

Cândice Maria Boff

**MICRO-ORGANISMOS DO *PERMAFROST* COMO
INDICADORES PALEOCLIMÁTICOS DA ANTÁRTICA**

Dissertação submetida ao Programa de
Pós-Graduação em Biotecnologia e
Biotecnologias da Universidade Federal
de Santa Catarina para obtenção do
Grau de Mestre em Biotecnologia e
Biotecnologias

Orientador: Prof. Dr. Rubens Tadeu
Delgado Duarte

Florianópolis
2018

Ficha de identificação da obra elaborada pelo autor através do Programa de Geração Automática da Biblioteca Universitária da UFSC.

Boff, Cândice Maria

Micro-organismos do permafrost como indicadores paleoclimáticos da Antártica / Cândia Maria Boff ; orientador, Rubens Tadeu Delgado Duarte, 2018.
124 p.

Dissertação (mestrado) - Universidade Federal de Santa Catarina, Centro de Ciências Biológicas, Programa de Pós-Graduação em Biotecnologia e Biociências, Florianópolis, 2018.

Inclui referências.

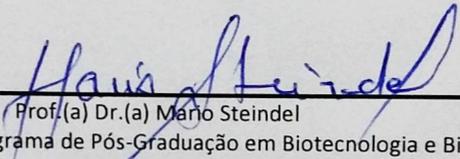
1. Biotecnologia e Biociências. 2. Metagenômica. 3. Paleoclima. 4. Antártica. I. Duarte, Rubens Tadeu Delgado. II. Universidade Federal de Santa Catarina. Programa de Pós-Graduação em Biotecnologia e Biociências. III. Título.

"Micro-organismos do *permafrost* como indicadores paleoclimáticos da Antártica"

Por

Cândice Maria Boff

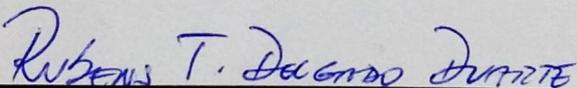
Dissertação julgada e aprovada em sua forma final pelos membros titulares da Banca Examinadora (13/PPGBTC/2018) do Programa de Pós-Graduação em Biotecnologia e Biociências - UFSC.



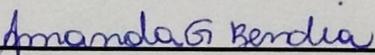
Prof. (a) Dr. (a) Mário Steindel

Coordenador(a) do Programa de Pós-Graduação em Biotecnologia e Biociências

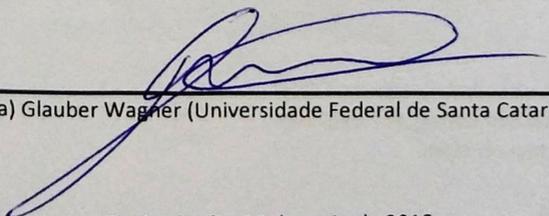
Banca examinadora:



Dr. (a) Rubens Tadeu Delgado Duarte (Universidade Federal de Santa Catarina)
Orientador(a)



Dr. (a) Amanda Gonçalves Bendia (Universidade de São Paulo)



Dr. (a) Glauber Wagner (Universidade Federal de Santa Catarina)

Florianópolis, 18 de maio de 2018.

Este trabalho é dedicado à vovó Maria,
que com tão pouco estudo, me ensinou
mais do que a Academia.

In memoriam.

AGRADECIMENTOS

Agradeço, em primeiro lugar, às minhas duas mães, a de primeiro e a de segundo grau. Pela educação que me deram, sempre me incentivando a estudar e buscar a minha formação acadêmica, mas antes de tudo, pelo que me ensinaram para a vida: a ser forte, independente, a enfrentar os obstáculos com serenidade e sabedoria. Obrigada por terem sido as mães, mulheres e amigas exemplares que são, por representarem meu porto seguro, meu colo, meu espelho para sempre ser alguém melhor.

