1 (0/1)	2	1 (1/0)	1 (0/1)	-	-
AXO1947	4.67	468	24 (2/22)	2 (0/2)	-	-	94 (61/31)	32 (11/21)	30 (15/15)	5 (0/5)	4 (0/4)	43 (1/42)	19 (9/9)	160 (118/42)	4 (0/4)	28 (23/5)	-	16 (4/12)	1 (1/0)	-	3 (1/0)	1 (1/0)	-	3 (1/0)	1 (1/0)	-	-	-
KACC	4.94	920	114 (43/71)	14 (5/9)	1 (0/0)	7 (4/3)	105 (44/61)	20 (2/15)	66 (19/47)	21 (10/11)	1 (0/1)	103 (4/98)	86 (38/47)	155 (54/100)	7 (1/6)	54 (23/31)	124 (19/104)	5 (0/5)	1 (0/1)	-	30 (17/9)	1 (1/0)	-	30 (17/9)	1 (1/0)	-	-	-
PXO86	5.02	914	98 (53/45)	14 (6/8)	-	8 (3/5)	107 (41/65)	16 (5/11)	97 (53/44)	16 (10/6)	2 (0/2)	104 (1/102)	97 (42/55)	154 (67/85)	7 (1/6)	39 (21/18)	111 (39/72)	5 (0/5)	1 (0/1)	1 (0/1)	1 (0/1)	1 (0/1)	-	30 (17/9)	1 (1/0)	-	1 (1/0)	-
MAFF	4.94	881	110 (53/55)	11 (8/3)	-	7 (4/3)	102 (43/59)	9 (5/4)	69 (17/52)	18 (5/13)	2 (0/2)	92 (1/91)	112 (46/66)	144 (64/79)	7 (0/7)	50 (25/25)	105 (32/73)	5 (0/5)	1 (0/1)	1 (0/1)	1 (0/1)	1 (0/1)	-	31 (17/10)	1 (1/0)	-	-	-
PXO83	5.03	914	95 (54/41)	13 (7/6)	8 (3/5)	108 (40/67)	10 (6/4)	101 (53/46)	16 (10/6)	2 (0/2)	104 (1/102)	96 (41/55)	159 (68/89)	7 (1/6)	39 (21/18)	112 (40/72)	5 (0/5)	1 (0/1)	1 (0/1)	1 (0/1)	1 (0/1)	1 (0/1)	-	30 (18/8)	1 (1/0)	-	1 (1/0)	-
PXO98A	5.24	926	102 (70/32)	13 (6/7)	2 (0/0)	13 (6/7)	108 (48/60)	22 (6/15)	66 (22/44)	16 (10/6)	2 (0/2)	104 (5/99)	100 (40/60)	154 (53/97)	6 (0/6)	48 (32/16)	132 (47/85)	3 (0/3)	1 (0/1)	1 (0/1)	1 (0/1)	1 (0/1)	-	26 (19/4)	1 (1/0)	-	1 (1/0)	-

NC - não classificado
sgr - subgrupo

As cepas oeste-africanas CFBP7331, CFBP7341 e CFBP7342 apresentam um aumento no número total de IS, resultante da amplificação das famílias IS30 e IS630, e dos subgrupos IS1031 e IS427 (família IS5) e IS3ssgrIS407 (família IS3).

Das seis cepas disponíveis do patovar *oryzae* (Xoo), observamos uma diferença importante entre as cinco cepas asiáticas (MAFF, KACC, PXO83, PXO86 e PXO99A) e AXO1947, proveniente de Camarões. Primeiramente, Xoo AXO1947 apresenta 406 inserções no genoma enquanto as asiáticas apresentam entre 785 a 832. Assim, a maioria das famílias em AXO1947 apresentam um número menor de elementos, sendo que as diferenças mais dramáticas são observadas nas famílias ISL3, IS5ssgrIS427, IS1595ssgrIS1595 e IS701. Esta última está presente em mais de 100 cópias nas cepas asiáticas e ausente em AXO1947. As famílias IS3ssgrIS51 e IS481, e IS5ssgrIS5 são exceções: estão em maior número em AXO1947, com uma alta proporção de inserções com ORFs completas.

Interessantemente, a diferença de perfil de distribuição entre as cepas Xoo asiáticas e africana é semelhante à diferença observada entre os patovares Xoo e Xoc. em ambas comparações, Xoo se distingue pela amplificação das famílias ISL3, IS701, IS5ssgrIS427 e IS1595ssgrIS1595.

Enquanto o número de ORF completas infere sobre o potencial de mobilidade dos elementos, ORFs parciais inferem sobre eventos de amplificação que ocorreram em momento anteriores. Os fragmentos remanescentes, ainda que quebrados, podem representar substratos para mobilidade ou recombinação homóloga. Observamos que as linhagens africanas de ambos patovares Xoo e Xoc, apresentam um maior número de ORFs completas do que parciais, enquanto linhagens asiáticas o número de ORFs incompletas ultrapassa o número de ORFs completas.

3.3.3 Repertório de genes de reparo em *X. oryzae*

Para avaliar a interferência de IS nos genes de reparo em *X. oryzae*, fizemos primeiramente um levantamento sobre o repertório de genes de reparo presentes em 11 linhagens representativas do gênero *Xanthomonas*, ampliando o universo anteriormente descrito por Martins-Pinheiro et al. (2004).

Processos de reparo de DNA podem ser divididos em três vias principais: reparo de quebra da fita dupla (DSBR, *double strand break repair*), reparo de quebra de fita simples (SSBR, *single strand break repair*) e ainda reparo por síntese de translesão (TLS, *translesion synthesis*). Os genes de reparo das vias de reparo por junção não-homóloga (NHEJ, *non-homologous end joining*) e de recombinação ilegítima dependente e independente de pequena homologia (SHDR e SHIR, respectivamente) também foram considerados para esta análise.

Dos 101 genes de reparo classificados pelo banco de dados KEGG como parte do 'Processamento da Informação Gênica', 26 genes estão ausentes, em comum nas 11 linhagens analisadas (blocos cinza nas Figuras 3.1, e ApêndiceE e F). A catalogação da presença e ausência dos genes de reparo analisados sugere que *Xanthomonas* não apresenta grande variação do repertório de genes de reparo de DNA e, de uma maneira geral, a ausência de um gene de reparo é universal para o grupo.

Em comparação às outras espécies, *X. albilineans* é a espécie que apresenta mais variação, principalmente pela ausência de glicosolases de DNA (*nei*, *mpg*, *ada-akIA*, *alkA*) e fatores de resposta SOS (*dnaE2* e *imuB*). *X. albilineans* e *X. oryzae* não apresentam os genes *cho*, *ku* e *ligD*. O gene *dam* é o único gene que está presente em *X. albilineans* e ausente/fragmentado nas outras 10 linhagens.

Figura 3.1 - Proximidade entre genes e elementos IS em diferentes espécies de *Xanthomonas* - via de reparo por síntese trans-lesão e outras vias acessórias. A distância entre os gene de reparo e IS foram medidos em numero de ORFs. Magenta, IS é imediatamente vizinho ao gene de reparo; vermelho, IS está de 1 a 5 genes de distância do gene de reparo; laranja, 6 a 10 genes de distância; amarelo, 11 a 15 genes de distância; azul, mais de 15 genes de distância. Cinza representa ausência do gene.

Os genes *umuC* e *umuD* apresentam uma distribuição irregular no grupo, ausentes em 7 das 11 linhagens identificadas, sendo que a presença não é espécie-específica: somente duas das quatro cepas de *X. campestris* apresentam os genes; em *X. oryzae*, estão presentes no patovar *oryzae* MAFF e ausentes no patovar *oryzicola*.

3.3.4 *X. oryzae* apresenta ISs próximos a genes de reparo

Avaliamos se elementos IS interferem na modulação de genes do sistema de reparo através da proximidade entre estes dois componentes. Cada gene de reparo foi localizado na anotação genômica das 11 linhagens de *Xanthomonas* e foi examinada sua proximidade a um elemento de transposição. Os eventos de proximidade foram então classificados em função da distância (em número de genes) gene de reparo-IS, assinalando uma cor a cada evento, de acordo com a distância de genes observada na anotação (Figuras 3.1, Apêndice E e Apêndice F).

O resultado permitiu uma avaliação visual dos eventos de proximidade. A maior frequência de eventos ‘magenta’ e ‘vermelho’ sugere que os elementos ISs presentes em *Xanthomonas oryzae* têm um elevado potencial de interferir no processo de reparo de DNA.

Dadas as diferenças do perfil de distribuição de ISs entre *Xoo* e *Xoc*, concentramos nossa atenção nesta espécie e comparamos os eventos de proximidade dos genes de reparo das principais vias entre as cepas *Xoc* BLS256 e *Xoo* MAFF311018, indicando a classificação por cor destes eventos dentro das principais vias de reparo (Figura 3.2).

Na via de reparo por excisão de base (BER, *base excision repair*), notamos a predominância de genes de reparo a 5 ou menos genes de distância de um elemento IS em ambos patovares (eventos magenta e vermelho). A DNA glicosilase *mutM* (ou *fpg*) é o único gene que apresenta ISs adjacentes comuns aos dois patovares. *Xoo* ainda apresenta o gene *nei* adjacente a um elemento.

Na via de reparo por excisão nucleotídica (NER, *nucleotide excision repair*), o gene *uvrB*, essencial para interação com *uvrA* e para o complexo proteína-DNA, possui uma inserção de IS em sua vizinhança imediata também em ambos patovares. O gene *uvrD* está adjacente a um IS somente em *Xoo*. Os genes *uvrC* e *dpoI* (ou *polA*) também representam genes com potencial de modulação por ISs.

Comparado a *Xoc*, a via de reparo de mal emparelhamento de bases (MMR, do inglês, *mismatch repair*) em *Xoo* parece estar sob maior influência de ISs. A ausência de *mutH* (central à via) e a presença de um IS adjacente a *mutS* são eventos comuns aos dois patovares. Na via de reparo por recombinação homóloga (HR, *homologous recombination*), a proximidade de ISs aos genes *recG*, *recO*, *priA* em *Xoo* e a *ssb* (em ambos *Xoo* e *Xoc*) podem representar interferências importantes

Em ambos *Xoc* e *Xoo*, todas as vias de reparo apresentam eventos de alta proximidade de ISs, incluindo genes de reparo associados à fase inicial de detecção ou reconhecimento da lesão no DNA. O mal funcionamento destes genes, logo no início de uma via de reparo, faria com que o processo de reparo fosse precocemente desacelerado ou impedido, colocando o DNA em uma situação crítica. Este cenário favoreceria nossa hipótese do ‘efeito esparadrapo’.

Figura 3.2 - Comparação dos eventos de proximidade de genes envolvidos nas principais vias de reparo de DNA e elementos IS entre Xoo e Xoc.

○ genes de reparo e IS - vizinhos imediatos
○ genes de reparo e IS - de 1 a 5 genes de distância
○ genes de reparo e IS - de 6 a 10 genes de distância
○ genes de reparo e IS - de 11 a 15 genes de distância
○ genes de reparo e IS - mais de 15 genes de distância

○ genes de reparo em *Xanthomonas oryzae pv. oryzae*
○ genes de reparo em *Xanthomonas oryzae pv. oryzae* MAFF311018

No entanto, a fim de avaliar com maior segurança o potencial de interferência de um elemento na via de reparo de DNA, estabelecemos para cada gene, a natureza do evento proximidade detectado.

3.3.5 Avaliação gene a gene dos eventos de proximidade IS-gene de reparo

O objetivo desta análise foi estabelecer, para cada evento de proximidade gene de reparo-IS em *Xoo* e *Xoc*: a) qual o tipo de ISs, b) se o elemento está completo ou parcial, c) a posição do IS em relação ao gene (a jusante ou a montante), d) a orientação do elemento em relação ao gene, e por fim d) a distância (em pares de base) entre o elemento e o gene de reparo.

A disponibilização no Genbank de 9 genomas de *Xoc* (Wilkins et al., 2015) e 1 genoma de *Xoo* (Huguet-Tapia et al., 2016) permitiu avaliar se os eventos de proximidade são conservados num total de 16 cepas de *X. oryzae*. Dessa forma, expandimos nossa análise de proximidade gene-IS e re-analisamos todos os 75 genes do sistema de reparo presentes nestas cepas.

Assim, descrevemos os detalhes de cada evento de proximidade classificado como 'magenta' e 'vermelho' da Figura 3.2, utilizando esta figura como referência para representação das novas informações (Figuras 3.3, 3.4, 3.5 e 3.6). Eventos anteriormente identificados como eventos de proximidade mas que não foram detectados nesta análise detalhada (ou seja, quando não foi encontrado nenhum elemento a 2500pb jusante ou a montante) foram considerados resultados falso-positivos e indicados com * nas Figuras.

Na **via de reparo por excisão de bases** (BER, Figura 3.3), diversas DNA glicosilases nas cepas de *Xoo* estão potencialmente sob potencial modulação de elementos IS. Nas cepas asiáticas (PXO83, PXO86, PXO99A, MAFF e KACC), o gene *mpg* apresenta o ISXo5 a 93pb a montante do gene e o gene *nei* apresenta um elemento IS427 a 148pb a jusante. Nas cepas PXO83 e PXO86, o gene *tag* apresenta o elemento IS1112b justaposto ao códon de terminação do gene.

Foram encontrados três ortólogos da endonuclease *xthA* nos genomas de *X. oryzae*, segundo o banco de dados KEGG. O evento de proximidade detectado anteriormente pela análise por anotação se refere à uma das cópias que se encontra quebrada (*xthA2*).

A glicosilase *mutM* apresenta um contexto genômico heterogêneo em *Xoo* e *Xoc*. Em *Xoc*, o ISXo4 parcial está conservado a 270pb a montante de *mutM*. A jusante de *mutM*, ISXo15 encontra-se a 152pb nas cepas *Xoc* asiáticas, sendo que em BLS256 e CFBP2286 este IS está interrompido. Nas linhagens *Xoc* africanas (CFBP7331, CFBP7341 e CFBP7342) há a inserção do IS1112a anteriormente ao elemento ISXo15, posicionado a 67bp a jusante do gene e caracterizando alto potencial de modulação deste gene pelo elemento.

Em *Xoo*, as cepas asiáticas apresentam diversos ISs conservados à montante de *mutM*, sendo ISXo15 parcial o elemento mais próximo ao gene (173pb de distância). A jusante de *mutM*, ISXo2 encontra-se a 284pb do final do gene em PXO86, PXO83, MAFF e KACC. Em PXO99A, o elemento IS1112a é o elemento mais próximo a jusante (245pb de distância). A cepa *Xoo* AXO1947 apresenta um único elemento ISXac5, a montante de *mutM*.

Na **via de reparo por excisão nucleotídica** (NER, Figura 3.4), verificamos que o evento de proximidade detectado no gene *uvrB* se refere à três tipos de inserção: presença do elemento ISXo17, conservadas a 394pb jusante do gene nas cepas de *Xoc* (com exceção de CFBP7342) e *Xoo* AXO1947; presença do elemento ISXoo13 parcial a 296pb jusante nas linhagens filipinas (PXO99A, PXO86 e PXO83); e à presença de ISXac5 parcial a 351bp a jusante de *uvrB* em KACC e MAFF.

Embora não representado na via principal de NER, o gene associado *rpoZ* apresenta o ISXo3 a 147pb a montante nas cepas asiáticas de *Xoo*, caracterizando um evento de proximidade de grande potencial de modulação do gene. A cepa *Xoo* AXO1947, juntamente com todas as cepas de *Xoc* não compartilham este evento de proximidade.

Os genes *mfd*, *uvrC*, *uvrD* e *rpoA* apresentam ISs próximos, porém os eventos de inserção estão distantes dos genes.

Na **via de reparo de bases mal emparelhadas** (MMR, Figura 3.7), diversos genes de reparo apresentam ISs em suas imediações genômicas. No entanto, a maioria destes eventos trata-se de inserções localizadas a certa distância dos genes. São os casos de *uvrD*, *xseA*, *ssb*, *holB*, *dnaE*, *dnaX* e *xseB*.

Somente os genes *dnaQ* e *mutS* apresentam inserções de elementos com potencial de modulação nesta via, e este evento de proximidade apresenta variação entre as cepas de *X. oryzae*. O elemento IS427 está aproximadamente 130pb a jusante de *dnaQ* somente nas três linhagens africanas de *Xoc*, com orientação oposta ao gene. Nas cepas *Xoo* KACC e *Xoo* AXO1947, diferentes tipos de elementos também estão a jusante de *dnaQ*, sendo que somente em KACC o elemento está em direção oposta ao gene.

Todas as cepas de *X. oryzae* apresentam ISs nas imediações do gene *mutS*. O IS407 está parcial e encontra-se a montante de *mutS*, conservado nas 10 cepas de *Xoc* analisadas e em *Xoo* AXO1947.

O elemento IS407 também está presente no contexto genômico das cinco cepas restantes de *Xoo*. Em KACC e MAFF, duas cópias completas do elemento, direcionadas opostamente entre si, flanqueiam o gene *mutS*; PXO99A apresenta um perfil semelhante a KACC e MAFF a montante, porém não apresenta o elemento completo a jusante; e em PXO83 e PXO86, os elementos IS407 a montante e a jusante de *mutS* encontram-se parciais.

Figura 3.4 – Proximidade entre ISs e genes da via de reparo por excisão nucleotídica em cepas de *X. oryzae*. O gene de reparo está indicado em amarelo nas cepas Xoc e verde, nas cepas Xoo. Em azul, elementos de transposição. Elementos parciais estão indicados por barras brancas no início ou fim do elemento A distância entre o gene de reparo e um elemento IS está indicada em parênteses.

Figura 3.6 – Proximidade entre ISs e genes da via de reparo por recombinação homóloga em cepas de *X. oryzae*. O gene de reparo está indicado em amarelo nas cepas Xoc e verde, nas cepas Xoo. Em azul, elementos de transposição. Elementos parciais estão indicados por barras brancas no início ou fim do elemento A distância entre o gene de reparo e um elemento IS está indicada em parênteses.

Dos eventos de proximidade anteriormente detectados na **via reparo por recombinação homóloga** (HR, Figura 3.6), os genes *ssb* e *polA* estão adjacentes a elementos IS, porém distantes em mais de 700 pares de base.

O evento de proximidade detectado para o gene *priA* se refere a um IS (ISXo4) a jusante do gene (403bp), parcial e na mesma orientação, exceto para a cepa *Xoo* AXO1947, na qual o mesmo tipo de elemento, também a jusante e parcial, está mais próximo 243pb, em direção oposta ao gene.

Também detectado na análise de anotação, o gene *recJ* também apresenta elementos adjacentes a jusante nas cepas de *Xoo* (exceto AXO1947). Porém, somente nas linhagens Filipinas (PXO83, PXO86 e PXO99A) o elemento poderia representar um evento de interferência. Ainda que em direção oposta ao gene, consideramos 465pb de distância um evento com menor probabilidade de modulação do gene pelo elemento.

O operon *recCBD* apresenta ISs no contexto genômico de algumas cepas de *Xoc* (linhagens africanas) e *Xoo* (AXO1947, PXO86 e PXO83). Os elementos sempre estão localizados distantemente do operon e igualmente aos outros eventos de proximidade desta via, com baixa probabilidade de interferência. Três genes (*hoIC*, *recG* e *recO*) anteriormente identificados como eventos de proximidade representam resultados falso-positivos. A análise detalhada não identificou nenhum elemento em 2500pb a jusante ou a montante destes genes, seja em *Xoo* ou *Xoc* (indicados com * na Figura 3.6).

Um último gene analisado da via de reparo por recombinação homóloga, *radA*, não está elencado como um gene principal nesta via. No entanto, *radA* apresenta uma diversidade de elementos em sua vizinhança genômica (Figura 3.7).

O elemento ISXo17 (pertencente ao grupo IS3ssgrIS407) está presente a montante de todas as cepas de *X. oryzae*, com distância e orientação variáveis. Em PXO99A, MAFF, KACC, PXO83 e PXO86 (linhagens asiáticas), ISXo17 está somente a 60pb, a montante e na mesma orientação de *radA*. A linhagem africana *Xoo* AXO1947 apresenta ISXo17 a montante, porém pouco mais distante do que nas linhagens asiáticas, a 284pb e em orientação oposta ao gene.

Nas linhagens africanas de *Xoc*, ISXo17 está orientado em direção ao gene, à aproximadamente 230pb a montante; nas linhagens asiáticas a orientação é oposta a *radA* e varia entre 145 (BLS256, CFBP2286, RS105, L8, B8-12) a 237pb (BXORI) de distância.

A jusante de *radA*, o tipo de IS, orientação e distância também variam entre as cepas. As *Xoo* asiáticas apresentam elementos a 300pb de distância, enquanto *Xoo* AXO1947 não apresenta nenhum IS a jusante de *radA*. Em *Xoc*, ambos elementos ISXo15 e ISXo5 estão orientados em direção a *radA* a 134pb de distância, exceto nas cepas CFBP7331 e CFBP7341, nas quais o IS mais próximo está a mais de 700bp de distância, em direção oposta ao gene.

Figura 3.7 - Proximidade de elementos IS ao gene *radA* em *X. oryzae*. O gene de reparo está indicado em amarelo nas cepas Xoc e verde, nas cepas Xoo. Em azul, elementos de transposição. Elementos parciais estão indicados por barras brancas no início ou fim do elemento. A distância entre o gene de reparo e um elemento IS está indicada em parênteses.

Além dos eventos de proximidade detectados nas principais vias de reparo descritas acima, a topoisomerase *gyrA* interessantemente apresenta elementos IS à sua proximidade (Apêndice G). Nas cepas africanas de Xoc e BXORI, ISXo17 está 154pb a jusante, em direção oposta ao gene; nas cepas asiáticas de Xoo o mesmo elemento está a 71pb a jusante do gene, porém nestas cepas ISXo17 está quebrado e na mesma direção que *gyrA*.

Além dos eventos citados acima, a cepa AXO1947 ainda apresenta elementos ISs inseridos próximos aos genes *phrB* (Reparo Direto) e *fis* (reparação ilegítima SHIDR) (Apêndice G). O elemento ISXo3 está imediatamente a jusante de *phrB* e o elemento IS481 está apenas 48pb a jusante de *fis*, ambos em orientação oposta aos genes.

3.4 Discussão

O repertório de genes associados ao sistema de reparo no gênero *Xanthomonas* é, de uma maneira geral, conservado. A espécie que apresenta maior variação em seu repertório de genes de reparo é *X. albilineans*, espécie fitopatogênica, restrita ao sistema

vascular de cana-de-açúcar, e que apresenta um genoma reduzido dentro do grupo *Xanthomonas* (3.77Gb na cepas GPE-PC73, Tabela 3.1).

A perda de genes de reparo é uma característica comum a bactérias com genomas reduzidos, processo geralmente associado a um alto grau de especialização de nicho e restrição de tecido de colonização do organismo (Moran, Wernegreen, 2000). A variação do repertório em *X. albilineans* é representada majoritariamente pela perda de glicosilases de DNA (*mpg*, *nei*, *ada-alkA* e *alkA*) e alguns genes acessórios (*imuB*, *cho*, *dnaE2*). *X. albilineans* preserva genes de mecanismos de recombinação, e de todas as outras principais vias de reparo.

Do conjunto de 77 genes do sistema de reparo de DNA presentes no gênero *Xanthomonas*, encontramos as seguintes evidências que particularizam *X. oryzae* em relação as demais espécies do gênero: (i) ausência de genes *ku* e *ligD* funcionais e (ii) o enriquecimento do número de eventos de proximidade dos genes de reparo a um elemento IS.

Diferentemente de outras espécies de *Xanthomonas*, identificamos a ausência do gene *ku*, e o gene *ligD* encontra-se modificado a ponto de ser considerado um pseudogene em *X. oryzae*. Ambos *ku* e *ligD* participam de um dos principais mecanismos de reparo de quebra da fita dupla por junção de pontas não-homólogas (NHEJ, *non-homologous end joining*), que ao contrário da recombinação homóloga, independe de um fragmento homólogo como molde para o reparo. Embora muitos grupos bacterianos não apresentem o sistema NHEJ naturalmente (Shuman, Glickman, 2007), é intrigante a observação que somente *X. oryzae* não apresente essa alternativa de reparo dentro do gênero.

Sob a perspectiva biotecnológica, o sistema NHEJ é um dos mecanismos de resolução da quebra de fita gerada na edição genômica por CRISPR-Cas (Selle, Barrangou, 2015). A falta destes genes em *X. oryzae* implica que a aplicação desta tecnologia neste sistema dependa de recombinação homóloga como resolução (ou da expressão heteróloga de *ku* e *ligD*).

A espécie *X. oryzae* apresenta amplificação do número de elementos IS em comparação às linhagens filogeneticamente próximas. Todas as 16 cepas desta espécie apresentam no mínimo 2 vezes mais ISs no genoma, em relação às outras 38 linhagens de *Xanthomonas*, ou ainda, às 15 linhagens de Xanthomonadaceae investigadas.

Ao serem mobilizados, elementos IS podem ser inseridos dentro de sequências codificantes, o que usualmente interrompe e inativa genes. Este tipo de impacto é rapidamente submetido a forças de seleção, em função da essencialidade do gene ou do impacto em *fitness* que este gene apresenta no estilo de vida do organismo. Além da

interrupção de sequências codificantes, a presença de ISs na vizinhança gênica pode modular a expressão gênica de genes adjacentes a inserção, aumentando ou diminuindo a expressão gênica (Siguier et al., 2015). Essa regulação de genes adjacentes pode se dar através da adição de elementos regulatórios do elemento inserido (Darmon, Leach, 2014; Fernández et al., 2007; Turton et al., 2006; Zhang, Saier, 2009); pela transcrição continuada da transposase codificada no elemento para além do códon de terminação da enzima transposase (Luque et al., 2006), ou ainda, pela mudança na topologia do DNA na região regulatória que impactar os genes a jusantes (Singh et al., 1995).

Avaliamos detalhadamente os eventos de proximidade de cada gene de reparo nas 16 cepas de *X. oryzae* com o intuito de acumular evidências sobre o impacto da amplificação de ISs no sistema de reparo ou ainda, se o sistema de reparo em *X. oryzae* possa estar associado à amplificação dos elementos IS.

A proximidade entre ISs e os genes *radA*, *mutM* e *mutS* é comum a todas as cepas de *X. oryzae*. MutM (ou *fpg*) é uma proteína que atua no sistema de reparo por excisão de bases (BER). O gene *mutM* é uma glicosilase de DNA envolvida no reparo de lesões de oxidação e de purinas e pirimidinas, incluindo lesões do tipo 7,8 dihidro-8-oxo-dGuanina (oxoG). Este tipo de lesão, causada pela exposição do DNA à espécies reativas de oxigênio, são abundantes no ambiente celular, altamente mutagênicas, cujo reparo é crítico para a estabilidade genômica. MutM remove especificamente o pareamento de 8oxo-G com a base C (8oxo-G:C)(Li, 2010).

