

**UNIVERSIDADE ESTADUAL PAULISTA - UNESP
CÂMPUS DE JABOTICABAL**

**ESTUDO DO “*MOBILE-METAGENOME*” A PARTIR DE
BIBLIOTECA METAGENÔMICA PROVENIENTE DE SOLO
COM CULTIVO DE CANA-DE-AÇÚCAR E SOLO DE
FLORESTA**

**Alessandra dos Santos Pinto
Bióloga**

2015

**UNIVERSIDADE ESTADUAL PAULISTA - UNESP
CÂMPUS DE JABOTICABAL**

**ESTUDO DO “*MOBILE-METAGENOME*” A PARTIR DE
BIBLIOTECA METAGENÔMICA PROVENIENTE DE SOLO
COM CULTIVO DE CANA-DE-AÇÚCAR E SOLO DE
FLORESTA**

Alessandra dos Santos Pinto

Orientador: Prof. Dr. Alessandro de Mello Varani

Co-orientadora: Dra. Elisângela Soares Gomes

Dissertação apresentada à Faculdade de Ciências Agrárias e Veterinárias – Unesp, Câmpus de Jaboticabal, como parte das exigências para a obtenção do título de Mestre em Microbiologia Agropecuária.

2015

P659e Pinto, Alessandra dos Santos
Estudo do “*Mobile-metagenome*” a partir de biblioteca metagenômica proveniente de solo com cultivo de cana-de-açúcar e solo de floresta / Alessandra dos Santos Pinto. – – Jaboticabal, 2015 x, 69 p. ; 29 cm

Dissertação (mestrado) - Universidade Estadual Paulista, Faculdade de Ciências Agrárias e Veterinárias, 2015
Orientador: Alessandro de Mello Varani
Co-orientadora: Elisângela Soares Gomes
Banca examinadora: Marcos Túlio Oliveira, Mariana Carina Frigieri
Bibliografia

1. Elementos do tipo IS. 2. Metagenoma.. 3. Bioinformática. I. Título. II. Jaboticabal-Faculdade de Ciências Agrárias e Veterinárias.

CDU 576.8:633.61

Ficha catalográfica elaborada pela Seção Técnica de Aquisição e Tratamento da Informação – Serviço Técnico de Biblioteca e Documentação - UNESP, Câmpus de Jaboticabal.

DADOS CURRICULARES DO AUTOR

Alessandra dos Santos Pinto, nascida no município de Matão, estado de São Paulo, em 07 de outubro de 1989. Formada em Ciências Biológicas (Licenciatura e Bacharelado) pelo Centro Universitário de Araraquara-UNIARA, no ano de 2010. Em julho de 2013 concluiu o Curso de Especialização em Educação Ambiental na Universidade de São Paulo – USP, obtendo o título de Especialista em Educação Ambiental. Em agosto de 2013 iniciou o Mestrado em Microbiologia Agropecuária na Faculdade de Ciências Agrárias e Veterinárias da Universidade Estadual Paulista “Julio de Mesquita Filho” (UNESP- FCAV).

“Não sou nada, sou apenas um instrumento, um pequeno lápis nas mãos do Senhor, com o qual Ele escreve aquilo que deseja. Por mais imperfeitos que sejamos Ele escreve magnificamente.”

Beata Madre Teresa de Calcutá

Dedico

Ao **Mestre dos mestres: Deus;**

À minha **família**, pelo incentivo e amor incondicional.

AGRADECIMENTOS

Agradeço à Deus, por ser Este Pai que me sustenta, Meu refúgio e Minha proteção. O Meu Deus em que eu deposito toda a minha confiança e nada temo. A Luz que me guia a cada instante.

Aos meus pais, Valditudes e Ana Maria, por me ensinarem a ter fé em Deus e na vida, por acreditarem em minhas capacidades, e pelas vezes que renunciaram aos seus sonhos para que eu pudesse realizar o meu;

Às minhas irmãs, Natália e Noely, por me impulsionarem a prosseguir com fé diante das dificuldades e também pelo carinho, paciência e incentivo constantes;

Ao meu orientador, Prof. Dr. Alessandro de Mello Varani, por toda confiança depositada, demonstrando sempre apoio incondicional e compreensão constante, incentivando sempre a superar meus limites, a lutar com coragem pela conquista dos meus objetivos;

À minha co-orientadora, Dra. Elisângela Soares Gomes, pelo incentivo em todos os momentos e por transmitir a cada palavra a certeza que este projeto seria possível;

À Profa. Dra. Eliana Gertrudes Macedo Lemos, por acreditar neste trabalho e por ceder o laboratório para que essa pesquisa fosse desenvolvida;

Aos amigos, colegas, estudantes e funcionários do Laboratório de Bioquímica de Micro-organismos e Plantas (LBMP), pessoas que tornaram mais fácil e mais alegre a caminhada;

E a todos os colegas e professores da pós-graduação pelo convívio e aprendizado.

SUMÁRIO

	Página
RESUMO	iii
ABSTRACT	iv
LISTA DE ABREVIATURAS	v
LISTA DE TABELAS	vi
LISTA DE FIGURAS	vii
1 INTRODUÇÃO	1
1.1 Objetivos	3
1.1.1. Objetivo geral	3
1.1.2 Objetivos específicos	3
2 REVISÃO DE LITERATURA	4
2.1 Os Elementos Genéticos Móveis (EGMs)	4
2.2 Impactos dos Elementos Genéticos Móveis	10
2.3 Transposons do tipo Sequências de Inserção (IS)	12
2.4 Diversidade microbiana do solo	16
2.5 Biblioteca metagenômica	17
2.6 Prospecção gênica em bibliotecas metagenômicas através da Bioinformática	18
2.7 MG-RAST e ISsaga 2.0	19
3 MATERIAL E MÉTODOS	22
3.1 Coleta das amostras	22
3.2 Extração e quantificação do DNA metagenômico	23
3.3 Sequenciamento do DNA metagenômico	24
3.4 Análise <i>in silico</i>	24
3.5 Anotação genômica	25
4 RESULTADOS E DISCUSSÃO	26
4.1 Primeira Frente de Investigação MG-RAST (Análise de <i>reads</i>)	26

4.2	Classificação transposases MG-RAST	27
4.3	Classificação taxonômica para <i>reads</i> (MG-RAST)	29
4.4	Segunda frente de investigação: Classificação MG-RAST – Metagenoma Montado (<i>scaffolds</i>).....	33
4.5	Classificação ISSaga 2.0 – Metagenoma montado (<i>Scaffolds</i>).....	36
4.6	Classificação transposases ISSaga 2.0	37
4.7	Classificação taxonômica ISSaga 2.0	39
4.8	Comparações dos resultados do MG-RAST (<i>reads</i>) contra o banco de dados ISfinder	42
4.9	Comparação plataforma MG-RAST/ISfinder X ISSaga 2.0	46
4.10	Caracterização do Contexto Genômico	47
5	CONCLUSÕES	53
6	PERSPECTIVAS	54
7	REFERÊNCIAS	55
	APÊNDICES	63

ESTUDO DO “MOBILE-METAGENOME” A PARTIR DE BIBLIOTECA METAGENÔMICA PROVENIENTE DE SOLO COM CULTIVO DE CANA-DE-AÇÚCAR E SOLO DE FLORESTA

RESUMO: Os Elementos Genéticos Móveis (EGMs) são segmentos de DNA presentes em todos os domínios, e que apresentam a capacidade de mobilização em um genoma. São considerados agentes chaves na evolução e diversificação dos seres vivos. Sequências de Inserção (IS) são os EGMs mais abundantes dos procariotos. IS são elementos compactos, que codificam o gene responsável pela mobilização, a transposase. Estudos de genômica comparativa indicam que a transposase é o gene mais abundante presente em banco de dados públicos de genomas e metagenomas. O presente estudo avaliou a abundância e diversidade de transposases a partir de dados de sequenciamento de alto rendimento provenientes de duas bibliotecas metagenômicas, a primeira originada de solo com cultivo de cana-de-açúcar e a segunda de solo de floresta. Os metagenomas foram analisados através de duas frentes de investigação. Na primeira frente de investigação foram identificadas as transposases em um conjunto de 267.879.464 *reads paired-end* (2x100bp) através da ferramenta MG-RAST. Na segunda frente de investigação, os *reads* de cada metagenoma foram montados (*de novo assembly*) e posteriormente anotados através da plataforma MG-RAST e da ferramenta ISSaga 2.0. A plataforma MG-RAST identificou 902.982 (*reads*) com função atribuída para transposases, enquanto que a ferramenta ISSaga 2.0 recuperou 592 (*scaffolds*) contendo quadros aberto de leitura (ORFs) codificando para transposases. Um total de 120 (22%) *scaffolds* e 767.534 (85%) *reads* anotados foram classificados em grupos taxonômicos distintos, onde o filo Proteobacteria (48% - MG-RAST e 34,3% - ISSaga 2.0) foi predominante. Análises comparativas e biocuração indicam que 695.296 (77%) dos *reads* foram incorretamente atribuídos com função transposase pela ferramenta MG-RAST. Dentre as famílias de transposases classificadas pelo ISSaga 2.0 destacam-se as famílias: IS110 (76/13,9%), IS3 (74/13,6%), e em particular 65 *scaffolds* (11,9%) provenientes da família ISNCY (*not classified yet*). Um outro resultado notável, é que nenhum IS previamente descrito na literatura foi identificado nestes dados de metagenomas. Este estudo revelou: (a) MG-RAST apresenta deficiências para identificação de transposases; (b) A ferramenta ISSaga 2.0 oferece uma metodologia precisa para identificação de transposases em dados de metagenomas; (c) A análise precisa de transposases em dados de metagenomas forneceu a identificação de novos ISs. Juntos, esses resultados poderão auxiliar na descoberta de novas famílias de transposases e que em última instância poderá auxiliar a revelar novos mecanismos de transposição.

Palavras-chave: elementos do tipo IS; metagenoma; bioinformática; anotação genômica; genômica comparativa.

"MOBILE-METAGENOME" STUDY OF METAGENOMIC LIBRARY ORIGINATING FROM SUGARCANE CULTIVATION AND FOREST SOIL

ABSTRACT: Mobile Genetic Elements (MGE) are segments of DNA present in the biological domains that have the ability to move around within a genome. They are considered key player in the evolution and diversification of living things. Insertion Sequences (IS) are the most abundant MGEs of prokaryotes. IS are compact elements that encode the gene responsible for the mobilization, the transposase. Comparative genomic studies indicate that the transposase is the most abundant gene found in public database of genomes and metagenomes. This study evaluated the abundance and diversity of transposases from high-throughput sequencing data originated from two metagenomic libraries, one from soil with sugarcane cultivation and the other from forest soil. The metagenomes were analyzed through two research fronts. In the first one, transposases were identified in a set of 267. 879. 464 reads paired-end (2x100bp) by MG RAST tool. In the second front of research, the reads of each metagenome were fitted (*de novo* assembly) and subsequently recorded by ISSaga 2.0 tool and platform MG-RAST. The platform MG- RAST identified 902.982 (reads) with allocated transposase function, whereas the 2.0 ISSaga tool recovered 592 (scaffolds) containing open reading frames (ORFs) coding for transposase. A total of 120 (22%) scaffolds and 767.534 (85%) reads noted were classified in different taxonomic groups resulting the phylum Proteobacteria (48% - MG-RAST and 34.3% - Issaga 2.0). Comparative analysis and biocuration indicate that 695.296 (77%) of the reads were incorrectly assigned with transposase function by MG-RAST tool. Among the families of transposase classified by Issaga 2.0, feature the families: IS110 (76 / 13.9%), IS3 (74 / 13.6%), and particularly 65 scaffolds (11.9%) from ISNCY family (not classified yet). Another noticeable result is that no IS previously described in the literature has been identified in these metagenomic data. This study revealed: (a) MG-RAST presents deficiencies to transposases identification; (b) ISSaga 2.0 tool provides a methodology for accurate identification of transposase metagenomic data; (c) The accurate analysis of transposases in metagenomic data provided the identification of new ISs. Together, these results may assist in the discovery of new families of transposases which, as a last resort, might help reveal new transposition mechanism.

Keywords: the IS type elements; metagenome; bioinformatics; genomic annotation; comparative genomics.

LISTA DE ABREVIATURAS

EGMs	Elementos Genéticos Móveis
TGL	Transferência Gênica Lateral
DNA	Ácido Desoxirribonucléico
RNA	Ácido Ribonucléico
IS	Sequência de Inserção
ISNCY	<i>Not classified yet</i> – Não classificado ainda
LTRs	Longas Terminações Repetidas
IRs	Repetições Terminais Invertidas
DRs	Repetições Diretas
ORF	Fase de leitura aberta
%	Porcentagem
RNA-r	Ácido Ribonucléico Ribossomal
SCW	<i>Sugarcane Wet</i> – Solo de cultivo cana-de-açúcar período chuvoso
SCD	<i>Sugarcane Dry</i> – Solo de cultivo cana-de-açúcar período de seca
FW	<i>Forest Wet</i> – Solo de floresta em período chuvoso
FD	<i>Forest Dry</i> – Solo de floresta em período de seca
pb	Pares de bases

LISTA DE TABELAS

	Página
Tabela 1. Dados meteorológicos mensais fornecidos pela estação convencional no ano de 2010, gentilmente cedidas pelo departamento de Ciências Exatas - FCAV/UNESP Jaboticabal-SP.	23
Tabela 2. Distribuição taxonômica da amostra SCW, de acordo com o programa ISsaga 2.0.....	39
Tabela 3. Distribuição taxonômica da amostra SCD, de acordo com o programa ISsaga 2.0.....	40
Tabela 4. Distribuição taxonômica da amostra FW, de acordo com o programa ISsaga 2.0.....	41
Tabela 5. Distribuição taxonômica da amostra FD, de acordo com o programa ISsaga 2.0.....	41

LISTA DE FIGURAS

	Página
Figura 1. Mecanismos de Transferência Gênica Lateral em organismos procarióticos: 1- Processo de transdução; 2-Processo de Conjugação; 3-Processo de transposição. Fonte: FROST et al., 2005.	9
Figura 2. Comparação genômica linear de <i>B. Pertussis</i> , <i>B. bronchiseptica</i> e <i>B. paraptussis</i> . Na parte superior, em <i>B. Pertussis</i> , os triângulos pretos representam ISs. No centro, em <i>B. bronchiseptica</i> , as caixas rosa representam profagos. Na parte inferior, em <i>B. paraptussis</i> , os triângulos pretos representam ISs. As linhas vermelhas entre os genomas representam semelhanças no DNA entre duas sequências (<i>Blastn matches</i>). Fonte: PARKHIL et al., 2003.....	11
Figura 3. Modelo canônico de um IS. DR (Repetições Diretas), IR (Repetições Invertidas).	12
Figura 4. Distribuição das famílias de ISs de acordo com o domínio protéico no banco de dados ISfinder. O histograma demonstra o número de ISs por família, depositadas no banco de dados até o ano de 2013. As linhas horizontais indicam o número e o tamanho relativo de diferentes subgrupos dentro de uma família. Fonte: SIGUIER; GOURBEYRE; CHANDLER, 2014.	13
Figura 5. Sequências de inserção e seus derivados. As setas de cor azul escuro correspondem ao gene da transposase, as setas de coloração alaranjada correspondem aos genes passageiros, as setas de cor azul claro, nas extremidades, são as repetições terminais invertidas. Fonte: SIGUIER; GOURBEYRE; CHANDLER, 2014.	14
Figura 6. Conjunto de dados utilizados na prospecção de transposases. “A”- Quantidade de <i>reads</i> gerados para cada amostra metagenômica “B”- Quantidade de <i>reads</i> identificados como transposases de acordo com a plataforma MG-RAST. SCW (<i>Sugarcane Wet</i>), SCD (<i>Sugarcane Dry</i>), FW (<i>Forest Wet</i>) e FD (<i>Forest Dry</i>). ...	26
Figura 7. Classificação de transposases (<i>reads</i>) mais representativas atribuídas pela plataforma MG-RAST. Valores totais de hits encontrados para cada uma das amostras foram 1496 (SCW – <i>Sugarcane Wet</i>), 1691 (SCD – <i>Sugarcane Dry</i>), 1599 (FW – <i>Forest Wet</i>) e 1639 (FD – <i>Forest Dry</i>).	28
Figura 8. Porcentagem dos gêneros taxonômicos mais abundantes, identificados a partir dos <i>reads</i> classificados como transposases, pela plataforma MG-RAST para as	

amostras SCW (<i>Sugarcane Wet</i>), SCD (<i>Sugarcane Dry</i>), FW (<i>Forest Wet</i>) e FD (<i>Forest Dry</i>).	32
Figura 9. Classificação de transposases (<i>scaffolds</i>) mais representativas atribuídas pela plataforma MG-RAST para as amostras SCW (<i>Sugarcane Wet</i>), SCD (<i>Sugarcane Dry</i>), FW (<i>Forest Wet</i>) e FD (<i>Forest Dry</i>).	34
Figura 10. Porcentagem dos gêneros taxonômicos mais abundantes, identificados a partir dos <i>scaffolds</i> classificados como transposases, pela plataforma MG-RAST para as amostras SCW (<i>Sugarcane Wet</i>), SCD (<i>Sugarcane Dry</i>), FW (<i>Forest Wet</i>) e FD (<i>Forest Dry</i>).	35
Figura 11. “A”- Quantidade de <i>scaffolds</i> gerados para cada amostra metagenômica. “B”- Quantidade de sequências de transposases identificadas em <i>scaffolds</i> através da plataforma ISSaga 2.0. SCW (<i>Sugarcane Wet</i>), SCD (<i>Sugarcane Dry</i>), FW (<i>Forest Wet</i>) e FD (<i>Forest Dry</i>).	36
Figura 12. Classificação de transposases (<i>scaffolds</i>) mais representativas atribuídas pela plataforma ISSaga 2.0 para as amostras SCW (<i>Sugarcane Wet</i>), SCD (<i>Sugarcane Dry</i>), FW (<i>Forest Wet</i>) e FD (<i>Forest Dry</i>).	38
Figura 13. “A”- Quantidade de transposases identificadas pela plataforma MG-RAST. “B”- Transposases identificadas pelo MG-RAST após filtragem pelo banco de dados ISfinder. SCW (<i>Sugarcane Wet</i>), SCD (<i>Sugarcane Dry</i>), FW (<i>Forest Wet</i>) e FD (<i>Forest Dry</i>).	42
Figura 14. Classificação de transposases (<i>reads</i>) mais representativas atribuídas pela plataforma MG-RAST após filtragem através do banco de dados de sequências de inserção ISfinder para as amostras SCW (<i>Sugarcane Wet</i>), SCD (<i>Sugarcane Dry</i>), FW (<i>Forest Wet</i>) e FD (<i>Forest Dry</i>).	45
Figura 15. Representação esquemática do contexto genômico de ISs pertencentes à amostra SCW (<i>Sugarcane Wet</i>). A escolha das sequências obedece à critérios citados no texto.	49
Figura 16. Representação esquemática do contexto genômico de ISs pertencentes à amostra SCD (<i>Sugarcane Dry</i>). A escolha das sequências obedece à critérios citados no texto.	50
Figura 17. Representação esquemática do contexto genômico de ISs pertencentes à amostra FW (<i>Forest Wet</i>). A escolha das sequências obedece a critérios citados no texto.	51

Figura 18. Representação esquemática do contexto genômico de ISs pertencentes à amostra FD (*Forest Dry*). A escolha das sequências obedece a critérios citados no texto.**52**

1 INTRODUÇÃO

Na década de 1940, Bárbara McClintock mudou o conceito de que os genes são estáticos, descrevendo a existência de Elementos Genéticos Móveis (EGMs). Os efeitos da transposição foram observados através da mudança na coloração do pericarpo dos grãos de milho e a quebra do braço curto do cromossomo nove, visualizado através de microscopia óptica. Bárbara McClintock estudava dois *loci* (*Ds* e *Ac*) dominantes que estavam associados ao padrão herdável e instável da coloração mosaico em pericarpos de milho. McClintock descobriu que o alelo *Ds* causava as quebras cromossômicas e que *Ds* e *Ac* poderiam se transpor, ou seja, mudar de posição dentro do cromossomo. Seus estudos revelaram que estes elementos (*Ds* e *Ac*), regulam os genes situados em suas proximidades, inibindo ou modulando sua expressão (FESCHOTTE et al., 2002).

Os EGMs são segmentos de DNA, com capacidade de mobilização dentro de um genoma ou ainda entre genomas de diferentes células. Podem ser adquiridos através de processos de transferência gênica lateral (TGL), ou seja, processos em que ocorre a troca de material genético entre células ou genomas de espécies não relacionadas. Possuem importante papel na evolução da variabilidade genética nos domínios biológicos Archaea, Bacteria e Eukarya (BINNEWIES et al., 2006; FROST et al., 2005).

Os procariotos são organismos unicelulares que englobam os domínios Bacteria e Archaea. Exibem uma imensa diversidade biológica e são amplamente distribuídos na biosfera terrestre (BINNEWIES et al., 2006). O estudo desta grande diversidade microbiológica tem crescido nos últimos anos, tendo como foco a evolução, principalmente nos processos mediados pelos Elementos Genéticos Móveis (EGMs) (FROST, 2005). Estes estudos estão fornecendo contribuições fascinantes e importantes para a nossa compreensão sobre a evolução da vida em nosso planeta.

Atualmente as análises genômicas e evolutivas relacionadas com a diversidade microbiana de diversos ambientes têm crescido exponencialmente. Historicamente, estes estudos tinham como fundamento a suposição de que através do cultivo, a maioria dos micro-organismos de um determinado ambiente seria

identificada. Entretanto, foi demonstrado, que os organismos cultivados correspondem apenas a uma pequena fração da diversidade de espécies em uma comunidade microbiana (LING et al., 2015). Neste contexto, a Metagenômica, (denominação dada ao conjunto de genomas da microbiota total encontrada em um determinado ambiente) tornou-se uma ferramenta importante para a caracterização de micro-organismos, já que independe do cultivo em laboratórios. As técnicas empregadas pela metagenômica partem do princípio do isolamento do DNA, portanto, possibilitando a descoberta das complexas interações das comunidades microbianas (REIGSTAD et al., 2011).