Aos meus colegas amigos família da graduação, pela quase década de amizade irrevelante. Pelas aventuras na UFSC e fora dela, pela cumplicidade, amor e apoio incondicionais que nos mantém unidos até hoje. Andinho, Chun, Dani, Elaiza, JG, Gug, Grosi, Laurinha, Mari, Pri, Panda, Tabatinha, Tomás, obrigada por saber que, mesmo nem sempre tão presentes, tenho vocês do meu lado, como um pedaço de mim, da minha memória, da minha personalidade, que sempre estará comigo.

Ao Tulio, a pessoa mais inteligente e impressionante que eu já tive a sorte e o prazer de conhecer. Obrigada por me ajudar na interpretação de dados com a sua inteligência e capacidade notáveis, por estar ao meu lado nos momentos mais difíceis e cruciais deste trabalho, suportando meu cansaço, meu mau-humor, meus colapsos emocionais e me levantando todas as vezes em que achei que não tinha mais forças. Mas principalmente, obrigada por ter entrado na minha vida e fazer parte dela agora.

Aos meus companheiros de mestrado, Chun (sim, de novo!), Mari e Gabriel, por compartilhar comigo essa ideia louca que a gente inventou de fazer Mestrado, por ter em quem se apoiar pra não endoidar (muito) de vez. Pelos almoços, pelas conversas, pelo apoio moral e pelos comentários ~~maldosos~~ ácidos (pra manter o bom humor), muito obrigada!

Aos meus *roomies* Akira, Henrique, Fifys e Patty, pelo companheirismo e carinho compartilhados nesses dois anos de convivência familiar, pelas conversas, pelas risadas, pelas sessões de Netflix, pelos lanches e Cocas compartilhados e pelos *sunsets*, muito obrigada!

Ao meu orientador, Prof. Dr. Rubens Duarte, que primeiro me apresentou ao mundo da Bioinformática e aceitou me guiar por esse caminho tortuoso, obrigada por todos os conhecimentos (e também pela calma e paciência!) transmitidos.

Aos meus colegas de trabalho do Biotério Central, em especial à Dona Jô, pela imensa sensibilidade, compreensão, paciência, incentivo e apoio a concluir minha formação, muito obrigada!

À equipe do Instituto Pasteur do Uruguai, que me proporcionou a oportunidade de realizar um curso prático de análise metagenômica no ano passado, onde aprendi muitas coisas imprescindíveis para este trabalho.

Muito obrigada!

“I used to think the top environmental problems were biodiversity loss, ecosystem collapse and climate change. I thought that, with 30 years of good science, we could address those problems. But I was wrong. The top environmental problems are selfishness, greed and apathy... and to deal with those we need a cultural and spiritual transformation – and we scientists don’t know how to do that.”

James Gustave Speth
(American Environmental Lawyer)

RESUMO

O *permafrost* (solo permanentemente congelado por pelo menos dois anos consecutivos) é um importante componente da criosfera, e apesar de um ambiente extremo (temperaturas abaixo de zero, pouca disponibilidade de água líquida e nutrientes), serve de abrigo para uma diversa e abundante comunidade de micro-organismos. As camadas do *permafrost* geralmente possuem uma estratificação vertical quanto à idade do congelamento, que pode atingir milhares a milhões de anos. Suas características físico-químicas constituem condições ideais para preservação de material biológico, incluindo moléculas de DNA, que podem ser recuperadas do ambiente, sequenciadas e analisadas – uma estratégia de estudo de diversidade denominada metagenômica. Metagenomas conservados no *permafrost* de diferentes idades de congelamento representam um registro paleoclimático que pode revelar como a comunidade microbiana respondeu a mudanças climáticas no passado. A Península Antártica passou por um evento de aquecimento, conhecido como o Ótimo Climático do Holoceno Médio, entre 4,5 e 2,8 mil anos antes do presente. O presente trabalho avaliou o conteúdo taxonômico e funcional de metagenomas do sedimento marinho moderno e de cinco estratos de *permafrost* da Península Antártica, correspondentes ao período de 7,5 a 2,25 mil anos atrás. Foram observadas alterações na abundância de diversas categorias metabólicas entre os estratos, demonstrando os impactos do Ótimo Climático do Holoceno Médio para as comunidades microbianas. Os resultados sugerem que o aumento de temperatura deu origem a um ambiente anóxico e rico em matéria orgânica, favorecendo populações copiotróficas e com maior capacidade de ajuste do metabolismo. Os metagenomas correspondentes ao período anterior ao Ótimo Climático apresentaram maior abundância de produtos gênicos relacionados ao metabolismo de lipídios e compostos aromáticos, ácidos nucleicos e transporte de membrana, indicando micro-organismos especializados na manutenção do metabolismo a baixas temperaturas. Estas populações foram substituídas por uma comunidade de metabolismo mais diversificado e generalista nos estratos do período mais quente, representados por uma maior variabilidade de funções metabólicas e abundância de genes relacionados à regulação da transcrição. Estes resultados demonstram que as comunidades microbianas do *permafrost* podem servir como indicadores de mudanças paleoclimáticas.