Mesmo com grande diversidade de ISs ao redor de *mutM*, *X. oryzae* parece preservar a região promotora do gene. A presença conservada de ISXo4 parcial a montante da região promotora de *mutM* em *Xoc* pode se referir a um evento ancestral de modulação. Assim, como o elemento ISXo15 parcial a montante das cepas de *Xoo*. Adicionalmente, nas cepas africanas de *Xoc*, IS1112a a jusante e em direção oposta a *mutM*, a somente 67pb de distância.

Interessantemente, Onda et al. (1999) descreve o envolvimento potencial de *mutM* na supressão de recombinação ilegítima com pequena homologia (SHDIR) induzida por stress oxidativo, no qual mutantes de *mutM* aumentaram de 5 a 12 vezes a taxa de recombinação ilegítima do fago lambda quando exposto a peróxido de hidrogênio, indutor de dano oxidativo. Neste contexto, *mutM* torna-se um alvo interessante para posterior investigação, uma vez que a produção de espécies reativas de oxigênio (ROS) é uma das primeiras respostas celulares da planta a invasão de patógenos (O'Brien et al., 2012; Torres et al., 2006).

O contexto de *mutS* nas cepas de *X. oryzae* conta com a presença conservada do elemento IS407, subgrupo da família IS3, ora completo, ora parcial a jusante e a montante de *mutS*. Membros da família IS3 até hoje investigados quanto ao mecanismo de transposição se mobilizam através de um intermediário circular de DNA e sua transposição aumenta o número de cópias do elemento no genoma (transposição replicativa). Se IS407 apresenta o mesmo mecanismo de transposição que os membros de sua família permanece em aberto.

O gene *mutS*, juntamente com *mutL* e *mutH*, é um gene fundamental no reparo de mal emparelhamento de bases (MMR) e sua proteína atua no primeiro passo desta via, no reconhecimento de bases mal emparelhadas. MutS interage fisicamente a MutL, e a formação deste complexo recruta e ativa MutH, endonuclease responsável pela clivagem da fita não-metilada e que, portanto, discrimina a fita molde e a nova fita sintetizada com necessidade de reparo). *Xanthomonas oryzae*, assim como todas as outras *Xanthomonas*, apresentam todos os genes da via MMR, exceto *mutH*.

A inativação de genes da via de mal emparelhamento de bases (MMR) resulta em um fenótipo *mutator*. A deleção do gene *mutS* aumenta significativamente a taxa de mutação e diminui a estringência de homologia em recombinações genéticas (Li, 2008; Oliver et al., 2002). Dessa forma, linhagens deficientes em MMR, incluindo mutantes de *mutS*, elevam suas taxas de recombinação homeóloga (recombinação entre sequências não-idênticas de DNA). Este tipo de mutação pode representar um mecanismo adaptativo no contexto de associações entre patógenos e hospedeiros ou de ocupação de um novo nicho (Denamur, Matic, 2006; Giraud et al., 2001; Oliver et al., 2002).

Elementos IS3ssgrIS407 constituem uma das famílias mais frequentes em *X. oryzae* (Tabela 3.2), com um alta frequência de elementos completos. Podemos especular a efemeridade de *mutS* principalmente nas cepas KACC, MAFF, PXO86 e PXO83 de *Xoo*, onde o elemento (ou parte dele) está localizado ambos a montante e jusante do gene, compondo um transposon composto que inclui *mutS* como gene passageiro. De acordo com este raciocínio, a transitoriedade de *mutS* nestas cepas poderia levar a um aumento da taxa de recombinação, consequentemente a uma maior dinâmica de mobilização de elementos como um todo, com potencial aumento do número de ISs por este mecanismo. Neste cenário, as cepas *Xoo* PXO99A, *Xoo* AXO1947 e *Xoc*, que apresentam IS407 somente a montante de *mutS*, perderam essa capacidade.

O gene *radA* também apresenta um contexto genômico que denota mobilidade, com a presença do elemento ISXo17 (também pertencente ao grupo IS3ssgrIS407) a montante

de todas as cepas de *X. oryzae*. As proteínas RadA e RecG possuem um efeito sinérgico no reparo de quebra de fita por recombinação. Em *E. coli* mutantes para *recG*, a ausência de RadA elimina a atividade residual de recombinação. Embora sua função seja redundante a *recG* e sua participação seja complementar no reparo recombinacional, *radA* é um gene universal no domínio Bacteria. Nas linhagens de asiáticas de *Xoo*, ISXo17 está a 60pb e na mesma direção que o gene *radA* e portanto, com alta probabilidade de interferir na expressão do gene. Permanece em aberto se este posicionamento de ISXo17 impede expressão apropriada de *radA* ou se ISXo17 substitui e modula a região regulatória do gene.

Por fim, a proximidade entre um elemento IS e *dnaQ* está distribuído irregularmente na espécie *X. oryzae*. As três linhagens africanas de *Xoc*, linhagem africana de *Xoo* e *Xoo* KACC, todas apresentam um IS adjacente a jusante de *dnaQ*, porém se tratam de eventos distintos de inserção (Figura 3.5). O gene *dnaQ* codifica a subunidade epsilon da polimerase III de DNA, compondo a via de reparo de mal emparelhamento de bases (MMR). Em *E. coli*, *dnaQ* confere a atividade de revisão (*proofreading*) da holoenzima e excisa bases mispareadas durante a replicação do DNA. Linhagens mutantes de *dnaQ* comprometem a indução da resposta SOS a danos, implicando *dnaQ* em algum aspecto da estabilidade da forquilha de replicação (Pohlhaus et al., 2008).

A investigação realizada neste Capítulo também detecta muitos exemplos de eventos de proximidade entre ISs e genes de reparo que não são compartilhados entre todas as cepas de *X. oryzae*. Observamos diversas instâncias nas cepas de *Xoo* que mostram que elementos são capazes de introduzir variações em um processo celular dentro de uma população ou entre populações distintas.

Nas cepas asiáticas de *Xoo* (*i.e.* exceto AXO1947), ISs estão proximamente localizados aos genes *rpoZ* (147pb a montante), *gyrA* (71pb a jusante) e *mpg* (92pb a montante). O gene *rpoZ* codifica a subunidade ω do complexo enzimático da polimerase de RNA (RNAP) que exerce um papel central no processo de transcrição e por este motivo, caracteriza uma molécula alvo na regulação do metabolismo primário. Especificamente, o gene *rpoZ* auxilia na manutenção da conformação da subunidade β (β') e recrutamento das outras subunidades para montagem do complexo enzimático funcional. Além do papel estrutural que exerce no complexo RNAP - conservado em todos os domínios da vida - a subunidade ω parece participar na reprogramação rápida da transcrição mediante mudanças na disponibilidade de nutrientes (*stringent response*) e dessa forma, na adaptação do RNAP e ajuste metabólico do organismo em durante deprivação nutricional (Mathew, Chatterji, 2006).

O gene *gyrA* codifica uma das duas subunidades da enzima DNA girase. A enzima DNA girase é uma topoisomerase do tipo II responsável pelo alívio da tensão estrutural da molécula de DNA através de quebras duplas e subsequente reparo. Além da sua função na replicação de DNA, a DNA girase é capaz de influenciar a expressão gênica, uma vez que o superenrolamento de moléculas de DNA está ligado ao processo de transcrição. Interessantemente, o superenrolamento (modulado pela DNA girase) gerado localmente durante o processo de transcrição é responsável por *explosões transcricionais*. Essa modulação rápida da expressão gênica resulta na diversidade celular de uma população isogênica e aumenta a sobrevivência da população face a constante flutuação de condições ambientais (Chong et al., 2014).

O gene *mpg* (metil-purina DNA glicosilase) é mais uma enzima da família glicosilase de DNA que excisa bases purínicas metiladas pela exposição do DNA a agentes alquilantes, com função complementar aos genes *ada*, *alkA*, *ogt* e *tag*.

Ainda dentro das linhagens asiáticas de *Xoo*, porém exclusivamente em *Xoo* PXO83 e *Xoo* PXO86, o gene *tag* apresenta um IS imediatamente a jusante (1pb de distância), em orientação oposta ao gene (Figura 3.3). O gene *tag* codifica uma DNA glicosilase (DNA-3-metiladenina glicosilase I), classe de enzimas que iniciam o processo de reparo de DNA por excisão de base (BER), catalisando a retirada da base alterada. O gene *tag*, em particular, excisa seletivamente bases alquiladas 3-meA e 3-meG, função também realizada por *alkA* em *E. coli* (Wyatt et al., 1999).

A cepa africana *Xoo* AXO1947, proveniente de Camarões, é distinta das outras cepas *Xoo* asiáticas em muitos aspectos abordados neste Capítulo. *Xoo* AXO1947 apresenta uma quantidade menor de elementos IS em seu genoma, o padrão de distribuição das famílias elementos difere das outras cepas de *Xoo* e ao mesmo tempo que não compartilha diversos eventos de proximidade com as cepas asiáticas, apresenta eventos particulares a sua linhagem, como os detectados em *phrB* (fotoliase) e *fis* (regulador de transcrição).

PhrB é uma enzima fotoliase que atua no reparo de dímeros de pirimidina (lesão de DNA) causado por irradiação ultravioleta em *E. coli*. Interessantemente, o ortólogo de *phrB* em *Neisseria gonorrhoeae* - organismo que normalmente não sofre com danos causados por UV devido ao seu estilo de vida (patógeno humano obrigatório) - não atua em fotoativação, mas no superenrolamento da molécula de DNA e resistência a danos oxidativos (Cahoon et al., 2011).

A proteína FIS de ligação a DNA, é capaz de interferir na expressão gênica de diferentes maneiras, dependendo da posição de seu sítio de ligação em relação ao promotor

do gene alvo. Também contribui para a estrutura nucleóide na célula bacteriana. A abundância na fase exponencial de crescimento bacteriano é uma das evidências da magnitude de sua influência celular (Ball, Johnson, 1991; Finkel, Johnson, 1992). Em *S. typhimurium*, FIS controla expressão de ambos genes *housekeeping* e fatores de virulência, localizados nas ilhas de patogenicidade. Em *E. coli* deficiente em *fis* mostra taxas reduzidas de recombinação (Dorman et al., 2016).

A dissonância observada na cepa africana representa mais uma evidência de uma história evolutiva de AXO1947 distinta das cepas asiáticas, que pode ser atribuída a recente pressão de seleção para adaptação de *X. oryzae* ao cultivares de arroz nesta região (Gonzalez et al., 2007; Soto-Suárez et al., 2010).

Comparada a Ásia, agricultura moderna de arroz no Oeste africano é relativamente recente, intensificada a partir de 1960, com os primeiros registros de escaldadura da folha causada por Xoo em 1980. O estriamento da folha causado por Xoc é atualmente considerada uma doença emergente. Assim, a reconfiguração da estrutura populacional do patógeno e dinâmica genômica através da agricultura moderna de arroz é um evento relativamente recente no continente africano (Gonzalez et al., 2007; McDonald, Linde, 2002; Wonni, 2016).

Adicionalmente, a produção de arroz na África engloba o cultivo *Oryza glaberrima*, a única espécie além de *O. sativa* a ser domesticada e cultivada. Novas variedades foram desenvolvidas com o intuito de associar a resistência de *O. glaberrima* a estresses regionais e diversas doenças prevalentes no continente africano com a alta produtividade das linhagens de *O. sativa* asiáticas. Dessa forma, a estrutura genômico-populacional de *X. oryzae* também deve responder ao fluxo gênico encontrado no hospedeiro (Wonni, 2016).

A amplificação de elementos IS aumenta a complexidade de variação genética em um processo que governa a rotina celular, como o sistema de reparo de DNA. Observamos que, a partir de um mesmo repertório de genes, cepas distintas de *X. oryzae* podem responder diferentemente a uma perturbação (dano ou stress) como consequência da diversidade de interferência que elementos IS podem impor neste sistema, em alguns casos, com alto potencial de modulação gênica. Sob uma perspectiva evolutiva, há exemplos no qual o processo de diversificação e adaptação é melhor explicado através da regulação de genes compartilhados do que pela presença do conjunto gênico do adquirido (Philippe et al., 2007; Vital et al., 2015).

O ‘efeito esparadrapo’ proposto em nossa hipótese alternativa recrutaria genes de reparo para garantir a integridade genômica com o consequente aumentando o número de ISs nas linhagens de *X. oryzae*. Nossos resultados não permitem atribuir um efeito esparadrapo no genoma de *X. oryzae*, no qual o processo transpositivo complementa a

ineficiência do sistema de reparo encontrado neste organismo. No entanto, evidenciamos a diversidade de inserção de elementos IS em *X. oryzae* no sistema de reparo. Estes eventos emergem como componentes importantes de diversificação e o trabalho realizado neste capítulo propõe um conjunto de observações a serem investigadas experimentalmente.

REFERÊNCIAS*

- Ball CA, Johnson RC. Efficient excision of phage lambda from the *Escherichia coli* chromosome requires the Fis protein. *J Bacteriol.* 1991;173(13):4027–31.
- Bloom LB. Loading Clamps for DNA Replication and Repair. *DNA Repair.* 2009;8(5):570–8.
- Cahoon LA, Stohl EA, Seifert HS. The *Neisseria gonorrhoeae* photolyase orthologue phrB is required for proper DNA supercoiling but does not function in photo-reactivation. *Mol Microbiol.* 2011;79(3):729–42.
- Chain PSG, Carniel E, Larimer FW, Lamerdin J, Stoutland PO, Regala WM, et al., Insights into the evolution of *Yersinia pestis* through whole-genome comparison with *Yersinia pseudotuberculosis*. *Proc Natl Acad Sci USA.* 2004;101(38):13826–31.
- Chong S, Chen C, Ge H, Xie XS. Mechanism of Transcriptional Bursting in Bacteria. *Cell.* 2014;158(2):314–26.
- Darmon E, Leach DRF. Bacterial Genome Instability. *Microbiol Mol Biol Rev.* 2014;78(1):1–39.
- Denamur E, Matic I. Evolution of mutation rates in bacteria. *Mol Microbiol.* 2006;60(4):820–7.
- Díaz-Maldonado H, Gómez MJ, Moreno-Paz M, Martín-Úriz PS, Amils R, Parro V, et al., Transposase interaction with the β sliding clamp: effects on insertion sequence proliferation and transposition rate. *Sci Rep.* 2015;5:13329.
- Dorman CJ, Colgan A, Dorman MJ. Bacterial pathogen gene regulation: a DNA-structure-centred view of a protein-dominated domain. *Clin Sci.* 2016;130(14):1165–77.
- Eichenbaum Z, Livneh Z. UV Light Induces IS10 Transposition in *Escherichia coli*. *Genetics.* 1998;149(3):1173–81.
- Fernández A, Gil E, Cartelle M, Pérez A, Beceiro A, Mallo S, et al., Interspecies spread of CTX-M-32 extended-spectrum β -lactamase and the role of the insertion sequence IS1 in down-regulating blaCTX-M gene expression. *J Antimicrob Chemother.* 2007;59(5):841–7.
- Finkel SE, Johnson RC. The Fis protein: it's not just for DNA inversion anymore. *Mol Microbiol.* 1992;6(22):3257–65.
- Giraud A, Matic I, Tenaillon O, Clara A, Radman M, Fons M, et al. Costs and Benefits of High Mutation Rates: Adaptive Evolution of Bacteria in the Mouse Gut. *Science.* 2001;291(5513):2606–8.
- Gonzalez C, Szurek B, Manceau C, Mathieu T, Séré Y, Verdier V. Molecular and Pathotypic Characterization of New *Xanthomonas oryzae* Strains from West Africa. *Mol Plant Microbe Interact.* 2007;20(5):534–46.
- Huguet-Tapia JC, Peng Z, Yang B, Yin Z, Liu S, White FF. Complete Genome Sequence of the African Strain AXO1947 of *Xanthomonas oryzae* pv. *oryzae*. *Genome Announc.* 2016;4(1).

*De acordo com:

International Committee of Medical Journal Editors. [Internet]. Uniform requirements for manuscripts submitted to Biomedical Journal: sample references. [updated 2011 Jul 15]. Available from: <http://www.icmje.org>

- Kuan CT, Liu SK, Tessman I. Excision and transposition of Tn5 as an SOS activity in *Escherichia coli*. *Genetics*. 1991;128(1):45–57.
- Lavatine L, He S, Caumont-Sarcos A, Guynet C, Marty B, Chandler M, et al., Single strand transposition at the host replication fork. *Nucleic Acids Res*. 2016;gkw661.
- Li G-M. Mechanisms and functions of DNA mismatch repair. *Cell Res*. 2008;18(1):85–98.
- Li G-M. Novel molecular insights into the mechanism of GO removal by MutM. *Cell Res*. 2010;20(2):116–8.
- Luque I, Andújar A, Jia L, Zabulon G, De Marsac NT, Flores E, et al., Regulated expression of glutamyl-tRNA synthetase is directed by a mobile genetic element in the cyanobacterium *Tolypothrix* sp. PCC 7601. *Mol Microbiol*. 2006;60(5):1276–88.
- Martins-Pinheiro M, Galhardo RS, Lage C, Lima-Bessa KM, Aires KA, Menck CF. Different patterns of evolution for duplicated DNA repair genes in bacteria of the Xanthomonadales group. *BMC Evol Biol*. 2004;4:29.
- Mathew R, Chatterji D. The evolving story of the omega subunit of bacterial RNA polymerase. *Trends Microbiol*. 2006;14(10):450–5.
- McDonald BA, Linde and C. Pathogen Population Genetics, Evolutionary Potential, and Durable Resistance. *Annu Rev Phytopathol*. 2002;40(1):349–79.
- Monteiro-Vitorello CB, De Oliveira MC, Zerillo MM, Varani AM, Civerolo E, Sluys M-AV. Xylella and Xanthomonas Mobil'omics. *OMICS J Integr Biol*. 2005;9(2):146–59.
- Moran NA, Wernegreen JJ. Lifestyle evolution in symbiotic bacteria: insights from genomics. *Trends Ecol Evol*. 2000;15(8):321–6.
- Nagy Z, Chandler M. Regulation of transposition in bacteria. *Res Microbiol*. 2004;155(5):387–98.
- O'Brien JA, Daudi A, Butt VS, Bolwell GP. Reactive oxygen species and their role in plant defence and cell wall metabolism. *Planta*. 2012;236(3):765–79.
- Oliver A, Baquero F, Blázquez J. The mismatch repair system (mutS, mutL and uvrD genes) in *Pseudomonas aeruginosa*: molecular characterization of naturally occurring mutants. *Mol Microbiol*. 2002;43(6):1641–50.
- Onda M, Hanada K, Kawachi H, Ikeda H. *Escherichia coli* MutM Suppresses Illegitimate Recombination Induced by Oxidative Stress. *Genetics*. 1999;151(2):439–46.
- Parkhill J, Sebahia M, Preston A, Murphy LD, Thomson N, Harris DE, et al., Comparative analysis of the genome sequences of *Bordetella pertussis*, *Bordetella parapertussis* and *Bordetella bronchiseptica*. *Nat Genet*. 2003;35(1):32–40.
- Pasternak C, Ton-Hoang B, Coste G, Bailone A, Chandler M, Sommer S. Irradiation-Induced *Deinococcus radiodurans* Genome Fragmentation Triggers Transposition of a Single Resident Insertion Sequence. *PLOS Genet*. 2010;6(1):e1000799.
- Philippe N, Crozat E, Lenski RE, Schneider D. Evolution of global regulatory networks during a long-term experiment with *Escherichia coli*. *BioEssays*. 2007;29(9):846–60.

Pieretti I, Royer M, Barbe V, Carrere S, Koebnik R, Cociancich S, et al., The complete genome sequence of *Xanthomonas albilineans* provides new insights into the reductive genome evolution of the xylem-limited Xanthomonadaceae. *BMC Genomics*. 2009;10:616.

Pohlhaus JR, Long DT, O'Reilly E, Kreuzer KN. The ϵ Subunit of DNA Polymerase III Is Involved in the Nalidixic Acid-Induced SOS Response in *Escherichia coli*. *J Bacteriol*. 2008;190(15):5239–47.

Ryan RP, Vorhölter F-J, Potnis N, Jones JB, Van Sluys M-A, Bogdanove AJ, et al., Pathogenomics of *Xanthomonas*: understanding bacterium–plant interactions. *Nat Rev Microbiol*. 2011;9(5):344–55.

Selle K, Barrangou R. Harnessing CRISPR–Cas systems for bacterial genome editing. *Trends Microbiol*. 2015;23(4):225–32.

Shi Q, Parks AR, Potter BD, Safir IJ, Luo Y, Forster BM, et al., DNA Damage Differentially Activates Regional Chromosomal Loci for Tn7 Transposition in *Escherichia coli*. *Genetics*. 2008;179(3):1237–50.

Siguier P, Filée J, Chandler M. Insertion sequences in prokaryotic genomes. *Curr Opin Microbiol*. 2006a;9(5):526–31.

Siguier P, Gournayre E, Varani A, Ton-Hoang B, Chandler M. Everyman's Guide to Bacterial Insertion Sequences. *Microbiol Spectr*. 2015;3(2).

Siguier P, Perochon J, Lestrade L, Mahillon J, Chandler M. ISfinder: the reference centre for bacterial insertion sequences. *Nucleic Acids Res*. 2006b;34:D32–6.

Singh J, Mukerji M, Mahadevan S. Transcriptional activation of the *Escherichia coli* *bgl* operon: negative regulation by DNA structural elements near the promoter. *Mol Microbiol*. 1995;17(6):1085–92.

Soto-Suárez M, González C, Piégu B, Tohme J, Verdier V. Genomic comparison between *Xanthomonas oryzae* pv. *oryzae* and *Xanthomonas oryzae* pv. *oryzicola*, using suppression-subtractive hybridization. *FEMS Microbiol Lett*. 2010;308(1):16–23.

Torres MA, Jones JDG, Dangl JL. Reactive Oxygen Species Signaling in Response to Pathogens. *Plant Physiol*. 2006;141(2):373–8.

Turton JF, Ward ME, Woodford N, Kaufmann ME, Pike R, Livermore DM, et al., The role of ISAb1 in expression of OXA carbapenemase genes in *Acinetobacter baumannii*. *FEMS Microbiol Lett*. 2006;258(1):72–7.

Varani A, Carvalho L, Zerillo MM, Monteiro-Vitorello CB. Elementos de Transposição: Classificação e Mecanismos de Mobilização. In: *Elementos de Transposição Diversidade, evolução aplicações e impacto nos genomas dos seres vivos*. FIOCRUZ; 2015. p. 11–43.

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

Vital M, Chai B, Østman B, Cole J, Konstantinidis KT, Tiedje JM. Gene expression analysis of *E. coli* strains provides insights into the role of gene regulation in diversification. *ISME J*. 2015;9(5):1130–40.

Wilkins KE, Booher NJ, Wang L, Bogdanove AJ. TAL effectors and activation of predicted host targets distinguish Asian from African strains of the rice pathogen *Xanthomonas oryzae* pv. *oryzicola* while strict conservation suggests universal importance of five TAL effectors. *Plant Biot Interact*. 2015;536.

Wonni I. Evaluation of Elite Rice Varieties Unmasks New Sources of Bacterial Blight and Leaf Streak Resistance for Africa. *J Rice Res*. 2016 22;4(1):8.

Wyatt MD, Allan JM, Lau AY, Ellenberger TE, Samson LD. 3-methyladenine DNA glycosylases: structure, function, and biological importance. *BioEssays News Rev Mol Cell Dev Biol*. 1999;21(8):668–76.

Zhang Z, Saier MH. A Novel Mechanism of Transposon-Mediated Gene Activation. *PLOS Genet*. 2009;5(10):e1000689.

4 UM NOVO ELEMENTO DO TIPO MITE EM *XANTHOMONAS ORYZAE*

4.1 Introdução

Elementos de transposição são amplamente distribuídos nos domínios da vida e são agentes importantes na evolução de genomas. Em genomas bacterianos, a diversidade de elementos é representada principalmente pela categoria de elementos de transposição da classe II (ou transposons de DNA), que inclui elementos do tipo IS e derivados não-autônomos, como os MITEs.

Miniature Inverted-repeats Transposable Elements (MITEs) são classificados como não-autônomos porque apresentam repetições terminais invertidas (TIRs), como um IS, mas não codificam transposases ou qualquer polipeptídeo em sua região interna (Figura 1.3D, Capítulo 1). De modo geral, apresentam-se como sequências de nucleotídeos curtas que variam de 200-300 bp. São incapazes de se movimentar autonomamente e portanto, se dependem da atividade *in trans* de transposases presentes nas cópias completas de IS co-residentes no genoma. Transposases específicas reconhecem as repetições terminais invertidas dos MITEs.

Assim como os elementos ISs, MITEs também podem impactar de diversas maneiras o genoma onde são encontrados. Podem interromper genes ou podem servir como substrato para recombinação, provocando deleções ou rearranjos cromossômicos, devido a presença geralmente em múltiplas cópias em um mesmo genoma. Apesar de haver exemplos de MITEs interrompendo regiões codificantes, eles são frequentemente localizados em regiões intergênicas, e também podem influenciar de diversas maneiras a expressão gênica no genoma que reside se inseridos nas regiões 5' ou 3' de regiões codificantes (Darmon, Leach, 2014).

MITEs são diversos e distribuídos pelo domínio Bacteria e o número de trabalhos que mostram *in silico* e experimentalmente o impacto deste elementos em diversos grupos bacterianos aumenta gradativamente (Darmon, Leach, 2014; Delilhas, 2011). Talvez o elemento MITE bacteriano melhor investigado seja o elemento Correia, encontrado nos genomas de *Neisseria* spp. Seu impacto nestes organismos patógenos acontece principalmente no nível transcricional da célula, ao serem co-transcritos com genes adjacentes às suas inserções. A incorporação do transcrito do elemento Correia aos transcritos dos genes adjacentes torna os mensageiros de RNA (mRNA) alvos para RNase III. O resultado, seja estabilizando ou desestabilizando a molécula de mRNA, pode alterar expressão gênica. Ainda, elementos Correia codificam sequências promotoras que sugestivamente podem contribuir com a variação da expressão gênica dos genes adjacentes à inserção em indivíduos dentro de uma população clonal (*phase variation*) (Siddique et al., 2011).

A presença de MITEs não é restrita a genomas bacterianos. Elementos do tipo MITE também estão presentes abundantemente em genomas eucariotos, e impactam genomas animais (Fattash et al., 2015; Han et al., 2010; Hikosaka et al., 2011; Luchetti, 2015) e de plantas agronomicamente importantes (Lu et al., 2012; Nouroz et al., 2015; Oki et al., 2008; Vaschetto 2016; Yaakov et al., 2013). Em *Brassica* sp., pelo menos 30 mil elementos do tipo MITE, pertencentes a 15 diferentes famílias, estão dispersos pelos genomas de diferentes espécies. Estes elementos são ativos e o polimorfismo entre diferentes genótipos indica o envolvimento destes elementos na diversificação de *Brassica* (Nouroz et al., 2015). Em *O. sativa*, mais de 178 mil MITEs estão agrupados em 338 famílias distintas, e 25% do total de pequenos RNAs (sRNAs) produzidos em arroz são derivados destes elementos. No genoma de arroz, a presença de MITEs está associada a expressão reduzida dos genes onde estes elementos estão proximamente inseridos (Lu et al., 2012).