Diante desta realidade, o estudo dos elementos genéticos móveis é de grande importância, já que pesquisas revelam serem os principais agentes de transmissão de, por exemplo, resistência a antibiótico, fatores de virulência, patogenicidade e adaptação ao meio ambiente, além de promover impactos no genoma, tais quais rearranjos e inversões genômicas, interrupção de genes e modificações de regiões promotoras. No contexto da microbiologia agropecuária, este estudo poderá contribuir em vários aspectos, sendo esses relacionados com o avanço do conhecimento na área destes elementos genéticos móveis, que poderá levar a uma melhor compreensão dos processos evolutivos existentes na microbiota do solo e de sua participação na transmissão de características que podem conferir vantagens para os micro-organismos, tais como potenciais genes de resistência ou adaptação ao meio, dentre outros.

1.1 Objetivos

1.1.1 Objetivo Geral

O presente estudo teve como objetivo avaliar a diversidade de EGMs do tipo Sequências de Inserção (IS) em organismos procariotos presentes em uma biblioteca metagenômica, proveniente de solo com cultivo de cana-de-açúcar e solo de floresta através de duas frentes de investigação. Além disso, comparar as ferramentas de bioinformática utilizadas no estudo de dados metagenômicos, e verificar possibilidades de identificação de novas sequências de inserção (ISs).

1.1.2 Objetivos Específicos

1. Prospectar em biblioteca metagenômica *“in silico”* elementos genéticos móveis do tipo IS de organismos procariotos de amostras provenientes de solo com cultivo de cana-de-açúcar e solo de floresta;
2. Realizar a anotação dos genes encontrados;
3. Comparar as ferramentas de bioinformática utilizadas no estudo de dados metagenômicos;
4. Avaliar o potencial destes dados para identificação de novas famílias de transposases;
5. Analisar o Contexto Genômico das transposases.

2 REVISÃO DE LITERATURA

2.1 Os Elementos Genéticos Móveis (EGMs)

Os elementos genéticos móveis (EGMs) são caracterizados como segmentos de DNA que possuem a capacidade de se movimentar entre genomas distintos ou em um mesmo genoma e inserir-se em novas regiões, mediados por proteínas codificadas pelo próprio elemento (FROST et al., 2005). Estes EGMs podem ser divididos em dois grupos principais: aqueles que são transmissíveis de célula para célula, como plasmídeos e bacteriófagos, e aqueles que não podem submeter-se a transferência por conta própria, mas que podem ser transportados quando integrados em elementos do primeiro grupo (CRAIG et al., 2002). São definidos como EGMs que podem fazer parte do genoma microbiano: os transposons, bacteriófagos (CAMPBELL, et al., 1979), plasmídeos, íntegrans e as ilhas genômicas (FROST et al., 2005). Embora as atividades de transposição dependam de diversos fatores, cada elemento particular apresenta mecanismos de transposição distintos (BIRKENBIHL; VIELMETTER, 1989).

De acordo com Finnegan (1989), os EGMs do tipo transposons, são classificados em duas classes, relacionadas à sua forma de transposição: Classe I – (retrotransposons) e Classe II – (transposons de DNA). Os elementos da Classe I, chamados coletivamente de retrotransposons, realizam a transposição através da síntese de uma molécula de RNA mensageiro que é utilizada como molde para síntese de uma nova cópia de DNA através de um transcriptase reversa codificada pelo próprio retroelemento (FINNEGAN, 1989). Estes elementos podem ser subdivididos em dois grupos: Retrotransposons com e sem LTR (Longas Terminações Repetidas), ambos apresentando transposição replicativa. Os Retrotransposons com LTRs codificam uma grande quantidade de proteínas importantes para sua transposição através de um único RNA mensageiro, apresentando as regiões promotoras e terminadoras de transcrição do RNA mensageiro dentro das LTRs. Os Retrotransposons sem LTR, normalmente, possuem em suas extremidades regiões não codificantes e uma cauda poli-A,

enquanto que em sua região codificante encontram-se geralmente, uma endonuclease e uma transcriptase reversa, entre outros genes (FINNEGAN, 1989).

Os elementos da Classe II (Transposos de DNA), por sua vez, têm como características comuns: sequências nucleotídicas muito similares nas extremidades (repetições terminais invertidas – IRs) carregam genes que codificam enzimas capazes de transportá-los, criam ao se inserir no DNA pequenas duplicações no sítio alvo de inserção (repetições diretas – DRs) e existem em cópias múltiplas no genoma. Podem apresentar transposição conservativa ou replicativa, mediada pela enzima transposase. A transposição conservativa mantém o número de cópias do elemento no genoma, por ser uma excisão completa, com posterior inserção em outra região do genoma. A enzima transposase, que é constituinte dos transposons, é responsável por catalisar seu movimento e realizar a transposição através da quebra e re-ligação da cadeia fosfodiéster da molécula de DNA. As sequências de inserção (ISs) são transposons que compreendem normalmente um único gene que codifica para a transposase. Os transposons conjugativos carregam um ou mais genes além daqueles responsáveis pela sua transposição, possuem a capacidade de inserção e excisão de um cromossomo e podem ser transpostos através da conjugação bacteriana. Não possuem uma forma extracromossomal replicativa como os plasmídeos, e por este motivo, integram-se ao cromossomo hospedeiro (KAPER; HACKER, 1999).

Os elementos da Classe II, do tipo sequência de inserção, são amplamente encontrados nos domínios biológicos Bactéria e Archea, enquanto os elementos classe I, os retrotransposons, são encontrados basicamente em Eukarya (CRUZ, 2012).

Além dos transposons, há outros EGMs, os quais estão descritos abaixo:

Os plasmídeos são elementos de DNA extracromossômico que se replicam de maneira autônoma e autocontrolada (MARQUES, 2012). Em geral, carregam genes não essenciais à sobrevivência da bactéria, mas muitos destes genes conferem algumas propriedades que promovem vantagens adaptativas, e, portanto, aumentam sua sobrevivência em condições específicas. Entre as características mais comuns conferidas por plasmídeos, estão: resistência (a antibióticos, metais pesados, enzimas, etc); alteração do metabolismo energético (genes que permitem

a utilização de fontes alternativas de nutrientes, catabolismo e anabolismo); patogenicidade e simbiose (produção de endo/exotoxinas transporte de enzimas, produção de cápsula, nodulação); disseminação e perpetuação (conjugação, replicação, imunidade) (MARQUES, 2012).

Os íntegrans são elementos que estão relacionados com a resistência antimicrobiana. Não são capazes de se mobilizar por conta própria, porém, são comumente encontrados em plasmídeos e transposons, elementos responsáveis por sua mobilização (CAMBRAY, et al., 2010) .

Os bacteriófagos (também chamados simplesmente de fagos) são vírus que infectam bactérias, apresentam genomas compostos por DNA ou RNA. De maneira geral, o material genético do fago é envolvido por uma estrutura proteica, chamada capsídeo, alguns possuem duas partes bem definidas, como a cabeça e a cauda, enquanto outros são filamentosos. A infecção da bactéria pelo vírus se inicia com a adsorção deste a receptores na superfície bacteriana. Ao se ligarem aos receptores, os bacteriófagos podem injetar o material genético para dentro da bactéria. Uma vez dentro do citoplasma, sua replicação é iniciada e pode gerar novas partículas virais, ou integrar-se no cromossomo bacteriano. (MARQUES, 2012). São veículos de transferência de genes através do processo de transdução (FROST et al., 2005).

As ilhas genômicas são regiões cromossômicas, adquiridas por transferência horizontal, flanqueadas por repetições diretas (BELLANGER et al., 2013). Podem ser classificadas de acordo com as vantagens que agregam à bactéria, sendo subdivididas principalmente em:

- *Ilhas adaptativas*: promovem aumento na adaptação do organismo ao meio;
- *Ilhas de patogenicidade*: possuem genes cujos produtos contribuem para a virulência da bactéria;
- *Ilhas de simbiose*: possuem genes relacionados à simbiose (associação entre dois seres vivos em que ambos são beneficiados, necessitando um ao outro para sobreviver) com organismos;
- *Ilhas de resistência*: possuem genes responsáveis pela degradação de antibióticos (MARQUES, 2012).

Sabe-se atualmente que os mecanismos de TGL desempenham um papel fundamental na evolução bacteriana, como gerador da variabilidade genética.

Bactérias e Árqueas exibem uma extraordinária diversidade fenotípica, propriedades metabólicas, estruturas celulares e estilos de vida variáveis, mesmo entre organismos da mesma espécie (OCHMAN, 2000).

Existem três formas de TGL: transformação, transdução e conjugação (Figura 1). Transformação foi o primeiro mecanismo de TGL descoberto. Este mecanismo envolve a transferência de DNA entre bactérias relacionadas. A troca de material genético ocorre através da lise celular de uma bactéria doadora, para uma bactéria receptora apta a receber os segmentos disponibilizados no meio (FROST et al., 2005)

A transdução ocorre através da ação de bacteriófagos durante sua replicação dentro da bactéria, onde acidentalmente podem apropriar segmentos de DNA hospedeiro em seu capsídeo e inserir em outro hospedeiro após a lise celular da bactéria. (FROST et al., 2005). Os fagos, ao infectarem uma bactéria podem iniciar sua replicação imediatamente e entrar no ciclo lítico, ou podem integrar-se ao genoma, dando início ao ciclo lisogênico. O ciclo lítico ocorre quando os bacteriófagos se replicam em grande quantidade, e acabam lisando as bactérias infectadas para liberar as partículas virais, com a finalidade de infectar novos hospedeiros. O ciclo lisogênico é o processo de incorporação do fago no genoma bacteriano. Os bacteriófagos que realizam o ciclo lisogênico são chamados de fagos temperados, e o DNA do fago integrado é chamado de profago (MARQUES, 2012).

Na conjugação, a transferência ocorre através do contato célula-a-célula, entre uma bactéria doadora e uma bactéria receptora e é característico dos mecanismos de propagação de plasmídeos capazes de realizar sua replicação e mobilização independente, no cromossomo hospedeiro (THOMAS; NIELSEN, 2005). É possível que genes relacionados à virulência e patogenicidade sejam integrados no cromossomo ou plasmídeo através da ação mediadora dos EGMs (FURUYA; LOWY, 2006). Os plasmídeos conjugativos têm sido citados como o vetor particularmente comum e muito bem sucedido na transferência de genes (NORMAN et al., 2009).

O processo de conjugação se inicia com a presença de um plasmídeo em uma bactéria, que irá expressar os genes necessários para a montagem de uma

fímbria especializada, o chamado *pilus* sexual. Esta bactéria, chamada de doadora, então se torna apta a transferir o plasmídeo para uma bactéria que não possua um *pilus* sexual, chamada de bactéria receptora. O *pilus* constitui um sistema de secreção tipo IV, que transmite proteínas efetoras entre várias espécies bacterianas, e é composto de várias proteínas que formam um canal que atravessa a membrana externa, o periplasma e a membrana interna. Estas proteínas são codificadas por genes que formam em geral um único óperon. O *pilus* sexual reconhece receptores específicos na bactéria receptora, ligando-se a ela, e formando assim uma conexão entre as duas células. A seguir o *pilus* conjugativo é encurtado, resultando na aproximação das células. Uma cópia do plasmídeo é feita por replicação do DNA e é passada da doadora para a receptora através de um canal especializado. Assim, a bactéria doadora não perde o plasmídeo e a bactéria receptora se torna uma potencial doadora para outras que não o possuem (MARQUES, 2012).

De acordo com Silva e colaboradores (2004), a incidência de eventos de transferência horizontal é significativamente maior entre os elementos da Classe II do que entre os elementos da Classe I. O motivo principal é a forma de transposição de DNA a DNA dos elementos da Classe II. A principal fonte de variabilidade genética realmente atuante sobre os elementos genéticos móveis é a transferência gênica lateral (Figura 1) (SILVA; KIDWELL, 2000).

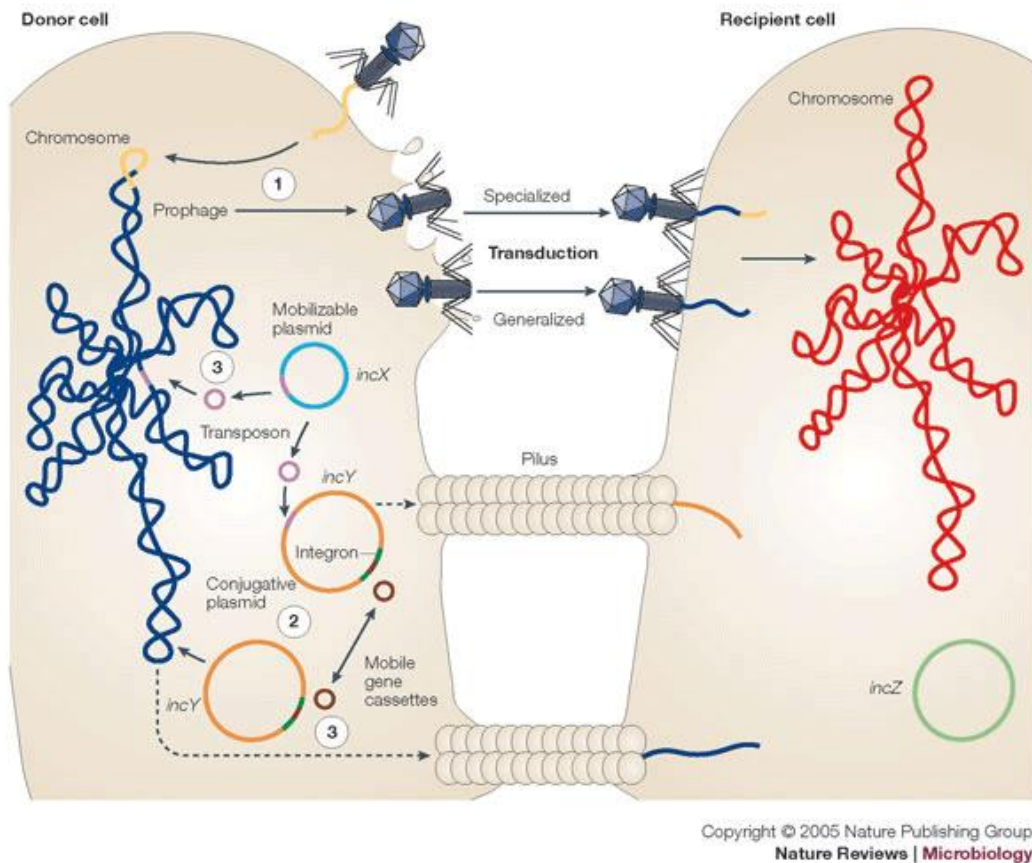


Figura 1. Mecanismos de Transferência Gênica Lateral em organismos procarióticos: 1- Processo de transdução; 2-Processo de Conjugação; 3-Processo de transposição. Fonte: FROST et al., 2005.

A regulação dos elementos transponíveis no genoma ocorre de três diferentes formas, por auto-regulação (o próprio elemento restringe seu movimento de acordo com sua abundância no genoma), controlada por fatores do hospedeiro ou eliminação através da ação da seleção natural (ALMEIDA; CARARETO, 2005).

Outro aspecto, pouco estudado sobre o impacto dos EGMs, é a domesticação (EGMs perdem a capacidade de transpor-se devido à perda ou degeneração de parte de suas sequências estruturais, como por exemplo, inversões repetidas). Neste mecanismo o hospedeiro pode utilizar as propriedades deste elemento para executar suas funções celulares. Este processo foi descrito em eucariotos (SINZELLE, et al., 2009), sendo raro em procariotos. Este fenômeno faz dos elementos transponíveis importantes no processo evolutivo do genoma hospedeiro, pois, estes elementos domesticados também apresentam altos níveis de expressão

dos genes, o que é incomum para um elemento transponível (ALMEIDA; CARARETO, 2005).

2.2 Impactos dos EGMs

Um exemplo da influência dos EGMs é o recente estudo realizado com a bactéria patogênica humana *Clostridium difficile* demonstrando a presença de grande quantidade de EGMs responsáveis pela patogenicidade e virulência. Os principais fatores de virulência desta bactéria estão contidos em ilhas genômicas e bacteriófagos que modulam a expressão gênica de toxinas e genes de resistência a antibióticos. (MULLANY; ALLAN; ROBERTS, 2015).

O estudo realizado por Varani e colaboradores (2013) sobre o papel dos fagos em bactérias fitopatogênicas, corrobora sobre o potencial destes elementos na diferenciação dos genomas. Um exemplo particular é a bactéria *Xylella fastidiosa*, foi observado que fagos foram inseridos entre os genes que codificam proteínas nos genomas de estirpes da América do Norte (Temecula1, GB514, M23, M23) estão ausentes nas estirpes da América do Sul (9a5c), promovendo assim, a diferenciação destes genomas

Outro exemplo é a bactéria *Pectobacterium atrosepticum*, um patógeno de batata, o qual foi identificado dois profagos em seu genoma, denominados ECA29 e ECA41, os quais possuem o papel biológico totalmente decifrado contribuindo para a patogenicidade (VARANI et al., 2013).

No trabalho realizado por Parkhil e colaboradores (2003), as espécies bacterianas: *Bordetella bronchiseptica*, *Bordetella parapertussis* e *Bordetella pertussis* foram analisadas. Estas espécies são responsáveis por colonizarem as vias respiratórias de mamíferos, causando doenças. Percebe-se uma variação da patogenicidade destas espécies e este estudo sugeriu que *B. bronchiseptica* é ancestral das outras duas espécies. As análises realizadas demonstraram que ISs foram responsáveis pela diferenciação genômica, sugerindo-se que a recombinação entre os ISs foi a principal causa dos rearranjos genômicos nas espécies, que

também ocasionaram diferenças quanto a virulência (PARKHIL et al., 2003) (Figura 2).

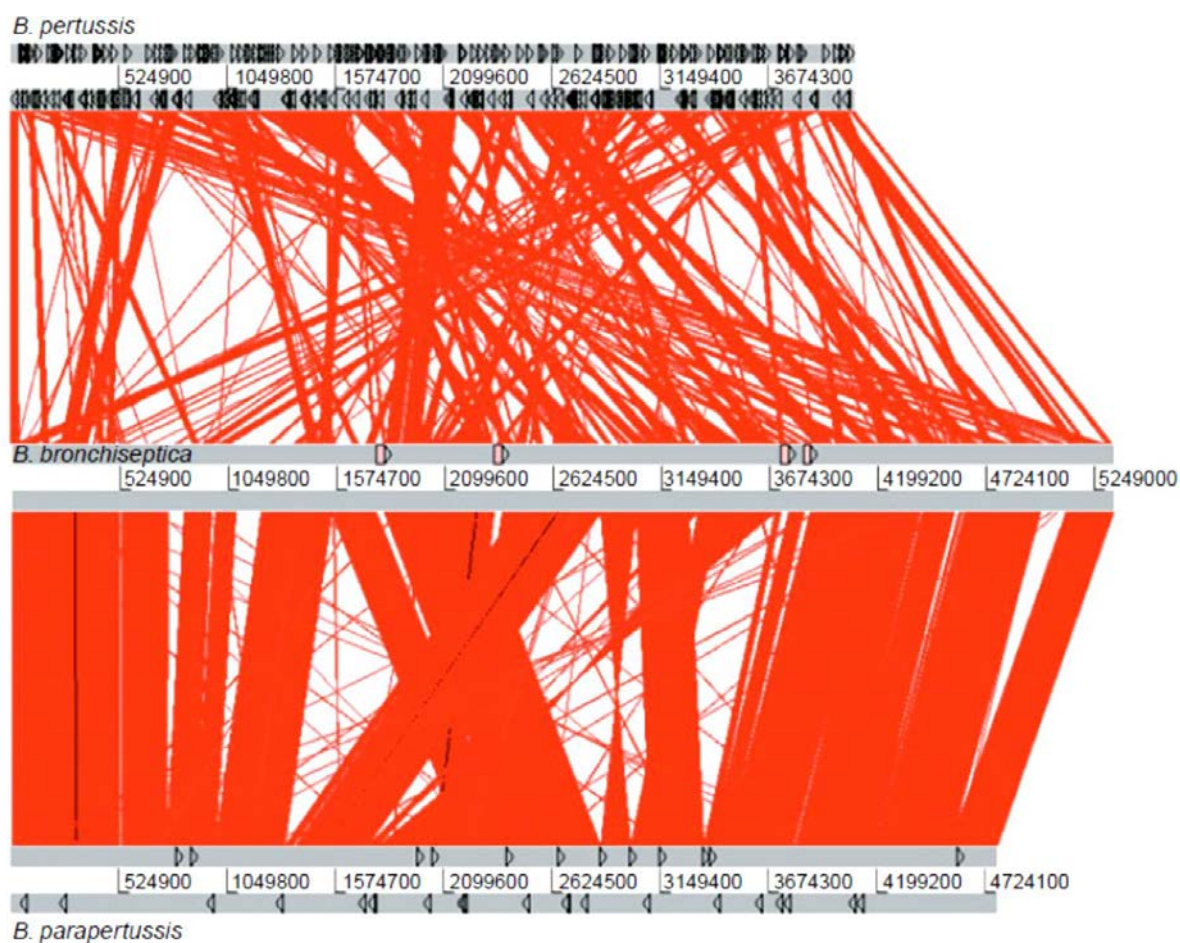


Figura 2. Comparação genômica linear de *B. pertussis*, *B. bronchiseptica* e *B. parapertussis*. Na parte superior, em *B. pertussis*, os triângulos pretos representam ISs. No centro, em *B. bronchiseptica*, as caixas rosa representam profagos. Na parte inferior, em *B. parapertussis*, os triângulos pretos representam ISs. As linhas vermelhas entre os genomas representam semelhanças no DNA entre duas sequências (*Blastn matches*). Fonte: PARKHIL et al., 2003.

2.3 Transposons do tipo Sequências de Inserção (IS)

Sequências de inserção (IS) são elementos genéticos móveis, caracterizados por uma estrutura simples e compacta, na maioria das vezes constituída por um único gene que codifica a enzima transposase, proteína responsável por sua própria mobilidade, (MAHILON; CHANDLER, 1998) com comprimento entre 0,7 e 3,5 Kb. São flanqueadas por Repetições Terminais Invertidas (IRs) que variam entre 5-200 pb, estas terminações contém as regiões promotoras completas ou parciais do elemento e são reconhecidas no momento de sua transposição, além disso, muitas ISs geram duplicações (DRs- *Repetições Diretas*) no sítio alvo, com comprimento entre 2 e 14 pb (CHANDLER; MAHILON, 2002) . A sua multiplicação pode causar variação no tamanho do genoma, ativar e inativar a expressão de genes adjacentes e realizar vários tipos de rearranjos genômicos, como deleções, inserções, translocações entre grandes regiões e também modificações pontuais ao longo das sequências (TREANGEN et al., 2009) (Figura 3).