Palavras-chave: Metagenômica. Paleoclima. Holoceno.

ABSTRACT

Permafrost (permanently frozen soil for at least two consecutive years) is an important feature of the cryosphere, and despite being an extreme environment (below-zero temperatures, low liquid water content and nutrient depleted), it hosts an abundant and diverse microbial community. Permafrost shows depth stratification associated with its age, which can reach thousands to millions of years. Its physico-chemical conditions are ideal to maintain biological material preserved, including DNA molecules that can be recovered, sequenced and analyzed directly from the environment – a strategy for microbial diversity studies called metagenomics. Metagenomes preserved in permafrost layers from different ages represent a paleoclimate record that can reveal how the microbial community responded to climate events in the past. The Antarctic Peninsula underwent a warming event known as the mid-Holocene Climate Optimum, between 4.5 and 2.8 thousand years before present. This work has evaluated the taxonomic and functional content of metagenomes of modern marine sediment and five permafrost layers of the Antarctic Peninsula, corresponding to the period from 7.5 to 2.25 thousand years ago. Changes were observed in the abundance of several metabolic categories between layers, showing the impacts of this mid-Holocene warmth on the microbial communities. Results suggest that the temperature rising has generated an anoxic, organic matter rich environment, favoring copiotrophic populations with increased regulatory capacities. Metagenomes corresponding to the prior period showed greater abundances of gene products related to the metabolism of lipids and aromatic compounds, nucleic acids and membrane transport, suggesting a microbial community specialized in maintaining growth and metabolism at low temperatures. These populations were replaced by microorganisms with a more diverse and generalist metabolic structure in the warmer samples, represented by a greater variability in metabolic functions and greater abundance of genes related to transcription regulation. These results have shown that the microbial communities of permafrost ancient layers can be used as proxy for past paleoclimatic changes.

Keywords: Metagenomics. Paleoclimate. Holocene.

LISTA DE FIGURAS

Figura 1. Dinâmica de circulação de massas de ar e correntes oceânicas influenciadas pela Antártica	30
Figura 2. Localização de algumas Estações Meteorológicas Automáticas na Antártica	35
Figura 3. Mudanças observadas na dinâmica de geleiras da Península Antártica ao longo do tempo por latitude	37
Figura 4. Registros paleoclimáticos da Terminação I, inferidos a partir de dados de testemunhos de gelo	39
Figura 5. Localização dos testemunhos de gelo a partir dos quais condições paleoclimáticas da Antártica podem ser inferidas	40
Figura 6. Representação esquemática de mudanças no nível do mar na Ilha Rei George, arquipélago das Shetlands do Sul	41
Figura 7. Consenso de temperatura para a Antártica ao longo dos últimos 7 mil anos	43
Figura 8. Características do <i>permafrost</i> geradas por processos de crioturbação	45
Figura 9. Distribuição taxonômica de filamentos de bactérias no <i>permafrost</i> de diversas localidades	48
Figura 10. Procedimento de montagem de genomas a partir de <i>short reads</i> provenientes das plataformas de sequenciamento	52
Figura 11. Representação esquemática do conceito de “profundidade da cobertura”	55
Figura 12. Diagrama de fluxo representando as etapas básicas da análise de dados metagenômicos e exemplos de ferramentas disponíveis atualmente	57
Figura 13. a) Mapa da Antártica indicando a localização da Ilha Rei George; b) Mapa da Ilha Rei George indicando o ponto de coleta das amostras de <i>permafrost</i>	60
Figura 14. a) Fotografia da área de coleta próxima à estação russa Bellingshausen, Ilha Rei George, Antártica; b) Equipamento de perfuração à seco utilizado na coleta das amostras de <i>permafrost</i> ; c) Detalhe do testemunhador utilizado na perfuração; d) Exemplos de algumas amostras de <i>permafrost</i> coletadas; e) Procedimento de descontaminação da superfície do <i>permafrost</i>	61