Neste capítulo, descrevemos um novo elemento do tipo MITE, aqui denominado MITE-Xo, encontrado exclusivamente nas linhagens de *Xanthomonas oryzae*. A localização das cópias foi efetuada inicialmente dentro de um genoma, *Xoc* BLS256, e então, estendemos a procura do elementos aos 15 outros genomas disponíveis de *X. oryzae*. Detectamos a presença de transcritos deste elemento em cepas de *X. oryzae* que possuem transcritomas publicamente disponíveis e, finalmente, avaliamos o potencial funcional deste MITE na interação de *X. oryzae* e seu hospedeiro *Oryza sativa*.

4.2 Materiais e Métodos

Os genomas de *X. oryzae* utilizados estão publicamente disponíveis no banco de dados NCBI (Tabela 4.1).

A representação gráfica do alinhamento múltiplo das sequências de repetições diretas (DR) foi criada pelo servidor Weblogo (<http://weblogo.berkeley.edu/>) (Crooks et al., 2004).

Para análise de sintenia, fizemos uma análise de similaridade (BLAST/NCBI) com cada cópia do MITE e sua vizinhança gênica. A partir das coordenadas de cada cópia de MITE-Xo encontrada, expandimos a região genômica até encontrar um gene de função definida. Evitamos ancorar o fragmento em proteínas hipotéticas, em genes que possuem mobilidade - como elementos de transposição - e pseudogenes. Toda a região genômica entre dois genes conhecidos, que inclui o MITE internamente, foi utilizada como ‘query’ em análise de similaridade (BLAST) contra um banco de dados composto pelos 16 genomas de *X. oryzae*.

O software SeaView (Gouy et al., 2010) foi utilizado para gerar alinhamentos, utilizando o algoritmo MAFFT (Kato et al., 2002). IQ-TREE (Nguyen et al., 2015) foi utilizado para construção de árvores filogenéticas. IQ-TREE realiza teste de modelo de

substituição e automaticamente aplica o modelo selecionado para construção da melhor árvore; testes de suporte *bootstrap*, com 1000 réplicas.

O software kallisto (Bray et al., 2016) foi utilizado para quantificação de *reads* de RNASeq sobre as cópias de MITE-Xo.

O gráfico de caixa (*boxplot*) com valores de TPM para as cópias de MITE-Xo, genes ribossomais e *housekeeping* foi criado pelo servidor web BoxplotR (<http://boxplot.tyerslab.com/>). Valores de TPM foram convertidos para escala logarítmica ($\log_{10}$). Cópias de MITEs que apresentaram valor de TPM iguais a 0 foram excluídos da análise. Hastes inferiores e superiores se estendem conforme definição de Tukey (1.5 x faixa interquartil).

Análise do *motif* de 18 nucleotídeos contra os genomas de *Oryza sativa* foi realizada no servidor Gramene (<http://www.gramene.org/>), pelo algoritmo BLASTN, limite de *e-value* 10, com filtro para regiões de baixa complexidade desativado, sensibilidade "normal" e banco de dados 'sequências genômicas' (genomic sequences).

Anotações gênicas do genoma do arroz foram adquiridas na última versão do genoma de *Oryza sativa* ssp. *japonica* cv. Nipponbare, encontrada no projeto anotação de genoma do arroz, pela Michigan State University (<http://rice.plantbiology.msu.edu/>).

A predição de estruturas secundária de RNA foram feitas pelo servidor web RNAfold disponível em (<http://rna.tbi.univie.ac.at/cgi-bin/RNAfold.cgi>) (Lorenz et al., 2011).

4.3 Resultados

4.3.1 Descrição do elemento MITE -Xo

Encontramos um novo elemento do tipo MITE nos genomas de *X. oryzae*, aqui denominado MITE-Xo, e analisamos a presença deste elemento em todas as 16 cepas de *X. oryzae* sequenciadas, disponíveis no banco de dados NCBI (www.ncbi.nlm.nih.gov) até final de agosto de 2016 - 6 cepas do patovar *oryzae* (Xoo) e 10 cepas dos patovar *oryzicola* (Xoc).

Os elementos MITE-Xo possuem uma organização típica de um MITE. Possuem entre 210-212 pares de base, repetições terminais invertidas (TIRs, *terminal inverted repeats*) apropriadamente orientadas nas extremidades do elemento (Figura 4.1A), e não codificam a transposase responsável por sua mobilidade. Os elementos MITE-Xo não codificam nenhuma proteína e não apresentam semelhança a nenhum domínio proteico em sua região interna.

Figura 4.1 - Descrição dos elementos MITE-Xo encontrados em *X. oryzae*. A) Sequência do MITE-Xo. Repetições terminais invertidas (TIR) estão indicadas por setas vermelhas. B) Sequências das TIRs encontradas nos MITEs-Xo1, MITE-Xo2, e ISs que potencialmente mobilizam estes elementos, ISXo2 e ISXo16. Diferenças nucleotídicas estão apontadas em vermelho. C) Logotipo que indica a conservação do sítio de inserção dos MITEs. Repetições diretas, DR.

Os 16 genomas de *X. oryzae* totalizam 435 cópias de MITEs-Xo (360 cópias de MITE-Xo1 e 75 cópias de MITE-Xo2). A comparação nucleotídica entre as diferentes cópias de MITE-Xo permitiu observar duas variantes deste elemento, aqui classificados como **MITE-Xo1** e **MITE-Xo2**, que formam dois clados distintos, separados com alto suporte filogeneticamente (Figura 4.2). As duas variantes possuem alta identidade nas repetições terminais invertidas mas divergem em sua região interna. Cópias de uma mesma variante no genoma de uma mesma cepa podem apresentar polimorfismos (Apêndice H).

A frequência de MITEs-Xo é semelhante na maioria das cepas de *X. oryzae* analisadas. Nas cepas de *Xoc*, o número total de cópias de MITE-Xo varia entre 27 e 33 cópias por cepa, conservadas em sequência e tamanho, e a variante MITE-Xo1 representa a maioria das cópias do elemento (no total, 239 MITEs-Xo1 e 57 MITEs-Xo2 em *Xoc*). Nas cepas de *Xoo*, o número total de cópias varia de 23 a 30 cópias por cepa, e a variante MITE-Xo1 também representa a maioria de cópias nestas linhagens (no total, 121 MITEs-Xo1 e 18 MITEs-Xo2) (Tabela 4.1, Lista completa de coordenadas de MITEs-Xo de todas as cepas de *X. oryzae* em Apêndice I).

Tabela 4.1 - Identificação das cepas de *X. oryzae* utilizadas e o número total de cópias de MITEs-Xo completas e parciais

Cepa	NCBI ID	Origem	Ano de coleta	Total de MITEs-Xo (completo/parcial)	MITE-Xo1*	MITE-Xo2*	ISXo2 completo	ISXo16 completo
<i>Xanthomonas oryzae</i> pv. <i>oryzicola</i>								
BLS256	CP003057.2	Filipinas	1984	30 (27/3)	21 (3)	6 (0)	4	5
CFBP2286	CP011962.1	Malásia	1964	31 (28/3)	22 (3)	6 (0)	4	5
CFBP2286 (plasmid)	CP011963.1	Malásia	1964	-	-	-	-	-
BLS279	CP011956.1	Filipinas	1984	28 (25/3)	19 (3)	6 (0)	5	4
B812	CP011955.1	China	2007	28 (25/3)	19 (3)	6 (0)	5	4
L8	CP011960.1	China	< 1995	29 (26/3)	20 (3)	6 (0)	5	4
RS105	CP011961.1	China	1992	28 (25/3)	19 (3)	6 (0)	5	4
BXOR1	CP011957.1	Índia	1996	34 (28/6)	23 (4)	5 (1)	5	4
CFBP7331	CP011958.1	Mali	2003	31 (26/5)	21 (4)	5 (1)	3	5
CFBP7341	CP011959.1	Burkina Faso	2009	31 (26/5)	21 (4)	5 (1)	3	4
CFBP7342	CP007221.1	Burkina Faso	2009	27 (23/4)	20 (4)	3 (0)	5	2
<i>Xanthomonas oryzae</i> pv. <i>oryzae</i>								
KACC10331	AE013598.1	Korea	-	26 (22/4)	19 (4)	3 (0)	6	1
MAFF311018	AP008229.1	Japão	-	26 (22/4)	19 (4)	3 (0)	6	2
PXO99A	CP000967.2	Filipinas	1980	30 (26/4)	24 (3)	2 (1)	13	1
PXO83	CP012947.1	Filipinas	1976	23 (19/4)	16(4)	3 (0)	10	1
PXO86	CP007166.1	Filipinas	1990	25 (21/4)	18 (4)	3 (0)	11	1
AXO1947	CP013666.1	Camarões	2015	4 (4/0)	1 (0)	3 (0)	1	1

*completo (parcial)

A cepa *Xoo* AXO1947, proveniente de Camarões (Oeste-Africano), é a única cepa que desvia da tendência descrita acima. Apresenta somente nove cópias deste elemento, sendo somente 1 cópia completa de MITE-Xo1 e 3 cópias completas de MITE-Xo2 (Tabela 4.1). Não foi encontrado MITEs-Xo nas demais espécies de *Xanthomonas* sequenciadas até o final de agosto de 2016.

A origem de elementos MITE pode ser atribuída à derivação de um elemento IS completo, cuja transposase foi perdida através de eventos de deleção na região interna do elemento (Jiang et al., 2004). As regiões internas de MITE-Xo1 e MITE-Xo2 não apresentam semelhança com quaisquer outro elemento IS, impossibilitando a identificação de um possível elemento do qual MITE-Xo tenha sido derivado.

Considerados não-autônomos, a mobilização *in trans* dos MITEs está condicionada à atividade de elementos cujas transposases possuem o mesmo sítio de reconhecimento. As repetições invertidas (TIRs) de MITE-Xo são constituídas por 24pb (MITE-Xo2) e 27pb

(MITE-Xo1) (Figura 4.1B). As duas variantes possuem alta identidade nucleotídica entre si (IRL, esquerda e IRL, direita em Figura 4.1B) e são semelhantes às TIRs encontradas nos elementos do grupo IS1595ssgrIS1595, especificamente de ISXo2 e ISXo16. As diferentes cepas de *X. oryzae* apresentam de duas a quatorze cópias inteiras de ISXo2 e ISXo16 (Tabela 4.1) e indicam o potencial de mobilização dos MITEs nestes genomas pela atividade destes elementos.

Figura 4.2 - Árvore filogenética de todas as cópias de MITEs-Xo encontradas nas 16 cepas de *X. oryzae*. Sombra cinza representa o clado de MITE-Xo1 e sombra azul representa o clado de MITE-Xo2. MITE-Xo1 e MITE-Xo2 formam clados distintos com robustes (valor de *bootstrap* igual a 100). Linha em vermelho representam cópias de MITEs-Xo encontradas em *Xoo*; em preto, encontradas em *Xoc*.

Assim como elementos IS, o evento de transposição de um MITE pode deixar ‘cicatrices’, duplicando alguns pares de base no sítio de inserção, conhecidas como repetições diretas (DRs, *direct repeats*). Reconstituímos e analisamos os sítios de inserção de 28 MITEs-Xo inteiros (exclusão das cópias quebradas) encontrados na cepa *Xoc* CFBP2286. Das 28 cópias completas do elemento, 16 apresentam DR composta por 8pb, 3 apresentam DR com 9pb, 1 apresenta DR com 10pb (8 cópias não apresentam DR). As

repetições diretas resultantes do processo de inserção dos MITEs-Xo não são conservadas entre as cópias, no entanto, podemos observar a tendência de MITE-Xo de ser inserido em dinucleotídeos A/T (Figura 4.1C).

4.3.2 *Estudo de sintenia entre as cópias de MITE-Xo em 16 genomas de X. oryzae.*

Elementos MITEs-Xo são exclusivos de *X. oryzae* e estão presentes em todas as cepas de *Xoo* e *Xoc* analisadas. Realizamos uma análise de sintenia comparativa das cópias de MITEs-Xo nas diferentes cepas de *X. oryzae* com o objetivo de compreender a associação entre as cópias de MITEs nos 16 genomas de *X. oryzae*. Nos questionamos neste momento se a amplificação destes elementos nas diferentes cepas foi resultado de eventos ancestrais à *X. oryzae* ou se cada cepa amplificou estes elementos convergentemente.

Assim, foram utilizadas as 31 cópias de MITEs-Xo (MITE-Xo1 e MITE-Xo2) presentes em *Xoc* CFBP2286 como referência para análise de sintenia. A partir das coordenadas de cada cópia do elemento neste genoma, estendemos sua vizinhança genômica até o encontro de genes com função conhecida. Esta região ampliada, que inclui o MITE, foi então procurada nas demais cepas de *Xoo* e *Xoc*. Quando um MITE-Xo foi localizado em um mesmo contexto genômico em mais de uma cepa (*i.e.*, apresentou o mesmo conjunto de genes a montante e a jusante do MITE-Xo, em diferentes cepas), assinalamos todas as cópias correspondentes deste mesmo contexto à um mesmo **grupo sintênico**.

O resultado da análise de sintenia dos elementos MITE-Xo está descrito a seguir separadamente para as variantes MITE-Xo1 e MITE-Xo2. Assinalamos todas as cópias de MITE-Xo a um grupo sintênico (Figura 4.3 e 4.6). Cópias exclusivas a uma única cepa também foram assinaladas a um grupo sintênico a fim de facilitar a classificação de MITEs-Xo em eventuais novos sequenciamentos genômicos de *X. oryzae*.

4.3.2.1 *MITE-Xo1*

Foram identificados um total de 360 cópias da variante MITE-Xo1 nos dezesseis genomas de *X. oryzae* (239 cópias nas 10 cepas de *Xoc* e 121 cópias nas 6 cepas de *Xoo*). Em *Xoc*, dezesseis grupos sintênicos representam 160 cópias de MITEs-Xo1, conservadas em todas as dez cepas de *Xoc* analisadas (grupos sintênicos MITEXo1-1 ao MITEXo1-16, Figura 4.3). As 79 cópias restantes estão distribuídas entre as cepas, embora seja possível identificar, por exemplo, grupos sintênicos às linhagens asiáticas (MITEXo1-20 e -24). As três linhagens africanas CFBP7331, CFBP7341 e CFBP7342 compartilham dois grupos sintênicos (MITEXo1- 28 e -31). Interessantemente, BXORI e CFBP7342 apresentam MITEs-Xo1 compartilhados com linhagens asiáticas (MITEXo1- 20 e -24) e também com as

outras duas linhagens africanas (MITEXo1- 29 e -30). Não foi identificado nenhum contexto genômico de MITE-Xo1 compartilhado entre *Xoo* e *Xoc*.

Os elementos MITE-Xo1 estão localizados majoritariamente em regiões intergênicas de *Xoc*. (Apêndice J). Destacamos alguns contextos genômicos de MITE-Xo1 em *Xoc*:

- MITEXo1-3 localizada junto à região promotora de operon de GA₄, investigado no Capítulo 2 desta tese. Esse operon foi transferido horizontalmente a *Xoc*, sendo exclusivo a este patovar. As cópias de MITES-Xo1 presentes nas linhagens Oeste-africanas e BXORI estão parciais (Figura 4.4A).

- MITE-Xo15 interrompe a sequência codificante de uma proteína que contém domínio N-terminal de ligação à peptidoglicanos e YkuD (domínio catalítico L, D transpeptidase) na porção C-terminal, potencialmente envolvida na biologia de parede celular. Essa proteína não está interrompida em *Xoo* (Figura 4.4B).

- MITEXo1-24 trata-se de um cópia de MITE-Xo1 sempre presente a montante do gene de virulência *avrRxo1*. *AvrRxo1* é efetor de virulência do tipo III forma juntamente com sua proteína chaperona (*Arc1*), um sistema toxina-antitoxina. É responsável pela resposta de hipersensibilidade em plantas que carregam o gene de resistência *Rxo1* (Figura 4.4C).

Figura 4.4 - Representação do contexto genômico de três grupos sintênicos de MITE-Xo1 em *Xoc*. Grupos sintênicos A) miteXo1-3, B) miteXo1-15 e C) miteXo1-24. Figuras fazem parte do Apêndice J. Blocos em verde indicam genes com função conhecida; em vermelho, grupos sintênicos de MITES-Xo1; em bege, proteínas hipotéticas; em azul, elementos IS. Linha vertical pontilhada indica elementos/genes parciais e contorno pontilhado, genes pseudogenizados; parênteses indica a posição do fragmento em outro local do genoma.

Em *Xoo*, 17 grupos sintênicos (MITEXo1-49 ao -59) correspondem a 102 das 121 cópias encontradas em todas as 6 cepas analisadas. O restante é representado principalmente por expansões exclusivas a cada cepa (Figura 4.3).

Construímos uma árvore filogenética com as cópias inteiras de MITE-Xo1 de ambos *Xoo* e *Xoc* (total de 305 elementos inteiros) a fim de avaliar se os grupos sintênicos descritos também podem ser detectados filogeneticamente (Figura 4.5). Os nomes dos taxons correspondem às cópias identificadas em todas as cepas, de acordo com Apêndice I.

Primeiramente, identificamos a formação de dois principais clados, A e B, com alto valor de suporte *bootstrap*. Essa topografia separa robustamente os MITEs-Xo1 encontrados em *Xoc* das cópias encontradas em *Xoo*, sugerindo que amplificação desta variante pode ter sido ancestral a cada patovar, mas não à espécie *X. oryzae*.

O **Clado A**, que agrupa 207 cópias de MITE-Xo1 provenientes de *Xoc*, contém majoritariamente as cópias de MITE-Xo1 encontradas neste patovar. O ramo colapsado (em laranja, na Figura 4.5) inclui 185 cópias do elemento em *Xoc*. Neste ramo, as cópias não se agrupam de acordo com a sintenia e os grupos formados não apresentam suporte por *bootstrap*. Somente os grupos sintênicos MITEXo1-3 e MITEXo1-8 são apropriadamente agrupados, com valor de bootstrap acima de 70 (não mostrado, dentro do ramo colapsado).

O Clado A também inclui 14 cópias de *Xoo* (indicadas em vermelho na árvore da Figura 4.5). As cópias KACC_mite12, PX86_mite13, PXO83_mite12, MAFF_mite15, e duas cópias de PXO99A (PXO99A_mite16 e PXO99A_mite17) são agrupados com alto valor de suporte. De fato, estas cópias pertencem ao grupo sintênico *miteXo1-49*. As duas cópias de PXO99A que compõem este grupo sintênico representam uma região genômica duplicada (que inclui o MITE-Xo1) neste genoma.

Outro grupo menor, composto por KACC-mite10 e MAFF-mite13, representa o grupo sintênico *miteXo1-62*, formado somente por estas duas cepas. Seis cópias não se agrupam a nenhum clado, e em concordância com a análise de sintenia, representam cópias exclusivas a seus respectivos genomas (KACC_mite18 compõe MITEXo1-75; MAFF_mite6, MITEXo1-65; PXO99A_mite19, MITEXo1-71; e AXO1947_mite5, MITEXo1-78). As cópias PXO83_mite22 e PXO86_mite24 compõem o grupo sintênico MITEXo1-77, mas não estão agrupadas na árvore.

O **Clado B** é composto exclusivamente por cópias de MITE-Xo1 encontrados no patovar *Xoo*, exceto por 1 cópia de *Xoc* que se agrupa a este clado (BXORI-mite29). Diferentemente do observado no Clado A, a topografia do Clado B apresenta muitos agrupamentos com alto valor de suporte *bootstrap*. E esses agrupamentos são constituídos por cópias de um mesmo grupo sintênico (Figura 4.5). Ou seja, o grupo sintênico *miteXo1-50*, composto pelas cópias de MITEs-Xo1 em KACC, MAFF, PXO86, PXO83 e

Figura 4.5 - Árvore filogenética das cópias de MITE-Xo1 em *Xoc* e *Xoo*. Sombras verdes indicam Clado A e Clado B. Taxons em vermelho indicam cópias de MITEs-Xo1 encontradas em *Xoo*; em preto, *Xoc*. Ramo colapsado em laranja inclui 185 cópias de *Xoc* MITE-Xo1. Ramos cortados estão representados com 1/4 do tamanho.

PXO99A e apresentam o mesmo conjunto gênico a montante e a jusante do elemento nestas cepas e estão agrupado na árvore filogenética com valor de *bootstrap* igual a 100.

Alguns clados, como *miteXo1-43*, são compostos pelos MITEs-Xo1 que constituem os grupos sintênicos e por cópias adicionais do elemento. Estas cópias adicionais (indicadas por 'E' na árvore da Figura 4.5) estão localizadas em contextos genômicos exclusivos em sua cepa de origem. Hipotetizamos que estas cópias adicionais representam um evento de transposição replicativa a partir do elemento do grupo sintênico que compartilha. Por exemplo, dentro do grupo sintênico MITE-Xo1-54, há dois elementos provenientes da cepa PXO99A (PXO99A_mite21 e PXO99A_mite22). Enquanto PXO_mite21 compartilha o contexto genômico com as outras cepas, PXO99A_mite22 encontra-se numa região exclusiva (não compartilhada com nenhuma outra cepa). Concluímos que PXO99A_mite22 (E) derivou de PXO99A_mite21 por transposição replicativa. Identificamos este tipo de evento nos também nos clados dos grupos sintênicos *miteXo1-45*, *-47*, *-52* e *-54*.

Os grupos sintênicos *miteXo1-44* e *miteXo1-60* formam um só agrupamento na árvore, com mais 4 diferentes cópias adicionais proveniente de PXO99A e 1 de KACC10331. Três das quatro cópias de PXO99A fazem parte de uma região cromossômica triplicada neste genoma. Isto é, não só o elemento MITE-Xo1 está duplicado, mas toda a região que o contém (indicadas por E§ na árvore). A quarta cópia de PXO99A (PXO99A_mite1) está em um contexto genômico distinto. Curiosamente, a cepa KACC10331 não possui um elemento no contexto genômico descrito para *miteXo1-60*. No entanto, KACC10331 apresenta um elemento adicional (KACC_mite15) que se agrupa com *miteXo1-60* na árvore, com alto valor de suporte. Assim, pode-se especular que KACC_mite15 esteve anteriormente no contexto genômico de *miteXo1-60*, mas se transpôs de maneira conservativa ao novo contexto identificado aqui. É necessário melhor investigação para determinar se esse evento representa a mobilização de KACC_mite15.

Por fim, três grupos sintênicos não se agrupam na árvore filogenética construída - *miteXo1-53*, *miteXo1-56* e *miteXo1-61*. Os grupos sintênicos *miteXo1-57*, *miteXo1-58*, *miteXo1-59*, *miteXo1-63* e *miteXo1-64* não estão inclusos na árvore por serem elementos parciais/incompletos.

4.3.2.2 MITE-Xo2

A análise de sintenia também foi realizada com as 75 cópias de MITEs-Xo2 encontradas nas 16 cepas de *X. oryzae* (57 cópias em *Xoc* e 18 cópias em *Xoo*) (Figura 4.6). Nesta variante, as cópias do elemento foram classificadas em onze grupo sintênicos distintos.

O grupo sintênico *miteXo2-1* (Figura 4.7A) se refere a uma cópia do elementos MITE-Xo2 que apresenta o mesmo contexto genômico em ambos patovares *Xoo* e *Xoc*, e portanto, universal em *X. oryzae*.

Figura 4.6 - Representação da evolução da sintenia de MITE-Xo2 sob a hipótese de Acctran. Retângulos sólidos representam o surgimento da cópia de MITE, retângulos vazios representam eventos de perda no local genômico. Ao lado da árvore filogenética, a presença (círculos sólidos) e ausência (círculos pontilhados) dos MITEs em cada cepa de *X. oryzae* para os 11 grupos sintênicos identificados. Círculos cortados representam cópias parciais/incompletas.

Além de *miteXo2-1*, *miteXo2-2* é encontrado em ambos patovares *Xoo* e *Xoc*, mas não é conservado na espécie *X. oryzae* (Figura 4.6 e 4.7B). Este grupo sintênico está presente em 5 das 6 cepas em *Xoo* (exceto AXO1947) e em 7 das 10 cepas de *Xoc* (ausente nas cepas CFBP7342, CFBP7341 e CFBP7331). Interessantemente, as linhagens que não apresentam este MITE neste contexto genômico são todas de origem africana.

As cópias referentes aos grupos sintênicos *miteXo2-3*, *miteXo2-4*, *miteXo2-5* e *miteXo2-6* são restritas a *Xoc* (Figura 4C-F). Os grupos sintênicos *miteXo2-3* e *miteXo2-5* estão presentes nas 10 cepas de *Xoc*; *miteXo2-6* está somente nas *Xoc* asiáticas; e CFBP7342 é a única cepas de *Xoc* que não apresenta a cópia do grupo sintênico *miteXo2-4*.

Estes seis grupos sintênicos apresentados englobam 64 das 75 cópias da variante MITE-Xo2 em *X. oryzae*. As 11 cópias restantes foram divididas em cinco diferentes grupos: *miteXo2-7* é restrito às cepas de origem asiática de *Xoo*, (*i.e.*, exceto *Xoo* AXO1947); *miteXo2-8* estão restritas às cepas *Xoc* CFBP7341 e CFBP7331; as cópias de *miteXo2-9* estão presentes em *Xoc* CFBP7341, CFBP7331 e a cepa indiana *Xoc* BXOR1; e, finalmente, os grupos sintênicos *miteXo2-10* e *miteXo2-11* são restritos à cepa *Xoo* AXO1947, e uma única cópia de MITE-Xo2 formam estes grupos.

Figura 4.7 - Representação do contexto genômico de dos grupos sintênicos MITE-Xo2 em *Xoc*. Blocos em verde indicam genes com função conhecida; em bege, proteínas hipotéticas (HyPr); em azul, elementos IS; em vermelho, MITEs-Xo2. Linha pontilhada indicam elementos parciais; contorno pontilhado indica genes em processo de pseudogenização. Grupos sintênicos MITE-Xo2-1 e MITE-Xo2 são compartilhados entre *Xoc* e *Xoo*.

A variante MITE-Xo2 também é majoritariamente localizada em regiões intergênicas. De todas as cópias da variante MITE-Xo2, somente as cópias do grupo sintênico *miteXo2-6* interrompe um gene (Figura 4.7F). Esta cópia é encontrada somente nas linhagens da China, Filipinas e Malásia (BLS256, BLS279, CFBP2286, B8-12, L8 E RS105) e interrompe o gene *cheY*. Este gene está associado a sinalização da movimentação quimiotática (Paolo et al., 2016), e encontra-se completo nas linhagens Oeste-africanas.

Construímos uma árvore filogenética com as cópias de MITEs-Xo a fim de avaliar se os eventos de sintenia descritos são detectados filogeneticamente e inferir com maior suporte a relação ancestral das cópias nas diferentes cepas de *X. oryzae* (Figura 4.8). Podemos observar que o conjunto de cópias de um determinado grupo sintênico também se agrupam em clados na filogenia, com alto valor de suporte.