Figura 3. Modelo canônico de um IS. DR (Repetições Diretas), IR (Repetições Invertidas). (IMAGEM CEDIDA POR MICHAEL CHANDLER)

Estudos recentes sugerem que a transposase é o gene mais abundante presente nos bancos de dados públicos de genomas e metagenomas procarióticos e eucarióticos (AZIZ et al., 2010). As sequências de inserção são classificadas em aproximadamente 30 famílias, baseando-se em variadas características compartilhadas (CHANDLER; MAHILLON, 2002), incluindo: 1- tamanho e sequência de repetições terminais invertidas (IRs); 2- tamanho e sequência de repetições diretas (DRs); 3- organização de suas ORFs; 4- sequências alvo em que se inserem. Mas o principal fator de classificação é de acordo com o domínio catalítico das

transposases, enzimas que catalisam seu movimento (SIGUIER; GOURBEYRE; CHANDLER, 2014).

As transposases se diferenciam com base em seu domínio proteico, que pode ser DDE (Ácido Aspártico, Ácido Aspártico e Ácido Glutâmico), DEDD (Ácido Aspártico, Ácido Glutâmico Ácido Aspártico e Ácido Aspártico), HuH (Tirosina recombinase) e Ser (Serina). A maioria das famílias codificam transposases do tipo DDE, ou seja, possuem a tríade de aminoácidos conservados Asp, Asp, Glu (HAREN et al., 1999) (Figura 4).

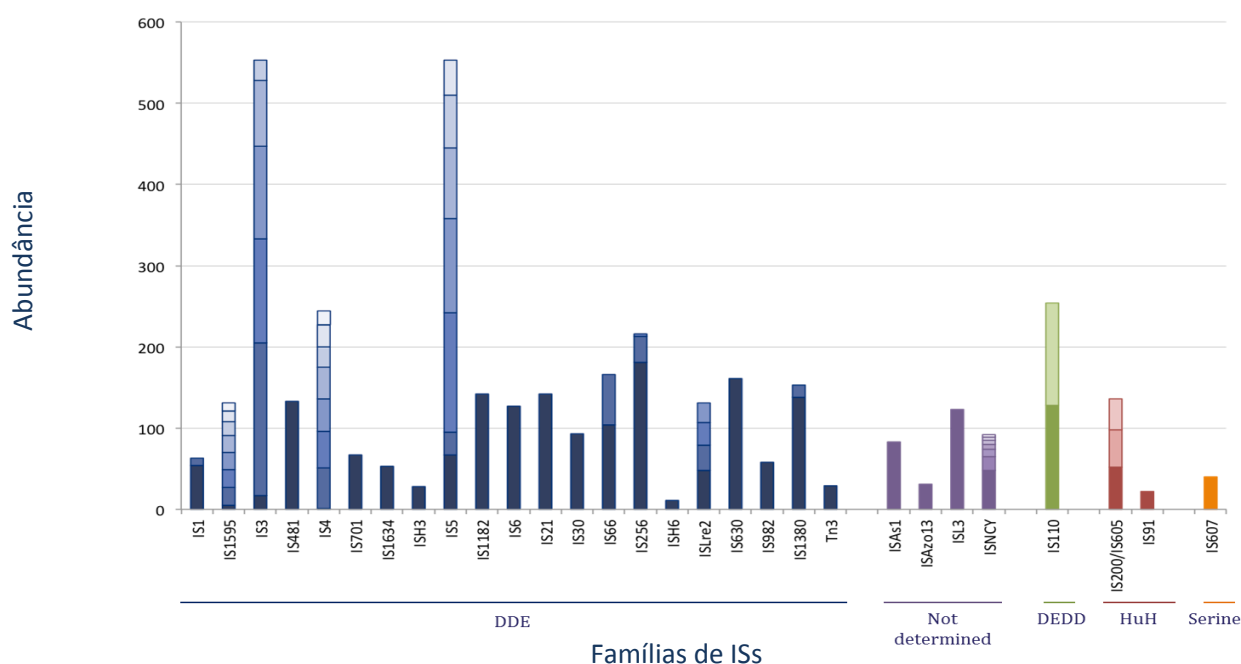


Figura 4. Distribuição das famílias de ISs de acordo com o domínio protéico no banco de dados ISfinder. O histograma demonstra o número de ISs por família, depositadas no banco de dados até o ano de 2013. As linhas horizontais indicam o número e o tamanho relativo de diferentes subgrupos dentro de uma família. Fonte: SIGUIER; GOURBEYRE; CHANDLER, 2014.

Muitas sequências de inserção também incluem genes acessórios envolvidos na regulação de sua transposição, sendo específicos para cada família e servindo também na classificação destes elementos (SIGUIER; GOURBEYRE; CHANDLER, 2014) (Figura 5). Muitos destes elementos são raramente anotados, conhecidos como MITEs (sequências que perderam sua transposase, permanecendo no

genoma apenas suas inversões terminais repetidas e repetições diretas) e MICs (sequências que adquiriram genes passageiros, e também perderam sua transposase) (Figura 5), e podem ser ativados por uma transposase intacta relacionada a uma sequência de inserção presente na célula, e assim, permanecem fornecendo a possibilidade de rearranjos genômicos. Além disso, sua presença pode prover cicatrizes de eventos de recombinação anteriores e sua análise detalhada é capaz de fornecer informações de grande importância para a compreensão dos diferentes genomas de uma única espécie ou espécies distintas (SIGUIER et al., 2012).

Também é possível encontrar ISs portadores de genes passageiros, os quais incluem genes reguladores de transcrição, metiltransferases e genes que conferem resistência a antibióticos. Podem ser encontrados ao lado do gene que codifica para a transposase e normalmente estão presentes em baixo número de cópias no genoma, muitas vezes em cópias únicas (deixando a dúvida se estão ativos ou não) (SIGUIER; GOURBEYRE; CHANDLER, 2014).

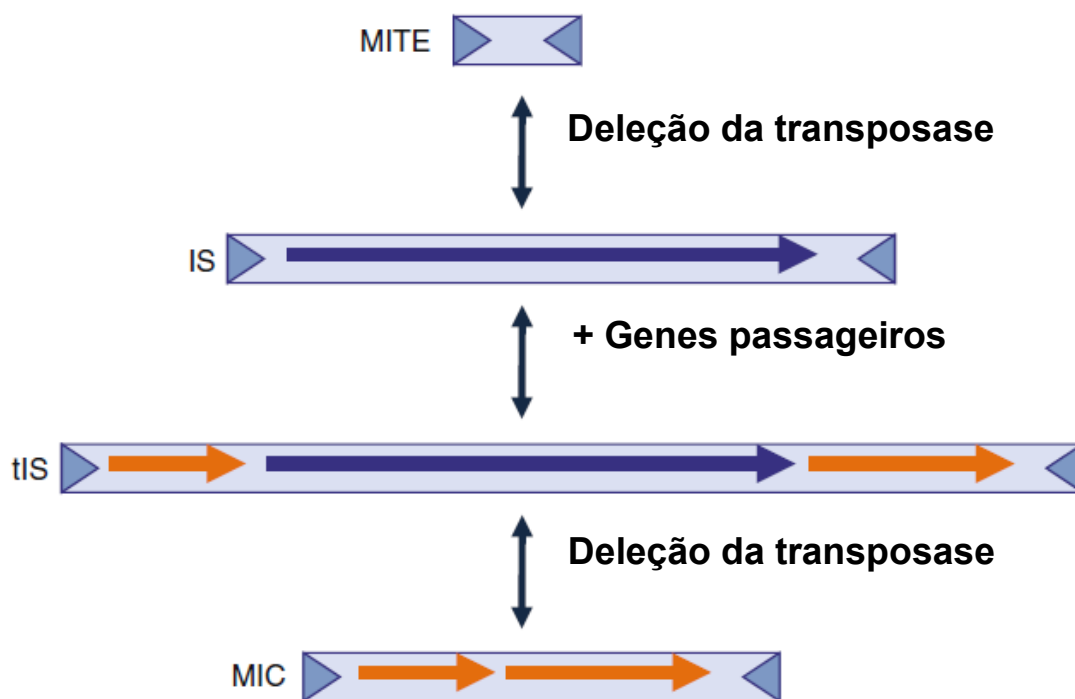


Figura 5. Sequências de inserção e seus derivados. As setas de cor azul escuro correspondem ao gene da transposase, as setas de coloração alaranjada correspondem aos genes passageiros, as setas de cor azul claro, nas extremidades, são as repetições terminais invertidas. Fonte: SIGUIER; GOURBEYRE; CHANDLER, 2014.

Os elementos transponíveis são raramente encontrados em sítios essenciais, ou seja, regiões que codificam genes importantes à sobrevivência do organismo, como as regiões regulatórias. As sequências da família IS4, normalmente se inserem na mesma região do óperon da galactosidase, as sequências da família IS1, prevalentes em regiões ricas nos nucleotídeos AT (ALMEIDA; CARARETO, 2005).

Fazendo um balanço, os ISs podem ser vistos de duas formas que se contrariam: em pequenas populações podem provocar mutações ou deleções que geram conseqüências negativas para seu nicho-ecológico, ou como agentes que promovem evolução adaptativa, pois em associação com outros elementos, elas podem mediar a transferência de informação genética, entre genomas ou em um mesmo genoma (PARKHIL et al., 2003) dessa forma, também desempenham papel no transporte de caracteres adaptativos, tais como por exemplo a resistência aos antibióticos (BOUTOILLE et al., 2004).

Podem ser encontradas no interior de plasmídeos, variando sua ocorrência de 1- 40% de acordo com o tamanho do plasmídeo. Uma explicação para isto é que os plasmídeos menores possuem apenas genes essenciais, enquanto que os plasmídeos maiores são formados por genes responsáveis pela transferência gênica e assim podem adquirir transposons, sequências de inserção e genes acessórios durante o trajeto de um genoma a outro. (SIGUIER et al., 2006). Inseridos em plasmídeos os transposons tornam-se transferíveis para outras bactérias (BOERLIN; REID-SMITH, 2008).

A presença de ISs no genoma, permite a recombinação homóloga entre elementos independentes, desde que cada elemento tenha uma cópia da mesma IS. Este mecanismo foi observado em *Escherichia coli* durante o processo de conjugação, quando o fator F sexual (proveniente do plasmídeo), integra-se ao cromossomo por recombinação entre as cópias de IS2 ou IS3 (THOMAS; NIELSEN, 2005).

Há um banco de dados público específico para armazenar ISs, conhecido como ISfinder (<https://www-is.biotoul.fr/>), que possui cerca de 4.000 sequências de inserção anotadas, o que representa apenas uma pequena quantidade do que ainda há para ser pesquisado, estimativas da equipe do ISfinder indicam mais de 30.000 ISs diferentes apenas nos genomas completos sequenciados disponíveis nas bases

de dados públicas, além de conter informações sobre a classificação de famílias de ISs conhecidas (SIGUIER et al., 2012).

2.4 Diversidade microbiana do solo

Os micro-organismos constituem a forma de vida mais abundante e diversificada do planeta Terra (WITHMAN; COLEMAN; WIEBE, 1998). É possível encontrar no solo micro-organismos pertencentes aos três domínios biológicos, sendo que a porção mais representativa é composta pelos domínios *Bacteria* e *Archea* (WARD, 1998; HUNTER-CEVERA, 1998).

A quantidade e a diversidade de micro-organismos podem variar de acordo com os diferentes tipos e condições do solo, prevalecendo na maioria das vezes, as bactérias (WHITMAN et al. 1998). Estudos comprovam que a diversidade de micro-organismos presente no solo é tão vasta quanto desconhecida. Um grama de solo pode apresentar cerca de 10 bilhões de micro-organismos das mais variadas espécies (ROSSELÓ-MORA; AMANN, 2001; DANIEL, 2005).

A principal fonte de matéria orgânica do solo é proveniente de resíduos vegetais, os micro-organismos são os principais responsáveis por realizar sua decomposição, sendo capazes de influenciar na dinâmica da formação do húmus e também na estocagem de carbono e nutrientes (MATSUOKA; MENDES; LOUREIRO, 2003).

As atividades microbianas estão baseadas em sua notável diversidade metabólica e plasticidade genética (KURTBOKE et al., 2004) o que os torna uma importante fonte de recursos genéticos para o avanço da biotecnologia e para o desenvolvimento sustentável (BORNEMAN et al., 1996).

Atualmente, metodologias moleculares, tais quais como a metagenômica, têm contribuído massivamente no estudo da diversidade microbiana dos solos (PEDRINHO et al., 2009).

2.5 Biblioteca Metagenômica

A partir da década de 1980 microbiologistas ambientais perceberam que somente uma pequena parcela de bactérias do solo, variando entre 0,1% a 1,0%, era capaz de formar colônias em meios de cultivo (TRINGE; RUBIN, 2005; TORSVIK, et al. 2002). Estima-se que menos de 10% dos micro-organismos existentes no planeta tenham sido caracterizados e descritos (STALEY, 1998), 90 % dos micro-organismos existentes não foram cultivados em laboratório, pois poucos são cultiváveis (ATLAS; BARTHA, 1997).

A metagenômica é uma nova tecnologia criada como forma de facilitar o acesso ao estudo dos micro-organismos sem necessitar técnicas de cultivo. A extração de ácidos nucléicos totais e análise das sequências-alvo nas amostras mistas de DNA de diversos micro-organismos vêm sendo adotada como metodologia padrão para o estudo da diversidade microbiana do solo (HUGENHOLTZ et al., 1998).

A partir da amplificação, clonagem e análise de segmentos de DNA microbiano do solo, pode-se avaliar a fisiologia e a função de micro-organismos na natureza. No solo, estudos envolvendo as sequências 16S do RNA/DNA ribossomal com análises metagenômicas, se constituíram nos primeiros passos para a ligação da fisiologia e função de sua microbiota total (RONDON, et al, 1999).

Através das bibliotecas metagenômicas, é possível explorar a diversidade genética de micro-organismos em determinado meio e novos genes que produzem produtos biotecnológicos estão sendo encontrados (HENNE et al., 2000; HONG et al., 2007), além de oferecer oportunidades de compreender a interação de comunidades microbianas e contribuir no estudo da conservação do solo (MOCALI; BENEDETTI, 2010).

Um exemplo foi o estudo realizado por Bingyong Mao e colaboradores (2015), cujo interesse foi analisar os efeitos de fruto-oligossacarídeos (um tipo de pró-biótico) sobre as bactérias do intestino de camundongos, investigadas através da análise metagenômica de sequências 16S RNA-r. A observação mais surpreendente foi uma estirpe da bactéria *Olsenella*, confirmada pela primeira vez utilizando fruto-

oligossacarídeos.

Atualmente também tem se realizado, grande quantidade de trabalhos comparando-se diferentes ecossistemas, pois, entre toda a comunidade microbiana existente, os procariotos, ainda são uma incógnita para a ciência. Um exemplo é o estudo realizado por Christopher Staley e colaboradores (2015), cujo foco, foi a análise da resistência das comunidades bacterianas a metais pesados no Rio Mississippi, Estados Unidos.

2.6 Prospecção gênica em bibliotecas metagenômicas através da Bioinformática

Com o advento da metagenômica, uma grande quantidade de micro-organismos foi descoberta, porém, a capacidade para produção de informações era muito limitada no início.

A primeira tecnologia de sequenciamento do DNA foi desenvolvida por Sanger e colaboradores (1977), a qual se baseava no método por síntese enzimática de uma fita complementar, possibilitando assim, a evolução dos primeiros projetos de sequenciamento de genomas.

Com o passar dos anos, novas tecnologias de sequenciamento de DNA foram desenvolvidas visando o sequenciamento em larga escala, são os chamados “sequenciadores de nova geração” (CARVALHO, SILVA, 2010).

O sequenciador *Illumina*, assim como também, a plataforma de Sanger, baseia-se no conceito de “sequenciamento por síntese”, diferenciando-se pela capacidade de gerar maior quantidade de informação. Esta técnica utiliza a enzima DNA polimerase e nucleotídeos terminadores marcados com diferentes fluoróforos. A novidade da plataforma *Illumina* é a clonagem *in vitro* dos fragmentos em uma plataforma sólida de vidro, processo também conhecido como PCR de fase sólida (TURCATTI, et al., 2008; CARVALHO, SILVA, 2010).

Atualmente, em decorrência desses avanços em técnicas de seqüenciamento mais eficientes, os genomas de milhares de espécies se encontram em bancos de

dados e ferramentas (ALTSCHUL et al., 1990).

O maior desafio atual é fazer com que além de acumular sequências, seja possível entender o funcionamento de novos genes e proteínas para seus nichos e papel nos ciclos globais (STREIT; SCHMITZ, 2004). A acumulação maciça de genomas bacterianos sequenciados está oferecendo oportunidades para o estudo da organização, e compreensão destes genomas. Esta área de investigação depende da comparação genômica e anotação precisa dos genomas a serem comparados. Além da anotação de ORFs (*Open Reading Frame*), ou seja, sequências codificadoras, uma correta anotação deve incluir também outros elementos essenciais, como elementos genéticos móveis (EGM) (CRAIG, et al., 2002).

Atualmente, sabe-se que grande parte dos genomas depositados nos bancos de dados públicos não apresenta informação concisa e completa a respeito da anotação dos EGMs, principalmente em relação a ISs e profagos. A informação de anotação se limita apenas a uma breve descrição da função proteica, sendo geralmente atribuída de forma automática por programas de computador, sem uma posterior etapa de validação manual ou curadoria (MOCALI; BENEDETTI, 2010).

Sendo assim, o sucesso da utilização da bioinformática em projetos de metagenômica de solo depende de uma combinação de vários aspectos como, seleção de amostra adequada, métodos eficientes de extração de DNA, clonagem, estratégias eficientes de rastreamento e sequenciamento, juntamente com a melhoria do sistema de gerenciamento de dados, principalmente valorizando-se a correta anotação de sequências através de uma precisa curadoria e novas ferramentas da bioinformática (MOCALI; BENEDETTI, 2010).

2.7 MG-RAST e ISsaga 2.0

A plataforma de análise de dados metagenômicos MG-RAST, utilizada na primeira frente de investigação deste trabalho, possibilita uma análise automatizada de metagenomas, fornecendo conhecimentos quantitativos sobre populações

microbianas, através da comparação de sequências com bases de dados já conhecidas. A plataforma MG-RAST foi lançada em 2007 e possui mais de 12.000 usuários registrados (MEYER, et al., 2008).

A plataforma MG-RAST atribui características funcionais das sequências metagenômicas comparando bancos de dados de proteína e nucleotídeos. Resumos filogenéticos e funcionais dos metagenomas são gerados, além disso, há ferramentas para análises comparativas de metagenomas. O servidor é implementado em Perl, incluindo o SEED, BLAST, SQLite e Sun Grid Engine, como componentes (MEYER, et al., 2008).

A plataforma ISSaga 2.0, utilizada na segunda frente de investigação deste trabalho, é uma ferramenta desenvolvida para alta qualidade de anotação de ISs, sendo possível o processamento de amostras de genomas e metagenomas e comparação de sequências com o banco de dados ISfinder. Esta ferramenta foi desenvolvida especificamente para análise de ISs, fornecendo também informações sobre o contexto genômico destes elementos (VARANI et al.; 2011).

O ISSaga é uma das seções do ISfinder, definido como "*Genomes*", que também inclui o ISbrowser, uma ferramenta de visualização genomas de ISs, que atualmente contém mais de 40 genomas anotados. É um serviço baseado na web que inclui um conjunto de métodos para a identificação de ISs e está disponível gratuitamente para a comunidade acadêmica. Quanto maior a quantidade de ISs adicionados ao ISfinder, maior será a eficiência de anotação de genomas subsequentes e, portanto, o banco de dados depende de contribuições da comunidade científica (VARANI et al.; 2011).

O sistema de anotação semi-automático do ISSaga, utiliza o Blast, algoritmo em dois módulos: proteína e nucleotídeos. Cada módulo consiste em um grupo de programas como o *BioPerl*, *Bourne Shell* e linguagem PHP, executados no http *Apache manager* (versão 2.2.12), juntamente com uma base de dados implementada pela MySQL (versão 5.1.37) (VARANI et al.; 2011).

No presente trabalho, foram utilizadas quatro amostras metagenômicas provenientes de um projeto anterior desenvolvido no Laboratório de Bioquímica de Micro-organismos e Plantas da Faculdade de Ciências Agrárias e Veterinárias de Jaboticabal, pela Dra. Eliamar Aparecida Nascibém Pedrinho. Com estas amostras,

foi realizada a prospecção metagenômica de EGMs, do tipo IS, utilizando as plataformas de análise de dados MG-RAST e ISsaga 2.0, com a finalidade de descrever as famílias de ISs encontradas, identificar novos elementos e compreender os processos evolutivos existentes na microbiota do solo provenientes de ambientes e estações climáticas distintas.

3 MATERIAL E MÉTODOS

A prospecção metagenômica foi realizada em quatro amostras provenientes do Laboratório de Bioquímica de Micro-organismos e Plantas da Faculdade de Ciências Agrárias e Veterinárias de Jaboticabal, coordenado pela Prof.^a Dra. Eliana Gertrudes de Macedo Lemos. Estas amostras são provenientes de um trabalho anterior, realizado pela Dra. Eliamar Aparecida Nascibém Pedrinho, no trabalho intitulado “Diversidade Microbiana e Construção de um ecochip para avaliações ambientais”, processo FAPESP (n.º 09/54274-9) vinculado ao projeto temático (08/58114-3). Os subitens a seguir (3.1, 3.2, 3.3) são etapas integrantes do trabalho realizado pela Dra. Eliamar. Os subitens (3.4 e 3.5) são etapas realizadas no presente trabalho.