Figura 15. Representatividade das sequências atribuídas a famílias ou domínios proteicos de função desconhecida ou hipotética em cada metagenoma	70
Figura 16. Representatividade de cada subgrupo de OFUs definidos pela seleção de dados a partir dos valores de distribuição entre as amostras	72
Figura 17. Representatividade (em termos percentuais) das 25 categorias funcionais do SEED <i>Subsystems</i> em cada amostra do <i>permafrost</i>	74
Figura 18. Representatividade (em termos percentuais) das 25 categorias funcionais do SEED <i>Subsystems</i> entre as amostras de <i>permafrost</i>	75
Figura 19. Abundância relativa das diferentes subcategorias relacionadas ao metabolismo de aminoácidos e proteínas	76
Figura 20. Abundância relativa de rotas metabólicas relacionadas ao metabolismo de proteínas	76
Figura 21. Abundância relativa das diferentes subcategorias relacionadas ao metabolismo de carboidratos	77
Figura 22. Abundância relativa de rotas metabólicas relacionadas ao metabolismo de carboidratos	77
Figura 23. Abundância relativa das diferentes subcategorias relacionadas ao metabolismo de lipídios	78
Figura 24. Abundância relativa de rotas metabólicas relacionadas ao metabolismo de lipídios	78
Figura 25. Abundância relativa das OFUs relacionadas ao metabolismo de compostos aromáticos	79
Figura 26. Abundância relativa de rotas metabólicas relacionadas à degradação de compostos xenobióticos	79
Figura 27. Abundância relativa das diferentes subcategorias relacionadas ao metabolismo de cofatores, vitaminas, grupos prostéticos, pigmentos e metabolismo secundário	80
Figura 28. Abundância relativa de rotas metabólicas relacionadas ao metabolismo secundário e metabolismo de cofatores e vitaminas	80
Figura 29. Abundância relativa de OFUs relacionadas ao metabolismo de fosfolipídios, fosfolipases e proteínas de membrana	81

Figura 30. Abundância relativa das diferentes subcategorias de transporte de membrana	81
Figura 31. Abundância relativa de rotas metabólicas relacionadas à biossíntese e metabolismo de glicanas	82
Figura 32. Abundância relativa de OFUs relacionadas à fotossíntese e respiração	82
Figura 33. Abundância relativa de OFUs relacionadas à respiração aeróbica e anaeróbica	83
Figura 34. Abundância relativa de classes respiratórias anaeróbicas	83
Figura 35. Abundância relativa de rotas metabólicas relacionadas ao metabolismo energético	84
Figura 36. Abundância relativa das subcategorias relacionadas ao metabolismo de ácidos nucleicos	84
Figura 37. Detalhamento das subcategorias relacionadas ao metabolismo de ácidos nucleicos	85
Figura 38. Abundância relativa de rotas metabólicas relacionadas ao metabolismo de ácidos nucleicos, derivadas do KEGG <i>Pathway "Nucleotide Metabolism"</i>	85
Figura 39. Abundância relativa de OFUs relacionadas à regulação da transcrição	86
Figura 40. Abundância relativa de OFUs relacionadas à regulação e sinalização celular	86
Figura 41. Abundância relativa de rotas metabólicas relacionadas à transdução de sinal	87
Figura 42. Abundância relativa de OFUs relacionadas aos processos de recombinação, competência e conjugação	87
Figura 43. Diversidade de OFUs relacionadas ao processamento de elementos transponíveis, DNA de fagos e Módulos toxina/antitoxina	88
Figura 44. Abundância relativa de OFUs relacionadas à resistência a antibióticos e outras substâncias	88
Figura 45. Abundância relativa de OFUs relacionadas à produção de antibióticos e toxinas e sistemas de defesa	89