Os dois grupos sintênicos *miteXo2-1* e *miteXo2-2*, cujas cópias são compartilhadas entre *Xoo* e *Xoc*, se agrupam coerentemente na árvore. As cópias de *Xoo* se separam das cópias de *Xoc*, mas conjuntamente formam um clado com alto valor de suporte (valor de *bootstrap* 99).

Os grupos sintênicos *miteXo2-4* (exclusivo a *Xoc*) e *miteXo2-5* (exclusivo a *Xoc* asiáticas) também se agrupam com alto suporte. Os grupos sintênicos *miteXo2-4* e *miteXo2-5* foram agrupadas em um só clado. Esta topografia sugere que estas duas cópias podem ter derivado uma da outra, dentro de um genoma. Esse evento, no entanto, deve ter ocorrido em um ancestral hipotético de *Xoc*, porque todas as cepas analisadas contém as duas cópias dos dois grupo sintênicos (exceto a cepa africana *Xoc* CFBP7342).

O *miteXo2-7*, que forma um clado distinto de todos os outros clados, parece se tratar de uma cópia de MITE-Xo2 ancestral ao patovar *Xoo*. Expansões destes elementos particulares a um conjunto de cepas também aparecem em agrupamentos distintos na árvore: *miteXo2-9* nas cepas *Xoc* CFBP7331 e *Xoc* CFBP7341; e *miteXo2-10* e *miteXo2-11* em *Xoo* AXO1947.

Dada a coerência entre os dados de sintenia e a árvore filogenética, plotamos os eventos de sintenia de MITE-Xo2 na árvore de espécie de *X. oryzae* (Figura 4.6). Assumimos que a probabilidade de uma cópia de MITE ser herdada verticalmente é maior do que a probabilidade de um elemento ser inserido no exato mesmo local genômico das diferente cepas, por convergência. A evolução do estado de caráter (neste caso, a sintenia do elemento) foi otimizada sob a hipótese de Acctran (*accelerated transformation*).

4.3.3 Expressão de MITE-Xo em linhagens de *Xanthomonas oryzae*

Investigamos se os elementos MITE-Xo estão sendo expressos em *X. oryzae*. Anteriormente ao sequenciamento de segunda geração, a verificação da transcrição de elementos de transposição dependia de experimentos desenhados especificamente com esse objetivo. Essa dificuldade foi amenizada pela utilização de sequenciamento de transcrito total (*deep-sequencing*), que independe de uma plataforma fixa de genes para hibridização.

Figura 4.8 - Árvore filogenética de MITEs-Xo2 em *X. oryzae*. Bootstraps com valor menor que 70 foram excluídos dos ramos. OTUs são os MITEs-Xo2 designados pelo nome da cepa e respectiva coordenada genômica. Sombreamento rosa delimita os clados coerentes com os grupos sintênicos identificados. Foi escolhido o elemento geneticamente mais distante como de todas as cópias como grupo externo (L8[3103083..3103292]). A árvore mostra somente ramos com valores de *bootstrap* maiores que 70.

Existem atualmente duas publicações que executaram e analisaram transcritomas de *Xanthomonas oryzae*. Seo et al. (2008) utilizou microarranjos de oligonucleotídeos (*spotted microarray*) para identificar genes diferencialmente expressos em dois meios de cultura distintos – meio de cultura rico PSB (*peptone sucrose broth*, usado para cultivo do organismo em laboratório) e o meio mínimo XOM2, capaz de induzir a expressão do conjunto gênico *hrp* (*hypersensitive response and pathogenicity*).

O estudo de Zhang et al. (2013) utilizou transcritoma total por sequenciamento de segunda geração a fim de investigar os padrões de expressão de três cepas diferentes de *Xoo* (PXO99A, PXO86 e GD1358) que possuem distintos graus de virulência em diferentes variedades de arroz.

Utilizamos os resultados publicamente disponíveis destes dois estudos para avaliar o nível de transcrição do elemento MITE-Xo nos genomas de *X. oryzae*.

4.3.3.1 Análise de expressão de MITE-Xo por microarranjo de oligonucleotídeos

Para analisar a diferença de expressão de *Xoo* PXO99A e *Xoc* BLS256 em diferentes meios de cultura, Seo e colaboradores (2008) utilizaram análise de expressão por microarranjo de oligonucleotídeos. Foi desenvolvida uma lâmina de microarranjo que combina oligonucleotídeos (de 50 a 70 nucleotídeos) específicos para *Xoo*, oligonucleotídeos específicos para *Xoc* e oligonucleotídeos comuns aos dois patógenos, utilizando como referência os genomas sequenciados de BLS256 e *oryzae* KACC10331 disponíveis à época. Como controle, foram utilizados oligonucleotídeos específicos para higromicina fosfotransferase (gene *hph*), ausente dos genomas de *X. oryzae* (Seo et al., 2008).

O oligonucleotídeo (ou sonda) identificada como XOCORF3496 possui 60 nucleotídeos e apresenta alta identidade com o elemento MITE-Xo, particularmente com a variante MITE-Xo2 (Figura 4.9).

Em *Xoc* BLS56, somente uma cópia de MITE-Xo2 apresentam 100% de identidade com a sonda XOCORF3496 enquanto as outras cópias variam de 92 a 98% de identidade, suficiente para a ocorrência de hibridização cruzada com alto sinal de fluorescência.

Para sondas de hibridização compostas por 60 nucleotídeos, similaridade global de 70% permite a hibridização não específica (Chou et al., 2004). A partir de cinco diferenças nucleotídicas distribuídas aleatoriamente entre a sonda e o transcrito, a hibridização ocorre com sinal de fluorescência reduzida a menos de 50% e a partir de 18 diferenças, a hibridização é reduzida totalmente (Hughes et al., 2001). As cópias de MITE-Xo1 possuem identidade somente em parte da sonda, com identidades que variam entre 57 e 66% de identidade (considerando os 60 nucleotídeos da sonda).

		80	100	120	
	XOCOR3496c	ACGTGCTGGCTGAAATAACGTGCGACCGGCTCTCGCGTCAGCAACGCGAAAG - CGCACAT			
MITE-Xo2	BLS256_mite22	TGACGTCTGCTGGCTGAAATAACGTGCGACCGGCTCTCGCGTCAGCAACGCGAAAG - CGCACAT	CAT	60 (100%)	
	BLS256_mite11	TGACGTCTGCTGGCTGAAATAACGTGCGACCGGCTCTCGCGTCAGCAACGCGAAAG - CGCACAT	CAT	59 (98%)	
	BLS256_mite28	TGACGTCTGCTGGCTGAAATAACGTGCGACCGGCTCTCGCGTCAGCAACGCGAAAG - CGCACAT	CAT	59 (98%)	
	BLS256_mite14	TGACGTCTGCTGGCTGAAATAACGTGCGACCGGCTCTCGCGTCAGCAACGCGAAAG - CGCACAT	CAT	58 (97%)	
	BLS256_mite30	TGACGTCTGCTGGCTGAAATAACGTGCGACCGGCTCTCGCGTCAGCAACGCGAAAG - CGCACAT	CAT	58 (97%)	
	BLS256_mite26	TGACGTCTAGCTGAAATAACGTGCGACCGGCTCTCGCGTCAGCGACGCGAAAG - CGCACAT	CAT	56 (93%)	
MITE-Xo1	BLS256_mite2	TGGGGTCTTGGCTGAAAGAACTTCCACCGTCTCCTCGCGTCAGCAGCGCAAGAGTCAGACGTCAT	CAT	38 (63%)	
	BLS256_mite3	TGGGGTCTTGGCTGAAAGAACTTTCGACCGTCTCCTCGCGTCAGCAGCGCAAGAGTCAGACGTCAT	CAT	39 (65%)	
	BLS256_mite4	TGGGGTCTTGGCTGAAAGAACTTTCGACCGGCTCTCCTCGCGTCAGCAGCGCAAGAGTCAGACGTCAT	CAT	39 (65%)	
	BLS256_mite5	TGGGGTCTTGGCTGAAAGAACTTTCGACCGTCTCCTCGCGTCAGCAGCGCAAGAGTCAGACGTCAT	CAT	39 (65%)	
	BLS256_mite6	TGGGGTCTTGGCTGAAAGAACTTTCGACCGTCTCCTCGCGTCAGCAGCGCAAGAGTCAGACGTCAT	CAT	39 (65%)	
	BLS256_mite7	TGGGGTCTTGGCTGAAAGAACTTCCACCGTCTCCTCGCGTCAGCAGCGCAAGAGTCAGACGTCAT	CAT	38 (63%)	
	BLS256_mite8	TGGGGTCTTGGCTGAAAGAACTTCGACCGTCTCCTCGCGTCAGCAGCGCAAGAGTCAGACGTCAT	CAT	40 (66%)	
	BLS256_mite9	TGGGGTCTTGGCTGAAAGAACTTCCACCGTCTCCTCGCGTCAGCAGCGCAAGAGTCAGACGTCAT	CAT	38 (63%)	
	BLS256_mite10	TGGGGTCTTGGCTGAAAGAACTTTCGACCGTCTCCTCGCGTCAGCAGCGCAAGAGTCAGACGTCAT	CAT	39 (65%)	
	BLS256_mite12	TGGGGTCTTGGCTGAAAGAACTTTCGACCGTCTCCTCGCGTCAGCAGCGCAAGAGTCAGACGTCAT	CAT	39 (65%)	
	BLS256_mite13	TGGGGTCTTGGCTGAAAGAACTTTCGACCGTCTCCTCGCGTCAGCAGCGCTAGAGTCAGACGTCAT	CAT	37 (61%)	
	BLS256_mite15	TGGGGTCTTGGCTGAAAGAACTTCGACCGTCTCCTCGCGTCAGCAGCGCAAGAGTCAGACGTCAT	CAT	40 (66%)	
	BLS256_mite16	TGGGGTCTTGGCTGAAAGAACTTTCGACCGTCTCCTCGCGTCAGCAGCGCAAGAGTCAGACGTCAT	CAT	38 (63%)	
	BLS256_mite17	TGGGGTCTTGGCTGAAAGAACTTTCGACCGTCTCCTCGCGTCAGCAGCGCAAGAGTCAGACGTCAT	CAT	39 (65%)	
	BLS256_mite18	TGGGGTCTTGGCTGAAAGAACTTTCGACCGTCTCCTCGCGTCAGCAGCGCAAGAGTCAGACGTCAT	CAT	39 (65%)	
	BLS256_mite20	TGGGGTCTTGGCTGAAAGAACTTTCGACCGTCTCCTCGCGTCAGCAGCGCAAGAGTCAGACGTCAT	CAT	39 (65%)	
	BLS256_mite21	TGGGGTCTTGGCTGAAAGAACTTCGACCGTCTCCTCGCGTCAGCAGCGCAAGAGTCAGACGTCAT	CAT	40 (66%)	
	BLS256_mite23	TGGGGTCTTGGCTGAAAGAACTTTCGACCGTCTCCTCGCGTCAGCAGCGCAAGAGTCAGACGTCAT	CAT	38 (63%)	
	BLS256_mite24	TGGGGTCTTGGCTGAAAGAACTTTCGACCGTCTCCTCGCGTCAGCAGCGCAAGAGTCAGACGTCAT	CAT	39 (65%)	
	BLS256_mite25	TGGGGTCTTGGCTGAAAGAACTTCGACCGTCTCCTCGCGTCAGCAGCGCAAGAGTCAGACGTCAT	CAT	40 (66%)	
	BLS256_mite27	TGGGGTCTTGGCTGAAAGAACTTCGACCGTCTCCTCGCGTCAGCAGCAAGAGTCAGACGTCAT	CAT	34 (57%)	
	BLS256_mite29	TGGGGTCTTGGCTGAAAGAACTTCGACCGTCTCCTCGCGTCAGCAGCAAGAGTCAGACGTCAT	CAT	35 (58%)	

Figura 4.9 – Comparação entre as sequências dos MITEs-Xo em Xoc BLS256 e a sonda XOCORF3496. A identidade nucleotídica (e porcentagem de identidade) para cada cópia está descrita ao lado de cada sequência (em vermelho).

Ainda que consideremos a hibridização cruzada de transcritos de MITEs-Xo1 e XOCORF3496, a contribuição destes transcritos para a intensidade de fluorescência será reduzida. Assim, a intensidade de fluorescência detectada para esta sonda foi utilizada para inferir o nível de expressão de MITE-Xo2 nas duas cepas de *X. oryzae* investigadas neste estudo.

A sonda XOCORF3496 foi incluída na lâmina microarranjo porque o sítio genômico do qual XOCORF3496 foi derivado está erroneamente anotado como gene *exsF* (proteína regulatória/gene de regulação do sistema de dois componentes). Além de MITE-Xo, XOCORF3496 não possui identidade com nenhum outro local genômico de *Xoo* ou *Xoc*.

Foram realizados experimentos de hibridização separadamente para *Xoc* BLS256 e *Xoo* PXO99A, comparando a intensidade de fluorescência em XOCORF3496 de cada cepa separadamente, nos dois tratamentos. Para cada uma das cepas, foram realizadas 3 réplicas biológicas e, para cada réplica biológica foi realizada a hibridização com *dye-swap* (para investigar a influência do fluoróforo na eficiência de hibridização). Dessa forma, *Xoo* PXO99A e *Xoc* BLS256 possuem resultados de hibridização de seis microarranjos, cada uma (Figura 4.10).

Nosso objetivo ao analisar os dados desta publicação foi, em primeiro lugar, investigar o grau de expressão de transcritos de MITEs-Xo2 nestes experimentos. Comparamos a intensidade de fluorescência de XOCORF3496 com dois genes *housekeeping* – *lepA* (fator de alongação 4) e *atpD* (subunidade beta de ATP-sintase).

Como controle, foi utilizado *spots* vazios e as sondas de higromicina (632 *spots* vazios e 100 sondas de higromicina, por lâmina de microarranjo). Cada gene de interesse foi representado em duplicata na lâmina de microarranjo, assim, os genes investigados possuem duas medidas de intensidade de fluorescência, (1) e (2), para cada experimento de hibridização.

Figura 4.10 - Intensidade de fluorescência de MITEs-Xo2 por microarranjo. Intensidade de fluorescência de hibridização de transcritos provenientes de A) *Xoc* BLS256 e B) *Xoo* PXO99A. Referente à expressão de MITEs-Xo2 pela sonda XOCORF3496, genes *housekeeping* e controle. Em vermelho, intensidade de fluorescência detectada em cultura em meio mínimo XOM2; em verde, detectada em meio de cultura rico em nutrientes (PSB).

De uma maneira geral, *Xoo* apresenta uma maior intensidade de fluorescência que *Xoc*, para todas as sondas investigadas. Independente das condições de meio de cultura testados, PSB ou XOM2, a intensidade total de fluorescência detectada para XOCORF3496 é sempre maior do que a fluorescência detectada para o gene *housekeeping atpD*, para ambos *Xoo* e *Xoc*. Quando comparado ao gene *housekeeping lepA*, XOCORF3496 tem uma intensidade de fluorescência maior em *Xoc*, mas menor em *Xoo* (Figura 4.10). A partir destes dados, podemos inferir que MITE-Xo2 está sendo expresso, em níveis celulares não basais em ambos patovares de *X. oryzae*.

O objetivo principal dos autores foi identificar genes que são modulados no meio de cultura XOM2 (Seo et al., 2008). O meio mínimo XOM2 induz a expressão do cluster gênico *hrp* (Tsuge et al., 2002), cujas proteínas são secretadas pelo sistema de secreção do tipo III, e contribuem na virulência de *X. oryzae* em plantas hospedeiras ou na resposta hipersensibilidade em não-hospedeiros. Devido a grande variação de intensidade entre as réplicas biológicas e a falta de reproducibilidade das diferenças de expressão de XOCORF3496 entre os dois tratamentos, não é possível afirmar se a expressão de MITE-Xo2 está ou não está associada à expressão de genes do cluster *hrp*, e de certa forma, à patogênese.

4.3.4 Análise de expressão de MITE-Xo por RNA seq

Para investigar a expressão diferencial entre três cepas de *Xoo*, que exibem graus de virulência em 30 variedades distintas de arroz, Zhang et al. (2013) analisaram o transcrito total de *Xoo* PXO86, PXO99A (proveniente das Filipinas) e GD1358 (proveniente da China) cultivados no meio mínimo XOM2.

A partir dos dados de transcrito total disponíveis publicamente deste estudo, avaliamos se os MITEs-Xo são transcritos nestes organismos e acessamos qual o nível de transcrição destes elementos nos genomas de *Xoo*. Restringimos nossa análise aos genomas de PXO99A e PXO86 porque a cepa GD1358 não possui sequenciamento completo do genoma.

MITEs-Xo não estão anotados apropriadamente nos genomas de *X. oryzae*. Assim, partimos dos dados cru de sequenciamento, e analisamos a abundância de transcritos relativos ao MITEs-Xo (30 cópias em PXO99A e 25 cópias em PXO86), utilizando o software kallisto (Schmidtke et al., 2013) (Apêndices K e L).

Para comparação, quantificamos a abundância de transcritos pertencentes a duas outras categorias gênicas: 57 genes de proteínas ribossomais, geralmente expressos em altos níveis celular, e de 14 genes *housekeeping* (Figura 4.11).

Figura 4.11 - Quantificação de transcritos de MITes em Xoo PXO99A e Xoo PXO86. Caixas em azul indicam a variação e distribuição das cópias de MITes-Xo; em vermelho, de genes de proteínas ribossomais; em verde, de genes *housekeeping*. Os limites das caixas indicam percentis 25 e 75; linhas centrais nas caixas indicam medianas; *whiskers* indicam 1.5 x faixa interquartil. Dados estão plotados como círculos e *outliers* estão representados pelos círculos fora das caixas. MITes com TPM igual a 0 foram excluídos.

Nosso resultado mostra que as três categorias gênicas apresentam abundância de transcritos semelhantes em ambos Xoo PXO86 e Xoo PXO99A, com alta correspondência de valor de TPM (transcritos por milhão) para os genes homólogos (Figura 4.1 e Apêndices K e L). Como esperado, a categoria de proteínas ribossomais apresenta a maior abundância de transcritos. Genes da categoria *housekeeping* apresentam uma mediana de TPM menor do que genes de proteínas ribossomais, com uma variação estreita, porém com alguns genes (*outliers*) de alta expressão. As cópias de MITes-Xo apresentam uma ampla variação de índices de TPM em ambas cepas, indicando que enquanto algumas cópias do elemento apresentam uma baixa abundância de transcritos, outras apresentam uma altíssima abundância de transcritos, com valores de TPM maiores do que algumas proteínas ribossomais.

Assim como os dados de microarranjo, os dados de transcriptomas provenientes deste estudo identificou que MITes-Xo são expressos em grandes quantidades na célula.

4.3.5 Potencial caráter funcional dos MITes-Xo

O alto número de cópias de MITes-Xo, a presença de cópias de MITes-Xo conservadas entre cepas e entre patovares, e o alto nível de expressão destes elementos demonstrados pela análise de microarranjo e RNA-seq gerou questionamentos sobre o possível caráter funcional dos MITes-Xo em *X. oryzae*.

MITEs podem exercer uma atividade regulatória dentro do organismo que os codificam. Esses elementos podem conter regiões regulatórias capazes de modular genes adjacentes a sua inserção ou ainda, influenciar a estabilidade do RNA mensageiro ao serem co-transcritos com outros genes. As propriedades regulatórias que MITEs-Xo podem ter em *X. oryzae* ainda não foram exploradas e devem ser verificadas experimentalmente.

Há um crescente acúmulo de evidências de que pequenos RNAs (sRNAs) são agentes importantes na interação patógeno-hospedeiro, geralmente elucidados por estudos que investigam o enriquecimento de pequenos RNAs na resposta ao processo de infecção, seja no organismo patógeno ou no hospedeiro (Chaloner et al., 2016; Weiberg et al., 2014).

Em bactérias fitopatógenas, foi identificado sRNAs que atuam na regulação intracelular de genes de seus virulência. Em *Erwinia carotovora* ssp. *carotovora*, causadora da podridão mole, RsmB (RNA de ligação à proteína) afeta o sistema de *quorum-sensing*, produção de enzima extracelulares e virulência através da modulação da atividade da proteína repressora RsmA (Liu et al., 1998); os sRNAs RprA e RyhA são fundamentais para virulência em *Erwinia amylovora* (Zeng et al., 2013); em *Xanthomonas campestris* ssp. *vesicatoria*, o sRNA sX13 modula mais de 60 mRNAs, além de ter um papel central na promoção da expressão de genes de virulência (Schmidtke et al., 2013).

O tráfego de sRNA também acontece entre diferentes organismos muito distantes evolutivamente. Esse fenômeno foi predominantemente descrito entre organismos que estabelecem uma associação de patogenicidade, de simbiose ou parasítica. (Knip et al., 2014).

A interação entre *Botrytis cinerea*, fungo que causa a podridão (ou mofo) cinzenta, e *A. thaliana* é um exemplo inequívoco da capacidade de sRNA produzidas pelo fungo cruzarem a barreira entre Reinos e influenciar genes alvo na planta hospedeira (Weiberg et al., 2013). Os pequenos RNAs (Bc-sRNAs) são processados em *B. cinerea*, porém atuam sobre genes envolvidos na resposta imune da planta. Neste processo, além de modularem a expressão de genes de defesa, sequestram o mecanismo de RNAi da planta hospedeira, se ligando a AGO1, e suprimem o sistema imune da planta de forma sistêmica. Interessantemente, dois dos três sRNAs capazes de tal interferência inter-organísmica são derivados da região LTR de elementos de transposição em *B. cinerea* (Weiberg et al., 2013).

A fim de explorar o potencial dos MITEs-Xo transcritos no fitopatógeno *X. oryzae* como mecanismo de interferência em genes de seu hospedeiro, avaliamos primeiramente se a sequência de MITE-Xo está presente no genoma *Oryza sativa*.

Através da análise de similaridade entre cópias de MITEs-Xo de Xoc BLS256 contra o genoma de *O. sativa* ssp. *japonica*, foi identificado um fragmento (*motif*) de 18 nucleotídeos que ocorre repetidamente em diferentes localidades cromossômicas do arroz. A região de similaridade está sempre restrito às coordenadas 88 e 105 de MITE-Xo1.

Das 24 cópias de MITE-Xo1 presentes em *Xoc* BLS256, o *motif* é idêntico em 13 cópias, enquanto as outras 9 cópias diferem em 1 ou 2 nucleotídeos. As 6 cópias da variante MITE-Xo2 diferem de 3 a 4 nucleotídeos, em relação a MITE-Xo1 (Figura 4.12). A estrutura secundária do MITE-Xo1 e a localização do *motif* está descrito em Apêndice N. Comparativamente, somente duas são cópias de MITE-Xo1 são 100% idênticas em PXO99A; a maioria das cópias apresentam um nucleotídeo distinto (Figura 4.12).

A

Figura 4.12 - Motivo de 18 nucleotídeos nas cópias de MITEs-Xo em *Xoo* e *Xoc*. Conservação do motivo de 18 nucleotídeos nas cópias de MITEs-Xo em A) *Xoc* BLS256 e B) *Xoo* PXO99A. Sombra azul delimita os motivo de 18 nucleotídeos. Diferenças nucleotídicas em relação ao motivo estão indicadas em vermelho.

Este *motif* de 18 nucleotídeos está presente 109 vezes no genoma de *O. sativa* ssp. japonica, distribuídos em 10 de seus 12 cromossomos, sendo que 82 regiões são 100% idênticas ao *motif* e 27 apresentam um nucleotídeo de diferença.

Trinta dos 109 motivos encontrados em *O. sativa* compõem regiões intrônicas de genes. Os genes identificados codificam são putativamente expressos e resultam em uma grande diversidade de proteínas que inclui RNA não codificante (nc-RNA), domínios de regiões de transmembrana, metiltransferase e elementos de transposição (Tabela 4.2). Considerando uma atuação hipotética deste *motif* no genoma de *Oryza sativa*, destacamos

a seguir seis loci, dos quais cinco possuem papéis diretos ou indiretos na interação planta-patógeno.

1. A proteína codificada pelo locus Os05g36080 possui identidade com proteína WRKY17. A classe de fator de transcrição WRKY tem um papel importante na regulação de resposta a stress e de defesa da planta e a família WRKY17, em particular, é fortemente induzida em *Arabidopsis* quando infectada com *Pseudomonas syringae* pv *tomato*. WRKY17, juntamente com WRKY11, atua como regulador negativo do defesa basal, evitando a expressão excessiva, e prejudicial, de genes de defesa (Journot-Catalino et al., 2006). Por outro lado, a deleção de *wrky17* aumentou susceptibilidade de *A. thaliana* à infectividade a *Agrobacterium* (Lacroix, Citovsky, 2013) e apresentou um aumento no número de galhas quando parasitadas pelo nemátodo *Meloidogyne* spp. (Teixeira et al., 2016).
2. O locus Os08g16120A codifica o gene de resistência RPM1, central no sistema imune de plantas. A proteína RPM1 pertence a classe NLR (Nucleotide-binding Leucine-Rich Repeats) caracterizada pela presença de um domínio proteico central de ligação a nucleotídeos (NB-ARC) e o domínio LRR (leucine-rich repeat) carboxi-terminal. A presença do *motif* CC (coiled-coil) amino-terminal subclassifica a proteína identificada como tipo-CC NLR (Jacob et al., 2013). Proteínas NLR são encontradas no citosol da célula e agem como receptores intracelulares que detectam diretamente proteínas efetoras de patógenos. A ativação de NLR desencadeia um conjunto de respostas associadas a morte celular programada, como observado na resposta de defesa por hipersensibilidade (Qi, Innes, 2013).
3. A subunidade do complexo Exo70 (locus Os02g36619) é uma das oito subunidades proteicas responsáveis pela formação do exocisto, que funcionam em um complexo de formação de vesícula e tráfego de macromoléculas. Em plantas, Exo70 representa uma família multigênica sendo representada em 23 parálogos em *A. thaliana* e 41 parálogos em *O. sativa* (Chong et al., 2010). O complexo exocisto também está envolvido na primeira linha de defesa da célula vegetal ao ataque de patógenos, no rápido reforço da parede celular (deposição de papila) (Collinge 2009; Underwood 2012). Mutantes de Exo70B2 e Exo70H1 em *Arabidopsis* mostraram maior susceptibilidade a infecção de *Pseudomonas syringae* pv. *maculicola* e do fungo *Blumeria graminis* f. sp. *hordei* (Pečenkova et al., 2011).
4. Repetições de ankirina, como a identificada no locus Os03g47720, constituem uma família proteica grande, geralmente encontrada em associação com uma diversidade de outros domínios proteicos. Proteínas com repetições de ankirina estão envolvidas em funções celulares muito distintas, como no desenvolvimento celular

(morfogênese de folha, embriogênese, biogênese de membrana do cloroplasto, diferenciação plastidial em *Arabidopsis*) e na resposta a stresses biótico e abióticos (Huang et al., 2009). Em arroz, são encontrados 175 genes contendo repetições de ankirina, das quais 22 contém a repetição tetratricopeptídica (ANK-TPR, semelhante ao gene em Os03g47720), comparado a 4 proteínas presentes em *Solanum lycopersicum* e 1 em *A. thaliana* (Yuan et al., 2013).