3.1 Coleta das amostras

As coletas foram realizadas na cidade de Santa Rita do Passa Quatro, interior do Estado de São Paulo, em duas áreas distintas: a primeira em área de cultivo de cana-de-açúcar localizado em uma propriedade particular (Fazenda Coqueiro), (21°42'30.67"S 47°34'56.89"W) e a outra amostra foi coletada em uma Floresta Nativa (Gleba Capetinga Leste do Parque Estadual de Vassununga), (21°42'34.62"S 47°34'30.04"W). Duas coletas foram realizadas em períodos sazonais distintos, a primeira ocorreu no mês de Fevereiro (período chuvoso) e a segunda em Novembro (período de seca) (Tabela 1).

As amostras coletadas de solo com cultivo de cana-de-açúcar foram denominadas SCW (*Sugarcane Wet*) e SCD (*Sugarcane Dry*), e as amostras coletadas de solo de floresta foram denominadas FW (*Forest Wet*) e FD (*Forest Dry*).

Tabela 1. Dados meteorológicos mensais fornecidos pela estação convencional no ano de 2010, gentilmente cedidas pelo departamento de Ciências Exatas - FCAV/UNESP Jaboticabal-SP.

Períodos de coleta		
Dados meteorológicos	Fevereiro (SCW e FW)	Novembro (SCD e FD)
Temperatura máxima (°C)	32	31
Temperatura mínima (°C)	20,4	18,7
Temperatura média (°C)	25,3	24,1
Precipitação (mm)	150,7	100

As amostragens de solo foram obtidas em pontos distintos de 0-20 cm de profundidade dentro de um quadrante, todos a partir dos vértices, distanciando 1m, 10m e 100m totalizando 10 pontos, sendo coletadas em tubos de P.V.C., os quais foram identificados, acondicionados, armazenados em caixas térmicas com gelo e enviados para o Laboratório de Bioquímica de Micro-organismos e Plantas (LBMP) para a extração das amostras de DNA metagenômico sendo ambos situados na Faculdade de Ciências Agrárias e Veterinárias, UNESP, Campus Jaboticabal.

3.2 Extração e quantificação do DNA metagenômico

A extração do DNA total presente nas amostras dos solos foi realizada utilizando o Kit FastDNA® SPIN Kit (BIO 101-QUANTUM BIOTECHNOLOGIES - Catálogo n.º #6560-200). A quantificação foi feita através de eletroforese em gel de agarose 1% (p/v), contendo brometo de etídio (0,5 µg/mL). A eletroforese foi realizada em uma cuba modelo Horizon 58 com uma alíquota de 3µL de DNA adicionada de 3 µL de tampão de carregamento [0,025% (p/v) de azul de bromofenol e 50% de glicerol (v/v)] e conduzida em tampão TBE 1X (Tris 89 mM, Ácido Bórico 89 mM e EDTA 2,5 mM, pH 8,3), a 100V. Foi aplicado ao gel uma alíquota de plasmídeo pGEM (APPLIED BIOSYSTEMS) com concentração conhecida (50 ng/µL)

em diferentes volumes para comparar a intensidade de fluorescência emitida pelo brometo de etídio e cálculo da concentração do material. A visualização do DNA foi realizada sob luz UV e a imagem foi documentada em um fotodocumentador (Bio-Rad). A concentração de DNA também foi estimada em espectrofotômetro Nano-Drop ND-1000 (ThermoScientific).

3.3 Sequenciamento do DNA metagenômico

Para dosar a concentração dos DNAs utilizou-se Qubit (Invitrogen) e a concentração foi compatível com o descrito no manual TruSeq™DNASamplePreparation V2 Guide. Após o preparo das bibliotecas as mesmas foram seqüenciadas utilizando-se plataforma Illumina®, equipamento HiScanSQ, com bibliotecas do tipo “*paired-end*” com 2x100 pb. O delineamento realizado foi feito calculando-se o número de *reads* gerados e houve uma cobertura satisfatória para cada uma das amostras submetidas às análises.

3.4 Análise *in silico*

Inicialmente a análise foi realizada através da comparação das seqüências das amostras já processadas com base de dados de seqüências conhecidas, através da plataforma de análise de dados de metagenomas MG-RAST (<http://metagenomics.anl.gov/>) (MEYER et al., 2008), primeira frente de investigação (Metagenoma Total).

Através do upload dos dados provenientes do sequenciamento, a plataforma MG-RAST realizou o controle de qualidade, a anotação automatizada e a análise das amostras de DNA metagenômico.

As amostras montadas (*de novo assembly*) pelo pacote CLC Genomics Workbench, foram analisadas através da plataforma MG-RAST e também pela

plataforma ISSaga 2.0 ([http:// http://issaga.biotoul.fr/ISSaga/issaga_index.php](http://issaga.biotoul.fr/ISSaga/issaga_index.php)) (VARANI et al., 2011), segunda frente de investigação (Metagenoma Montado). Em ambas as abordagens os dados sobre a diversidade e classificação das famílias de ISs foram fornecidos.

As amostras do Metagenoma Total, em uma segunda análise, foram filtradas contra o banco de dados de sequências de inserção ISfinder, a prospecção ocorreu de forma exaustiva via programa BLAST (*Basic Local Alignment Search Tool*) foi otimizado com o "word size = 7" e com o filtro de complexidade desativado.

3.5 Anotação genômica

Após a prospecção dos genes de interesse, foi realizada a caracterização do contexto genômico das sequências obtidas, utilizando-se da ferramenta Snap Gene Viewer 2.8.2 (www.snapgene.com/). As sequências escolhidas para caracterização obedeceram aos seguintes critérios: a-) *scaffolds* prospectados das amostras de Metagenoma Montado via plataforma ISSaga 2.0; b-) identificada sua classificação taxonômica; c-) estão codificando transposases completas.

4 RESULTADOS E DISCUSSÃO

O presente trabalho teve como objeto de estudo sequências geradas em estações climáticas diferentes, para os ambientes de cana-de-açúcar (SCW e SCD) e floresta nativa (FW e FD), obtidos por sequenciamento de nova geração.

4.1 Primeira Frente de Investigação MG-RAST (Análise de *reads*)

Para as amostras de solo de cana-de-açúcar, foram obtidos um total de 60.606.976 (SCW) e 79.208.498 (SCD) *reads*; enquanto que para amostras de floresta foram obtidos 63.679.540 (FW) e 64.384.450 (FD).

Dentre este conjunto de dados, os *reads* que codificam para transposases foram prospectados e classificados, empregando-se a ferramenta MG-RAST. Referente aos respectivos totais para cada uma das amostras foi classificado como transposases pertencentes as mais diferentes famílias: 101.981 (SCW/ 0,17%) e 298.036 (SCD/0,38%), 248.645 (FW /0,39%) e 254.320 (FD/ 0,40%) *reads* (Figura 6).

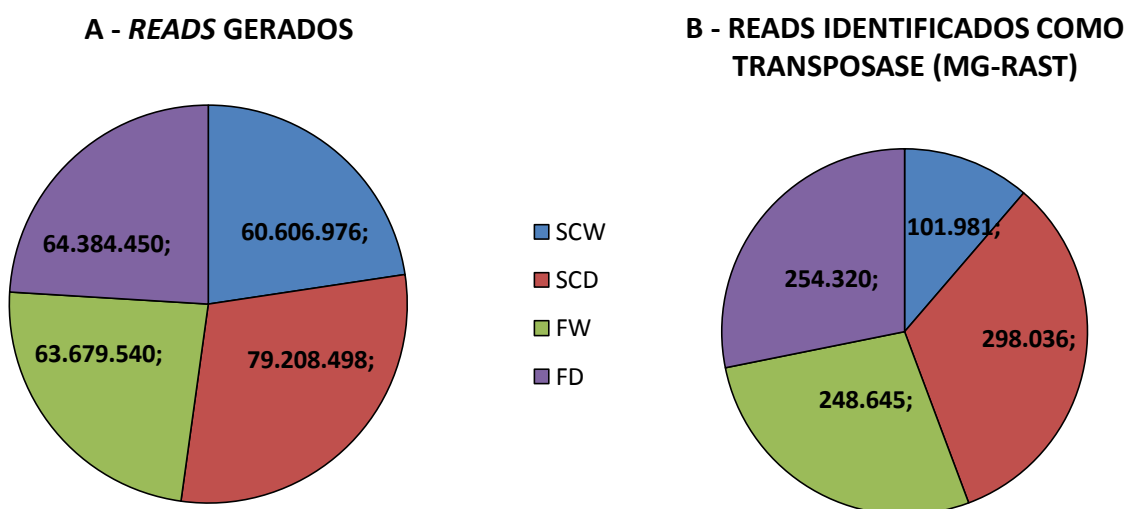


Figura 6. Conjunto de dados utilizados na prospecção de transposases. “A”- Quantidade de *reads* gerados para cada amostra metagenômica. “B”- Quantidade de *reads* identificados como transposases de acordo com a plataforma MG-RAST. SCW (*Sugarcane Wet*), SCD (*Sugarcane Dry*), FW (*Forest Wet*) e FD (*Forest Dry*).

4.2 Classificação Transposases MG-RAST

Conforme pode ser visualizado na figura 7, a plataforma MG-RAST utiliza um sistema próprio de classificação de transposases, empregando termos genéricos como “transposases” e “putative transposase”, entre outros, o qual não está de acordo com a classificação em famílias apresentada pela literatura (Figura 4).

O resultado encontrado sugere que a plataforma MG-RAST pode estar identificando sequencias parciais de transposases e classificando-as como completas, também é possível que esteja identificando genes com outras funções como se fossem transposases.

Os ISs podem gerar diversas cópias no genoma, principalmente quando sua transposição é de forma replicativa, através deste fato, é possível supor que a plataforma MG-RAST está identificando transposases repetidas em um mesmo genoma. Outra característica de análises envolvendo dados de sequências não montadas (*reads*), é que a própria técnica gera milhares de cópias curtas para uma mesma região da sequência, o que causa redundância para os resultados.

Para todas as amostras a nomenclatura “*transposase*” foi a que obteve maior abundância, em SCW (27,17%), SCD (25,91%), FW (29,71%) e FD (28,30%) (Figura 7) (APÊNDICE A - TABELA 1A).

A segunda classificação mais encontrada foi “*putative transposase*” em todas amostras analisadas, SCW (21,48%), SCD (22,15%), FW (21,77%) e FD (21,53%).

A classificação “*transposase IS4 family protein*”, também demonstrou grande abundância em todas as amostras, SCW (16,26%), SCD (15,0%), FW (14,15%) e FD (14,96%).

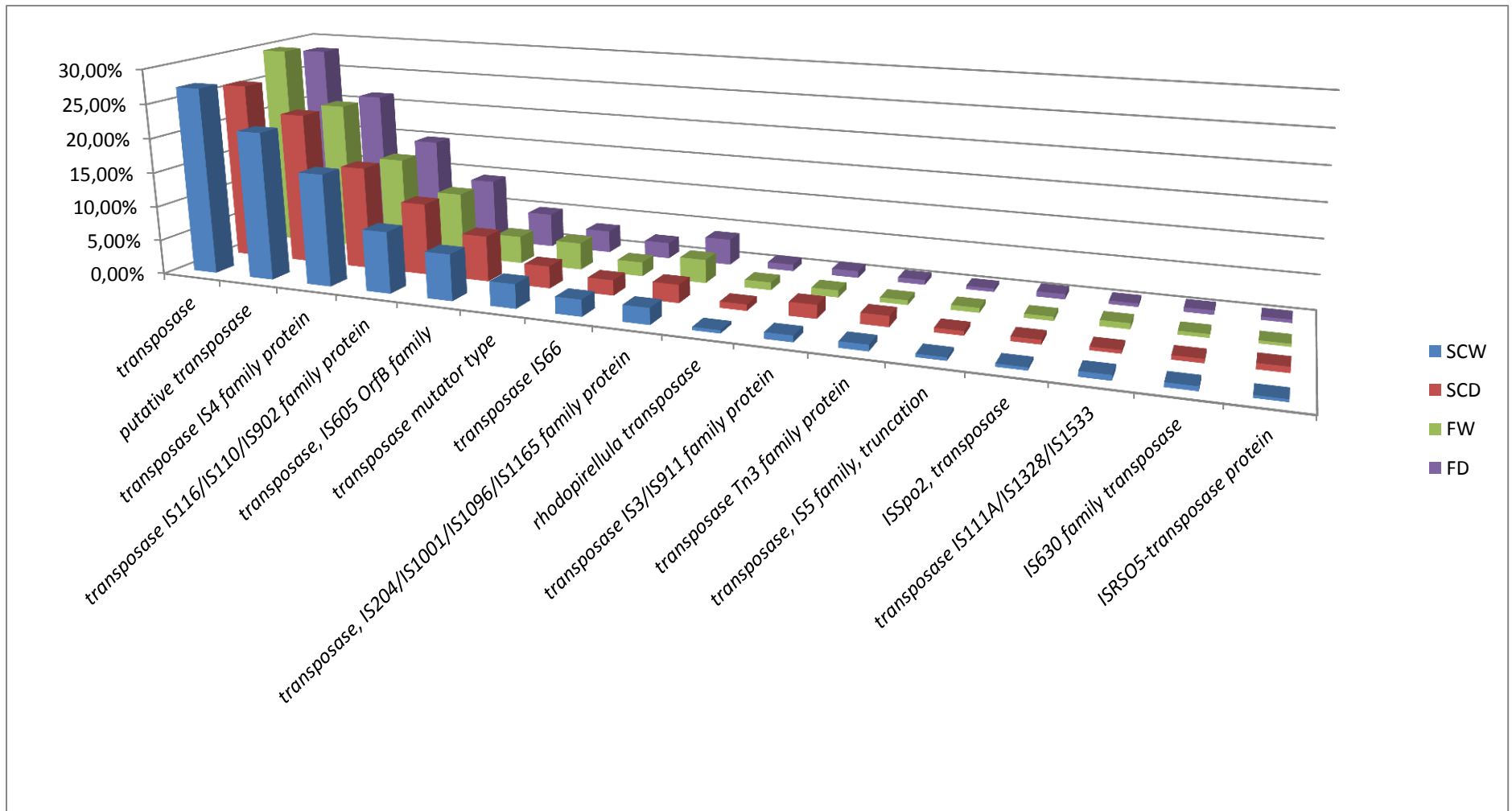


Figura 7. Classificação de transposases (*reads*) mais representativas atribuídas pela plataforma MG-RAST. Valores totais de hits encontrados para cada uma das amostras foram 1496 (SCW – *Sugarcane Wet*), 1691 (SCD – *Sugarcane Dry*), 1599 (FW – *Forest Wet*) e 1639 (FD – *Forest Dry*). (APÊNDICE A - TABELA 1A).

4.3 Classificação taxonômica para *reads* (MG-RAST)

Em relação à organização taxonômica destas transposases, o filo Proteobacteria foi o mais abundante em SCW e SCD, contendo 20.618 (52,6%) e 58.118 (48,9%), respectivamente. O segundo filo mais abundante presente nas amostras SCW e SCD foi das Actinobacteria com 5.113 (13,1%) e 21.405 (18%), respectivamente. Em FW e FD, os filios bacterianos mais abundantes também foram Proteobacteria com 48.734 (44,6%) e 51.411 (48,9%) transposases, respectivamente, e Actinobacteria com 31.893 (29,2%) e 23.258 (22,1%), respectivamente. A abundância de sequências para Proteobacteria e o Actinobacteria também foi observado no trabalho realizado por Taketani e colaboradores (2015), cujo interesse foi coletar amostras do solo de rizosfera, em estações climáticas diferentes, ao longo do semiárido brasileiro, utilizando a abordagem metagenômica.

De forma similar, Spain e colaboradores (2009) compararam a abundância relativa de Proteobacteria em diferentes solos, observando uma variação de 25% (solo de pradaria-EUA) a 50% (solo agrícola com plantação de milho-Brasil) do total de sequências analisadas. A proporção de Proteobacteria encontrada nas amostras confirma o resultado observado em solos brasileiros e também sob cultivo de cana-de-açúcar nos EUA, que foi de 48% (SPAIN et al., 2009).

É possível compreender estes resultados, tendo como base, que o domínio Bacteria é composto por 23 filios, alguns destes filios consistem em uma única ou poucas espécies, enquanto os filios Proteobacteria, Firmicutes e Actinobacteria incluem milhares de espécies, as quais correspondem a 90-95% de todas as espécies de bactérias conhecidas (NUNES, 2006).

O filo Proteobacteria, por sua vez, constitui-se o maior e mais diverso grupo dentre bactérias cultivadas, apresentando grande diversidade de morfologia e metabolismo. Possui cinco subdivisões: *Alphaproteobacteria*, *Betaproteobacteria*, *Deltaproteobacteria*, *Epsilonproteobacteria*, e *Gammaproteobacteria*. Essas bactérias estão relacionadas ao ciclo do nitrogênio, além de muitas serem consideradas patogênicas (NÜSSLEIN; TIEDJE, 1999). A presença deste filo foi

observada em todas as amostras de solo realizadas neste trabalho, confirmando assim a abundância destes micro-organismos no solo.

O filo Actinobacteria é representado por bactérias Gram-positivas, (GAO; GUPTA, 2005) as quais estão distribuídas em ecossistemas aquáticos e terrestres e possui organismos importantes na decomposição da matéria orgânica e na formação do húmus, dentre outros papéis (GOODFELLOW; WILLIAMS, 1983).

Nas amostras analisadas, 14,75% dos dados não foram identificados pela plataforma MG-RAST. Esse resultado é comumente encontrado em bibliotecas metagenômicas, considerando o elevado índice de novidades de sequências obtidas através desta abordagem. Uma grande quantidade destes dados devem ser melhor explorados através da correta anotação de suas sequências, além de alguns dados em que não é possível classificação por representarem filos inteiramente novos (BRUCE, 2010).

Os gêneros bacterianos mais abundantes em SCW são os gêneros *Azospirillum* e *Mesorhizobium*, com 2.102 (8,15%) e 1.885 (7,3%), respectivamente (Figura 8) (APÊNDICE A - TABELA 2A); enquanto que em SCD o gênero bacteriano mais abundante foi *Streptomyces* pertencente ao filo Actinobacteria, com 6.090 (6,93%) transposases, seguido do gênero *Mesorhizobium*, com 5.991 (6,82%) transposases. De acordo com Pisa et al. (2011), *Azospirillum* e *Mesorhizobium* são bactérias Gram-negativas relacionadas à fixação de nitrogênio e podem desempenhar importante papel no crescimento de plantas. *Streptomyces* são bactérias Gram-positivas filamentosas, abundante em solos, caracterizadas pela capacidade de suprimir o crescimento de outros micro-organismos através da produção de metabólitos secundários, dessa forma, possui destaque em estudos de bioprospecção, devido sua variabilidade na produção de produtos (PONTIROLI et al., 2013).

Em FW e FD, o gênero bacteriano mais abundante foi *Mycobacterium*, com 11.885 (14,48%) e 6.592 (8,36%) transposases, respectivamente. Bactérias do gênero *Mycobacterium* são Gram-positivas, encontradas nos mais diversos ambientes, saprófitas de vida livre ou potenciais patógenos humanos (PONTIROLI et al., 2013).

De acordo com Mahillon e Chandler (1998), a família IS3, é amplamente encontrada em diversas espécies de bactérias. Seu conteúdo G+C assemelha-se ao seu hospedeiro, o qual, em Micobactérias chega até mesmo a 70%.

Em trabalhos anteriores, foi observado que os genomas de bactérias classificadas como Proteobacterias, do gênero *Shigella*, *Yersinia* e *Bordetella* são enriquecidos com sequências de inserção (YANG et al., 2005;; PARKHILL et al., 2003).

No trabalho realizado por Touchon e Rocha (2007), sobre a abundância de sequências de inserção em genomas procarióticos, foram examinados 227 genomas bacterianos, chegando a conclusão que alguns genomas são particularmente ricos em ISs. Por exemplo, nos seguintes genomas foram encontrados mais de 60 ISs completos: *Sulfolobus sulfataricus* (Archea), *Bacillus halodurans* (Firmicutes), *Nitrobacter winogradskyi* (Proteobacteria), *Bordetella pertussis* (Proteobacteria), *Shigella sonnei* (Proteobacteria) e *Mycoplasma mycoides* (Tenericutes).

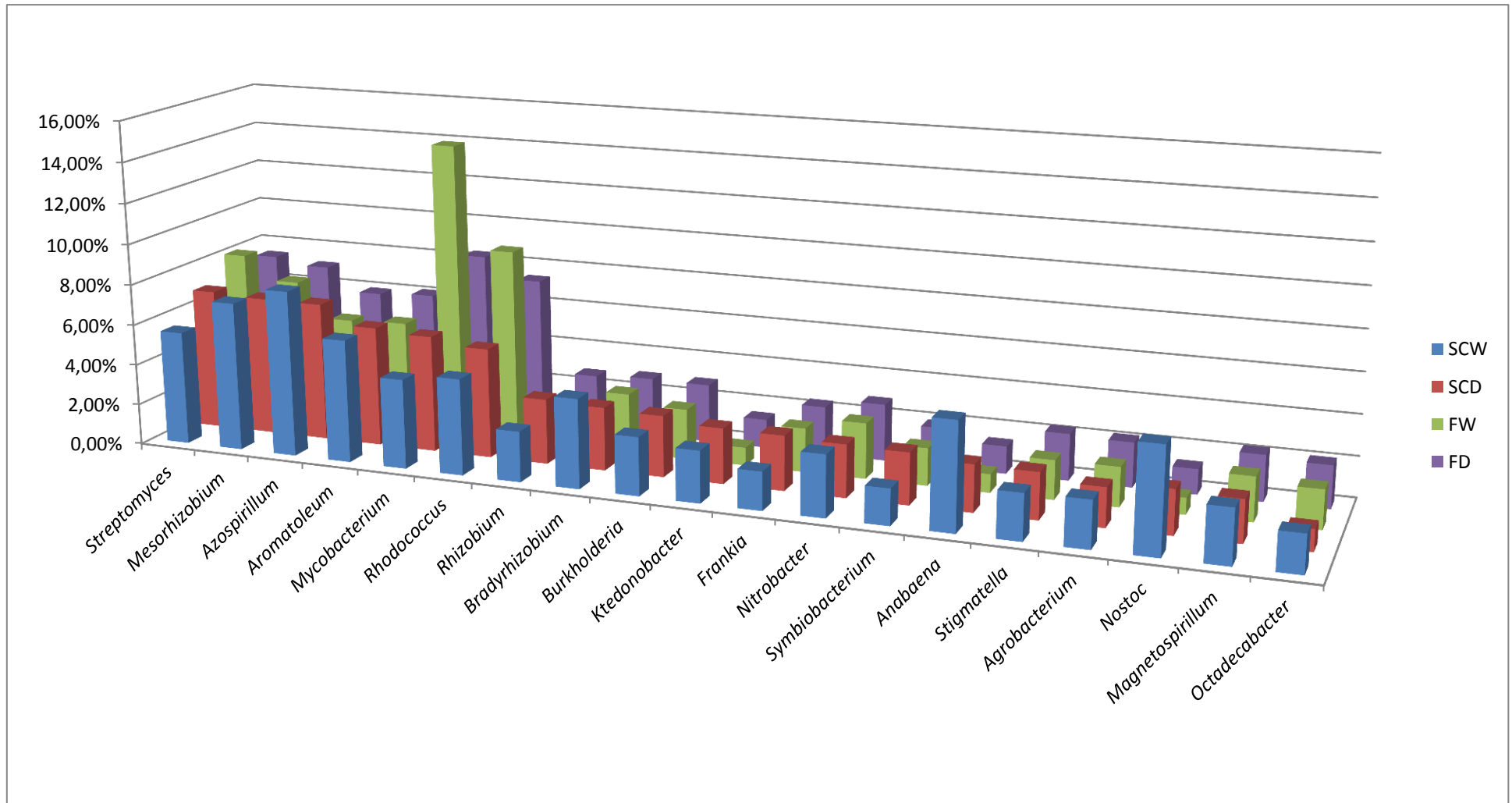


Figura 8. Porcentagem dos gêneros taxonômicos mais abundantes, identificados a partir dos *reads* classificados como transposases, pela plataforma MG-RAST para as amostras SCW (*Sugarcane Wet*), SCD (*Sugarcane Dry*), FW (*Forest Wet*) e FD (*Forest Dry*). (APÊNDICE A - TABELA 2A)

4.4 Segunda Frente de Investigação: Classificação MG-RAST – Metagenoma Montado (*Scaffolds*)

Os resultados observados através da análise do Metagenoma Montado (*scaffolds*) pela plataforma MG-RAST, é semelhante ao resultado encontrado para a análise do Metagenoma Total (*reads*).