Figura 46. Abundância relativa de OFUs relacionadas à resposta ao stress	89
Figura 47. Análise de Componentes Principais (PCA) entre os dados de abundância relativa de categorias funcionais do SEED <i>Subsystems</i> e os parâmetros físico-químicos das amostras de <i>permafrost</i>	91
Figura 48. Distribuição taxonômica de filós de Bacteria entre as amostras	92
Figura 49. Distribuição taxonômica de filós de Archaea entre as amostras	93
Figura 50. Abundância de alguns dos principais grupos de bactérias marinhas de estratégia generalista nos diferentes estratos do <i>permafrost</i>	93
Figura 51. Abundância de gêneros de arqueias metanogênicas nas amostras de <i>permafrost</i>	94
Figura 52. Abundância de gêneros de bactérias redutoras de sulfato nas amostras de <i>permafrost</i>	94
Figura 53. Abundância de gêneros de Cyanobacteria nas amostras de <i>permafrost</i>	95
Figura 54. Abundância de algas eucarióticas nos diferentes estratos do <i>permafrost</i>	95
Figura 55. Principais eventos paleoclimáticos dos últimos 21 mil anos na Península Antártica	99

LISTA DE TABELAS

Tabela 1. Tecnologias de monitoramento de algumas variáveis ambientais acompanhadas periodicamente no continente antártico	34
Tabela 2. Algumas das principais bases de dados utilizadas para anotação de genomas e metagenomas, de acordo com suas funcionalidades	54
Tabela 3. Amostras de sedimento marinho e <i>permafrost</i> utilizadas neste trabalho	60
Tabela 4. Resultados de controle de qualidade e anotação das sequências obtidos pelo <i>pipeline</i> automático do MG-RAST	67
Tabela 5. Resultados do pipeline de pré-processamento aplicado localmente	68
Tabela 6. Resultados do processo de anotação de sequências utilizando as ferramentas InterProScan e eggNOG- <i>mapper</i>	68
Tabela 7. Resultados do processo de anotação utilizando a ferramenta InterProScan, demonstrando o número de sequências anotadas de acordo com as bases de dados Pfam e KEGG Pathways, e o número de OFUs (Pfam Ids e KEGG Pathways) não redundantes identificados em cada amostra	69
Tabela 8. Número de OFUs compartilhadas entre cada par de amostras	70
Tabela 9. Resultados de abundância relativa para as 27 categorias funcionais do SEED <i>Subsystems</i> entre as amostras de <i>permafrost</i>	73
Tabela 10. Dados físico-químicos (metadados) obtidos para os estratos de <i>permafrost</i>	90

LISTA DE ABREVIATURAS E SIGLAS

CCA – Corrente Circumpolar Antártica

MAHS – Modo Anular do Hemisfério Sul

AP – Antes do Presente

IPCC – Painel Intergovernamental de Mudanças Climáticas (do inglês “*Intergovernmental Panel for Climate Change*”)

CID – Carbono Inorgânico Dissolvido

CTD – Instrumento oceanográfico para medição de condutividade, temperatura e profundidade (do inglês “*Conductivity, Temperature, Depth*”)

ONU – Organização das Nações Unidas

CFCs – Clorofluorcarbonetos

UV – Ultravioleta

ppm – Partes por milhão

RNAr 16S – gene codificante para a subunidade menor (16S) do ribossomo de procariotos, utilizado como marcador taxonômico.