5. O locus Os11g4470 codifica uma proteína de ligação à calmodulina, putativamente expressa. A calmodulina, quando ligada a Ca^{+2} , ativa proteínas alvo envolvidas processos celulares como regulação do metabolismo, transporte iônico, fosforilação e defosforilação proteica, metabolismo de fosfolípidos, entre outras. Porque calmodulina não possui atividade enzimática sozinha, a sinalização por calmodulina e os processos fisiológicos que desencadeia são derivados da atividade de proteínas reguladas por calmodulina, como proteínas de ligação à calmodulina. O aumento da concentração de cálcio intracelular do hospedeiro representa um das primeiras sinalizações da interação planta-patógenos.
6. Embora não esteja diretamente relacionado à defesa da planta, curiosamente um dos genes hipoteticamente alvos do *motif* é uma proteína Argonauta. Argonautas (AGO) são proteínas fundamentais no mecanismo de regulação gênica por sRNAs, fundamentais para o processo na formação do complexo de silenciamento gênico por RNA, RISC (*RNA-induced silencing complex*). Três domínios proteicos são fundamentais em sua composição: PAZ e Mid, que confere capacidade de ligação de RNA a proteína, e PIWI que confere a atividade catalítica, pela conservação da tríade de aminoácidos Asp-Asp-His (Hutvagner, Simard, 2008; Joshua-Tor, 2006). *O. sativa* possui 19 genes da família AGO. A proteína Argonauta (Ago) do locus Os07g28850 foi classificada como OsAGO18. De acordo com a filogenia das família gênica em arroz, OsAGO18 não se agrupa a nenhum dos quatro subgrupos formados na árvore filogenética (Kapoor et al., 2008). Proteínas AGO também desempenham outras funções. Em *O. sativa*, por exemplo, OsMEL1 regula meiose em células germinativas e no desenvolvimento de gametas femininos e masculinos através da modificação da cromatina; e OsPNH1 regula o desenvolvimento foliar e a manutenção do meristema apical (Nishimura et al., 2002).

Tabela 4.2 - Local cromossômico e anotações de genes potencialmente alvos do motif de 18 nucleotídeos de MITE-Xo1.

Local genômico		Anotação gênica		Orientação	Tamanho (nt)	%ID	Anotações sobre domínio proteicos ou funções putativas
Cr.	Coordenadas	IRGSP	MSU				
2	22113504-22113521	OS02G0575900	Os02g36619	Reverse	18	100	proteína Exo70 co complexo exocisto
2	30765520-30765537	-	Os02g50374	Forward	18	100	peptídeo-sinal / região transmembrana
2	30860713-30860730	-	Os02g50540	Forward	18	100	domínios seg
2	6309571-6309588	-	Os02g12120	Reverse	18	100	putativo Ty1-copia (retrotransposon)
2	8514926-8514943	-	Os02g15240	Reverse	18	100	putativo Mutator sub-class (transposon)
3	18060398-18060415	-	Os03g31634	Forward	18	100	domínios seg / família proteica PE-PGRS
3	27044612-27044629	OS03G0680600	Os03g47720	Forward	18	100	repetição de ankrina com repetição TPR, proteína putativa induzida por stress
3	27851081-27851098	OS03G0695425	Os03g48904	Reverse	18	100	repetição de ankrina/transcriptase reversa de retroelemento sem LTR
3	36411501-36411518	OS03G0862300	-	Forward	18	100	ncRNA
3	5466930-5466947	OS03G0204300	Os03g10700	Reverse	18	100	domínios seg / proteína SOUL de ligação heme
3	993667-993684	OS03G0118200	Os03g02670	Forward	18	100	transportador de UDP-galactose/transportador UAA/transmembrana
4	18496097-18496114	EPIOSAG00000052212	Os04g30940	Reverse	18	100	ncRNA
4	29333204-29333221	OS04G0581000	-	Forward	18	94.4	isopenicilina N sintase
4	4570278-4570295	-	Os04g08470	Reverse	18	100	domínios seg
5	21380117-21380134	-	Os05g36080	Reverse	18	100	WRKY17-like
5	26029804-26029821	-	Os05g44780	Reverse	18	100	metiltransferase DNA citosina (Zmet3)
5	26778903-26778920	OS05G0539150	Os05g46170	Forward	18	100	domínio F-box
6	7267292-7267309	OS06G0240001	-	Reverse	18	94.4	não-caracterizada, expressa
7	16898650-16898667	OS07G0471300	Os07g28850	Forward	18	100	proteína Argonauta (OsAGO18)
7	21522023-21522040	OS07G0543700	Os07g35979	Forward	18	100	domínios seg
7	23769529-23769546	OS07G0585300	-	Forward	18	94.4	domínios seg
7	29400792-29400809	OS07G0691200	Os07g49110	Reverse	18	100	D-alanina-D-alanina ligase, superfamília ATP-grasp
7	3916811-3916828	-	Os07g07820	Reverse	18	100	domínios sinal PHMM/Coil/Seg
8	28137235-28137252	OS08G0561800	Os08g44780	Forward	18	100	proteína SOUL de ligação heme
8	9828441-9828458	OS08G0261000	Os08g16120	Forward	18	100	proteína RPM1 de resistência, putativa, expressa
9	18278592-18278609	OS09G0477550	-	Reverse	18	100	domínios seg
10	14408014-14408031	OS10G0413101	Os10g27310	Reverse	18	100	domínios seg, expressa
11	27031803-27031820	-	Os11g44700	Reverse	18	100	proteína de ligação a calmodulina, putativa, expressa
11	6481878-6481895	-	Os11g11660	Forward	18	100	proteína expressa
11	7867456-7867473	OS11G0245900	Os11g14120	Forward	18	100	domínio de transmembrana/seg

Exploramos a universalidade do *motif* em diferentes genomas, ou seja, se a alta frequência do *motif* encontrada na espécie *O. sativa* é particular a esta espécie ou se trata de uma sequência comum a outros genomas. Através do banco de dados Gramene (<http://www.gramene.org/>) analisamos a presença do *motif* de 18 nucleotídeos derivado de MITE-Xo1 nas 11 espécies de *Oryza* atualmente sequenciadas (Figura 4.13, Apêndice M). Para comparação, foi incluída a espécie *Leersia perrieri*, que compartilha a tribo Oryzeae com *Oryza*; *Zea mays*, que compartilha a família Poaceae e *A. thaliana*, como um exemplo do dicotiledônia (Figura 4.13).

Figura 4.13 - Cladograma das espécies de *O. sativa* e presença do *motif* de 18 nucleotídeos. O tipo de genoma de *O. sativa* está indicado à frente de cada genoma. Em parênteses, estão apontados o número de *motifs* com alta identidade. Cladograma retirado de Geering et al. (2014). As espécies de *Z. mays* e *A. thaliana* foram adicionadas posteriormente. Linhas pontilhadas indicam uma representação não acurada da distância destas espécies ao gênero *Oryza*.

Os genomas de espécies de *Oryza* são classificados em 11 grupos, de AA a LL. *O. brachyantha* (FF) não apresenta o *motif* e *O. punctata* (genótipo BB) apresenta somente 35 *motifs* – menor frequência de *motifs* dentro gênero. No complexo de genomasn AA, *O. meridionalis* e *O. glaberrima* apresenta 92 e 93 *motif*, respectivamente. As outras sete espécies apresentam mais do que 100 vezes o *motif* no genoma, sendo que *O. longistaminata* apresenta o maior número de *motif* (148). As subespécies de *O. sativa*, ssp. *japonica* e ssp. *indica*, que representam as variedades mais cultivadas de arroz no mundo, também apresentam alta frequência do *motif*, presentes 109 e 129 vezes em seus

respectivos genomas. Nos grupos externos ao gênero *Oryza*, observamos que o *motif* de 18 nucleotídeos é ausente nas espécies *Z. mays* e *A. thaliana*, e está presente somente uma vez no genoma de *L. perrieri*.

4.4 Discussão

Neste capítulo descrevemos um novo elemento de transposição do tipo MITE, encontrados em todas as cepas de *X. oryzae* atualmente sequenciadas e analisamos *in silico* seus aspectos estruturais, filogenéticos, de expressão.

O elemento MITE-Xo possui todas as características estruturais de um elemento MITE – é pequeno, apresenta repetições terminais invertidas (que podem ser reconhecidas por outros elementos IS) e não codifica nenhuma proteína em sua região interna. MITEs-Xo são exclusivos da espécie *X. oryzae* e foram classificados em duas variantes distintas, MITE-Xo1 e MITE-Xo2.

Utilizamos a filogenia e o contexto genômico (sintenia) das cópias de MITEs-Xo como estratégias complementares para entender o cenário evolutivo do elemento. Os 16 genomas de *X. oryzae* (provenientes de 10 cepas de *Xoc* e 6 cepas de *Xoo*) compreendem um total de 435 cópias de MITEs-Xo, das quais 75 pertencem a variante MITE-Xo2 e 360 pertencem a MITE-Xo1. Nossas análises assumem que a probabilidade de uma cópia de MITE ser herdada verticalmente é maior do que a probabilidade de um elemento ser inserido no exato mesmo local genômico em diferentes cepas, independentemente.

A filogenia de de MITE-Xo1 infere, primeiramente, que esta variante expandiu em cada patovar *Xoo* e *Xoc* em eventos distintos, ou seja, as cópias de MITE-Xo1 não são ancestrais à espécie *X. oryzae*, mas a cada patovar. Também observamos uma alta concordância entre grupos sintênicos e clados filogenéticos das cópias de MITE-Xo1, presentes em *Xoo*. MITEs-Xo1 em *Xoo* que pertencem a um mesmo grupo sintênico são sempre agrupadas em clados com alto valor de suporte. A alta frequência desta observação nos levou a concluir que os elementos MITE-Xo1 se comportam filogeneticamente como genes e o agrupamento de cópias em um mesmo clado filogenético representam ancestralidade comum. Dessa forma, cópias de MITE-Xo1 provenientes de um mesmo genoma, quando agrupadas em um só clado, sugere que são derivadas uma da outra, e representa a transposição replicativa do elemento na cepa.

A correspondência entre grupos sintênicos de MITE-Xo1 e clados filogenéticos é bem menos frequente em *Xoc*. Em *Xoc*, com exceção de *miteXo1-4*, *miteXo1-8* e *miteXo1-11*, grupos sintênicos não são agrupados robustamente na árvore filogenética de MITE-Xo1. Ou ainda, cópias de um mesmo grupo sintênico estão separadas na árvore. Essa topografia pode indicar que as cópias de MITE-Xo1 de um grupo sintênico estão sob taxas de substituição distintas que as separa em clados filogenéticos distintos. Alternativamente,

estão em um mesmo local genômico por convergência, devido a tendência de se transporem em sítios AT, como descrito anteriormente. Essa hipótese devem ser testadas e permanece sob investigação.

Para a variante MITE-Xo2, cópias de um mesmo contexto genômico (sintênicas) são agrupadas de forma robusta na árvore filogenética. Duas cópias de MITE-Xo2 (*miteXo2-1* e *miteXo2-2*) são ancestrais à espécie *X. oryzae* e as expansões subsequentes de MITE-Xo2 é específica aos patovares *Xoo* e *Xoc*.

MITE-Xo2 expande mais em *Xoc* do que em *Xoo*. Em *Xoc*, Além das duas cópias ancestrais a *X. oryzae*, *miteXo2-3*, *miteXo2-4* e *miteXo2-5* são cópias ancestrais a *Xoc*, presente em todas as linhagens asiáticas e Oeste-africanas deste patovar. Os grupos sintênicos *miteXo2-4* e *miteXo2-5* são agrupados conjuntamente, e sugere que esta cópia são derivadas uma da outra. Um novo MITE-Xo2 ocorre ancestralmente ao clado asiático (*miteXo2-6*) e outro, ancestral ao clado Oeste-Africano/Índia (*miteXo2-8*, com reversão em CFBP7342). Além desses eventos, reversões e o surgimento de uma nova cópia de MITEs-Xo2 (*miteXo2-9*) ocorre somente em no clado Oeste-africano.

Em *Xoo*, os MITEs-Xo2 ancestrais representam 2 das 3 cópias do elemento nesta variedade. Somente mais um evento de inserção ocorre em *Xoo*, restrito às linhagens asiáticas. A cepa africana *Xoo* AXO1947 é uma exceção à todas a história de MITEs-Xo2 *Xoo*. De acordo com o cenário evolutivo apresentado, AXO1947 perde uma das cópias ancestrais (*miteXo2-2*) e apresenta duas outras cópias, em contextos genômicos distintos e exclusivas à esta cepa.

De uma forma geral, o perfil de MITEs-Xo é coerente com a história evolutiva da espécie *X. oryzae*. Cepas evolutivamente mais próximas possuem perfis de MITE-Xo semelhantes, indicando que MITEs-Xo podem servir como marcadores moleculares em *X. oryzae*. Também mostramos que a análise filogenética de elementos auxilia na identificação de MITEs conservados dentro da população e ainda pode auxiliar na identificação da tipo e frequência de transposição dentro de um genoma, ainda que o elemento tenha potencial de mobilidade, seja pequeno (210 nucleotídeos) e não codifique nenhuma proteína,

As cepas africanas tendem a ter uma dinâmica maior para ambas variantes de MITEs-Xo, visualizadas pelas aquisições de novas cópias e reversões de cópias ancestrais. A cepa *Xoc* BXORI, embora proveniente da Índia, se agrupa filogeneticamente com as cepas africanas e o perfil de MITEs é igualmente contrastante: ora compartilha eventos de inserção do elemento com as cepas asiáticas, ora com as cepas africanas.

As linhagens analisadas de *Xoc* foram isoladas dentro de um período de 45 anos (1964-2009), sendo que as linhagens asiáticas foram coletadas entre 1996 e 2007, enquanto as cepas africanas estão restritas a 2003 e 2009. Não acreditamos que a data de isolamento possa ser usada para explicar as diferenças observadas no perfil de MITEs-Xo

destas populações. Nossa conclusão é baseada nos dados de *Xoc* CFBP2286 (coletada em 1964) e *Xoc* B8-12 (coletada em 2007) que possuem perfis idênticos de MITEs-Xo2 e compartilham o maior número de cópias de MITE-Xo1 (21 inserções). Como discutido anteriormente no Capítulo 3, a diversidade de inserções de MITEs-Xo observada nas linhagens africanas de *Xoo* ou *Xoc* pode ser reflexo do processo de reestruturação genômico-populacional do patógeno frente a recente intensificação de cultivo de arroz e à diversidade de arroz cultivada no Oeste africano.

Cópias de MITE-Xo encontradas em um contexto genômico exclusivo em uma única cepa também foram assinaladas a um grupo sintênico, a fim de oferecer um arcabouço para avaliação do perfil de MITEs-Xo de novas linhagens de *X. oryzae* sequenciadas. O sequenciamento e montagem de genomas bacterianos torna-se cada vez mais rápido, facilitado pelas novas tecnologias de sequenciamento. Plataformas como SMRT (single-molecule read time sequencing, Pacific Biosciences) e MinION (Oxford Nanopore Technologies), que utilizam sequenciamento de longos fragmentos (média de 5 a 10 kilobases) elimina a ambiguidade de tamanho e posição de elementos repetitivos (Goodwin et al., 2016). Assim, sequenciamento de genomas bacterianos começa a ser um método de rotina e a variação genética entre populações cada vez mais e melhor explorada. Ao final da análise deste documento, genomas de mais oito novas linhagens de *Xoo* foram sequenciados a fim de avaliar a adaptação fenotípica de isolados das Filipinas (Quibod et al., 2016).

Para que um MITE desempenhe um papel funcional é necessário que seja transcrito no organismo em que está presente. Foi analisado transcritomas publicamente disponíveis de *X. oryzae*, de dois estudos distintos. Ambos utilizaram o meio sintético mínimo XOM2, de indução do cluster *hrp*, a fim de mimetizar um estado de expressão gênica na patogenicidade. A quantificação da intensidade total de fluorescência por microarranjo permitiu observar que MITE-Xo2 está sendo expresso celularmente em níveis não-negligíveis em *Xoo* e *Xoc*, embora não seja possível estabelecer um padrão de diferencial de expressão entre os meios de cultura utilizados.

O transcritoma total sequenciado por (Zhang et al., 2013) demonstra que ambos MITE-Xo estão sendo altamente expressos em duas cepas de *Xoo*. A alta similaridade entre as cópias de MITE-Xo não permite a quantificação acurada de transcritos para cada uma delas. Embora o algoritmo utilizado pelo software kallisto consiga distinguir *reads* entre cópias que apresentem um nucleotídeo de diferença, *reads* ambíguos - que podem ser assinalados a mais de uma cópia - são divididos em partes iguais entre as cópias. No entanto, o método consegue indicar a tendência de certas cópias serem mais ou menos expressas. De acordo com esta análise, algumas cópias de MITEs-Xo apresentam níveis de expressão comparáveis a genes ribossomais altamente expressos.

A presença de um alto número de cópias de MITEs-Xo presentes em *X. oryzae*, a identificação de cópias ancestrais à *X. oryzae* e subsequente expansão destes elementos em cada patovar e o alto índice de expressão dos elementos na célula nos motivaram a investigar o potencial funcional que MITEs-Xo podem desempenhar em seu hospedeiro arroz.

Elementos MITEs-Xo não codificam proteínas mas a estrutura secundária predita do transcrito de MITEs-Xo1 poderia, por exemplo, produzir *hairpins* e serem processados a sRNAs. MicroRNAs modulam a expressão gênica de processo celulares através do complexo de silenciamento induzido por RNA (RISC), se ligando a mRNAs e resultando na degradação do transcrito ou na inibição da tradução.

Em arroz, elementos de transposição (TEs) representam umas das diversas fontes de microRNA e a maior parte de pequenos RNA regulatórios (miRNA e siRNA) gerados por TEs são derivados de elementos do tipo MITE. Transcritos de MITEs residentes no genoma de *O. sativa*, por serem pequenos e manterem a estrutura palindrômica de repetições terminais invertidas, se dobras em *hairpins* similares a precursores de miRNAs. Os *hairpins* podem então ser processados (pelas enzimas Dicer) e resultar em pequenos RNAs maduros e funcionais (Zhang et al., 2011; Piriyaongsa, Jordan, 2008).

Dentro de um cenário hipotético no qual o transcrito de MITE-Xo1, proveniente de *X. oryzae*, seja capaz de chegar à célula do hospedeiro, examinamos os possíveis genes alvos que este elemento poderia atuar. Interessantemente, o *motif* de 18 nucleotídeos, localizados dentro de um *hairpin* da estrutura secundária de MITEXo1 (Apêndice N), ocorre em um número de vezes muito maior em *Oryza* do que em espécies não hospedeiras de *X. oryzae*, como o *L. perrieri*, milho (*Z. mays*) e *A. thaliana*.

O gênero *Oryza* engloba 23 espécies, 2 espécies cultivadas e 21 espécies selvagens, cujos genomas podem ser classificados de AA a LL. A espécie *O. sativa*, composta por milhares de variedades de cultivo, é cultivada por todo o mundo e corresponde a vasta maioria do arroz produzido e consumido no planeta. A espécie *O. glaberrima* foi domesticada posteriormente, atualmente cultivada somente em partes do Oeste Africano (<http://ricepedia.org/>) e é a única espécie além de *O. sativa* cultivada. Outras espécies, embora não utilizadas para produção de arroz, são conhecidamente infectadas por *X. oryzae*, como *O. longistaminata* e *O. barthii*.

A presença recorrente do *motif* também varia dentro do gênero *Oryza*, e aparece com maior frequência dentro dos cultivares pertencentes a mesma classe de genótipos (complexo *sativa*, genótipos AA, com distribuição tropical) do que nas duas espécies de genótipo mais distantes, *O. brachyanta* e *O. punctata*. Assim, a frequência do *motif* parece sofrer uma explosão no genomas AA de arroz. Esse resultado descaracteriza, até certo ponto, a estocasticidade da presença do *motif* nas espécies de *Oryza*.

Os possíveis genes alvo de MITE-Xo1 apresentam identidade com o *motif* na região intrônica do modelo gênico disponível nos bancos de dados Gramene e MSU Rice Genome Annotation Project. São genes envolvidos em diversos processos celulares, incluindo genes envolvidos diretamente na interação planta-patógeno como RPM1 e o fator de transcrição WRKY17. A análise *in silico* realizada com MITEs-Xo oferece evidências que este elemento pode ter uma atividade funcional, seja dentro do fitopatógeno *X. oryzae* ou do principal hospedeiro vegetal, *O. sativa*.

Será importante testar experimentalmente a existência de forças promotoras em MITE-Xo1. Ainda, pretendemos explorar experimentalmente o potencial de MITE-Xo1 ser processado e atuar como um pequeno RNA dentro da célula hospedeira. Existe um esforço internacional coletivo de criação, organização e disponibilidade de informações genéticas e genômicas para o avanço do melhoramento do arroz em face às mudanças climáticas e demandas de produção exigidas num futuro próximo. Assim, foi lançado projeto de sequenciamento genômico de 3,000 variedades de arroz cultivados hoje no mundo (The 3,000 rice genomes project, 2014). Com a disponibilização pública deste novo conjunto de informações, será interessante estender nosso entendimento sobre a frequência do *motif* e correlacionar com a suscetibilidade das variedades de arroz à *X. oryzae*.

REFERÊNCIAS*

- Bray NL, Pimentel H, Melsted P, Pachter L. Near-optimal probabilistic RNA-seq quantification. *Nat Biotechnol.* 2016;34(5):525–7.
- Chaloner T, van Kan JAL, Grant-Downton RT. RNA “Information Warfare” in Pathogenic and Mutualistic Interactions. *Trends Plant Sci.* 2016;21(9):738–48.
- Chong YT, Gidda SK, Sanford C, Parkinson J, Mullen RT, Goring DR. Characterization of the *Arabidopsis thaliana* exocyst complex gene families by phylogenetic, expression profiling, and subcellular localization studies. *New Phytol.* 2010;185(2):401–19.
- Chou C-C, Chen C-H, Lee T-T, Peck K. Optimization of probe length and the number of probes per gene for optimal microarray analysis of gene expression. *Nucleic Acids Res.* 2004;32(12):e99–e99.
- Collinge DB. Cell wall appositions: the first line of defence. *J Exp Bot.* 2009;60(2):351–2.
- Crooks GE, Hon G, Chandonia J-M, Brenner SE. WebLogo: A Sequence Logo Generator. *Genome Res.* 2004;14(6):1188–90.
- Darmon E, Leach DRF. Bacterial Genome Instability. *Microbiol Mol Biol Rev.* 2014;78(1):1–39.
- Delihias N. Impact of Small Repeat Sequences on Bacterial Genome Evolution. *Genome Biol Evol.* 2011;3:959–73.
- Fattash I, Lee C-N, Mo K, Yang G. Efficient transposition of the youngest miniature inverted repeat transposable element family of yellow fever mosquito in yeast. *FEBS J.* 2015;282(10):1829–40.
- Geering ADW, Maumus F, Copetti D, Choisne N, Zwickl DJ, Zytnicki M, et al., Endogenous florendoviruses are major components of plant genomes and hallmarks of virus evolution. *Nat Commun.* 2014;5:5269.
- Goodwin S, McPherson JD, McCombie WR. Coming of age: ten years of next-generation sequencing technologies. *Nat Rev Genet.* 2016;17(6):333–51.
- Gouy M, Guindon S, Gascuel O. SeaView Version 4: A Multiplatform Graphical User Interface for Sequence Alignment and Phylogenetic Tree Building. *Mol Biol Evol.* 2010;27(2):221–4.
- Han M-J, Shen Y-H, Gao Y-H, Chen L-Y, Xiang Z-H, Zhang Z. Burst expansion, distribution and diversification of MITEs in the silkworm genome. *BMC Genomics.* 2010;11:520.
- Hikosaka A, Nishimura K, Hikosaka-Katayama T, Kawahara A. Recent transposition activity of *Xenopus* T2 family miniature inverted-repeat transposable elements. *Mol Genet Genomics MGG.* 2011 Mar;285(3):219–24.

*De acordo com:

International Committee of Medical Journal Editors. [Internet]. Uniform requirements for manuscripts submitted to Biomedical Journal: sample references. [updated 2011 Jul 15]. Available from: <http://www.icmje.org>

- Huang J, Zhao X, Yu H, Ouyang Y, Wang L, Zhang Q. The ankyrin repeat gene family in rice: genome-wide identification, classification and expression profiling. *Plant Mol Biol*. 2009;71(3):207–26.
- Hughes TR, Mao M, Jones AR, Burchard J, Marton MJ, Shannon KW, et al., Expression profiling using microarrays fabricated by an ink-jet oligonucleotide synthesizer. *Nat Biotechnol*. 2001;19(4):342–7.
- Hutvagner G, Simard MJ. Argonaute proteins: key players in RNA silencing. *Nat Rev Mol Cell Biol*. 2008;9(1):22–32.
- Jacob F, Vernaldi S, Maekawa T. Evolution and Conservation of Plant NLR Functions. *Front Immunol*. 2013;4.
- Jiang N, Feschotte C, Zhang X, Wessler SR. Using rice to understand the origin and amplification of miniature inverted repeat transposable elements (MITEs). *Curr Opin Plant Biol*. 2004;7(2):115–9.
- Joshua-Tor L. The Argonautes. *Cold Spring Harb Symp Quant Biol*. 2006;71:67–72.
- Journot-Catalino N, Somssich IE, Roby D, Kroj T. The Transcription Factors WRKY11 and WRKY17 Act as Negative Regulators of Basal Resistance in *Arabidopsis thaliana*. *Plant Cell*. 2006;18(11):3289–302.
- Kapoor M, Arora R, Lama T, Nijhawan A, Khurana JP, Tyagi AK, et al., Genome-wide identification, organization and phylogenetic analysis of Dicer-like, Argonaute and RNA-dependent RNA Polymerase gene families and their expression analysis during reproductive development and stress in rice. *BMC Genomics*. 2008;9:451.
- Katoh K, Misawa K, Kuma K, Miyata T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res*. 2002;30(14):3059–66.
- Knip M, Constantin ME, Thordal-Christensen H. Trans-kingdom Cross-Talk: Small RNAs on the Move. *PLOS Genet*. 2014;10(9):e1004602.
- Lacroix B, Citovsky V. A mutation in negative regulator of basal resistance WRKY17 of *Arabidopsis* increases susceptibility to *Agrobacterium*-mediated genetic transformation. *F1000Research*. 2013;2.
- Liu Y, Cui Y, Mukherjee A, Chatterjee AK. Characterization of a novel RNA regulator of *Erwinia carotovora* ssp. *carotovora* that controls production of extracellular enzymes and secondary metabolites. *Mol Microbiol*. 1998;29(1):219–34.
- Lorenz R, Bernhart SH, Höner zu Siederdissen C, Tafer H, Flamm C, Stadler PF, et al., ViennaRNA Package 2.0. *Algorithms Mol Biol*. 2011;6:26.
- Lu C, Chen J, Zhang Y, Hu Q, Su W, Kuang H. Miniature Inverted–Repeat Transposable Elements (MITEs) Have Been Accumulated through Amplification Bursts and Play Important Roles in Gene Expression and Species Diversity in *Oryza sativa*. *Mol Biol Evol*. 2012;29(3):1005–17.
- Luchetti A. terMITEs: miniature inverted-repeat transposable elements (MITEs) in the termite genome (Blattodea: Termitoidae). *Mol Genet Genomics*. 2015;290(4):1499–509.

Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. *Mol Biol Evol.* 2015;32(1):268–74.

Nishimura A, Ito M, Kamiya N, Sato Y, Matsuoka M. OsPNH1 regulates leaf development and maintenance of the shoot apical meristem in rice. *Plant J.* 2002;30(2):189–201.

Nouroz F, Noreen S, Heslop-Harrison JS. Evolutionary genomics of miniature inverted-repeat transposable elements (MITEs) in Brassica. *Mol Genet Genomics.* 2015;290(6):2297–312.

Oki N, Yano K, Okumoto Y, Tsukiyama T, Teraishi M, Tanisaka T. A genome-wide view of miniature inverted-repeat transposable elements (MITEs) in rice, *Oryza sativa* ssp. japonica. *Genes Genet Syst.* 2008;83(4):321–9.

Paolo DD, Afanar O, Armitage JP, Berry RM. Single-molecule imaging of electroporated dye-labelled CheY in live *Escherichia coli*. *Phil Trans R Soc B.* 2016;371(1707):20150492.

Pečenková T, Hála M, Kulich I, Kocourková D, Drdová E, Fendrych M, et al., The role for the exocyst complex subunits Exo70B2 and Exo70H1 in the plant–pathogen interaction. *J Exp Bot.* 2011;62(6): 2107-2116.

Piriyapongsa J, Jordan IK. Dual coding of siRNAs and miRNAs by plant transposable elements. *RNA.* 2008;14(5):814–21.

Qi D, Innes RW. Recent Advances in Plant NLR Structure, Function, Localization, and Signaling. *Front Immunol.* 2013;4.

Quibod IL, Perez-Quintero A, Booher NJ, Dossa GS, Grande G, Szurek B, et al., Effector Diversification Contributes to *Xanthomonas oryzae* pv. *oryzae* Phenotypic Adaptation in a Semi-Isolated Environment. *Sci Rep.* 2016;6:34137.

Schmidtke C, Abendroth U, Brock J, Serrania J, Becker A, Bonas U. Small RNA sX13: A Multifaceted Regulator of Virulence in the Plant Pathogen *Xanthomonas*. *PLOS Pathog.* 2013;9(9):e1003626.

Seo Y-S, Sriariyanun M, Wang L, Pfeiff J, Phetsom J, Lin Y, et al., A two-genome microarray for the rice pathogens *Xanthomonas oryzae* pv. *oryzae* and *X. oryzae* pv. *oryzicola* and its use in the discovery of a difference in their regulation of hrp genes. *BMC Microbiol.* 2008;8:99.

Siddique A, Buisine N, Chalmers R. The Transposon-Like Correia Elements Encode Numerous Strong Promoters and Provide a Potential New Mechanism for Phase Variation in the *Meningococcus*. *PLOS Genet.* 2011;7(1):e1001277.

Teixeira MA, Wei L, Kaloshian I. Root-knot nematodes induce pattern-triggered immunity in *Arabidopsis thaliana* roots. *New Phytol.* 2016;211(1):276–87.

The 3,000 rice genomes project. The 3,000 rice genomes project. *GigaScience.* 2014;3(1):7.

Tsuge S, Furutani A, Fukunaka R, Oku T, Tsuno K, Ochiai H, et al., Expression of *Xanthomonas oryzae* pv. *oryzae* hrp Genes in XOM2, a Novel Synthetic Medium. *J Gen Plant Pathol.* 2002;68(4):363–71.

Underwood W. The plant cell wall: a dynamic barrier against pathogen invasion. *Plant Physiol.* 2012;3:85.

Vaschetto LM. Miniature Inverted-repeat Transposable Elements (MITEs) and their effects on the regulation of major genes in cereal grass genomes. *Mol Breed.* 2016;36(3):30.

Weiberg A, Wang M, Bellinger M, Jin H. Small RNAs: A New Paradigm in Plant-Microbe Interactions. *Annu Rev Phytopathol.* 2014;52(1):495–516.

Weiberg A, Wang M, Lin F-M, Zhao H, Zhang Z, Kaloshian I, et al., Fungal Small RNAs Suppress Plant Immunity by Hijacking Host RNA Interference Pathways. *Science.* 2013;342(6154):118–23.

Yaakov B, Ben-David S, Kashkush K. Genome-Wide Analysis of Stowaway-Like MITEs in Wheat Reveals High Sequence Conservation, Gene Association, and Genomic Diversification. *Plant Physiol.* 2013;161(1):486–96.

Yuan X, Zhang S, Qing X, Sun M, Liu S, Su H, et al., Superfamily of ankyrin repeat proteins in tomato. *Gene.* 2013;523(2):126–36.

Zeng Q, McNally RR, Sundin GW. Global Small RNA Chaperone Hfq and Regulatory Small RNAs Are Important Virulence Regulators in *Erwinia amylovora*. *J Bacteriol.* 2013 Apr 15;195(8):1706–17.

Zhang F, Du Z, Huang L, Cruz CV, Zhou Y, Li Z. Comparative Transcriptome Profiling Reveals Different Expression Patterns in *Xanthomonas oryzae* pv. *oryzae* Strains with Putative Virulence-Relevant Genes. *PLOS ONE.* 2013;8(5):e64267.

Zhang Y, Jiang W, Gao L. Evolution of MicroRNA Genes in *Oryza sativa* and *Arabidopsis thaliana*: An Update of the Inverted Duplication Model. *PLoS ONE*;6(12).

5 CONCLUSÃO

Investigamos a influência de elementos genéticos móveis em linhagens de *X. oryzae* sob três diferentes aspectos nesta tese. Primeiramente, investigamos um evento particular de transferência horizontal adquirido exclusivamente pelo patovar *Xoc*, porém ausente em *Xoo* e linhagens filogeneticamente próximas (Capítulo 2). A presença de elementos IS na vizinhança do *operon* de biossíntese de giberelina A₄ denota mobilidade desta característica singular à *Xoc*. A importância da produção do fitormônio GA₄ em *Xoc* é indicada pela alta prevalência do *operon* em diferentes populações, *i.e.* isolados coletados de plantas com sintomas de estriamento foliar bacteriano (BLS) de ampla distribuição geográfica e por intervalo de tempo de mais de quatro décadas.

De uma maneira mais ampla, analisamos o impacto potencial de elementos de transposição inseridos na vizinhança de genes do sistema fundamental do metabolismo bacteriano, o sistema de reparo de DNA (Capítulo 3). Diferentes cepas de *Xoo* e *Xoc* apresentam quantidades e perfis de elementos de transposição, representados por elementos do tipo IS, bastante distintos. Como esperado, observamos que o aumento do número de elementos, como identificado em *Xoo*, resulta no aumento no número de genes potencialmente influenciados pelos elementos. No entanto, a comparação entre cepas do mesmo patovar *Xoo*, mostrou que a influência de elementos é multifacetada: detectamos eventos de proximidade compartilhados por um patovar, eventos compartilhados entre populações de certa região geográfica, e ainda, eventos que são exclusivos à uma cepa, neste momento desprovidos de padrão.

Por fim, descrevemos um novo elemento encontrado em *X. oryzae* e analisamos sua organização, distribuição, filogenia, assim como sua atividade transcricional em *X. oryzae* e o potencial funcional na interação com seu hospedeiro (Capítulo 4).

Assim, esta tese elabora um arcabouço de múltiplas linhas de investigação e provê fundamentação para diversas avaliações experimentais futuras. A aquisição do *operon* de síntese de GA₄ é relevante e prevalente em *Xoc*, porém é necessário entender de forma mais precisa qual etapa molecular do processo de colonização por *Xoc* está sendo influenciado por esse hormônio. Ainda, considerando que pequenos incrementos em *fitness* observados experimentalmente representam grandes consequências em escala evolutiva, avaliar o impacto em *fitness* desta característica é fundamental para provar sua função adaptativa em *Xoc*. Uma pergunta intrigante que permanece neste tema é: por quê o patovar *Xoo* não adquiriu este *operon*, uma vez que pode ocorrer fluxo gênico entre esses dois patovares que co-habitam o mesmo hospedeiro e coexistem em escala biogeográfica?

A amplificação dos elementos IS na espécie *X. oryzae* não se associa à mudança no repertório de genes do sistema de reparo de DNA em relação às outras espécies de *Xanthomonas*. Das poucas diferenças encontradas, será interessante averiguar porque *X.*

oryzae é a única espécie dentro do gênero que não contém mecanismo de NHEJ de reparo haja visto a ausência do gene *ku*.

Dentre os genes que estão potencialmente sob influência de um IS, a detecção de seus níveis de expressão nas cepas que apresentam eventos de proximidade é um primeiro método de detecção de possíveis diferenças transcricionais influenciadas pelo elemento. Comum a todas as cepas de *X. oryzae*, *radA*, *mutM* e *mutS* representam três genes que são potencialmente influenciados pela presença de diversos ISs a montante e a jusante. Um experimento de evolução experimental relevante envolve a criação da linhagem mutante para estes genes e observação do número de elementos ao longo de gerações. Também nos questionamos se os elementos que circundam esse genes, são (ou apresentam indícios que foram) capazes de formar transposons compostos, conferindo mobilidade a estes genes de reparo.

A descrição de MITE-Xo gerou diversos questionamentos que suportam uma avaliação experimental. É indispensável investigar se MITEs-Xo codificam elementos regulatórios e influenciam a vizinhança gênica de suas diversas inserções no genoma de *X. oryzae*. Também abordamos aqui, em caráter especulativo, o processamento de MITEs-Xo em pequenos RNAs e sua atuação na interação com o hospedeiro. A pesquisa de pequenos RNA, ambos intracelular e inter-sistemas ocorre hoje através da análise de expressão diferencial, que apesar da produção de um altíssimo volume de dados, proporcionalmente gera uma quantidade bem menor de alvos testáveis. Será muito interessante entender se a partir de análise bioinformática anterior à experimental, como a descrita neste trabalho, é capaz de desvendar alvos de significância biológica.

Apesar de sua natureza autônoma (muitas vezes referida como egoística), deixamos evidente com este trabalho que a dinâmica da estrutura populacional pode ser apreciada pelo ângulo dos elementos de transposição. A comparação dos fenômenos observados entre linhagens africanas e linhagens asiáticas é recorrente por toda a tese. Em *Xoo*, a linhagem africana AXO1947 é muito distinta das cepas asiáticas na quantidade de elementos IS presentes e nos eventos de proximidade de ISs e genes de reparo. E somente um MITE-Xo nesta cepa é compartilhado com o restante de *Xoo* ou *Xoc*.

Em *Xoc*, o perfil de distribuição de MITEs-Xo nas linhagens africanas revelam uma maior heterogeneidade, com diversos grupos sintênicos exclusivos às linhagens africanas. Os eventos de proximidade de ISs a genes de reparo são mais frequentes nas linhagens *Xoc* africanas e, há ainda alguns casos o elemento próximo ao gene de reparo difere no grupo asiático e africano. O quanto a recente intensificação da agricultura moderna e a diversificação do hospedeiro (utilização de *O. glaberrima* e variedades híbridas de *O. sativa* e *O. glaberrima* na produção de arroz, particular ao continente africano) reconfigura a estrutura populacional do patógeno permanece sob investigação. Nosso trabalho sugere

que elementos de transposição, que podem atuar neste processo, também auxiliam na investigação da dinâmica de populações bacterianas.

O conceito de espécie, assim como o processo de especiação, são temas polêmicos no grupo Bacteria (e, na verdade, em qualquer um dos Domínios da vida) e seria insensato atribuir aqui um valor aos elementos de transposição neste processo em *X. oryzae*. No entanto, nosso conjunto de resultados reitera a importância dos elementos de transposição na variabilidade genética de populações, e dessa forma, infere sobre seu papel relevante na diversificação dessa linhagem.

A contribuição dos elementos móveis na diversificação de uma população clonal tem sido progressivamente explorada, principalmente devido o destaque deste tipo de componente genômico no surgimento de novos patógenos. Neste trabalho, descrevemos essa fonte de variabilidade tanto entre os patovares de *X. oryzae* como dentro de cada patovar. O resultado aqui apresentado é um avanço na caracterização da influência de elementos móveis na diversificação de genomas, uma vez que explora as consequências geradas pelos elementos de transposição e evidencia a complexidade desta variação em nível populacional.

APÊNDICES

A - Submissão do artigo referente aos dados de biogeografia do operon de biossíntese de GA₄

**An operon for production of bioactive gibberellin A4
phytohormone with wide distribution in the bacterial rice
leaf streak pathogen *Xanthomonas oryzae* pv. *oryzicola***

Journal:	<i>New Phytologist</i>
Manuscript ID	Draft
Manuscript Type:	MS - Regular Manuscript
Date Submitted by the Author:	n/a
Complete List of Authors:	Nagel, Raimund; Iowa State University, Biochemistry, Biophysics, & Molecular Biology Turrini, Paula; Universidade de São Paulo, Departamento de Botânica Nett, Ryan; Iowa State University, Biochemistry, Biophysics, & Molecular Biology Leach, Jan; Colorado State University, Bioagricultural Sciences and Pest Management; Verdier, Valerie; Institut de Recherche pour le Développement, UMR Résistance des Plantes aux Bioagresseurs Van Sluys, Marie-Anne; Universidade de São Paulo, GaTE Lab Departamento de Botânica Peters, Reuben; Iowa State University, Biochemistry, Biophysics, & Molecular Biology
Key Words:	Gibberellin biosynthesis, virulence factor, plant-microbe interactions, biogeographical distribution, cytochromes P450

SCHOLARONE™
Manuscripts

Manuscript submitted to New Phytologist for review

B - Géis de eletroforese das amplificações dos fragmentos que compõem o *operon* GA₄ em isolados de *X. oryzae*

B - Continuação

B - Continuação

B - Continuação

C - Isolados de *X. oryzae* utilizados para verificação da presença do operon GA₄

Isolados de <i>Xanthomonas</i>	Ano de Coleta	País	Cidade	Província	Hospedeiro	Fragmentos do operon GA ₄			
						1	2	3	4
<i>oryzae</i> pv. <i>oryzicola</i>									
MAI_03	2003	Mali	O. du Niger	Niono	<i>O. sativa</i>				
MAI_04	-	Mali	-	-	-				
MAI_05		Mali							
MAI_06	-	Mali	-	-	-				
MAI_07	2003	Mali	O. du Niger	Niono	<i>O. sativa</i> Jamajigi				
MAI_10 ^a	2003	Mali	O. du Niger	Niono	<i>O.sativa</i>				
MAI_11	2003	Mali	O. du Niger	Niono	<i>O.sativa</i>				
MAI_16	2009	Mali	O. du Niger	Molodo	<i>O.sativa</i>				
MAI_18	2009	Mali	O. du Niger	Kogoni	<i>O.longistaminata</i>				
MAI_19	2009	Mali	O. du Niger	Kogoni	<i>O.longistaminata</i>				
MAI_20	2009	Mali	O. du Niger	Kogoni	<i>O.longistaminata</i>				
MAI_21	2009	Mali	O. du Niger	Kogoni	<i>O.longistaminata</i>				
MAI_22	2009	Mali	O. du Niger	Kogoni	<i>O.barthii</i>				
MAI_23	2009	Mali	O. du Niger	Kogoni	<i>O.barthii</i>				
MAI_24	2009	Mali	O. du Niger	Kogoni	<i>O.barthii</i>				
MAI_25	2009	Mali	O. du Niger	Kogoni	<i>O.longistaminata</i>				
MAI_26	2009	Mali	O. du Niger	Macina	<i>O.sativa</i>				
MAI_28	2009	Mali	O. du Niger	Macina	<i>O.sativa</i>				
MAI_29	2009	Mali	O. du Niger	Macina	<i>O.sativa</i>				
MAI_30	2009	Mali	O. du Niger	Macina	<i>O.sativa</i>				
MAI_31	2009	Mali	O. du Niger	Macina	<i>O.sativa</i>				
MAI_32	2009	Mali	O. du Niger	Macina	<i>O.sativa</i>				
MAI_33	2009	Mali	O. du Niger	Macina	<i>O.sativa</i>				
MAI_34	2009	Mali	O. du Niger	Macina	<i>O.sativa</i>				
MAI_35	2009	Mali	O. du Niger	Macina	<i>O.sativa</i>				
MAI_36	2009	Mali	O. du Niger	Ndebougou	<i>O.barthii</i>				
MAI_37	2009	Mali	O. du Niger	Niono	<i>O.sativa</i>				
MAI_38	2009	Mali	Koulikoro	Dialakoroba	<i>O.sativa</i>				
MAI_39	2009	Mali	Sikasso	Sélingué	<i>O.sativa</i>				
MAI_40	2009	Mali	Sikasso	Sélingué	<i>O.sativa</i>				
MAI_41	2009	Mali	Sikasso	Sélingué	<i>O.sativa</i>				
MAI_42	2009	Mali	Sikasso	Sélingué	<i>O.sativa</i>				
MAI_43	2009	Mali	Sikasso	Sélingué	<i>O.sativa</i>				
MAI_44	2009	Mali	Sikasso	Sélingué	<i>O.sativa</i>				
MAI_46	2009	Mali	Koulikoro	Baguineda	<i>O.sativa</i>				
BAI_05 ^a	2009	B. Faso	HautsBassins	Vallée	<i>O.sativa</i>				
BAI_06	2009	B. Faso	HautsBassins	Vallée	<i>O.sativa</i>				
BAI_07	2009	B. Faso	EastCenter	Itenga	<i>O.longistaminata</i>				
BAI_08	2009	B. Faso	EastCenter	Itenga	<i>O.longistaminata</i>				
BAI_09	2009	B. Faso	EastCenter	Itenga	<i>O.longistaminata</i>				
BAI_10	2009	B. Faso	EastCenter	Itenga	<i>O.longistaminata</i>				
BAI_11 ^a	2009	B. Faso	EastCenter	Itenga	Wild rice				
BAI_12	2009	B. Faso	EastCenter	Bagré	<i>O.sativa</i>				
BAI_13	2009	B. Faso	EastCenter	Bagré	Weeds				
BAI_14	2009	B. Faso	EastCenter	Bagré	<i>O.sativa</i>				

C - Continuação

Isolados de <i>Xanthomonas</i>	Ano de Coleta	País	Cidade	Província	Hospedeiro	Fragmentos do operon GA ₄			
						1	2	3	4
<i>oryzae</i> pv. <i>oryzicola</i>									
BAI_15	2009	B. Faso	EastCenter	Bagré	<i>O.sativa</i>				
BAI_16	2009	B. Faso	EastCenter	Bagré	<i>O.sativa</i>				
BAI_17	2009	B. Faso	EastCenter	Bagré	<i>Pennisetum</i> sp.				
BAI_18	2009	B. Faso	EastCenter	Koulwooko	<i>O.sativa</i>				
BAI_19	2009	B. Faso	Cascades	Karfiguela	<i>O.sativa</i>				
BAI_20	2009	B. Faso	Cascades	Karfiguela	Weeds				
BAI_21	2009	B. Faso	Cascades	Karfiguela	<i>O.sativa</i>				
BAI_22	2009	B. Faso	Cascades	Karfiguela	<i>O.longistaminata</i>				
BAI_34	2009	B. Faso	Cascades	Karfiguela	<i>O.sativa</i>				
BAI_35	2010	B. Faso	Center	Nabadougou	<i>O.longistaminata</i>				
BAI_36	2010	B. Faso	B.Mouhoun	Niassan	<i>O.sativa</i>				
BAI_37	2010	B. Faso	EastCenter	Bagré	<i>O.sativa</i>				
BAI_38	2009	B. Faso	Cascades	Karfiguela	<i>O.sativa</i>				
BAI_39	2009	B. Faso	Cascades	Karfiguela	<i>O.sativa</i>				
BAI_40	2010	B. Faso	EastCenter	Bagré	<i>O.sativa</i>				
BAI_41	2010	B. Faso	Center	Nabadougou	<i>O.longistaminata</i>				
BAI_42	2010	B. Faso	EastCenter	Bagré	<i>O.sativa</i>				
BAI_43	2010	B. Faso	HautsBassins	Nienadonkele	<i>O.sativa</i>				
BAI_44	2010	B. Faso	HautsBassins	Nienadonkele	<i>O.sativa</i>				
BAI_49	2011	B. Faso	B. Mouhon	Niassan	<i>O. longistaminata</i>				
BAI_60	2011	B. Faso	B. Mouhon	Niassan	<i>O. sativa</i>				
BAI_82	2011	B. Faso	Cascades	Karfiguela	<i>O. sativa</i>				
BLS_001	1978	Philippines	Laguna	College	-				
BLS_030	1978	Philippines	Laguna	College	-				
BLS_046	1980	Philippines	Nueva Ecija	MRRTC	Peta				
BLS_098	1982	Philippines	Nueva Ecija	Talavera	IR52				
BLS_099	1982	Philippines	Bacolod	Sum-Ag	-				
BLS_102	1982	Philippines	Benguet	Baguio	-				
BLS_103	1982	Philippines	Benguet	Baguio	-				
BLS_105	1982	Philippines	Laguna	College	-				
BLS_106	1982	Philippines	Camarines Sur		-				
BLS_123	1983	Philippines	Laguna	College	-				
BLS_125	1983	Philippines	Laguna	College	-				
BLS_150	1983	Philippines	Laguna	College	-				
BLS_159	1983	Philippines	Laguna	College	-				
BLS_175	1983	Philippines	North Cotabato		-				
BLS_176	1983	Philippines	North Cotabato		-				
BLS_222	1983	Philippines	Laguna	College	-				
BLS_256 ^a	1984	Philippines	IRRI-IRTP	Los Baños	IR29692-99-3-2-1				
BLS_276	1984	Philippines	Laguna	College	-				
BLS_279 ^a	1984	Philippines	Laguna	College	-				
BLS_280	1984	Philippines	Laguna	College	-				
BLS_281	1984	Philippines	IRRI	Los Baños	-				
BLS_285	1984	Philippines	Laguna	College	-				
BLS_289	1984	Philippines	Camarines Norte ^c		-				
BLS_290	1984	Philippines	Leyte ^c		-				
BLS_291	1985	Philippines	Iloilo ^c		-				
BLS_294	1984	Philippines	Nueva Ecija ^c		-				

C - Continuação

Isolados de <i>Xanthomonas</i>	Ano de Coleta	País	Cidade	Província	Hospedeiro	Fragmentos do operon GA ₄			
						1	2	3	4
<i>oryzae</i> pv. <i>oryzicola</i>									
BLS_295	1984	Philippines	North Cotabato ^c		-				
BLS_303	1985	Philippines	Iloilo	San Dionisio	C732-14(upland)				
BLS_305	1985	Philippines	Iloilo	San Dionisio	C894-21				
BLS_348	1988	Philippines	Cagayan	Alilino	-				
BLS_354 ^b	1983	Philippines	Isabela	Cabatuan	-				
BLS_356	1988	Philippines	Isabela	San Fabian	IR68				
BLS_357	1988	Philippines	Isabela	San Manuel	-				
BLS_365	1988	Philippines	Isabela	Burgos					
BLS_372	1989	Philippines	Nueva Ecija	CLSU, Muñoz	C4				
BLS_377	1989	Philippines	Nueva Ecija	Guimba	-				
BLS_391	1989	Philippines	Nueva Ecija	Bagsak, Calinan					
BLS_413	1990	Philippines	Tarlac	La Paz	O. sativa IR72				
BLS_415	1990	Philippines	Tarlac	La Paz	O. sativa IR72				
BLS_417	1990	Philippines	Tarlac	La Paz	O. sativa IR72				
BLS_420	1990	Philippines	Tarlac	La Paz	O. sativa IR72				
RS105 ^a	1992	China							
B8-12 ^a	2007	China							
L8 ^a	< 1995	China							
Xoc China	1990	China	-	-	-				
CFBP2286 ^a	1964	Malaysia	na	na	<i>O. sativa</i>				
CFBP2287	-	Malaysia	-	-	-				
UPB497		Malaysia	-	-	-				
BXOR1 ^a	1996	India							
<i>oryzae</i> pv. <i>oryzae</i>									
BAI3	2003	Burkina	East center	Bagre					
MAFF311018	na	Japan							
37012	2003	Mali		Niono					
PXO113	-	Philippines							
Thai Xoo12	-	Thailand							
KACC10331 ^a		Korea							
PXO99A ^a	1980	Philippines							
PXO86 ^a	1977	Philippines							
PXO282 ^a	1990	Philippines							
PXO83 ^a	1976	Philippines	Nueva Ecija						
AXO1947 ^a	2015	Cameroon							
PXO602 ^a	2006	Philippines							
PXO71 ^a	1974	Philippines							
PXO236 ^a	1989	Philippines							
PXO145 ^a	1982	Philippines							
PXO211 ^a	1989	Philippines							
PXO524 ^a	1994	Philippines							
PXO563 ^a	1998	Philippines							
<i>campestris</i> pv. <i>leersiae</i>									
BAI23	2009	Burkina	Cascades	Karfiguela	Weeds				
<i>oryzae</i>									
US-X8-1A		USA	Lousiana		Rice				

^a Genoma publicado.^b Genoma sequenciado (comunicação pessoal com A. Bogdanove, Cornell University, EUA)^c Cidade de coleta desconhecida.

D – Lista de genes de reparo utilizados, de acordo com banco de dados KEGG.