A plataforma MG-RAST, não classificou os *scaffolds* identificados como transposases, em famílias de ISs de acordo com a literatura (Figura 4).

A nomenclatura “*transposase*” foi a prevalecente em todas as amostras, em SCW (27,65%), SCD (28,64%), FW (32,84%) e FD (29,73) (Figura 9) (APÊNDICE A - TABELA 3A).

A nomenclatura “*putative transposase*” foi a segunda classificação mais abundante, em SCW (18,30%), SCD (19,04%), FW (19,54%) e FD (18,37%).

Também foi realizada a classificação taxonômica dos *scaffolds* através da plataforma MG-RAST, sendo observado um resultado semelhante à análise dos *reads*. Como exemplo, o gênero *Mycobacterium* também foi predominante nas amostras FW (9,34%) e FD (6,41%) (Figura 10) (APÊNDICE A - TABELA 4A).

Na amostra SCW o resultado também foi semelhante, predominando os gêneros *Azospirillum* e *Mesorhizobium*, com 5,37% e 4,28%, respectivamente. Na amostra SCD predominou o gênero *Mesorhizobium*, com 5,38%.

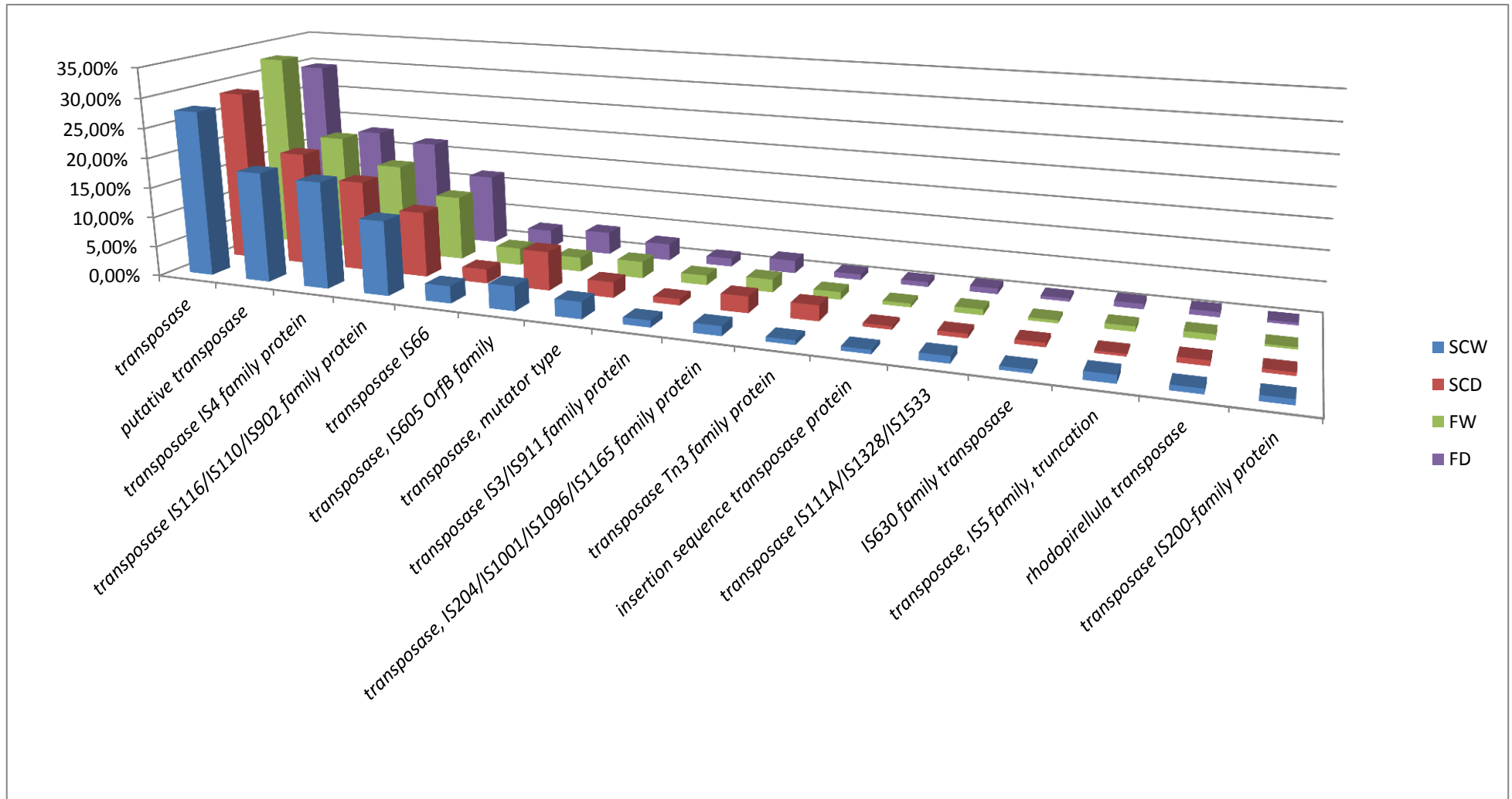


Figura 9. Classificação de transposases (*scaffolds*) mais representativas atribuídas pela plataforma MG-RAST para as amostras SCW (*Sugarcane Wet*), SCD (*Sugarcane Dry*), FW (*Forest Wet*) e FD (*Forest Dry*). (APÊNDICE A - TABELA 3A).

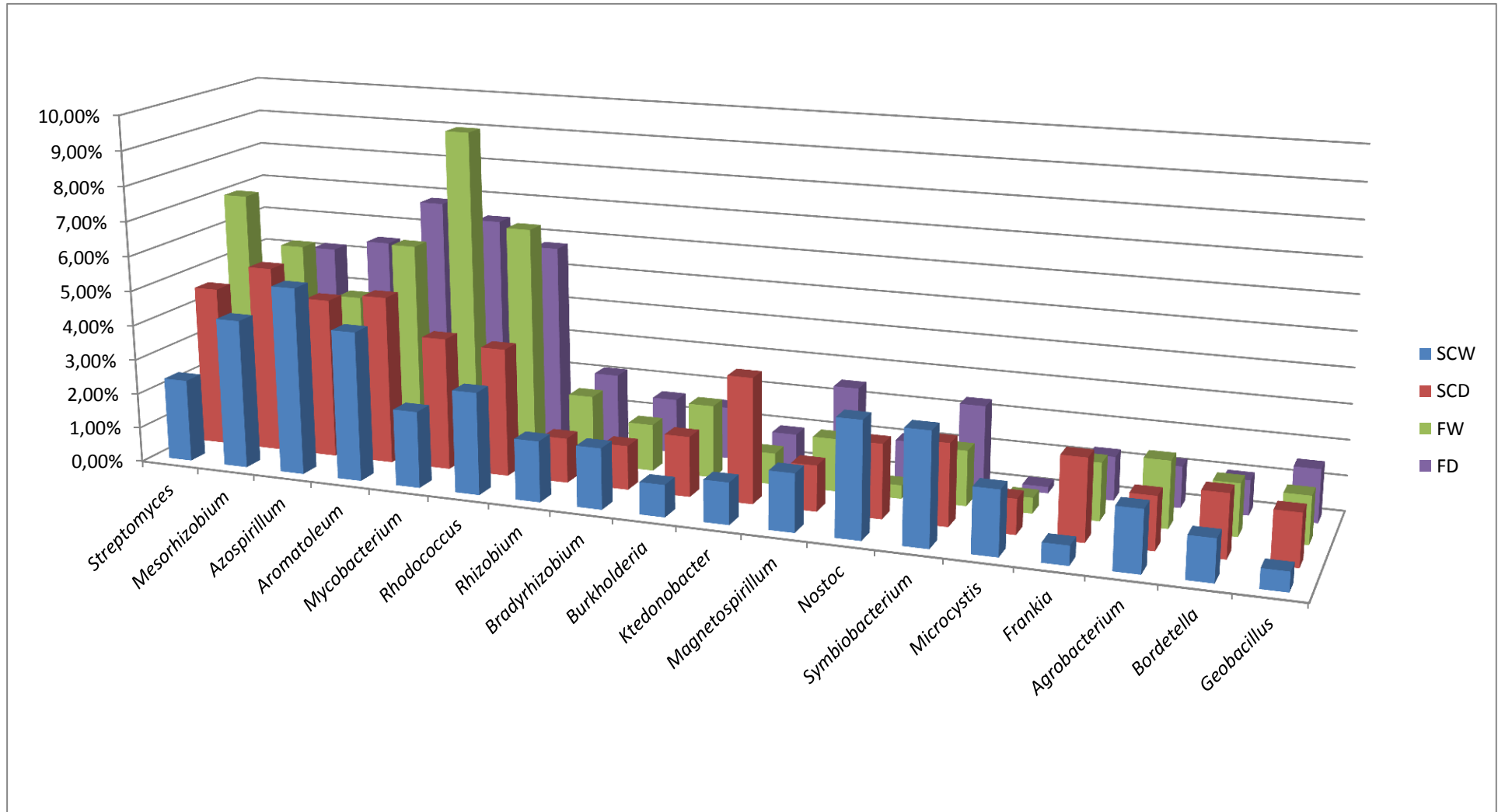


Figura 10. Porcentagem dos gêneros taxonômicos mais abundantes, identificados a partir dos *scaffolds* classificados como transposases, pela plataforma MG-RAST para as amostras SCW (*Sugarcane Wet*), SCD (*Sugarcane Dry*), FW (*Forest Wet*) e FD (*Forest Dry*). (APÊNDICE A - TABELA 4A).

4.5 Classificação ISSaga 2.0 - Metagenoma Montado (*Scaffolds*)

As análises de diversidade e classificação das famílias de transposases, utilizando as amostras montadas pelo pacote *CLC Genomics Workbench*, foram conduzidas também através da plataforma ISSaga 2.0. A amostra SCW forneceu 86.236 (17%) *scaffolds*, de um total de 94.973.352 pb, enquanto a amostra SCD forneceu 212.354 (42%) *scaffolds*, de um total de 243.420.202 pb. Na amostra FW foram fornecidos 100.775 (20%) *scaffolds*, de um total de 115.036.043 pb e a amostra FD forneceu 106.859 (21%) *scaffolds*, de um total de 130.318.256 pb (Figura 11).

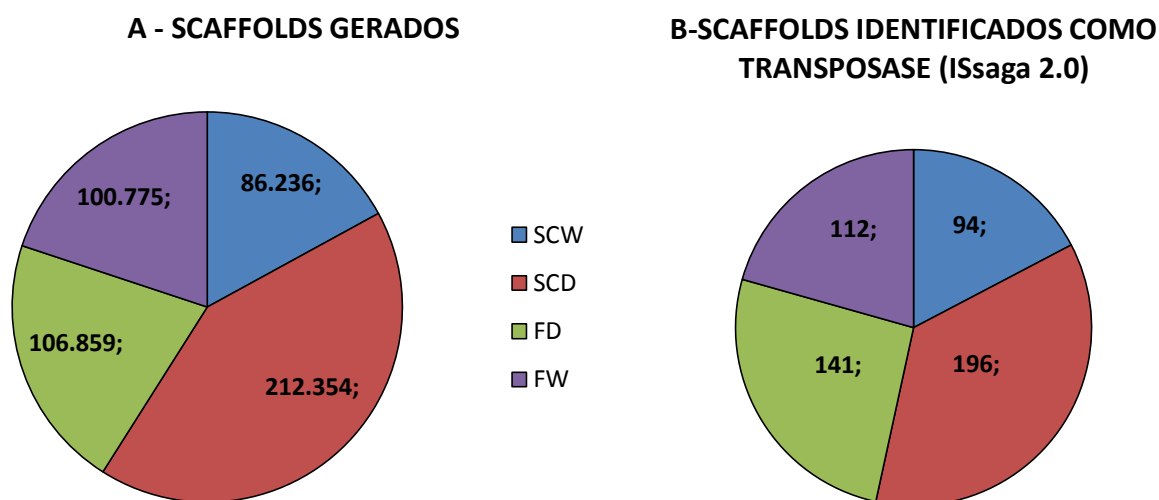


Figura 11. “A”- Quantidade de *scaffolds* gerados para cada amostra metagenômica. “B”- Quantidade de sequências de transposases identificadas em *scaffolds* através da plataforma ISSaga 2.0. SCW (*Sugarcane Wet*), SCD (*Sugarcane Dry*), FW (*Forest Wet*) e FD (*Forest Dry*).

4.6 Classificação transposases ISSaga 2.0

Para a amostra SCW foram encontradas 98 ORFs (*Open Reading Frame*), 12 completas, 9 parciais e 77 não categorizáveis (Figura 12) (APÊNDICE A - TABELA 5A). Foram preditas 94 ISs diferentes, prevalecendo a família de elementos não classificados, ISNCY (38,30%), indicando a possibilidade de identificação de novas famílias de transposases. Nesta família, incluem-se as ISs em que o domínio catalítico da transposase ainda não é claro, ou porque há uma série de potenciais resíduos catalíticos ou porque não há um número suficiente de exemplos para definir os resíduos altamente conservados, os membros deste grupo muitas vezes emergem como novas famílias de transposases (SIGUIER, 2015). Além disso, este resultado merece ser melhor explorado, já que, comparando-se esta amostra, com o banco de dados ISfinder, a família ISNCY não é a mais abundante no banco de dados, como foi na amostra (Figura 4).

Na amostra SCD foram estimadas 219 ORFs, 53 completas, 60 parciais e 106 não categorizáveis. Foram preditas 196 ISs, sendo 193 diferentes. Para a amostra FW, foram encontradas 119 ORFs, 14 completas, 38 parciais e 67 não categorizáveis. A predição indicou 112 ISs, sendo 109 diferentes. A família IS110 prevaleceu nas amostras SCD (16,33%) e FW (19,64%). Na amostra FD foram encontradas, 156 ORFs, 30 completas, 59 parciais e 67 não categorizáveis. Foram preditas 141 ISs diferentes, prevalecendo a família IS630 (15,49%).

A família IS5 é uma das mais encontradas no banco de dados ISfinder, enquanto nas amostras analisadas, o resultado não segue esta proporção, em SCW (5,32%), SCD (3,57%), FW (3,57%) e FD (7,75%).

Através destes resultados encontrados é possível perceber o potencial destes dados para identificação de novas famílias de transposases, além da necessidade de maior compreensão do papel funcional destas sequências.

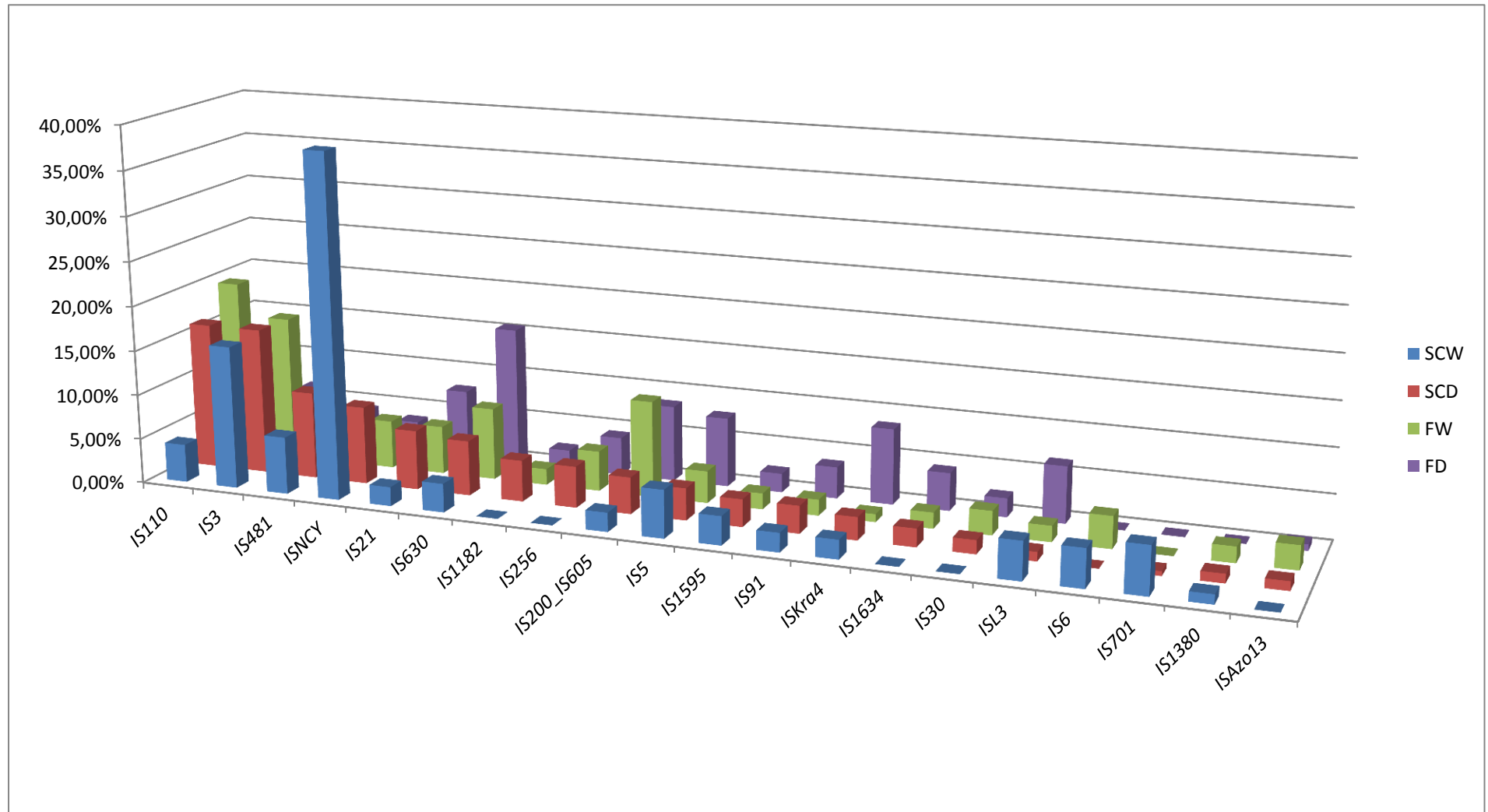


Figura 12. Classificação de transposases (*scaffolds*) mais representativas atribuídas pela plataforma ISSaga 2.0 para as amostras SCW (*Sugarcane Wet*), SCD (*Sugarcane Dry*), FW (*Forest Wet*) e FD (*Forest Dry*). (APÊNDICE A - TABELA 5A).

4.7 Classificação Taxonômica ISSaga 2.0

Quanto a distribuição taxonômica, o programa ISSaga 2.0 disponibiliza tabelas que demonstram as prováveis espécies encontradas em cada amostra. Porém, nas amostras observadas, a maior parte dos dados não foi possível classificar (Tabelas 2, 3, 4 e 5). O programa ISSaga 2.0 classificou 21,66% dos *scaffolds* em grupos taxonômicos, posteriormente, utilizando a plataforma Taxonomy Browser (<http://www.ncbi.nlm.nih.gov/Taxonomy/Browser/>), foi possível identificar os reinos a que pertencem estas espécies, sendo, o filo Proteobacteria (34,3 %) mais prevalente, seguido do filo Actinobacteria (27,08%).

Tabela 2. Distribuição taxonômica da amostra SCW, de acordo com o programa ISSaga 2.0

Espécie	Abundância
Não classificado	74
<i>Thermus thermophilus</i>	4
<i>Aromatoleum aromaticum</i>	3
<i>Allochromatium vinosum</i>	2
<i>Anaeromyxobacter Fw109</i>	2
<i>Opitutus terrae</i>	2
<i>Actinoplanes missouriensis</i>	1
<i>Anaerolinea thermophila</i>	1
<i>Candidatus koribacter</i>	1
<i>Leptothrix cholodnii</i>	1
<i>Mesorhizobium australicum</i>	2
<i>Polymorphum gilvum</i>	1
<i>Starkeya novela</i>	1

Tabela 3. Distribuição taxonômica da amostra SCD, de acordo com o programa ISSaga 2.0

Espécie	Abundância
Não classificado	187
<i>Delftia Cs1</i>	3
<i>Nocardia brasiliensis</i>	3
<i>Streptomyces coelicolor</i>	3
<i>Terriglobus roseus</i>	3
<i>Truepera radiovictrix</i>	2
<i>Rhodopseudomonas palustres</i>	2
<i>Actinoplanes friuliensis</i>	1
<i>Aminobacterium colombiense</i>	1
<i>Bradyrhizobium japonicum</i>	1
<i>Candidatus koribacter</i>	1
<i>Caulobacter K31</i>	1
<i>Ignavibacterium álbum</i>	1
<i>Marinithermus hydrothermalis</i>	1
<i>Methylocella silvestres</i>	1
<i>Mycobacterium smegmatis</i>	1
<i>Oceanithermus profundus</i>	1
<i>Rhizobium leguminosarum</i>	1
<i>Rhodococcus equi</i>	1
<i>Rubrobacter xylanophilus</i>	1
<i>Streptomyces coelicolor</i>	1
<i>Thermaerobacter marianensis</i>	1
<i>Thioalkalivibrio nitratireducens</i>	1
<i>Xanthobacter autotrophicus</i>	1

Tabela 4. Distribuição taxonômica da amostra FW, de acordo com o programa ISSaga 2.0

Espécie	Abundância
Não classificado	92
<i>Candidatus solibacter</i>	7
<i>Nocardiopsis dassonvillei</i>	3
<i>Xylanimonas cellulositytica</i>	3
<i>Actinoplanes missouriensis</i>	2
<i>Catenulispora acidiphila</i>	2
<i>Agromonas oligotrophica</i>	1
<i>Anaerolinea thermophila</i>	1
<i>Cellulomonas fimi</i>	1
<i>Frankia EAN1pec</i>	1
<i>Methylobacterium nodulans</i>	1
<i>Mycobacterium JDM601</i>	1
<i>Nocardiopsis alba</i>	1
<i>Sphaerobacter thermophilus</i>	1

Tabela 5. Distribuição taxonômica da amostra FD, de acordo com o programa ISSaga 2.0

Espécie	Abundância
Não classificado	139
<i>Stigmatella aurantiaca</i>	6
<i>Candidatus solibacter</i>	3
<i>Acidobacterium capsulatum</i>	2
<i>Opitutus terrae</i>	2
<i>Bordetella avium</i>	1
<i>Candidatus koribacter</i>	1
<i>Dinoroseobacter shibae</i>	1
<i>Thermus thermophilus</i>	1

4.8 Comparações dos resultados do MG-RAST (*reads*) contra o banco de dados ISfinder

As sequências obtidas com a função transposase pela base de dados MG-RAST, foram filtradas através do banco de dados ISfinder (<https://www-is.biotoul.fr/>). Através desta comparação foi possível observar que a plataforma MG-RAST, classificou com a função “transposase” sequências pertencentes a outras funções. Na amostra SCW, a plataforma MG-RAST havia identificado 101.981 sequências com a função transposase, quando comparado ao ISfinder, foi identificado apenas 23.398 sequências com esta função. O mesmo foi observado nas demais amostras, em SCD, a plataforma MG-RAST havia identificado 298.036 sequências com esta função, quando comparado ao ISfinder, o número de sequências diminuiu para 66.190. Na amostra FW, a plataforma MG-RAST havia identificado 248.645 sequências, comparando-se ao ISfinder, foram identificadas 58.439 sequências para a função “transposase”. Para a amostra FD, a plataforma MG-RAST identificou 254.320 transposases, enquanto comparando-se com o ISfinder, foram identificadas 57.568 sequências para esta função (Figura 13).

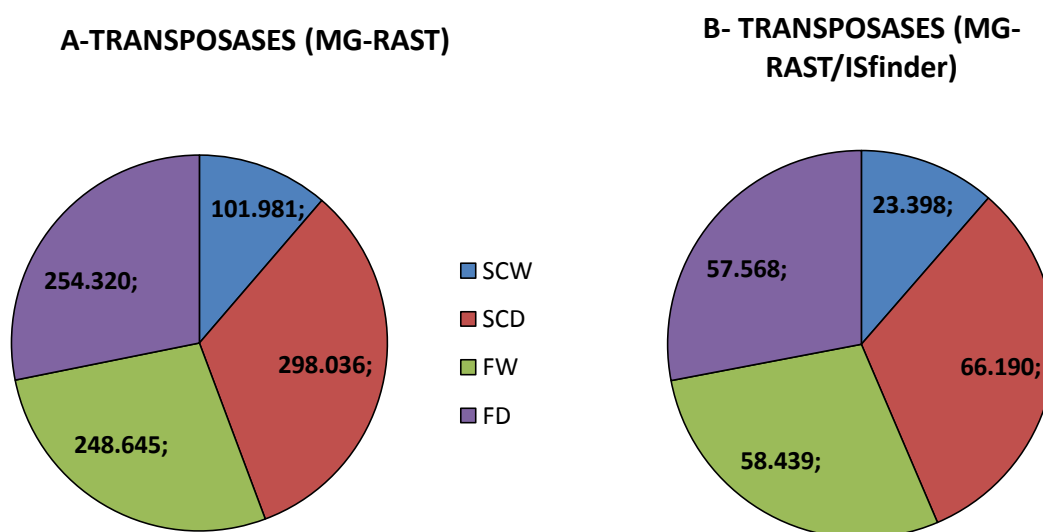


Figura 13. “A”- Quantidade de transposases identificadas pela plataforma MG-RAST. “B”- Transposases identificadas pelo MG-RAST após filtragem pelo banco de dados ISfinder. SCW (*Sugarcane Wet*), SCD (*Sugarcane Dry*), FW (*Forest Wet*) e FD (*Forest Dry*).

O resultado encontrado demonstrou a família IS110 como predominante em todas as amostras, SCW (14,09%), SCD (16,97%), FW (15,27%) e FD (15,99%), sua maior abundância foi observada na amostra SCD (Figura 14) (APÊNDICE A - TABELA 6A).

Este resultado pode ser explicado pelo fato de que a família IS110 é uma das mais prevalentes em procariotos, sendo a terceira família com maior abundância de sequências depositadas no banco de dados ISfinder, correspondendo a cerca de 250 sequências diferentes, distribuídas em aproximadamente 130 espécies de bactérias e arqueas (SIGUIER et al, 2015) (Figura 4).

A família IS110 possui sequências que variam entre 1136 e 1558 pb, mas a maioria possui uma média de 1450 pb. Todos os membros desta família não possui IRs, ou são muito pequenas (<7 pb) e geralmente não criam DRs. Esta família difere das demais por ser a única que possui o domínio catalítico DEDD (CHANDLER; MAHILON, 2002)..

A família IS630 foi a segunda com maior abundância na maioria das amostras, SCD (8,93%), FW (10,68%) e FD (11,58%), exceto na amostra SCW (10,91%), onde a família IS5, foi a segunda mais predominante. A maior quantidade de ISs pertencentes à família IS630, foi encontrada na amostra FD. No banco de dados ISfinder esta família também é abundante, possui 160 membros, pertencentes a cerca de 80 espécies de bactérias e arqueas (SIGUIER et al, 2015) (Figura 4).

As sequências de inserção pertencentes a esta família variam de 1100 a 1200 pb, movimentam-se pelo mecanismo “corta e cola” e possuem semelhanças significativas com o elemento transponível identificado em eucariotos Tc1/mariner. Além disso, análises demonstram que a família IS630 possui preferência pela sequência alvo 5'-CTAG-3', dessa forma, apresentam alta especificidade alvo (CHANDLER; MAHILON, 2002).

A família IS5 foi abundante em todas as amostras, SCW (11,53%), SCD (8,26%), FW (8,60%) e FD (8,99%), sendo a segunda mais encontrada na amostra SCW. Este resultado já era o esperado, já que esta família de ISs é a que possui maior quantidade de sequências depositadas no banco de dados ISfinder, juntamente com a família IS3, correspondendo a total de 564 membros (SIGUIER et al, 2015) (Figura 4). Esta família é muito heterogênea, o seu comprimento varia de

850 pb e 1643 pb, a principal característica que define este grupo é a semelhança em suas transposases. Esta família foi dividida em seis subfamílias (IS903, ISL2, ISH1, IS5, IS1031 e IS427) e os membros de cada família podem gerar DRs com comprimentos semelhantes (CHANDLER; MAHILON, 2002).

A família IS3, é conhecida por ser uma das mais abundantes no banco de dados, apresenta um total 554 membros, distribuídas em aproximadamente 270 diferentes espécies de bactérias (Figura 4). As amostras analisadas não seguem esta proporção do banco de dados. A proporção encontrada nas amostras foi: em SCW (6,88%), SCD (5,54%), FW (5,20%) e FD (5,24%). Também foi encontrada em todas as amostras, a família de elementos não classificados, ISNCY, SCW (3,59%), SCD (2,28%), FW (2,54%) e FD (1,97%). O que fornece a possibilidade de identificação de novas famílias de transposases. A amostra SCW apresentou maior abundância de elementos desta família.

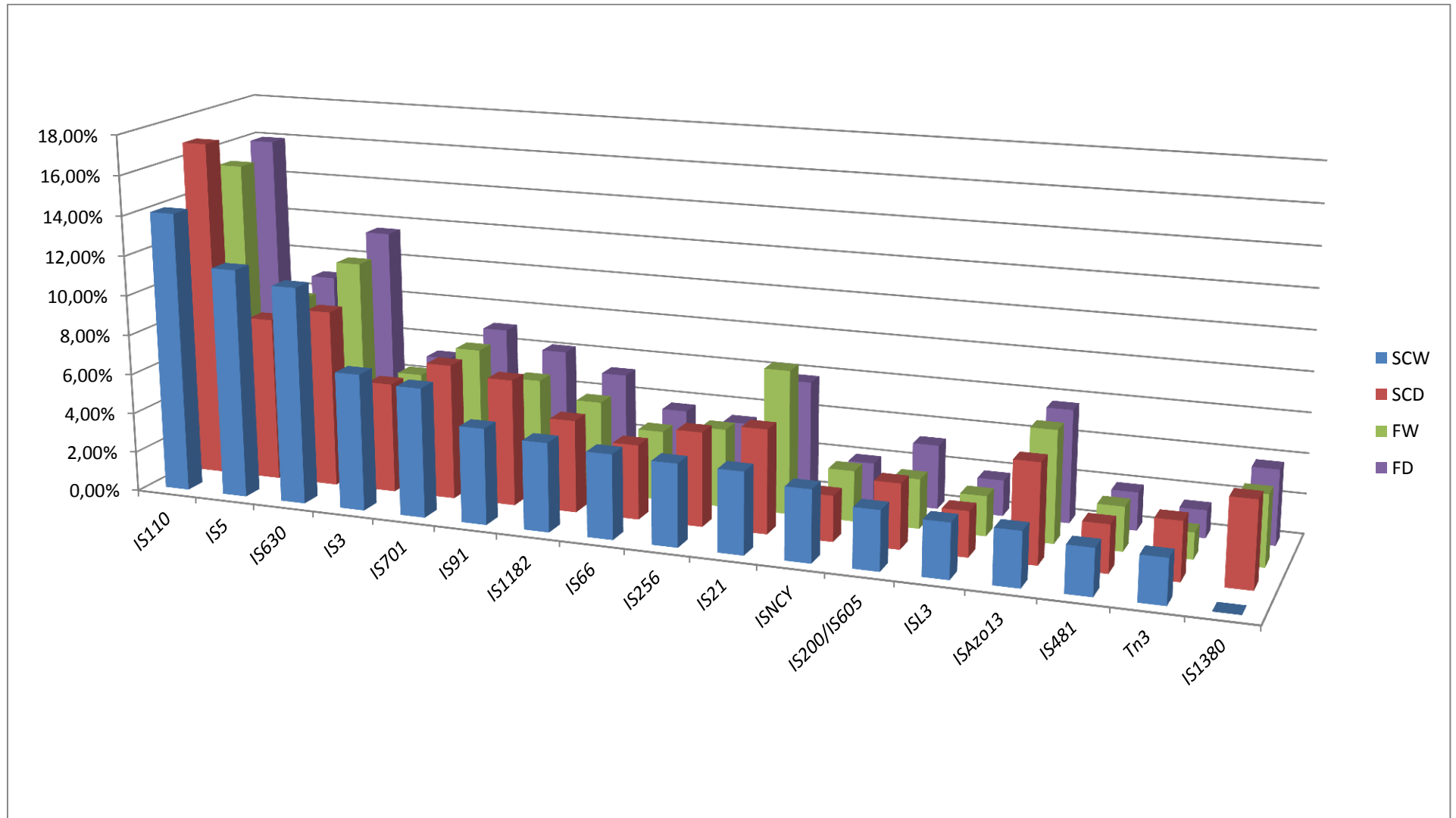


Figura 14. Classificação de transposases (*reads*) mais representativas atribuídas pela plataforma MG-RAST após filtragem através do banco de dados de sequências de inserção ISfinder para as amostras SCW (*Sugarcane Wet*), SCD (*Sugarcane Dry*), FW (*Forest Wet*) e FD (*Forest Dry*). (APÊNDICE A - TABELA 6A).

4.9 Comparação plataforma MG-RAST/ISfinder X ISSaga 2.0

MG-RAST e ISSaga 2.0 recuperaram 902.982 e 543 transposases, respectivamente. MG-RAST classificou 85,25% dos *reads* em grupos taxonômicos distintos, sendo o filo Proteobacteria o mais prevalente. ISSaga 2.0 classificou 21,66% dos *scaffolds* em grupos taxonômicos, onde o filo Proteobacteria também foi o mais prevalente. Quando os resultados do MG-RAST são confrontados com banco de dados ISfinder, observa-se que 77,23% das transposases preditas são predições falso positiva (Figura 13). Comparando-se os dados obtidos pelo programa ISSaga 2.0 (*scaffolds*), com os dados fornecidos pela plataforma MG-RAST (*reads*) e classificados através do banco de dados ISfinder, é possível observar que a família IS110 foi a mais abundante em SCD 16,97% (MG-RAST/ISfinder); 16,33% (ISSaga 2.0) e FW 15,27% (MG-RAST/ISfinder); 19,64% (ISSaga 2.0) para os dois métodos analisados, além de demonstrar grande abundância nas demais amostras, em SCW 14,09% (MG-RAST/ISfinder); 4,26% (ISSaga 2.0) e FD 15,99% (MG-RAST/ISfinder); 12,68% (ISSaga 2.0).

A família ISNCY foi encontrada em todas as amostras e em ambos os métodos comparados, sua maior abundância foi obtida na amostra SCW 3,59% (MG-RAST/ISfinder); 38,30% (ISSaga 2.0). Dessa forma, é possível confirmar, a existência de ISs não classificados no metagenoma de todas as amostras de solos analisados, abrindo-se uma grande perspectiva para a descoberta de novas famílias de ISs.

A família IS630, também demonstrou grande abundância, particularmente na amostra FD 11,58% (MG-RAST/ISfinder); 15,49% (ISSaga 2.0), sendo a mais encontrada pelo programa ISSaga 2.0 e a segunda mais encontrada pela plataforma MG-RAST nesta amostra.

Quanto à classificação taxonômica, a plataforma MG-RAST, fornece muitas possibilidades de identificação, quanto ao reino, filo, classe, ordem, família, gênero ou espécie taxonômica, classificando os “*reads*” encontrados em cada amostra. O programa ISSaga 2.0, fornece a classificação em nível de espécie, que pode ser extrapolado para qualquer outro nível taxonômico.

4.10 Caracterização do Contexto Genômico

Utilizando-se da ferramenta *Snap Gene Viewer 2.8.2* (www.snapgene.com/) foi realizada a caracterização do contexto genômico (Figura 15, 16, 17 e 18). As sequências escolhidas para caracterização obedecem aos seguintes critérios: são *scaffolds* prospectados das amostras de Metagenoma Montado via plataforma ISSaga 2.0; foi identificada sua classificação taxonômica; e contém transposases completas.

Observando-se o Scaffold 85 é possível observar a presença de dois ISs completos, pertencentes à família IS630. Esta família possui membros que variam de 1100 a 1200 pb e de acordo com Siguier e colaboradores (2015) dos 160 membros depositados no banco de dados ISfinder, 90 apresentam duas ORFs que codificam a transposase, sugerindo assim, a possibilidade de apenas um IS identificado neste scaffold. O Scaffold 1371 apresentou um único gene codificando a transposase da família IS630.

No Scaffold 226 foram identificados transposases pertencentes à família IS5 e também uma proteína chamada “Soj”. De acordo com o trabalho realizado por Lasocki e colaboradores (2007) em *Pseudomonas aeruginosa*, esta proteína está relacionada ao crescimento bacteriano, à segregação dos cromossomos e à motilidade.

Em todos os genes com ISs da família IS21 (Scaffold 221 e Scaffold 2.529), foi possível observar sempre a presença de duas sequências próximas identificadas como pertencentes a esta família. Este evento pode ser explicado pelo fato que membros da família IS21 codificam dois genes, um para transposase chamado *istA*, e um gene “auxiliar” chamado *istB* (SIGUIER et al., 2015).

No Scaffold 2.621 foram identificados dois ISs completos pertencentes à família IS91 e um IS não categorizável, pertencente à família IS110, além de duas proteínas próximas aos ISs. De acordo Siguier e colaboradores (2015), os ISs da família IS91 apresentam de 1.500 a 2.000 pb e possuem uma única ORF que codifica para a transposase. Os dois ISs identificados apresentam respectivamente, 879 pb e 1128 pb, sugerindo-se assim, que sejam um único IS.

Nos Scaffolds 581 e 1.895 foi identificado um IS completo pertencente à família *IS1634*, caracterizada principalmente por grandes transposases devido a um domínio de inserção β -Strand, também possuem uma única ORF que codifica para a transposase (SIGUIER et al., 2015). No Scaffold 581 também foi identificado a presença de três proteínas, entre elas a “*intA*”, uma enzima conhecida como “integrase”, presente em profagos, responsável por liga-se ao DNA da célula infectada.

Foi possível identificar no Scaffold 197 a presença de dois ISs pertencentes à família *ISKra4*, um é completo (1.206 pb) e o outro é não categorizável (402 pb). Os membros desta família variam de 1.250 pb a 3.700 pb e podem apresentar mais de uma ORF que codifica para a transposase, dessa forma, é possível que as duas ORFs encontradas façam parte de um único *ISKra4* (SIGUIER et al., 2015).

Nos Scaffolds 1.178 e 772 foram identificados ISs completos pertencentes à família *IS200_605*, os quais caracterizam-se por codificarem um ou dois genes para a transposase, e variam de 600 a 2.000 pb (SIGUIER et al., 2015).. É possível que no Scaffold 772 as duas ORFs identificadas pertençam a um único IS, já que a menor ORF possui apenas 351 pb .

Através desta análise do contexto genômico foi possível compreender melhor sobre as principais características de algumas famílias de ISs, abrindo novas perspectivas de estudos no campo dos EGMs, como verificar a ocorrência de TGL, transporte de genes passageiros pelos ISs, entre outros.

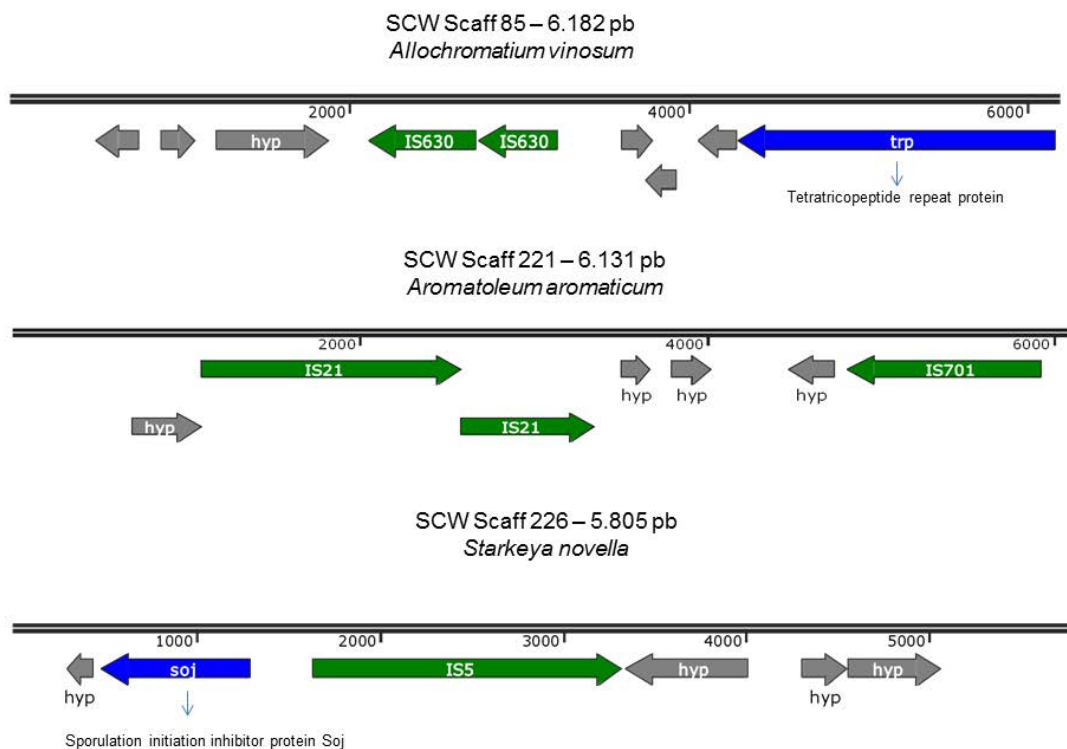


Figura 15. Representação esquemática do contexto genômico de ISs pertencentes à amostra SCW (*Sugarcane Wet*). A escolha das sequências obedece à critérios citados no texto.