Categoria KEGG			KEGG #	Nome do gene	Organismo	Gene referência		
						ID do gene	Tamanho (pb)	Tamanho (aa)
SSBR	MMR	DNA exonucleases	K03601	<i>xseA</i>	xax	XACM_2415	1338	446
SSBR	MMR	DNA exonucleases	K01141	<i>sbcB</i>	xax	XACM_1570	1446	482
SSBR	MMR	DNA exonucleases	K03602	<i>xseB</i>	xax	XACM_2702	261	87
SSBR	MMR	DNA polymerase III	K02337	<i>dnaE</i>	xax	XACM_1394	3591	1197
SSBR	MMR	DNA polymerase III	K02339	DPO3C,holC	xax	XACM_3452	426	142
SSBR	MMR	DNA polymerase III	K02340	<i>holA</i>	xax	XACM_2720	1029	343
SSBR	MMR	DNA polymerase III	K02338	<i>dnaN</i>	xax	XACM_0002	1101	367
SSBR	MMR	DNA polymerase III	K02341	<i>holB</i>	xax	XACM_1097	960	320
SSBR	MMR	DNA polymerase III	K02343	<i>dnaX</i>	xax	XACM_1076	2052	684
SSBR	MMR	DNA polymerase III	K02342	<i>dnaQ</i>	xax	XACM_1048	735	245
SSBR	MMR	Rrecognition factors	K03555	<i>mutS</i>	xax	XACM_1279	2622	874
SSBR	MMR	Molecular matchmaker	K03572	<i>mutL</i>	xax	XACM_2410	1878	626
SSBR	MMR	Other MMR factors	K06223	<i>dam</i>	xal	XALc_0178	849	283
SSBR	MMR	Other MMR factors	K03111	<i>ssb</i>	xax	XACM_2837	534	178
SSBR	MMR	Other MMR factors	K07458	<i>vsr</i>	psu	Psesu_1607	408	136
SSBR	NER	Global genome repair factors	K03702	<i>uvrB</i>	xax	XACM_2600	2022	674
SSBR	NER	TCR / RNAP	K03060	<i>rpoZ</i>	xax	XACM_3282	300	100
SSBR	NER	Global genome repair factors	K03657	<i>uvrD,pcrA</i>	xao	XAC29_20895	2187	729
SSBR	NER	TCR / RNAP	K03040	<i>rpoA</i>	xax	XACM_0977	999	333
SSBR	NER	Global genome repair factors	K03703	<i>uvrC</i>	xax	XACM_2200	1860	620
SSBR	NER	TCR / RNAP	K03043	<i>rpoB</i>	xax	XACM_0944	4152	1384
SSBR	NER	TCR / RNAP	K03046	<i>rpoC</i>	xax	XACM_0945	4215	1405
SSBR	NER	Global genome repair factors	K05984	<i>cho</i>	xac	XAC2558	855	285
SSBR	NER	Global genome repair factors	K03701	<i>uvrA</i>	xcc	XCC1148	2967	989
SSBR	NER	TCR / TRCF	K03723	<i>mfd</i>	xax	XACM_2804	3465	1155
SSBR	BER	DNA glycosylases	K10563	<i>mutM,fpg</i>	xcc	XCC4147	816	272
SSBR	BER	DNA glycosylases	K03652	<i>MPG</i>	xax	XACM_2428	621	207
SSBR	BER	RecJ	K07462	<i>recJ</i>	xax	XACM_1893	1740	580
SSBR	BER	DNA glycosylases	K05522	<i>nei</i>	xax	XACM_0609	753	251
SSBR	BER	DNA glycosylases	K13529	Ada-alkA	xax	XACM_2766	1464	488
SSBR	BER	DNA polymerase I	K02335	DPO1, polA	xax	XACM_3977	2802	934
SSBR	BER	DNA glycosylases	K03648	UNG,UDG	xax	XACM_3726	726	242
SSBR	BER	DNA glycosylases	K03575	<i>mutY</i>	xax	XACM_2534	1074	358
SSBR	BER	DNA glycosylases	K01246	<i>tag</i>	xax	XACM_2273	582	194
SSBR	BER	DNA glycosylases	K10773	NTH	xax	XACM_1552	693	231
SSBR	BER	AP endonucleases	K05982	<i>nfi</i>	xax	XACM_2816	714	238
SSBR	BER	DNA glycosylases	K01247	<i>alkA</i>	xax	XACM_1207	675	225
SSBR	-	Direct Repair	K00567	MGMT,ogt	xax	XACM_2767	543	181
SSBR	-	Direct Repair	K01669	<i>phrB</i>	xax	XACM_1467	1428	476
SSBR	-	Direct Repair	K03919	<i>alkB</i>	rhd	R2APBS1_1797	657	219
SSBR	BER	AP endocutlease	K01142	<i>xthA1</i>	xac	XAC4171	780	260
SSBR	BER	AP endocutlease	K01142	<i>xthA2</i>	xac	XAC2030	993	331
SSBR	BER	AP endocutlease	K01142	<i>exoA</i>	xac	XAC3902	801	267

D - Continuação

Categoria KEGG			KEGG #	Nome do gene	Organismo	Gene referência		
						ID do gene	Tamanho (pb)	Tamanho (aa)
DSBR	HR	RecFOR pathway	K04485	<i>sms,radA</i>	xax	XACM_1239	1392	464
DSBR	HR	RecBC pathway	K04066	<i>priA</i>	xax	XACM_3719	2223	741
DSBR	HR	RecFOR pathway	K03655	<i>recG</i>	xax	XACM_3279	2154	718
DSBR	HR	RecFOR pathway	K03584	<i>recO</i>	xax	XACM_1310	723	241
DSBR	HR	RecBC pathway	K03582	<i>recB</i>	xax	XACM_4219	4053	1351
DSBR	HR	RecBC pathway	K03583	<i>recC</i>	xax	XACM_4220	3414	1138
DSBR	HR	RecBC pathway	K03581	<i>recD</i>	xax	XACM_4218	2376	792
DSBR	HR	RecBC pathway	K03550	<i>ruvA</i>	xax	XACM_3069	453	151
DSBR	HR	RecBC pathway	K03551	<i>ruvB</i>	xax	XACM_3066	1041	347
DSBR	HR	RecBC pathway	K01159	<i>ruvC</i>	xax	XACM_3070	525	175
DSBR	HR	RecFOR pathway	K06187	<i>recR</i>	xax	XACM_1078	594	198
DSBR	HR	RecFOR pathway	K03631	<i>recN</i>	xax	XACM_1488	1665	555
DSBR	HR	RecFOR pathway	K03554	<i>rdgC</i>	xax	XACM_4097	906	302
DSBR	HR	RecBC pathway	K03553	<i>recA</i>	xax	XACM_1761	1035	345
DSBR	HR	RecFOR pathway	K03629	<i>recF</i>	xax	XACM_0004	1107	369
STL	-	Y-family DNA polymerases	K02346	<i>dinB</i>	xfu	XFF4834R_chr35190	1080	360
STL	-	Y-family DNA polymerases	K03502	<i>umuC</i>	xcc	XCC2964	1287	429
STL	-	Y-family DNA polymerases	K03503	<i>umuD</i>	xcc	XCC2963	480	160
STL	-	Other SOS response factors	K01356	<i>lexA</i>	xax	XACM_1166	606	202
STL	-	Other SOS response factors	K01357	<i>lexA2</i>	xax	XACM_1760	642	214
STL	-	Other SOS response factors	K14160	<i>imuA</i>	rhd	R2APBS1_1873	666	222
NHEJ	-	Two-component complex	K10979	<i>ku</i>	xax	XACM_0108	960	320
NHEJ	-	Two-component complex	K01971	<i>ligD</i>	xac	XAC2414	2619	873
SHDR	-	Facilitator	K03557	<i>fis</i>	xax	XACM_0514	273	91
SHDR	-	Facilitator	K04764	<i>ihfA, himA</i>	xax	XACM_2570	300	100
SHDR	-	Facilitator	K05788	<i>ihfB, himD</i>	xax	XACM_2305	312	104
SHDR	-	Supressor	K03654	<i>recQ</i>	xax	XACM_3044	1797	599
SHDR	-	Supressor	K03746	<i>hns</i>	xax	XACM_3463	462	154
SHDR	-	recET pathway	K01972	<i>ligA, ligB</i>	xax	XACM_1599	2502	834
SHDR	-	recET pathway	K07455	<i>recT</i>	xoz	BE73_14305	915	305
SHIR	-	Facilitator	K02469	<i>gyrA</i>	xax	XACM_1603	2700	900
SHIR	-	Facilitator	K02470	<i>gyrB</i>	xax	XACM_0005	2454	818
SHIR	-	Facilitator	K03168	<i>topA</i>	xax	XACM_3707	2493	831
SHIR	-	Supressor	K03530	<i>hupB</i>	xax	XACM_1039	273	91

[illegible]

[illegible]

G – Proximidade de ISs e genes de reparo de vias acessórias do sistema de reparo de DNA em cepas de *X. oryzae*

REPARO DIRETO

SHIIR

recombinação ilegítima dependente de pequena homologia

SHIDR

recombinação ilegítima independente de pequena homologia

SÍNTESE TRANSLEÇÃO

I - Lista completa das cópias de MITEs-Xo encontradas em *X. oryzae*

MITE ID	Coordenadas começo	fim	Tamanho (bp)	Status	Grupo sintênico
<i>Xanthomonas oryzae</i> pv. <i>oryzicola</i> CFBP2286					
CFBP2286_mite1	16911	16832	80	P	miteXo1-1
CFBP2286_mite2	45359	45148	212	C	miteXo1-21
CFBP2286_mite3	54854	54644	211	C	miteXo1-2
CFBP2286_mite4	87716	87505	212	C	miteXo1-3
CFBP2286_mite5	131770	131981	212	C	miteXo1-4
CFBP2286_mite6	239472	239261	212	C	miteXo1-18
CFBP2286_mite7	353339	353549	211	C	miteXo1-5
CFBP2286_mite8	386842	387053	212	C	miteXo1-6
CFBP2286_mite9	481115	481326	212	C	miteXo1-7
CFBP2286_mite10	1537367	1537578	212	C	miteXo1-8
CFBP2286_mite11	1556470	1556679	210	C	miteXo2-2
CFBP2286_mite12	1744497	1744707	211	C	miteXo1-9
CFBP2286_mite13	1962711	1962551	161	P	miteXo1-19
CFBP2286_mite14	2000051	1999842	210	C	miteXo2-4
CFBP2286_mite15	2297321	2297110	212	C	miteXo1-10
CFBP2286_mite16	2420829	2421040	212	C	miteXo1-11
CFBP2286_mite17	2527942	2528153	212	C	miteXo1-12
CFBP2286_mite18	2689121	2688910	212	C	miteXo1-20
CFBP2286_mite19	2752949	2752886	64	P	miteXo1-22
CFBP2286_mite20	3022907	3022696	212	C	miteXo1-25
CFBP2286_mite21	3155158	3155368	211	C	miteXo1-13
CFBP2286_mite22	3423317	3423528	212	C	miteXo1-23
CFBP2286_mite23	3512922	3513131	210	C	miteXo2-6
CFBP2286_mite24	3840385	3840174	212	C	miteXo1-14
CFBP2286_mite25	3849038	3849248	211	C	miteXo1-15
CFBP2286_mite26	4254143	4253932	212	C	miteXo1-16
CFBP2286_mite27	4288440	4288649	210	C	miteXo2-1
CFBP2286_mite28	4306532	4306322	211	C	miteXo1-24
CFBP2286_mite29	4408095	4407886	210	C	miteXo2-3
CFBP2286_mite30	4555438	4555227	212	C	miteXo1-17
CFBP2286_mite31	4700375	4700166	210	C	miteXo2-5
<i>Xanthomonas oryzae</i> pv. <i>oryzicola</i> BLS256					
BLS256_mite1	16911	16832	80	P	miteXo1-1
BLS256_mite2	45359	45148	212	C	miteXo1-21
BLS256_mite3	54854	54644	211	C	miteXo1-2
BLS256_mite4	87717	87506	212	C	miteXo1-3
BLS256_mite5	131771	131982	212	C	miteXo1-4
BLS256_mite6	236426	236215	212	C	miteXo1-18
BLS256_mite7	350282	350492	211	C	miteXo1-5
BLS256_mite8	383799	384010	212	C	miteXo1-6
BLS256_mite9	485952	486163	212	C	miteXo1-7
BLS256_mite10	1498777	1498988	212	C	miteXo1-8
BLS256_mite11	1517888	1518097	210	C	miteXo2-2
BLS256_mite12	1699080	1699290	211	C	miteXo1-9
BLS256_mite13	1917306	1917146	161	P	miteXo1-19
BLS256_mite14	1954646	1954437	210	C	miteXo2-4
BLS256_mite15	2256371	2256160	212	C	miteXo1-10
BLS256_mite16	2378710	2378921	212	C	miteXo1-11
BLS256_mite17	2485816	2486027	212	C	miteXo1-12
BLS256_mite18	2651196	2650985	212	C	miteXo1-20
BLS256_mite19	2713547	2713484	64	P	miteXo1-22
BLS256_mite20	3074047	3074257	211	C	miteXo1-13
BLS256_mite21	3342224	3342435	212	C	miteXo1-23
BLS256_mite22	3431845	3432054	210	C	miteXo2-6
BLS256_mite23	3718239	3718028	212	C	miteXo1-14
BLS256_mite24	3726892	3727102	211	C	miteXo1-15
BLS256_mite25	4131757	4131546	212	C	miteXo1-16
BLS256_mite26	4166054	4166263	210	C	miteXo2-1
BLS256_mite27	4184155	4183945	211	C	miteXo1-24
BLS256_mite28	4271936	4271727	210	C	miteXo2-3
BLS256_mite29	4419279	4419068	212	C	miteXo1-17
BLS256_mite30	4564220	4564011	210	C	miteXo2-5

I - Continuação

MITE ID	Coordenadas		Tamanho (bp)	Status	Grupo sintênico
	começo	fim			
Xanthomonas oryzae pv. oryzicola L8					
L8_mite1	18071	17992	80	P	miteXo1-1
L8_mite2	26937	26727	211	C	miteXo1-2
L8_mite3	280444	280653	210	C	miteXo2-5
L8_mite4	424592	424803	212	C	miteXo1-17
L8_mite5	572057	572266	210	C	miteXo2-3
L8_mite6	656462	656672	211	C	miteXo1-24
L8_mite7	1515525	1515736	212	C	miteXo1-8
L8_mite8	1535572	1535781	210	C	miteXo2-2
L8_mite9	1722278	1722488	211	C	miteXo1-9
L8_mite10	2127001	2127212	212	C	miteXo1-20
L8_mite11	2294675	2294464	212	C	miteXo1-12
L8_mite12	2395189	2394978	212	C	miteXo1-11
L8_mite13	2512463	2512674	212	C	miteXo1-10
L8_mite14	2819483	2819692	210	C	miteXo2-4
L8_mite15	2856998	2856946	53	P	miteXo1-26
L8_mite16	2860605	2860394	212	C	miteXo1-19
L8_mite17	3103083	3103292	210	C	miteXo1-27
L8_mite18	3133163	3133373	211	C	miteXo1-13
L8_mite19	3477075	3477284	210	C	miteXo2-6
L8_mite20	3775332	3775121	212	C	miteXo1-14
L8_mite21	3783954	3784164	211	C	miteXo1-15
L8_mite22	4180940	4180729	212	C	miteXo1-16
L8_mite23	4222913	4222703	210	C	miteXo2-1
L8_mite24	4367819	4367613	207	C	miteXo1-7
L8_mite25	4464227	4464016	212	C	miteXo1-6
L8_mite26	4497924	4497714	211	C	miteXo1-5
L8_mite27	4611894	4612105	212	C	miteXo1-18
L8_mite28	4710996	4710785	212	C	miteXo1-4
L8_mite29	4751307	4751518	212	C	miteXo1-3
Xanthomonas oryzae pv. oryzicola BLS279					
BLS279_mite1	18071	17992	80	P	miteXo1-1
BLS279_mite2	26937	26727	211	C	miteXo1-2
BLS279_mite3	280494	280703	210	C	miteXo2-5
BLS279_mite4	424644	424855	212	C	miteXo1-17
BLS279_mite5	572085	572294	210	C	miteXo2-3
BLS279_mite6	656483	656693	211	C	miteXo1-24
BLS279_mite7	1515054	1515265	212	C	miteXo1-8
BLS279_mite8	1534157	1534366	210	C	miteXo2-2
BLS279_mite9	1714103	1714313	211	C	miteXo1-9
BLS279_mite10	2118820	2119031	212	C	miteXo1-20
BLS279_mite11	2286470	2286259	212	C	miteXo1-12
BLS279_mite12	2386984	2386773	212	C	miteXo1-11
BLS279_mite13	2504240	2504451	212	C	miteXo1-10
BLS279_mite14	2807306	2807515	210	C	miteXo2-4
BLS279_mite15	2844821	2844769	53	P	miteXo1-26
BLS279_mite16	2848428	2848217	212	C	miteXo1-19
BLS279_mite17	3078788	3078998	211	C	miteXo1-13
BLS279_mite18	3421759	3421968	210	C	miteXo2-6
BLS279_mite19	3760700	3760489	212	C	miteXo1-14
BLS279_mite20	3769322	3769532	211	C	miteXo1-15
BLS279_mite21	4166343	4166132	212	C	miteXo1-16
BLS279_mite22	4208314	4208105	210	C	miteXo2-1
BLS279_mite23	4353228	4353022	207	C	miteXo1-7
BLS279_mite24	4449637	4449426	212	C	miteXo1-6
BLS279_mite25	4483327	4483117	211	C	miteXo1-5
BLS279_mite26	4598096	4598307	212	C	miteXo1-18
BLS279_mite27	4704718	4704507	212	C	miteXo1-4
BLS279_mite28	4745402	4745613	212	C	miteXo1-3

I - Continuação

MITE ID	Coordenadas		Tamanho	Status	Grupo
	começo	fim	(bp)		sintênico
<i>Xanthomonas oryzae</i> pv. <i>oryzicola</i> B8-12					
B8-12_mite1	18068	17989	80	P	miteXo1-1
B8-12_mite2	26934	26724	211	C	miteXo1-2
B8-12_mite3	280440	280649	210	C	miteXo2-5
B8-12_mite4	424589	424800	212	C	miteXo1-17
B8-12_mite5	572046	572255	210	C	miteXo2-3
B8-12_mite6	656452	656662	211	C	miteXo1-24
B8-12_mite7	1516149	1516360	212	C	miteXo1-8
B8-12_mite8	1536196	1536405	210	C	miteXo2-2
B8-12_mite9	1723113	1723323	211	C	miteXo1-9
B8-12_mite10	2129355	2129566	212	C	miteXo1-20
B8-12_mite11	2297035	2296824	212	C	miteXo1-12
B8-12_mite12	2397549	2397338	212	C	miteXo1-11
B8-12_mite13	2514823	2515034	212	C	miteXo1-10
B8-12_mite14	2817623	2817832	210	C	miteXo2-4
B8-12_mite15	2855138	2855086	53	P	miteXo1-26
B8-12_mite16	2858745	2858534	212	C	miteXo1-19
B8-12_mite17	3133525	3133735	211	C	miteXo1-13
B8-12_mite18	3475932	3476141	210	C	miteXo2-6
B8-12_mite19	3774180	3773969	212	C	miteXo1-14
B8-12_mite20	3782802	3783012	211	C	miteXo1-15
B8-12_mite21	4179795	4179584	212	C	miteXo1-16
B8-12_mite22	4221766	4221557	210	C	miteXo2-1
B8-12_mite23	4366673	4366467	207	C	miteXo1-7
B8-12_mite24	4463065	4462854	212	C	miteXo1-6
B8-12_mite25	4496741	4496531	211	C	miteXo1-5
B8-12_mite26	4609659	4609870	212	C	miteXo1-18
B8-12_mite27	4708785	4708574	212	C	miteXo1-4
B8-12_mite28	4749096	4749307	212	C	miteXo1-3
<i>Xanthomonas oryzae</i> pv. <i>oryzicola</i> RS105					
RS105_mite1	18071	17992	80	P	miteXo1-1
RS105_mite2	26937	26727	211	C	miteXo1-2
RS105_mite3	280438	280647	210	C	miteXo2-5
RS105_mite4	424584	424795	212	C	miteXo1-17
RS105_mite5	572041	572250	210	C	miteXo2-3
RS105_mite6	656447	656657	211	C	miteXo1-24
RS105_mite7	1516152	1516363	212	C	miteXo1-8
RS105_mite8	1536199	1536408	210	C	miteXo2-2
RS105_mite9	1708321	1708531	211	C	miteXo1-9
RS105_mite10	2114570	2114781	212	C	miteXo1-20
RS105_mite11	2282246	2282035	212	C	miteXo1-12
RS105_mite12	2382759	2382548	212	C	miteXo1-11
RS105_mite13	2500032	2500243	212	C	miteXo1-10
RS105_mite14	2802850	2803059	210	C	miteXo2-4
RS105_mite15	2840365	2840313	53	P	miteXo1-26
RS105_mite16	2843972	2843761	212	C	miteXo1-19
RS105_mite17	3118753	3118963	211	C	miteXo1-13
RS105_mite18	3461618	3461827	210	C	miteXo2-6
RS105_mite19	3759868	3759657	212	C	miteXo1-14
RS105_mite20	3768490	3768700	211	C	miteXo1-15
RS105_mite21	4165470	4165259	212	C	miteXo1-16
RS105_mite22	4207440	4207231	210	C	miteXo2-1
RS105_mite23	4352349	4352143	207	C	miteXo1-7
RS105_mite24	4448741	4448530	212	C	miteXo1-6
RS105_mite25	4482410	4482200	211	C	miteXo1-5
RS105_mite26	4595327	4595538	212	C	miteXo1-18
RS105_mite27	4694422	4694211	212	C	miteXo1-4
RS105_mite28	4734732	4734943	212	C	miteXo1-3

I - Continuação

MITE ID	Coordenadas		Tamanho (bp)	Status	Grupo sintênico
	começo	fim			
<i>Xanthomonas oryzae</i> pv. <i>oryzicola</i> CFBP7341					
CFBP7341_mite1	16911	16832	80	P	miteXo1-1
CFBP7341_mite2	25773	25562	212	C	miteXo1-2
CFBP7341_mite3	123947	124158	212	C	miteXo1-4
CFBP7341_mite4	344763	344567	197	P	miteXo2-8
CFBP7341_mite5	400223	400433	211	C	miteXo1-5
CFBP7341_mite6	433943	434154	212	C	miteXo1-6
CFBP7341_mite7	525123	525334	212	C	miteXo1-7
CFBP7341_mite8	659267	659476	210	C	miteXo2-3
CFBP7341_mite9	736887	736678	210	C	miteXo2-1
CFBP7341_mite10	772221	772430	210	C	miteXo1-16
CFBP7341_mite11	873755	873544	212	C	miteXo1-28
CFBP7341_mite12	1177892	1177682	211	C	miteXo1-15
CFBP7341_mite13	1186506	1186717	212	C	miteXo1-14
CFBP7341_mite14	1406867	1406790	78	P	miteXo1-32
CFBP7341_mite15	1819986	1819776	211	C	miteXo1-13
CFBP7341_mite16	2297604	2297815	212	C	miteXo1-33
CFBP7341_mite17	2409548	2409337	212	C	miteXo1-12
CFBP7341_mite18	2512774	2512563	212	C	miteXo1-11
CFBP7341_mite19	2568197	2568336	140	P	miteXo1-30
CFBP7341_mite20	2647687	2647898	212	C	miteXo1-10
CFBP7341_mite21	2956742	2956951	210	C	miteXo2-4
CFBP7341_mite22	3026452	3026661	210	C	miteXo2-9
CFBP7341_mite23	3027432	3027221	212	C	miteXo1-19
CFBP7341_mite24	3255998	3255788	211	C	miteXo1-9
CFBP7341_mite25	3454541	3454330	212	C	miteXo1-8
CFBP7341_mite26	3868814	3869025	212	C	miteXo1-31
CFBP7341_mite27	3917031	3917242	212	C	miteXo1-29
CFBP7341_mite28	4724431	4724222	210	C	miteXo2-5
CFBP7341_mite29	4812251	4812462	212	C	miteXo1-34
CFBP7341_mite30	4815962	4815752	211	C	miteXo1-35
CFBP7341_mite31	4958980	4959151	172	P	miteXo1-3
<i>Xanthomonas oryzae</i> pv. <i>oryzicola</i> CFBP7331					
CFBP7331_mite1	16911	16832	80	P	miteXo1-1
CFBP7331_mite2	25773	25562	212	C	miteXo1-2
CFBP7331_mite3	122737	122948	212	C	miteXo1-4
CFBP7331_mite4	331661	331465	197	P	miteXo2-8
CFBP7331_mite5	387121	387331	211	C	miteXo1-5
CFBP7331_mite6	421672	421883	212	C	miteXo1-6
CFBP7331_mite7	514056	514267	212	C	miteXo1-7
CFBP7331_mite8	775270	775061	210	C	miteXo2-1
CFBP7331_mite9	810604	810813	210	C	miteXo1-16
CFBP7331_mite10	911086	910875	212	C	miteXo1-28
CFBP7331_mite11	1216452	1216242	211	C	miteXo1-15
CFBP7331_mite12	1225066	1225277	212	C	miteXo1-14
CFBP7331_mite13	1447336	1447259	78	P	miteXo1-32
CFBP7331_mite14	1859227	1859017	211	C	miteXo1-13
CFBP7331_mite15	2335227	2335438	212	C	miteXo1-33
CFBP7331_mite16	2454634	2454423	212	C	miteXo1-12
CFBP7331_mite17	2556780	2556570	211	C	miteXo1-11
CFBP7331_mite18	2615647	2615786	140	P	miteXo1-30
CFBP7331_mite19	2698307	2698518	212	C	miteXo1-10
CFBP7331_mite20	3009187	3009396	210	C	miteXo2-4
CFBP7331_mite21	3074543	3074752	210	C	miteXo2-9
CFBP7331_mite22	3075523	3075312	212	C	miteXo1-19
CFBP7331_mite23	3304630	3304420	211	C	miteXo1-9
CFBP7331_mite24	3504127	3503916	212	C	miteXo1-8
CFBP7331_mite25	3907764	3907975	212	C	miteXo1-31
CFBP7331_mite26	3955981	3956192	212	C	miteXo1-29
CFBP7331_mite27	4370041	4369832	210	C	miteXo2-3
CFBP7331_mite28	4715905	4715696	210	C	miteXo2-5
CFBP7331_mite29	4804027	4804238	212	C	miteXo1-34
CFBP7331_mite30	4807738	4807528	211	C	miteXo1-35
CFBP7331_mite31	4949506	4949677	172	P	miteXo1-3