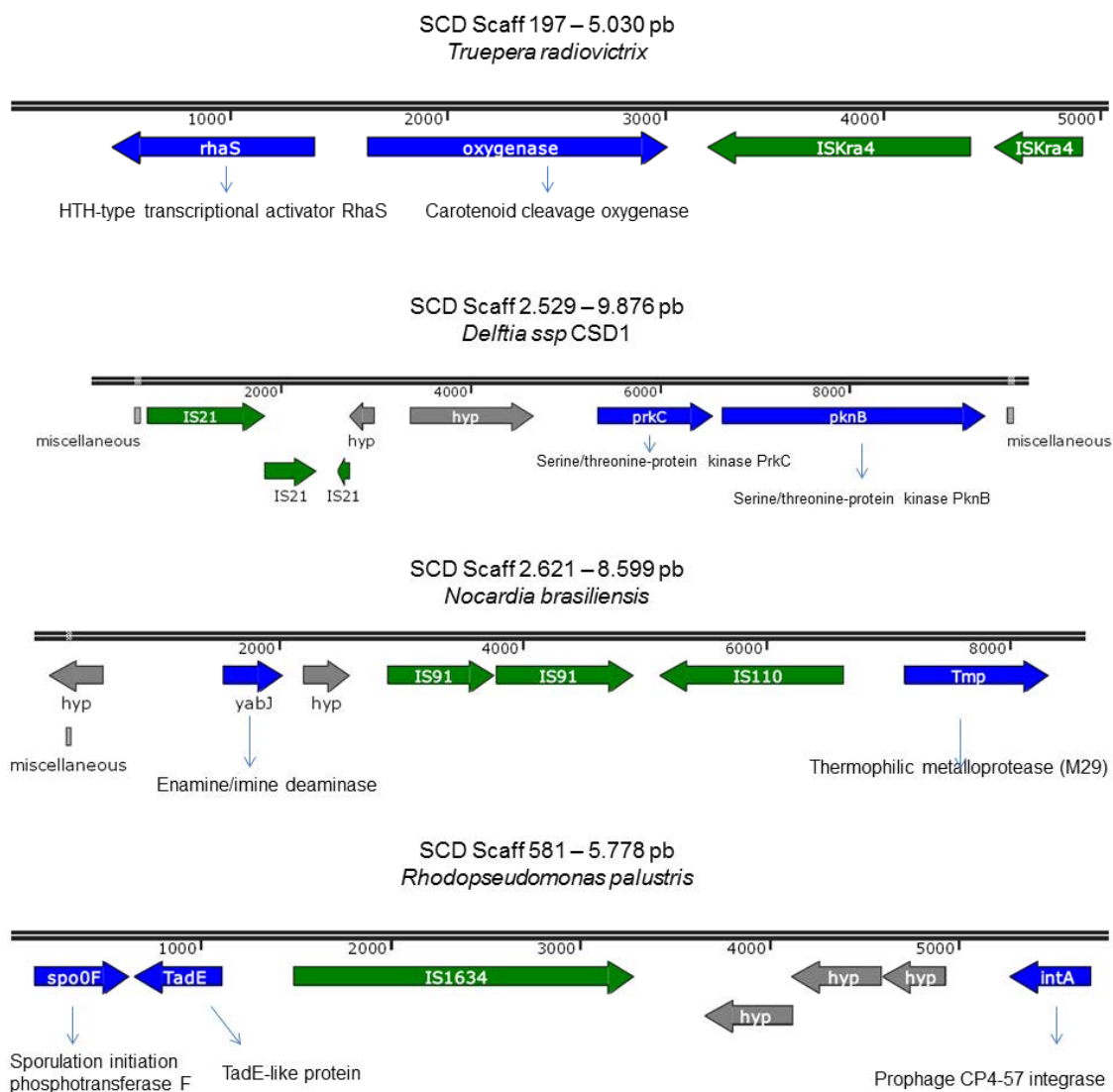


Figura 16. Representação esquemática do contexto genômico de ISs pertencentes à amostra SCD (*Sugarcane Dry*). A escolha das sequências obedece à critérios citados no texto.

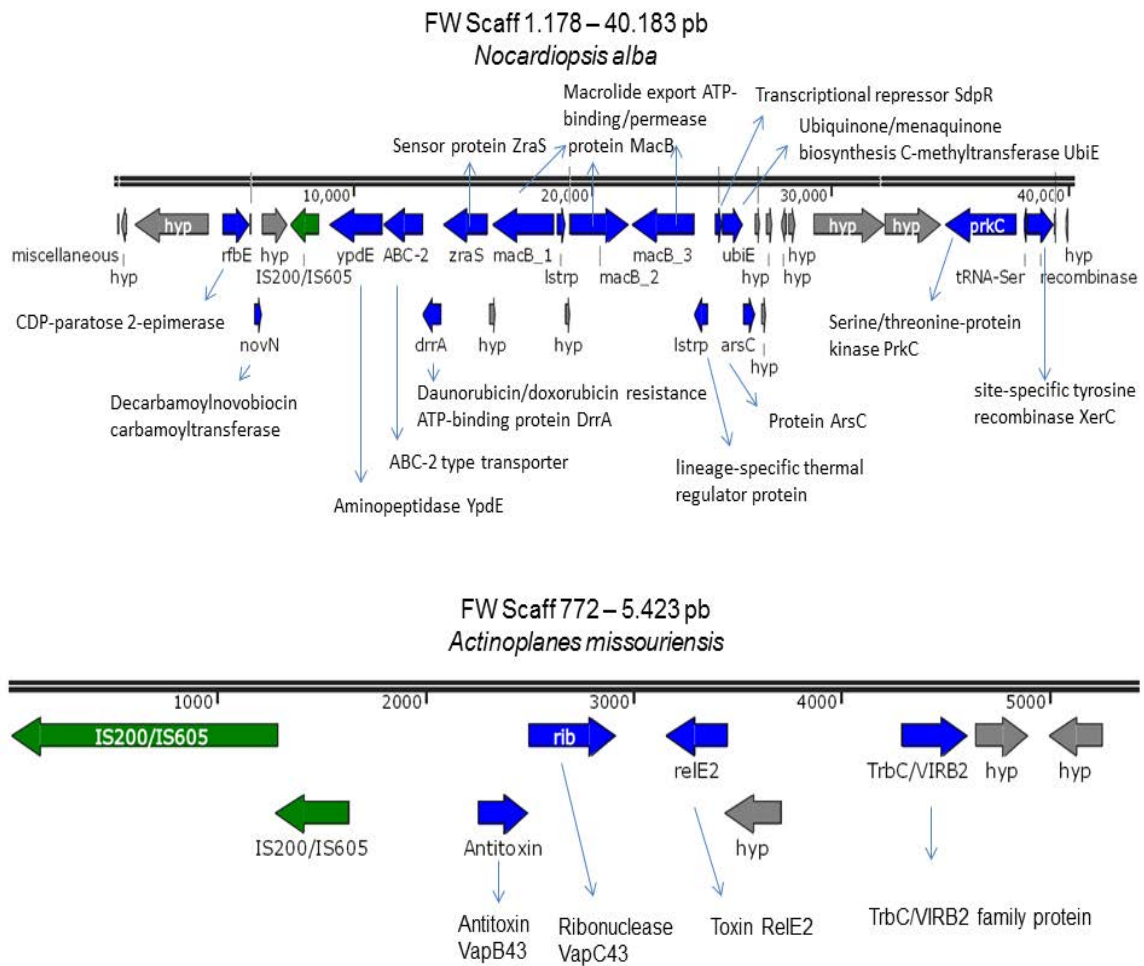


Figura 17. Representação esquemática do contexto genômico de ISs pertencentes à amostra FW (*Forest Wet*). A escolha das sequências obedece a critérios citados no texto.

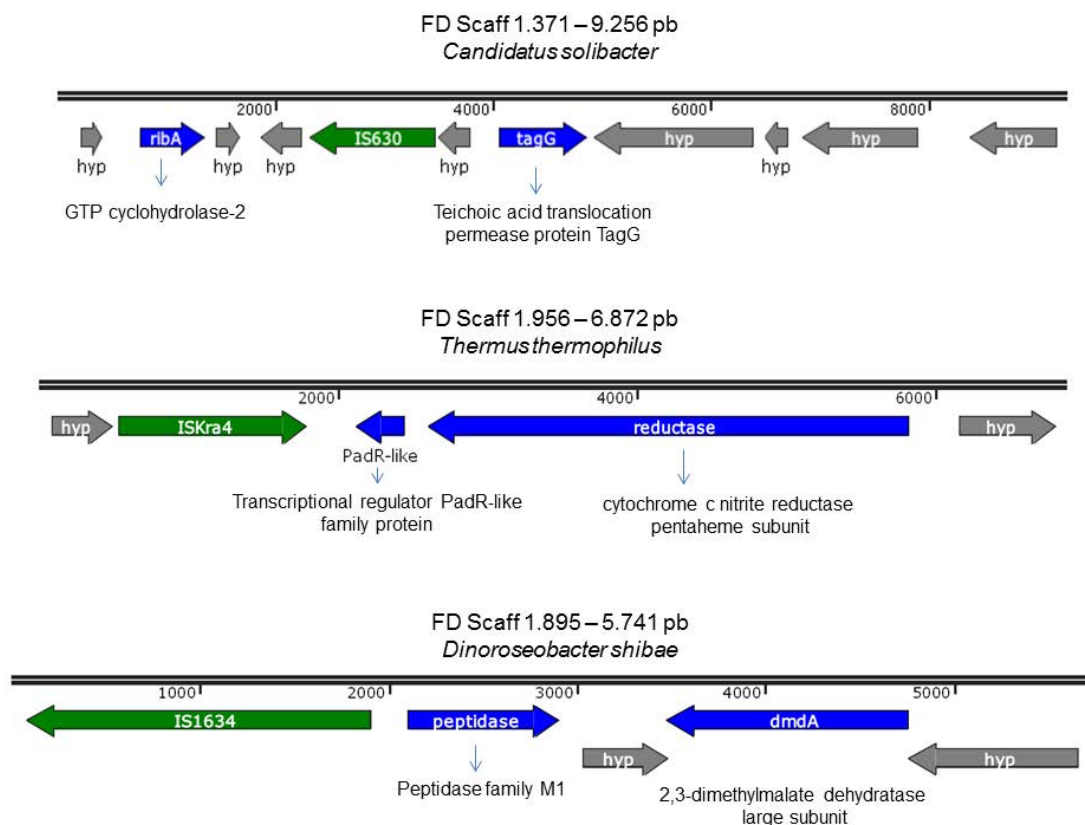


Figura 18. Representação esquemática do contexto genômico de ISs pertencentes à amostra FD (*Forest Dry*). A escolha das sequências obedece a critérios citados no texto.

5 CONCLUSÕES

Através da análise dos dados, foi possível perceber que a plataforma MG-RAST apresenta deficiências para identificação de transposases, enquanto a ferramenta ISsaga 2.0 oferece uma metodologia mais precisa para identificação de transposases em dados de metagenomas.

Também foi possível notar, a grande quantidade de ISs pertencentes à família ISNCY, confirmando o grande potencial destes dados para identificação de novas famílias de transposases. A família IS110 demonstrou grande abundância da família em todas as amostras, nas duas frentes de investigação. Além disso, é importante ressaltar que nenhum IS previamente descrito na literatura foi identificado.

Quanto à análise dos gêneros bacterianos mais abundantes, de acordo com a plataforma MG-RAST, em SCW foi o gênero *Azospirillum*, enquanto que em SCD foi o *Streptomyces* e em FW e FD, o gênero bacteriano mais abundante foi *Mycobacterium*.

A análise precisa de transposases em dados de metagenomas pode fornecer a identificação de novos ISs, auxiliando a descoberta de novas famílias de transposases e que em última instância pode auxiliar a revelar novos mecanismos de transposição.

6 PERSPECTIVAS

- Submissão de ISs à Base de dados ISfinder.
- Anotação de ISs completos, verificando a ocorrência de mecanismos de TGL.

7 REFERÊNCIAS

ALMEIDA, L. M.; CARARETO, C. M. A. Multiple events of horizontal transfer of the Minos transposable element between *Drosophila* species. **Molecular Phylogenetics And Evolution**, v. 35, n. 3, p.583-594, jun. 2005.

ALTSCHUL, S.F.; GISH, W.; MILLER, W.; MYERS, E.W.; LIPMAN, D.J. Basic local alignment search tool. **Journal of Molecular Biology**, Amsterdam, v. 215, p. 403-410, 1990.

ATLAS, R. M.; BARTHA, R.. **Microbial Ecology: fundamentals and applications**. 4. ed. Addison Wesley Longman. The Benjamin/cumming Publishing Company, Inc,. 694 p., 1997.

AZIZ, R.K.; BREITBART, M.; EDWARDS, R.A. Transposases are the most abundant, most ubiquitous genes in nature. **Nucleic Acids Res.**, v.13, p. 4.207-4.217, 2010.

BELLANGER, X; PAYOT, S; LEBLOND-BOURGET, N.; GUÉDON, G. Conjugative and mobilizable genomic islands in bacteria: evolution and diversity. **Fems Microbiol. Rev.**, v. 38, p.1-42, 2013.

BINNEWIES, T. T.; MOTRO, Y; HALLIN, P.F.; LUND, O.; LA T.; HAMPSON, D.J.; DUNN, D.; BELLGARD, M.; WASSENAAR, T. M.; USSERY, D.W.. Ten years of bacterial genome sequencing: comparative-genomics-based discoveries. **Functional & Integrative Genomics**, v. 6, p.165-185, 2006.

BIRKENBIHL, R.P.; VIELMETTER, W. Complete maps of IS1, IS2, IS3, IS4, IS5, IS30 and IS150 locations in *Escherichia coli* K12. **Molecular and General Genetics**, vol. 220, n.1, p.147-153, 1989.

BOERLIN, P.; REID-SMITH, R. J.. Antimicrobial resistance: its emergence and transmission. **Anim Health Res Rev.**, v. 9, n. 2, p.115-126, 2008.

BORNEMAN, J.; SKROCH, P.W.; O'SULLIVAN, K.M.; PALUS, J.A.; RUMJANEK, N.G.; JANSEN, J.L.; NIENHUIS, J.; TRIPLETT, E.W. Molecular Microbial diversity of an agricultural soil in Wisconsin. **Aplied and Environmental Microbiology**, vol. 62, n.6, p. 935-943, 1996.

BOUTOILLE, D.; CORVEC, S.; CAROFF, N.; GIRAUDEAU, C.; ESPAZE, E.; CAILLON, J.; PLÉSIAT, P.; REYNAUD, A. Detection of an IS21 insertion sequence in the mexR gene of *Pseudomonas aeruginosa* increasing L-lactam resistance. **Fems Microbiology Letters**, v. 230, n. 1, p.143-146, 2004.

BRUCE, T.; MARTINEZ I.B.; MAIA NETO, O.; VICENTE A.C.; KRUGER R.H.; THOMPSON, F.L. Bacterial Community Diversity in the Brazilian Atlantic Forest Soils. **Microbial Ecology**, v. 60, n. 4, p.840-849, 2010.

CAMBRAY, G.; GUEROUT, A.M.; MAZEL, D. Integrons. **Annual Reviews Genet.**, v. 44, p. 141–166, 2010.

CAMPBELL, A. STARLINGER, P.; BERG, D.E.; BOTSTEIN, D.; LEDERBERG, E.M.; NOVICK, R.P.; SZYBALSKI, W. Nomenclature of transposable elements in prokaryotes. **Plasmid**, vol. 2, n.3, p. 466-473, 1979

CARVALHO, M. C. C. G.; SILVA, D. C. G. Sequenciamento de DNA de nova geração e suas aplicações na genômica de plantas. **Cienc. Rural**, v. 40, n. 3, p.735-744, 2010

CHANDLER, M.; MAHILLON, J. Insertion Sequences revisited In CRAIG, N.L., CRAIGIE, R., GELLERT, M., LAMBOWITZ, A.M. (Eds.). *Mobile DNA II*. Washington D.C: **American Society for Microbiology**, p. 305–366, 2002.

CRAIG, N. L.; CRAIGIE, R.; GELLERT, M.; LAMBOWITZ A.M. **Mobile DNA II**. American Society for Microbiology, Washington, 1204p, 2002.

CRUZ, Edgar Andrés Ochoa. **Caracterização de transposases da família SChAT em cana de açúcar: estudo molecular e funcional**. 114 f. Tese (Doutorado) - Universidade de São Paulo, São Paulo, 2012.

DANIEL, R. The Metagenomics of soil. **Nature**, v. 3, p. 471-478, 2005.

FESCHOTTE, C.; JIANG, N.; WESSLER, S. R. Plant transposable elements: where genetics meets genomics. **Nature Reviews Genetics**, v.3, p.329-341, 2002.

FINNEGAN, D.J. Eukaryotic transposable elements and genome evolution. **Trends genet**, vol. 5, n.3, p. 103-107,1989.

FROST, L. S.; LEPLAE, R.; SUMMERS, A.O.; TOUSSAINT, A. Mobile genetic elements: the agents of open source evolution. **Nature Reviews Microbiology**, v. 3, n. 9, p.722-732, 2005.

FURUYA, E. Y.; LOWY, F. D. Antimicrobial-resistant bacteria in the community setting. **Nature Reviews Microbiology**, v. 1, n. 4, p.36-45, 2006.

GAO, B.; GUPTA, R.S. Conserved indels in protein sequences that are characteristic of the phylum Actinobacteria. **International Journal Of Systematic And Evolutionary Microbiology**, v. 55, n. 6, p.2401-2412, 2005.

GOODFELLOW, M; WILLIAMS, S T. Ecology of Actinomycetes. **Annu. Rev. Microbiol.**, v. 37, n. 1, p.189-216, 1983.

HAREN, L.; TON-HOANG, B.; CHANDLER, M.. INTEGRATING DNA: Transposases and Retroviral Integrases. **Annual Review of Microbiology**, v. 53, p.245-281, 1999.

HENNE, A.; SCHMITZ, R.A.; BOMEKE, M.; GOTTSCHALK, G.; DANIEL, R. Screening of Environmental DNA Libraries for the Presence of Genes Conferring Lipolytic Activity on Escherichia coli. **Appl Environ Microbiol.**, v. 66, n. 7, p.3113-3116, 2000.

HONG, K. S.; LIM, H.K.; CHUNG, E.J.; PARK, E.J.; LEE, M.H.; KIM, J.C.; CHOI, G.J.; CHO, K.Y.; LEE, S.W. Selection and characterization of forest soil metagenome genes encoding lipolytic enzymes. **J Microbiol Biotechnol.**, v. 17, n. 10, p.1655-1660, 2007.

HUGENHOLTZ, P.; GOEBEL, B. M.; PACE, N. R.. Impact of Culture-Independent Studies on the Emerging Phylogenetic View of Bacterial Diversity. **J. Bacteriol.**, v. 180, n. 18, p.4765-4774, 1998.

HUNTER-CEVERA, J.C. The value of microbial diversity. **Curr. Opin. Microbiol.**, v. 1, n. 3, p.278-285, 1998.

KAPER, J.B.; HACKER, J.; **Pathogenicity Islands and other mobile virulence elements**. ASM Press. Washington, 1999, 366p.

KURTBOKE, D.I.; SWINGS J.; STORMS V. **Microbial Genetic Resources and Biodiscovery**, United Kingdom, WFCC Publ., 400p, 2004.

LASOCKI, K.; BARTOSIK, A.A.; MIERZEJEWSKA, J.; THOMAS, C.M.; JAGURA-BURDZY, G. Deletion of the *parA* (*soj*) Homologue in *Pseudomonas aeruginosa* Causes ParB Instability and Affects Growth Rate, Chromosome Segregation, and Motility. **Journal Of Bacteriology**, v. 189, n. 15, p.5762-5772, 2007.

LING, L.L.; SCHNEIDER, T.; PEOPLES, A.J.; SPOERING, A.L.; ENGELS, I.; CONLON, B.P.; MUELLER, A.; SCHÄBERLE, T.F.; HUGHES, D.E.; EPSTEIN, S.; JONES, M.; LAZARIDES, L.; STEADMAN, V.A.; COHEN, D.R.; FELIX, C.R.; FETTERMAN, K.A.; MILLETT, W.P.; NITTI, A.G.; ZULLO, A.M.; CHEN, C.; LEWIS, K. A new antibiotic kills pathogens without detectable resistance. **Nature**. v. 517, p. 455-459, 2015.

MAHILON, J; CHANDLER, M. Insert sequences. **Microbiol. Mol. Biol. Rev.**, v.62, p. 725-774, 1998.

MAO, B.; LI, D.; ZHAO, J.; LIU, X.; GU, Z.; CHEN, Y.Q.; ZHANG, H.; CHEN, W. Metagenomic Insights into the Effects of Fructo-oligosaccharides (FOS) on the Composition of Fecal Microbiota in Mice. **J. Agric. Food Chem.**, v. 63, n. 3, p.856-863, 2015.

MARQUES, M. V.. **Biologia Molecular e Genética Bacteriana**. Ribeirão Preto: Sociedade Brasileira de Genética, 348 p., 2012.

MATSUOKA, M., MENDES, I.C.; LOUREIRO, M.F. Biomassa microbiana e atividade enzimática em solos sob vegetação nativa e sistemas agrícolas anuais e perenes na região de Primavera do Leste (MT). **Revista Brasileira de Ciência do Solo**, v.27, p. 425-433, 2003.

MEYER, F.; PAARMANN, D; SOUZA, M.D.; OLSON, R.; GLASS, E.M.; KUBA, M.; PACZIAN, T.; RODRIGUEZ, A.; STEVENS, R.; WILKE, A.; WILKENING, J.; EDWARDS, R.A. The metagenomics RAST server - a public resource for the automatic phylogenetic and functional analysis of metagenomes. **Bmc Bioinformatics**, v. 9, 2008.

MOCALI, S.; BENEDETTI, A. Exploring research frontiers in microbiology: the challenge of metagenomics in soil microbiology. **Research in Microbiology**, v. 161, n.6, p. 497-505, 2010.

MULLANY, P.; ALLAN, E.; ROBERTS, A. P. Mobile genetic elements in *Clostridium difficile* and their role in genome function. **Research In Microbiology**, v. 166, n. 4, p.361-367, 2015.

NORMAN, A., HANSEN, L.H., SORENSEN, S. J. Conjugative plasmids: vessels of the communal gene pool. **Philosophical Transactions of the Royal Society**, v. 364, n. 1527, p. 2275–2289, 2009.

NUNES, G. L. **Diversidade e Estrutura de comunidades de Bacteria e Archaea em solo de mangue contaminado com hidrocarbonetos de petróleo**. 2006. 85 f. Dissertação (Mestrado) - Curso de Mestrado em Microbiologia, Universidade de São Paulo, Piracicaba, 2006.

NÜSSLEIN, K.; TIEDJE, J. M. Soil bacterial community shift correlated with change from forest to pasture vegetation in a tropical soil. **Applied and Environmental Microbiology**, v. 65, n 8, p. 3622-3626, 1999.

OCHMAN, H.; LAWRENCE, J. G.; GROISMAN, E. A. Lateral gene transfer and the nature of bacterial innovation. **Nature**, v. 405, n. 6784, p.299-304, 2000.