I - Continuação

MITE ID	Coordenadas		Tamanho (bp)	Status	Grupo sintênico
	começo	fim			
<i>Xanthomonas oryzae</i> pv. <i>oryzicola</i> CFBP7342					
CFBP7342_mite1	18430	18351	80	P	miteXo1-1
CFBP7342_mite2	31226	31015	212	C	miteXo1-2
CFBP7342_mite3	285072	285281	210	C	miteXo2-5
CFBP7342_mite4	508327	508118	210	C	miteXo2-3
CFBP7342_mite5	713137	713348	212	C	miteXo1-24
CFBP7342_mite6	727653	727444	210	C	miteXo2-1
CFBP7342_mite7	763151	763362	212	C	miteXo1-16
CFBP7342_mite8	852049	851838	212	C	miteXo1-28
CFBP7342_mite9	1020436	1020225	212	C	miteXo1-31
CFBP7342_mite10	1619645	1619435	211	C	miteXo1-13
CFBP7342_mite11	2257604	2257393	212	C	miteXo1-10
CFBP7342_mite12	2380398	2380609	212	C	miteXo1-11
CFBP7342_mite13	2485240	2485451	212	C	miteXo1-12
CFBP7342_mite14	2658036	2657970	67	P	miteXo1-20 (a)
CFBP7342_mite15	2659240	2659090	151	P	miteXo1-20 (b)
CFBP7342_mite16	3100792	3100582	211	C	miteXo1-9
CFBP7342_mite17	3197623	3197834	212	C	miteXo1-36
CFBP7342_mite18	3314212	3314001	212	C	miteXo1-37
CFBP7342_mite19	3335370	3335159	212	C	miteXo1-8
CFBP7342_mite20	3843188	3842977	212	C	miteXo1-14
CFBP7342_mite21	3851802	3852012	211	C	miteXo1-15
CFBP7342_mite22	4020426	4020637	212	C	miteXo1-29
CFBP7342_mite23	4627257	4627046	212	C	miteXo1-7
CFBP7342_mite24	4721919	4721708	212	C	miteXo1-6
CFBP7342_mite25	4755632	4755422	211	C	miteXo1-5
CFBP7342_mite26	4992902	4993113	212	C	miteXo1-4
CFBP7342_mite27	5038484	5038655	172	P	miteXo1-3
<i>Xanthomonas oryzae</i> pv. <i>oryzicola</i> BXORI					
BXORI_mite1	16911	16832	80	P	miteXo1-1
BXORI_mite2	25773	25562	212	C	miteXo1-2
BXORI_mite3	276183	276392	210	C	miteXo2-5
BXORI_mite4	448563	448352	212	C	miteXo1-38
BXORI_mite5	566884	567093	210	C	miteXo2-3
BXORI_mite6	620140	620351	212	C	miteXo1-24
BXORI_mite7	632077	631868	210	C	miteXo2-1
BXORI_mite8	667430	667641	212	C	miteXo1-16
BXORI_mite9	1069330	1069120	211	C	miteXo1-15
BXORI_mite10	1077944	1078155	212	C	miteXo1-14
BXORI_mite11	1261658	1261869	212	C	miteXo1-39
BXORI_mite12	1709124	1708914	211	C	miteXo1-13
BXORI_mite13	2126294	2126360	67	P	miteXo1-20
BXORI_mite14	2274322	2274111	212	C	miteXo1-12
BXORI_mite15	2373218	2373007	212	C	miteXo1-11
BXORI_mite16	2412968	2413107	140	P	miteXo1-30
BXORI_mite17	2490523	2490734	212	C	miteXo1-10
BXORI_mite18	2795763	2795972	210	C	miteXo2-4
BXORI_mite19	2832835	2833046	212	C	miteXo1-19
BXORI_mite20	3051690	3051480	211	C	miteXo1-9
BXORI_mite21	3224971	3224762	210	C	miteXo2-2
BXORI_mite22	3244319	3244108	212	C	miteXo1-8
BXORI_mite23	3670541	3670752	212	C	miteXo1-29
BXORI_mite24	4004451	4004240	212	C	miteXo1-40
BXORI_mite25	4252551	4252340	212	C	miteXo1-7
BXORI_mite26	4343541	4343330	212	C	miteXo1-6
BXORI_mite27	4377241	4377031	211	C	miteXo1-5
BXORI_mite28	4428352	4428548	197	P	miteXo2-8
BXORI_mite29	4556168	4555959	210	C	miteXo1-40
BXORI_mite30	4574582	4574371	212	C	miteXo1-42
BXORI_mite31	4612604	4612393	212	C	miteXo1-4
BXORI_mite32	4649753	4649795	43	P	miteXo1-3a
BXORI_mite33	4650953	4651124	172	P	miteXo1-3b

I - Continuação

MITE ID	Coordenadas		Tamanho	Status	Grupo
	começo	fim	(bp)		sintênico
<i>Xanthomonas oryzae</i> pv. <i>oryzae</i> MAFF311018					
MAFF_mite1	143697	143610	88	P	miteXo1-57
MAFF_mite2	175495	175286	210	C	miteXo1-43
MAFF_mite3	305644	305435	210	C	miteXo1-44
MAFF_mite4	361498	361708	211	C	miteXo1-45
MAFF_mite5	468779	468570	210	C	miteXo2-7
MAFF_mite6	643801	644012	212	C	miteXo1-65
MAFF_mite7	1022322	1022532	211	C	miteXo1-46
MAFF_mite8	1300003	1300213	211	C	miteXo1-47
MAFF_mite9	1314315	1314105	211	C	miteXo1-66
MAFF_mite10	1476257	1476466	210	C	miteXo2-2
MAFF_mite11	1509451	1509660	210	C	miteXo1-60
MAFF_mite12	1745225	1745435	211	C	miteXo1-61
MAFF_mite13	1838872	1838661	212	C	miteXo1-62
MAFF_mite14	1869677	1869887	211	C	miteXo1-48
MAFF_mite15	2185361	2185150	212	C	miteXo1-49
MAFF_mite16	2321378	2321325	54	P	miteXo1-58
MAFF_mite17	2322593	2322434	160	P	miteXo1-59
MAFF_mite18	2628613	2628403	211	C	miteXo1-50
MAFF_mite19	2895546	2895756	211	C	miteXo1-51
MAFF_mite20	3137917	3137707	211	C	miteXo1-52
MAFF_mite21	3329320	3329110	211	C	miteXo1-53
MAFF_mite22	3405986	3406196	211	C	miteXo1-54
MAFF_mite23	3538192	3537982	211	C	miteXo1-55
MAFF_mite24	4451866	4451921	56	P	miteXo1-63
MAFF_mite25	4543008	4542799	210	C	miteXo2-1
MAFF_mite26	4867231	4867441	211	C	miteXo1-56
<i>Xanthomonas oryzae</i> pv. <i>oryzae</i> PXO86					
PXO86_mite1	151212	151125	88	P	miteXo1-57
PXO86_mite2	182083	181874	210	C	miteXo1-43
PXO86_mite3	543603	543812	210	C	miteXo2-1
PXO86_mite4	1490998	1491208	211	C	miteXo1-55
PXO86_mite5	1590575	1590365	211	C	miteXo1-67
PXO86_mite6	1626955	1626745	211	C	miteXo1-54
PXO86_mite7	1703334	1703544	211	C	miteXo1-53
PXO86_mite8	1883983	1884193	211	C	miteXo1-52
PXO86_mite9	2118934	2118724	211	C	miteXo1-51
PXO86_mite10	2485051	2484998	54	P	miteXo1-58
PXO86_mite11	2486266	2486107	160	P	miteXo1-59
PXO86_mite12	2826896	2826686	211	C	miteXo1-50
PXO86_mite13	2867080	2867291	212	C	miteXo1-49
PXO86_mite14	3186505	3186295	211	C	miteXo1-48
PXO86_mite15	3517397	3517188	210	C	miteXo1-60
PXO86_mite16	3551215	3551006	210	C	miteXo2-2
PXO86_mite17	3728762	3728552	211	C	miteXo1-47
PXO86_mite18	3857724	3857934	211	C	miteXo1-68
PXO86_mite19	4014840	4014630	211	C	miteXo1-46
PXO86_mite20	4501129	4501184	56	P	miteXo1-64
PXO86_mite21	4684334	4684543	210	C	miteXo2-7
PXO86_mite22	4791506	4791296	211	C	miteXo1-45
PXO86_mite23	4853080	4853289	210	C	miteXo1-44
PXO86_mite24	4931023	4930812	212	C	miteXo1-77
PXO86_mite25	4943581	4943791	211	C	miteXo1-56

I - Continuação

MITE ID	Coordenadas		Tamanho	Status	Grupo sintênico
	começo	fim	(bp)		
<i>Xanthomonas oryzae</i> pv. <i>oryzae</i> PXO99A					
PXO99A_mite1	139128	139337	210	C	miteXo1-69
PXO99A_mite2	147811	147724	88	P	miteXo1-57
PXO99A_mite3	180202	179993	210	C	miteXo1-43
PXO99A_mite4	544300	544091	210	C	miteXo1-70(d)
PXO99A_mite5	545439	545648	210	C	miteXo2-1
PXO99A_mite6	1343225	1343420	196	P	miteXo2-2
PXO99A_mite7	1389605	1389814	210	C	miteXo1-60
PXO99A_mite8	1560060	1560270	211	C	miteXo1-73
PXO99A_mite9	1728405	1728615	211	C	miteXo1-52
PXO99A_mite10	1966826	1966616	211	C	miteXo1-51
PXO99A_mite11	2096864	2097074	211	C	miteXo1-50
PXO99A_mite12	2097913	2098122	210	C	miteXo1-70(d)
PXO99A_mite13	2306796	2306587	210	C	miteXo1-70(d)
PXO99A_mite14	2424072	2424231	160	P	miteXo1-59
PXO99A_mite15	2425287	2425340	54	P	miteXo1-58
PXO99A_mite16	2709164	2709376	213	C	miteXo1-49 (d)
PXO99A_mite17	2921251	2921463	213	C	miteXo1-49 (d)
PXO99A_mite18	3233137	3232927	211	C	miteXo1-48
PXO99A_mite19	3264654	3264865	212	C	miteXo1-71
PXO99A_mite20	3422033	3421823	211	C	miteXo1-53
PXO99A_mite21	3496073	3496283	211	C	miteXo1-54
PXO99A_mite22	3535847	3536057	211	C	miteXo1-72
PXO99A_mite23	3548672	3548462	211	C	miteXo1-61
PXO99A_mite24	3673058	3672848	211	C	miteXo1-55
PXO99A_mite25	4045899	4045689	211	C	miteXo1-47
PXO99A_mite26	4324321	4324111	211	C	miteXo1-46
PXO99A_mite27	4914088	4914297	210	C	miteXo2-7
PXO99A_mite28	5019013	5018803	211	C	miteXo1-45
PXO99A_mite29	5073749	5073958	210	C	miteXo1-44
PXO99A_mite30	5161621	5161831	211	C	miteXo1-56
<i>Xanthomonas oryzae</i> pv. <i>oryzae</i> KACC10331					
KACC_mite1	161253	161166	88	P	miteXo1-57
KACC_mite2	193725	193516	210	C	miteXo1-43
KACC_mite3	317750	317541	210	C	miteXo1-44
KACC_mite4	373589	373799	211	C	miteXo1-45
KACC_mite5	480882	480673	210	C	miteXo2-7
KACC_mite6	1055013	1055223	211	C	miteXo1-46
KACC_mite7	1323781	1323991	211	C	miteXo1-47
KACC_mite8	1492561	1492770	210	C	miteXo2-2
KACC_mite9	1762923	1763133	211	C	miteXo1-61
KACC_mite10	1859718	1859510	209	C	miteXo1-62
KACC_mite11	1890544	1890754	211	C	miteXo1-48
KACC_mite12	2197802	2197591	212	C	miteXo1-49
KACC_mite13	2346527	2346474	54	P	miteXo1-58
KACC_mite14	2347742	2347583	160	P	miteXo1-59
KACC_mite15	2391434	2391643	210	C	miteXo1-74
KACC_mite16	2651458	2651248	211	C	miteXo1-50
KACC_mite17	2917270	2917480	211	C	miteXo1-51
KACC_mite18	2977575	2977364	212	C	miteXo1-75
KACC_mite19	3145225	3145015	211	C	miteXo1-52
KACC_mite20	3324326	3324116	211	C	miteXo1-53
KACC_mite21	3399481	3399691	211	C	miteXo1-54
KACC_mite22	3533713	3533503	211	C	miteXo1-55
KACC_mite23	4232008	4232218	211	C	miteXo1-76
KACC_mite24	4454434	4454489	56	P	miteXo1-63
KACC_mite25	4548110	4547901	210	C	miteXo2-1
KACC_mite26	4870992	4871202	211	C	miteXo1-56

I - Continuação

MITE ID	Coordenadas		Tamanho	Status	Grupo sintênico
	começo	fim	(bp)		
<i>Xanthomonas oryzae</i> pv. <i>oryzae</i> AXO1947					
AXO1947_mite1	274466	274504	39	P	miteXo1-79
AXO1947_mite2	874647	874685	39	P	miteXo1-80
AXO1947_mite3	931186	931395	210	C	miteXo2-10
AXO1947_mite4	975036	975074	39	P	miteXo1-81
AXO1947_mite5	2373915	2374126	212	C	miteXo1-78
AXO1947_mite6	3084142	3084180	39	P	miteXo1-82
AXO1947_mite7	3400676	3400880	205	C	miteXo2-11
AXO1947_mite8	4108708	4108746	39	P	miteXo1-83
AXO1947_mite9	4119247	4119456	210	C	miteXo2-1
<i>Xanthomonas oryzae</i> pv. <i>oryzae</i> PXO83					
PXO83_mite1	149084	148997	88	P	miteXo1-57
PXO83_mite2	179955	179746	211	C	miteXo1-43
PXO83_mite3	541956	542165	210	C	miteXo2-1
PXO83_mite4	1501931	1502141	212	C	miteXo1-55
PXO83_mite5	1636618	1636408	212	C	miteXo1-54
PXO83_mite6	1713008	1713218	212	C	miteXo1-53
PXO83_mite7	1892603	1892813	212	C	miteXo1-52
PXO83_mite8	2360232	2360391	161	P	miteXo1-59
PXO83_mite9	2361447	2361500	54	P	miteXo1-58
PXO83_mite10	2729312	2729522	212	C	miteXo1-51
PXO83_mite11	2840000	2839790	212	C	miteXo1-50
PXO83_mite12	2880184	2880395	211	C	miteXo1-49
PXO83_mite13	3199597	3199387	212	C	miteXo1-48
PXO83_mite14	3529793	3529584	211	C	miteXo1-60
PXO83_mite15	3564670	3564462	210	C	miteXo2-2
PXO83_mite16	3743238	3743028	212	C	miteXo1-47
PXO83_mite17	4024848	4024638	212	C	miteXo1-46
PXO83_mite18	4510006	4510061	56	P	miteXo1-64
PXO83_mite19	4693211	4693419	210	C	miteXo2-7
PXO83_mite20	4801854	4801644	212	C	miteXo1-45
PXO83_mite21	4863421	4863630	211	C	miteXo1-44
PXO83_mite22	4941348	4941137	212	C	miteXo1-77
PXO83_mite23	4953906	4954116	212	C	miteXo1-56

Em vermelho, cópias da variante MITE-Xo2

C, completo; P, parcial

(a) e (b) representam metades de um MITE interrompido

(d) região genômica duplicada

J - Principais grupos sintênicos de MITE-Xo1 conservados em *Xoc*. Blocos em verde indicam genes com função conhecida; em bege, proteínas hipotéticas (HyPr); em azul, elementos IS; em vermelho, MITes-Xo1. Linha pontilhada indicam elementos parciais e contorno pontilhado em genes indicam pseudogenização.

J - Continuação

J - Continuação

J - Continuação

J - Continuação

K - Quantificação de transcritos de MITEs e genes de Xoo PXO99A.

Locus/ID	Gene	Xoo PXO99A			
		tamanho	eff_length	est_counts	TPM
Proteínas ribossomais					
PXO_00889	rpmJ	126	64	1624	2441
PXO_03485	rpmH	141	72	3891	5213
PXO_03239	rpmG	168	83	597	701
PXO_04101	rpsU	216	106	2229	2034
PXO_00745	rpsR	231	116	1005	843
PXO_04517	rpsS	270	140	1295	896
PXO_04522	rpsJ	315	173	3530	1974
PXO_04510	rplX	318	175	1569	868
PXO_01386	rplT	360	207	1981	929
PXO_04495	rplQ	384	224	1336	578
PXO_00744	rplI	450	269	1227	443
PXO_04520	rplD	606	385	4004	1008
PXO_05534	rpmA	261	133	3850	2798
PXO_02147	rpsP	258	131	3286	2428
PXO_00891	rpsA	1686	1384	12043	844
PXO_04513	rpmC	186	89	677	734
PXO_02705	rpmF	189	90	1587	1701
PXO_04502	rpmD	192	93	844	884
PXO_01385	rpmI	198	96	1842	1856
PXO_03240	rpmB	219	108	1462	1313
PXO_02360	rpmE	231	116	5743	4818
PXO_05533	rplU	258	131	3843	2840
PXO_01306	rpsO	261	133	1901	1381
PXO_04512	rpsQ	270	140	1526	1056
PXO_00014	rpsT	270	140	2735	1892
PXO_04519	rplW	300	162	1708	1023
PXO_04508	-	306	167	1760	1022
PXO_04507	rpsH	327	181	3612	1932
PXO_04516	rplV	336	187	1295	670
PXO_04499	rpsM	357	204	2962	1407
PXO_04511	rplN	360	207	2159	1013
PXO_04504	rplR	360	207	3411	1600
PXO_04528	rpsL	375	218	3438	1527
PXO_04011	rpsI	393	230	2364	996
PXO_04498	rpsK	393	230	2930	1234
PXO_04514	rplP	414	244	1390	553
PXO_02150	rplS	417	247	3616	1422
PXO_04534	rplK	429	255	4560	1730
PXO_04010	rplM	429	255	4672	1772
PXO_04501	rplO	444	265	2045	748
PXO_04527	rpsG	468	282	3055	1048
PXO_04503	rpsE	543	340	3999	1140
PXO_04521	rplC	624	398	3711	903
PXO_04497	rpsD	627	401	4699	1137
PXO_04533	rplA	699	456	5452	1159
PXO_04515	rpsC	735	486	2736	546
PXO_01132	rpsB	789	532	3250	592
PXO_00746	-	432	257	2411	909
PXO_04506	-	432	257	3138	1184
PXO_04532	-	534	333	2683	782
PXO_04541	-	540	337	1207	347
PXO_04509	-	543	340	2417	689
PXO_04518	rplB	819	556	3574	623
PXO_02302	-	918	631	791	122
PXO_04786	-	183	87	1055	1172
PXO_04531	-	369	213	5089	2311
PXO_03005	-	408	240	367	148

K - Continuação

Locus/ID	Gene	Xoo PXO99A			
		tamanho	eff_length	est_counts	TPM
Genes Housekeeping					
PXO_03113	atpD	1407	1105	4244	372
PXO_04097	dnaG	1749	1447	765	51
PXO_00382	efp	552	347	2252	629
PXO_03568	glnA	1404	1102	2454	216
PXO_03480	gyrB	2445	2143	7340	332
PXO_01302	nusA	1488	1186	2402	196
PXO_01714	lepA	1791	1489	2805	183
PXO_02116	pgk	1176	874	1603	178
PXO_00174	pyrG	1665	1363	6916	492
PXO_04530	rpoB	4152	3850	14560	367
PXO_01185	dnaK	1926	1624	27917	1666
PXO_04069	rpoD	1875	1573	7296	450
PXO_00149	recA	1035	733	2841	376
PXO_03703	groEL	1641	1339	52020	3766
MITES					
PXO99A_mite1	miteXo1	210	103	2317	2184
PXO99A_mite2	miteXo1	88	45	13	27
PXO99A_mite3	miteXo1	210	103	93	87
PXO99A_mite4	miteXo1	210	103	2317	2184
PXO99A_mite5	miteXo2-4	210	103	67	63
PXO99A_mite6	miteXo2-1	196	95	4	4
PXO99A_mite7	miteXo1	210	103	2317	2184
PXO99A_mite8	miteXo1	211	103	224	210
PXO99A_mite9	miteXo1	211	103	224	210
PXO99A_mite10	miteXo1	211	103	22	20
PXO99A_mite11	miteXo1	210	103	2317	2184
PXO99A_mite12	miteXo1	211	103	276	259
PXO99A_mite13	miteXo1	210	103	2317	2184
PXO99A_mite14	miteXo1	160	79	20	25
PXO99A_mite15	miteXo1	54	28	21	74
PXO99A_mite16	miteXo1	213	104	5	5
PXO99A_mite17	miteXo1	213	104	5	5
PXO99A_mite18	miteXo1	211	103	269	252
PXO99A_mite19	miteXo1	212	104	42	39
PXO99A_mite20	miteXo1	211	103	131	123
PXO99A_mite21	miteXo1	211	103	81	76
PXO99A_mite22	miteXo1	211	103	81	76
PXO99A_mite23	miteXo1	211	103	10962	10271
PXO99A_mite24	miteXo1	211	103	259	243
PXO99A_mite25	miteXo1	211	103	296	277
PXO99A_mite26	miteXo1	211	103	81	76
PXO99A_mite27	miteXo2-7	210	103	1259	1187
PXO99A_mite28	miteXo1	211	103	42	39
PXO99A_mite29	miteXo1	210	103	2317	2184
PXO99A_mite30	miteXo1	211	103	555	520

L - Quantificação de transcritos de MITEs e genes de Xoo PXO86

Locus/ID	Gene	Xoo PXO86			
		tamanho	eff_length	est_counts	TPM
Proteínas ribossomais					
AZ54_11850	rpmJ	126	62	888	2130
AZ54_24460	rpmH	141	70	1524	3244
AZ54_24050	rpmG	168	78	341	654
AZ54_03155	rpsU	216	100	1287	1915
AZ54_13290	rpsR	231	112	632	841
AZ54_05820	rpsS	270	139	921	984
AZ54_05795	rpsJ	315	172	2267	1968
AZ54_05855	rplX	318	173	1029	885
AZ54_07940	rplT	360	202	1386	1025
AZ54_05930	rplQ	384	219	856	582
AZ54_13285	rplI	450	264	824	465
AZ54_05805	rplD	606	379	2782	1095
AZ54_16320	rpmA	261	132	2230	2511
AZ54_17855	rpsP	258	130	2143	2456
AZ54_11860	rpsA	1686	1382	7943	857
AZ54_05840	rpmC	186	83	531	950
AZ54_20430	rpmF	195	88	1155	1964
AZ54_05895	rpmD	192	86	638	1102
AZ54_07935	rpmI	198	89	1211	2026
AZ54_24045	rpmB	237	116	1000	1288
AZ54_18800	rpmE	243	120	2740	3403
AZ54_16325	rplU	321	175	3243	2757
AZ54_7760	rpsO	261	132	1174	1322
AZ54_05845	rpsQ	270	139	1072	1146
AZ54_16310	rpsT	270	139	1748	1868
AZ54_05810	rplW	300	161	1188	1101
AZ54_05865	-	306	165	1217	1098
AZ54_05875	rpsH	399	230	3135	2036
AZ54_05825	rplV	336	186	873	702
AZ54_05910	rpsM	357	200	1927	1436
AZ54_05850	rplN	369	209	1381	988
AZ54_05885	rplR	360	202	2221	1643
AZ54_05770	rpsL	375	213	2087	1461
AZ54_21755	rpsI	393	225	1422	941
AZ54_05915	rpsK	393	225	1879	1243
AZ54_05835	rplP	414	240	1045	650
AZ54_17840	rplS	408	235	2418	1532
AZ54_05740	rplK	429	250	3001	1787
AZ54_21760	rplM	429	250	2887	1719
AZ54_05900	rplO	438	257	1546	898
AZ54_05775	rpsG	468	278	1936	1040
AZ54_05890	rpsE	543	333	2894	1294
AZ54_05800	rplC	651	415	2527	909
AZ54_05920	rpsD	627	395	3205	1211
AZ54_05745	rplA	699	453	3571	1175
AZ54_05830	rpsC	735	481	1858	576
AZ54_14605	rpsB	804	536	2112	587
AZ54_13295	-	432	252	1655	978
AZ54_05880	-	528	322	2529	1170
AZ54_05750	-	534	328	1908	868
AZ54_05690	-	636	402	966	359
AZ54_05860	-	543	333	1670	747
AZ54_05815	rplB	828	556	2536	680
AZ54_18500	-	888	607	428	105
AZ54_16735	-	183	82	774	1402
AZ54_05755	-	369	209	3530	2524
AZ54_22195	-	417	242	227	140

L - Continuação

Locus/ID	Gene	Xoo PXO86			
		tamanho	eff_length	est_counts	TPM
<i>Genes Housekeeping</i>					
AZ54_21125	<i>atpD</i>	1407	1103	2896	392
AZ54_03130	<i>dnaG</i>	1749	1445	445	46
AZ54_10620	<i>efp</i>	567	350	1543	658
AZ54_01480	<i>glnA</i>	1410	1106	1581	213
AZ54_00020	<i>gyrB</i>	2445	2141	4843	337
AZ54_07740	<i>nusA</i>	1512	1208	1644	203
AZ54_15260	<i>lepA</i>	1806	1502	1807	179
AZ54_06735	<i>pgk</i>	1176	872	1125	192
AZ54_09165	<i>pyrG</i>	1665	1361	4768	522
AZ54_05760	<i>rpoB</i>	4152	3848	10427	404
AZ54_14330	<i>dnaK</i>	1926	1622	24943	2293
AZ54_21995	<i>rpoD</i>	1875	1571	5305	504
AZ54_09285	<i>recA</i>	1035	731	1784	364
AZ54_02530	<i>groEL</i>	1641	1337	47689	5319
<i>MITEs</i>					
PXO86_mite1	miteXo1	88	45	0	0
PXO86_mite2	miteXo1	210	96	35	55
PXO86_mite3	miteXo2-4	210	96	28	43
PXO86_mite4	miteXo1	211	97	1380	2123
PXO86_mite5	miteXo1	211	97	54	83
PXO86_mite6	miteXo1	211	97	95	146
PXO86_mite7	miteXo1	211	97	501	771
PXO86_mite8	miteXo1	211	97	164	253
PXO86_mite9	miteXo1	211	97	69	106
PXO86_mite10	miteXo1	54	28	9	49
PXO86_mite11	miteXo1	160	77	0	0
PXO86_mite12	miteXo1	211	97	178	274
PXO86_mite13	miteXo1	212	98	14	21
PXO86_mite14	miteXo1	211	97	697	1073
PXO86_mite15	miteXo1	210	96	3187	4934
PXO86_mite16	miteXo2-1	210	96	297	459
PXO86_mite17	miteXo1	211	97	54	83
PXO86_mite18	miteXo1	211	97	13	20
PXO86_mite19	miteXo1	211	97	34	52
PXO86_mite20	miteXo1	56	29	0	0
PXO86_mite21	miteXo2-7	210	96	290	448
PXO86_mite22	miteXo1	211	97	13	20
PXO86_mite23	miteXo1	210	96	3187	4934
PXO86_mite24	miteXo1	212	98	34	52
PXO86_mite25	miteXo1	211	97	1516	2332

M - Identificação das montagens dos genomas de arroz

Espécie	Identificação da montagem do genoma	Total de <i>motifs</i> encontrados*
<i>Oryza sativa</i> ssp. <i>japonica</i>	IRGSP-1.0	109
<i>Oryza rufipogon</i>	OR_W1943	122
<i>Oryza sativa</i> ssp. <i>indica</i>	ASM465v1	129
<i>Oryza nivara</i>	AWHD000000000	104
<i>Oryza glaberrima</i>	AGI1.1	92
<i>Oryza barthii</i>	O.barthii_v1	100
<i>Oryza glumaepatula</i>	ALNU020000000	107
<i>Oryza longistaminata</i>	O_longistaminata_v1.0	148
<i>Oryza meridionalis</i>	Oryza_meridionalis_v1.3	93
<i>Oryza punctata</i>	AVCL000000000	35
<i>Oryza brachyantha</i>	Oryza_brachyantha.v1.4b	0
<i>Leersia perrieri</i>	Lperr_V1.4	1
<i>Zea mays</i>	AGPv4	0
<i>Arabidopsis thaliana</i>	TAIR10	0

* regiões que apresentam 100% de identidade

N - Predição da estrutura secundária de MITE-Xo1 e MITE-Xo2. Predições foram calculadas com (A) MITE-Xo1, cópia BLS256_mite3 e (B) MITE-Xo2, cópia BLS256_mite11. Os *motif* de 18 nucleotídeos está destacado em amarelo, na cópia de MITE-Xo1.