PARKHILL, J. et al. Comparative analysis of the genome sequences of *Bordetella pertussis*, *Bordetella parapertussis* and *Bordetella bronchiseptica*. **Nature Genetics**, v. 35, n. 1, p.32-40, 2003.

PEDRINHO, E.A.N.; LEMOS E.G. M.; PEREIRA, R.M.; SCAQUITTO, D. C.; SILVEIRA, E. L.; VAL-MORAES, S. P.; ALVES, L. M. C.; WICKERT, E.; VALARINI, M. J. Avaliação do impacto do lodo de esgoto na microbiota do solo utilizando o gene 16S rRNA. **Arquivos do Instituto Biológico** (Impresso), v. 76, p. 447-452, 2009.

PISA, G.; MAGNANI, G.S.; WEBER, H.; SOUZA, E.M.; FAORO, H.; MONTEIRO, R.A.; DAROS, E.; BAURA, V.; BESPALHOK, J.P.; PEDROSA, F.O.; CRUZ, L.M. Diversity of 16S rRNA genes from bacteria of sugarcane rhizosphere soil. **Braz. J. Med. Biol. Res.**, v. 12, p. 1215-1221, 2011.

PONTIROLI, A.; KHERA, T.T.; OAKLEY B.B.; MASON, S.; DOWD, S.E.; TRAVIS, E.R.; ERENZO, G.; ASEFFA, A.; COURTENAY, O.; WELLINGTON, E.M.

Prospecting environmental mycobacteria: combined molecular approaches reveal unprecedented diversity. **PloS One**, v. 7, 2013.

REIGSTAD, L. J.; BARTOSSEK, R.; SCHLEPER, C. Preparation of High-Molecular Weight DNA and Metagenomic Libraries from Soils and Hot Springs. **Methods In Enzymology**, p.319-344, 2011.

RONDON, M. R.; GOODMAN, R.M.; HANDELSMAN, J. The Earth's bounty: assessing and accessing soil microbial diversity. **Trends Biotechnol.**, v. 17, n. 10, p.403-409, 1999.

ROSSELÓ-MORA, R.; AMANN, R. The species concept for prokaryotes. **FEMS Microbiology Review**, Amsterdam, v. 25, n. 1, p. 39-67, 2001.

SANGER, F.; NICKLEN, S.; COULSON, A.R. DNA sequencing with chain-terminating inhibitors. **Proc. Natl. Acad. Sci.**, v. 74, n.12, p. 5463-5467, 1977.

SIGUIER, PEROCHON, J.; LESTRADE, L.; MAHILLON, J.; CHANDLER, M. ISfinder: the reference centre for bacterial insertion sequences. **Nucleic Acids Research**, v. 34, p.32-36, 2006.

SIGUIER, P.; VARANI, A.; PEROCHON, J.; CHANDLER, M. Exploring Bacterial Insertion Sequences with ISfinder: Objectives, Uses, and Future Developments. **Methods In Molecular Biology**, p.91-103, 2012.

SIGUIER, P.; GOURBEYRE, E.; CHANDLER, M. Bacterial insertion sequences: their genomic impact and diversity. **Fems Microbiology Reviews**, v. 38, n. 5, p.865-891, 2014.

SIGUIER, P.; GOURBEYRE, E.; VARANI, A; TON-HOANG., B.; CHANDLER, M. Everyman's Guide to Bacterial Insertion Sequences. **Microbiology Spectrum**, v. 3, n. 2, 2015.

SILVA, J. C.; KIDWELL, M. G. Horizontal Transfer and Selection in the Evolution of P Elements. **Mol Biol Evol**, v. 10, n. 17, p.1542-1557, 2000.

SILVA, J. C.; LORETO, E.I.; CLARK, J.B.. Factors that affect the horizontal transfer of transposable elements. **Curr Issues Mol Biol.**, v. 1, n. 6, p.57-71, 2004.

SINZELLE, L.; IZSVÁK, Z.; IVICS, Z.. Molecular domestication of transposable elements: From detrimental parasites to useful host genes. **Cell. Mol. Life Sci.**, v. 66, n. 6, p.1073-1093, 2009.

SPAIN, A M.; KRUMHOLZ, L. R. E ELSHAHED, M. S. Abundance, composition, diversity and novelty of soil Proteobacteria. **ISME Journal.**, v. 3, n. 8, p. 992 1000, 2009.

STALEY, C.; JHONSON, D.; GOULD, T.J.; WANG, P.; PHILLIPS, J.; COTNER, J.B., SADOWSKY, M.J . Frequencies of heavy metal resistance are associated with land cover type in the Upper Mississippi River. **Science Of The Total Environment**, v. 511, p.461-468, 2015.

STALEY, J. **Microbial Diversity and the Biosphere**. 1998. Disponível em: <<http://www.bdt.org.br/oea/sib/staley>>. Acesso em: 20 mar. 2014.

STREIT W.R.; SCHMITZ R.A. Metagenomics-the key to the uncultured microbes. **Curr Opin Microbiol.**, v. 7, n. 5, p.492-498, 2004.

TAKETANI, R. G.; KAVAMURA, V.N.; MENDES, R.; MELO, I.S. Functional congruence of rhizosphere microbial communities associated to leguminous tree from Brazilian semiarid region. **Environmental Microbiology Reports**, v. 7, n. 1, p.95-101, 2015.

THOMAS, C. M.; NIELSEN, K. M. Mechanisms of, and Barriers to, Horizontal Gene Transfer between Bacteria. **Nature Reviews Microbiology**, v. 3, n. 9, p.711-721, 2005.

TORSVIK, V.; DAAE, F. L.; SANDAA, R. A.; OVREAS, L. Microbial diversity and function in soil: from genes to ecosystems. **Current Opinion in Microbiology**, Bergen, v. 5, n. 3, p. 240–245, 2002.

TOUCHON, M.; ROCHA, E. P. C.. Causes of Insertion Sequences Abundance in Prokaryotic Genomes. **Molecular Biology And Evolution**, v. 24, n. 4, p.969-981, 2007.

TREANGEN, T.J; DARLING A.E.; ACHAZ, G.; RAGAN, M.A.; MESSEGUER, X.; ROCHA, E.P . A novel heuristic for local multiple alignment of interspersed DNA repeats. **IEEE/ACM Trans Comput Biol Bioinform.**, v. 6, n. 2, p.180-189, 2009.

TRINGE, S. G.; RUBIN, E. M. Metagenomics: DNA sequencing of environmental samples. **Nature Reviews Genetics**, California, v. 6, n. 11, p. 805–814, 2005.

TURCATTI, G.I; ROMIEU, A.; FEDURCO, M.; TAIRI, A.P. A new class of cleavable fluorescent nucleotides: synthesis and optimization as reversible terminators for DNA sequencing by synthesis. **Nucleic Acids Research**, v. 36, n. 4, p.e25, 2007.

VARANI, A.M.; SIGUIER, P.; GOURBEYRE, E.; CHARNEAU, V.; CHANDLER, M. ISsaga is an ensemble of web-based methods for high throughput identification and semiautomatic annotation of insertion sequences in prokaryotic genomes. **Genome Biology**, v.12, n.3: R30, 2011.

VARANI, A. M.; MONTEIRO-VITORELLO, C.B.; NAKAYA, H.I.; VAN SLUYS, M.A. The Role of Prophage in Plant-Pathogenic Bacteria. **Annu. Rev. Phytopathol.** v. 51, n. 1, p.429-451, 2013.

WARD, D. M. A natural species concept for prokaryotes. **Curr. Opin. Microbiol.**, v. 1, n. 3, p.271-277, 1998.

WHITMAN, W. B.; COLEMAN, D. C.; WIEBE, W. J. Prokaryotes: The unseen majority. **Proc. Natl. Acad. Sci.**, v. 95, n. 12, p.6578-6583, 1998.

YANG, F.; YANG, J.; Zhang, X.; CHEN, L.; JIANG, Y.; YAN, Y.; TANG, X.; WANG, J.; XIONG, Z.; DONG, J.; XUE, Y.; ZHU, Y.; XU, X.; SUN, L.; CHEN, S.; NIE, H.; PENG, J.; XU, J.; WANG, Y.; YUAN, Z.; WEN, Y.; YAO, Z.; SHEN, Y.; QIANG, B.; HOU, Y.; YU, J.; JIN, Q. Genome dynamics and diversity of *Shigella* species, the etiologic agents of bacillary dysentery. **Nucleic Acids Res.**, v. 33, n. 19, p. 6445-6458, 2005.

APÊNDICES

APÊNDICE A

TABELA 1A. Classificação de transposases (*reads*) mais representativas atribuídas pela plataforma MG-RAST. Valores totais de hits encontrados para cada uma das amostras foram 1496 (SCW – *Sugarcane Wet*), 1691 (SCD – *Sugarcane Dry*), 1599 (FW – *Forest Wet*) e 1639 (FD – *Forest Dry*), em detalhes o número de *reads* identificados para cada classificação e abundância relativa.

	SCW		SCD		FW		FD	
	N.º reads	Abund.	N.º reads	Abund.	N.º reads	Abund.	N.º reads	Abund.
<i>transposase</i>	21876	27,17%	62447	25,91%	57340	29,71%	55606	28,30%
<i>putative transposase</i>	17290	21,48%	53386	22,15%	42015	21,77%	42310	21,53%
<i>transposase IS4 family protein</i>	13092	16,26%	36151	15,00%	27299	14,15%	29399	14,96%
<i>transposase IS116/IS110/IS902 family protein</i>	7167	8,90%	25267	10,48%	18651	9,67%	18540	9,44%
<i>transposase, IS605 OrfB family</i>	5379	6,68%	15985	6,63%	7754	4,02%	9872	5,02%
<i>transposase mutator type</i>	2817	3,50%	7728	3,21%	7648	3,96%	6410	3,26%
<i>transposase IS66</i>	1935	2,40%	5359	2,22%	4001	2,07%	4643	2,36%
<i>transposase, IS204/IS1001/IS1096/IS1165 family protein</i>	1897	2,36%	6316	2,62%	6626	3,43%	7522	3,83%
<i>rhodopirellula transposase</i>	368	0,46%	2173	0,90%	2179	1,13%	1939	0,99%
<i>transposase IS3/IS911 family protein</i>	738	0,92%	4737	1,97%	1983	1,03%	1952	0,99%
<i>transposase Tn3 family protein</i>	718	0,89%	3746	1,55%	1334	0,69%	1406	0,72%
<i>transposase, IS5 family, truncation</i>	371	0,46%	1552	0,64%	1358	0,70%	980	0,50%
<i>ISSpo2, transposase</i>	349	0,43%	1560	0,65%	1131	0,59%	1472	0,75%
<i>transposase IS111A/IS1328/IS1533</i>	594	0,74%	1178	0,49%	1625	0,84%	1005	0,51%
<i>IS630 family transposase</i>	500	0,62%	1389	0,58%	1029	0,53%	1251	0,64%
<i>ISRSO5-transposase protein</i>	297	0,37%	1963	0,81%	815	0,42%	937	0,48%

TABELA 2A. Porcentagem dos gêneros taxonômicos mais abundantes, identificados a partir dos *reads* classificados como transposases, pela plataforma MG-RAST para as amostras SCW (*Sugarcane Wet*), SCD (*Sugarcane Dry*), FW (*Forest Wet*) e FD (*Forest Dry*).

Gêneros	SCW		SCD		FW		FD	
	N.º <i>reads</i>	Abund.	N.º <i>reads</i>	Abund.	N.º <i>reads</i>	Abund.	N.º <i>reads</i>	Abund.
<i>Streptomyces</i>	1439	5,58%	6090	6,93%	6665	8,12%	5853	7,42%
<i>Mesorhizobium</i>	1885	7,30%	5991	6,82%	5728	6,98%	5600	7,10%
<i>Azospirillum</i>	2102	8,15%	5979	6,80%	4326	5,27%	4694	5,95%
<i>Aromatoleum</i>	1554	6,02%	5172	5,88%	4412	5,37%	4816	6,10%
<i>Mycobacterium</i>	1131	4,38%	5031	5,72%	11885	14,48%	6592	8,36%
<i>Rhodococcus</i>	1214	4,70%	4732	5,38%	7780	9,48%	5793	7,34%
<i>Rhizobium</i>	637	2,47%	2799	3,18%	1814	2,21%	2166	2,75%
<i>Bradyrhizobium</i>	1124	4,36%	2716	3,09%	2399	2,92%	2280	2,89%
<i>Burkholderia</i>	734	2,84%	2635	3,00%	2020	2,46%	2262	2,87%
<i>Ktedonobacter</i>	652	2,53%	2382	2,71%	724	0,88%	1114	1,41%
<i>Frankia</i>	482	1,87%	2355	2,68%	1766	2,15%	1859	2,36%
<i>Nitrobacter</i>	776	3,01%	2266	2,58%	2232	2,72%	2202	2,79%
<i>Symbiobacterium</i>	454	1,76%	2219	2,52%	1502	1,83%	1511	1,92%
<i>Anabaena</i>	1348	5,22%	2016	2,29%	742	0,90%	1066	1,35%
<i>Stigmatella</i>	579	2,24%	1993	2,27%	1568	1,91%	1800	2,28%
<i>Agrobacterium</i>	590	2,29%	1676	1,91%	1580	1,92%	1727	2,19%
<i>Nostoc</i>	1314	5,09%	1904	2,17%	661	0,81%	973	1,23%
<i>Magnetospirillum</i>	682	2,64%	1807	2,06%	1764	2,15%	1773	2,25%
<i>Octadecabacter</i>	478	1,85%	929	1,06%	1553	1,89%	1643	2,08%

TABELA 3A. Classificação de transposases (*scaffolds*) mais representativas atribuídas pela plataforma MG-RAST para as amostras SCW (*Sugarcane Wet*), SCD (*Sugarcane Dry*), FW (*Forest Wet*) e FD (*Forest Dry*), em detalhes o número de *scaffolds* e abundância relativa.

	SCW		SCD		FW		FD	
	N.º Scaff.	Abund.	N.º Scaff.	Abund.	N.º Scaff.	Abund.	N.º Scaff.	Abund.
<i>transposase</i>	831	27,65%	2895	28,64%	1998	32,84%	1855	29,73%
<i>putative transposase</i>	550	18,30%	1925	19,04%	1189	19,54%	1146	18,37%
<i>transposase IS4 family protein</i>	533	17,74%	1530	15,13%	934	15,35%	1066	17,09%
<i>transposase IS116/IS110/IS902 family protein</i>	371	12,35%	1103	10,91%	658	10,82%	743	11,91%
<i>transposase IS66</i>	85	2,83%	239	2,36%	175	2,88%	195	3,13%
<i>transposase, IS605 OrfB family</i>	121	4,03%	652	6,45%	146	2,40%	236	3,78%
<i>transposase, mutator type</i>	84	2,80%	262	2,59%	171	2,81%	175	2,80%
<i>transposase IS3/IS911 family protein</i>	34	1,13%	105	1,04%	99	1,63%	87	1,39%
<i>transposase, IS204/IS1001/IS1096/IS1165 family protein</i>	47	1,56%	274	2,71%	130	2,14%	134	2,15%
<i>transposase Tn3 family protein</i>	22	0,73%	257	2,54%	76	1,25%	58	0,93%
<i>insertion sequence transposase protein</i>	20	0,67%	55	0,54%	37	0,61%	49	0,79%
<i>transposase IS111A/IS1328/IS1533</i>	34	1,13%	72	0,71%	57	0,94%	58	0,93%
<i>IS630 family transposase</i>	17	0,57%	70	0,69%	30	0,49%	33	0,53%
<i>transposase, IS5 family, truncation</i>	39	1,30%	42	0,42%	52	0,85%	55	0,88%
<i>rhodopirellula transposase</i>	25	0,83%	86	0,85%	59	0,97%	52	0,83%
<i>transposase IS200-family protein</i>	26	0,87%	55	0,54%	21	0,35%	26	0,42%

TABELA 4A. Porcentagem dos gêneros taxonômicos mais abundantes, identificados a partir dos *scaffolds* classificados como transposases, pela plataforma MG-RAST para as amostras SCW (*Sugarcane Wet*), SCD (*Sugarcane Dry*), FW (*Forest Wet*) e FD (*Forest Dry*).

Gêneros	SCW		SCD		FW		FD	
	N.º reads	Abund.	N.º reads	Abund.	N.º reads	Abund.	N.º reads	Abund.
<i>Streptomyces</i>	26	2,37%	213	4,62%	219	7,01%	110	4,03%
<i>Mesorhizobium</i>	47	4,28%	248	5,38%	176	5,63%	141	5,17%
<i>Azospirillum</i>	59	5,37%	212	4,60%	133	4,25%	150	5,50%
<i>Aromatoleum</i>	47	4,28%	223	4,83%	185	5,92%	186	6,82%
<i>Mycobacterium</i>	24	2,19%	175	3,79%	292	9,34%	175	6,41%
<i>Rhodococcus</i>	32	2,91%	169	3,66%	209	6,69%	157	5,76%
<i>Rhizobium</i>	19	1,73%	59	1,28%	62	1,98%	58	2,13%
<i>Bradyrhizobium</i>	19	1,73%	58	1,26%	42	1,34%	43	1,58%
<i>Burkholderia</i>	10	0,91%	79	1,71%	65	2,08%	41	1,50%
<i>Ktedonobacter</i>	13	1,18%	164	3,56%	28	0,90%	25	0,92%
<i>Magnetospirillum</i>	18	1,64%	60	1,30%	47	1,50%	67	2,46%
<i>Nostoc</i>	36	3,28%	97	2,10%	12	0,38%	29	1,06%
<i>Symbiobacterium</i>	35	3,19%	107	2,32%	49	1,57%	63	2,31%
<i>Microcystis</i>	20	1,82%	46	1,00%	14	0,45%	5	0,18%
<i>Frankia</i>	6	0,55%	107	2,32%	51	1,63%	34	1,25%
<i>Agrobacterium</i>	19	1,73%	69	1,50%	59	1,89%	32	1,17%
<i>Bordetella</i>	13	1,18%	82	1,78%	46	1,47%	27	0,99%
<i>Geobacillus</i>	6	0,55%	68	1,47%	42	1,34%	41	1,50%

TABELA 5A. Classificação de transposases (*scaffolds*) mais representativas atribuídas pela plataforma ISSaga 2.0, em detalhes o número de *scaffolds* por amostra e abundância relativa para as amostras SCW (*Sugarcane Wet*), SCD (*Sugarcane Dry*), FW (*Forest Wet*) e FD (*Forest Dry*).

Família	SCW		SCD		FW		FD	
	N.º Scaff.	Abund.	N.º Scaff.	Abund.	N.º Scaff.	Abund.	N.º Scaff.	Abund.
IS110	4	4,26%	32	16,33%	22	19,64%	18	12,68%
IS3	15	15,96%	32	16,33%	18	16,07%	9	6,34%
IS481	6	6,38%	19	9,69%	6	5,36%	6	4,23%
ISNCY	36	38,30%	17	8,67%	7	5,36%	5	3,52%
IS21	2	2,13%	13	6,63%	6	5,36%	11	7,75%
IS630	3	3,19%	12	6,12%	9	8,04%	22	15,49%
IS1182	0	0,00%	9	4,59%	2	1,79%	3	2,11%
IS256	0	0,00%	9	4,59%	5	4,46%	6	4,23%
IS200_IS605	2	2,13%	8	4,08%	12	10,71%	12	8,45%
IS5	5	5,32%	7	3,57%	4	3,57%	11	7,75%
IS1595	3	3,19%	6	3,06%	2	1,79%	0	2,11%
IS91	2	2,13%	6	3,06%	2	1,79%	5	3,52%
ISKra4	2	2,13%	5	2,55%	0	0,89%	12	8,45%
IS1634	0	0,00%	4	2,04%	0	1,79%	6	4,23%
IS30	0	0,00%	3	1,53%	3	2,68%	3	2,11%
ISL3	4	4,26%	0	1,02%	0	1,79%	4	6,34%
IS6	4	4,26%	0	0,00%	4	3,57%	0	0,00%
IS701	5	5,32%	0	0,51%	0	0,00%	0	0,00%
IS1380	1	1,06%	0	1,02%	0	1,79%	0	0,00%
ISAzo13	0	0,00%	0	1,02%	3	2,68%	0	0,70%

TABELA 6A. Classificação de transposases (*reads*) mais representativas atribuídas pela plataforma MG-RAST após filtragem através do banco de dados de sequências de inserção ISfinder para as amostras SCW (*Sugarcane Wet*), SCD (*Sugarcane Dry*), FW (*Forest Wet*) e FD (*Forest Dry*), em detalhes o número de *reads* e a abundância relativa.

Família	SCW		SCD		FW		FD	
	N.º <i>reads</i>	Abund.	N.º <i>reads</i>	Abund.	N.º <i>reads</i>	Abund.	N.º <i>reads</i>	Abund.
IS110	3317	14,09%	11235	16,97%	8925	15,27%	9205	15,99%
IS5	2714	11,53%	5469	8,26%	5028	8,60%	5178	8,99%
IS630	2569	10,91%	5914	8,93%	6241	10,68%	6667	11,58%
IS3	1619	6,88%	3668	5,54%	3036	5,20%	3019	5,24%
IS701	1529	6,49%	4500	6,80%	3954	6,77%	4043	7,02%
IS91	1134	4,82%	4207	6,36%	3212	5,50%	3546	6,16%
IS1182	1047	4,45%	3066	4,63%	2737	4,68%	3018	5,24%
IS66	996	4,23%	2477	3,74%	2050	3,51%	2117	3,68%
IS256	971	4,12%	3123	4,72%	2308	3,95%	1926	3,35%
IS21	962	4,09%	3441	5,20%	4220	7,22%	3336	5,79%
ISNCY	845	3,59%	1512	2,28%	1482	2,54%	1134	1,97%
IS200/IS605	697	2,96%	2149	3,25%	1447	2,48%	1854	3,22%
ISL3	646	2,74%	1500	2,27%	1166	2,00%	1038	1,80%
ISAzo13	642	2,73%	3290	4,97%	3264	5,59%	3264	5,67%
IS481	546	2,32%	1572	2,37%	1290	2,21%	1101	1,91%
Tn3	528	2,24%	1928	2,91%	775	1,33%	798	1,39%
IS1380	0	0,00%	2837	4,29%	2066	3,54%	2158	3,75%