ORF – Janela Aberta de Leitura (do inglês “*Open Reading Frame*”)

ARN – Abundância Relativa Normalizada

PCA – Análise de Componentes Principais (do inglês “*Principal Component Analysis*”)

OTU – Unidade Taxonômica Operacional (do inglês “*Operational Taxonomic Unit*”)

OFU – Unidade Funcional Operacional (do inglês “*Operational Functional Unit*”)

SUMÁRIO

1	INTRODUÇÃO	23
1.1	OBETIVOS	25
1.1.1	<i>Objetivo Geral</i>	25
1.1.2	<i>Objetivos Específicos</i>	25
2	REVISÃO BIBLIOGRÁFICA	27
2.1	A ANTÁRTICA NO ECOSISTEMA GLOBAL	27
2.1.1	<i>O Continente dos Extremos</i>	27
2.1.2	<i>O Continente dos Micro-organismos</i>	28
2.1.3	<i>Regulação do Clima Global</i>	29
2.1.4	<i>Importância no Ciclo do Carbono</i>	31
2.2	MUDANÇAS CLIMÁTICAS	32
2.2.1	<i>Impactos sobre o Continente Antártico</i>	32
2.2.2	<i>A Península Antártica</i>	35
2.3	PALEOCLIMA	36
2.3.1	<i>O Holoceno</i>	38
2.3.2	<i>O Holoceno na Península Antártica</i>	40
2.4	PERMAFROST	43
2.4.1	<i>Diversidade Microbiana no Permafrost</i>	46
2.4.2	<i>O Permafrost no Contexto de Mudanças Climáticas</i>	47
2.5	METAGENÔMICA	49
2.5.1	<i>Short Reads x Montagem de Sequências</i>	50
2.5.2	<i>Anotação de metagenomas</i>	52
2.5.3	<i>Mapeamento de reads</i>	53
2.5.4	<i>Pipelines completos</i>	55
3	METODOLOGIA	59
3.1	COLETA E PROCEDIMENTOS DE BANCADA	59
3.2	ANÁLISES <i>IN SILICO</i>	62
3.2.1	<i>Hardware</i>	62
3.2.2	<i>Pré-processamento das sequências</i>	63
3.2.3	<i>Anotação das sequências do metagenoma</i>	63
3.2.4	<i>Mapeamento de reads</i>	64

3.2.5	<i>Pós-processamento</i>	64
3.2.6	<i>Seleção de dados</i>	65
3.2.7	<i>Estatística Multivariada</i>	66
4	RESULTADOS	67
4.1	PRÉ-PROCESSAMENTO	67
4.2	ANOTAÇÃO	68
4.3	ABUNDÂNCIA DIFERENCIAL	72
4.3.1	<i>Metabolismo de biomoléculas</i>	76
4.3.2	<i>Membrana Plasmática</i>	81
4.3.3	<i>Metabolismo Energético</i>	82
4.3.4	<i>Metabolismo de Ácidos Nucleicos</i>	84
4.3.5	<i>Regulação e Sinalização Celular</i>	86
4.3.6	<i>Promotores de Diversidade Genética</i>	87
4.3.7	<i>Resistência, Stress e Defesa</i>	88
4.4	ESTATÍSTICA MULTIVARIADA.....	90
4.5	TAXONOMIA	92
5	DISCUSSÃO	95
5.1	DISCUSSÃO METODOLÓGICA.....	95
5.2	DISCUSSÃO BIOLÓGICA	96
6	CONCLUSÕES	105
7	REFERÊNCIAS	107
8	APÊNDICES	107
8.1	APÊNDICE A – PROGRAMAS, COMANDOS E PARÂMETROS.....	117
8.2	APÊNDICE B – RESULTADOS DE ABUNDÂNCIA DIFERENCIAL INCLUINDO A AMOSTRA B2.....	123

1 INTRODUÇÃO

O *permafrost* é definido como o tipo de solo que se mantém permanentemente congelado (abaixo de 0°C) por pelo menos dois anos consecutivos (JANSSON; TAŞ, 2014). Ocorre principalmente em regiões próximas dos polos, sendo mais comum no Ártico, embora também seja encontrado na Antártica e em regiões montanhosas, como os Alpes europeus, Patagônia, cordilheira dos Andes e Alpes do Sul da Nova Zelândia (BOCKHEIM; MUNROE, 2014). Sua composição pode variar substancialmente de acordo com as características do local, podendo incluir, além dos próprios minerais que o compõem, grandes quantidades de matéria orgânica, como vestígios animais e vegetais, metabólitos, fragmentos celulares ou mesmo tecidos e células inteiras, dentre as quais destaca-se uma abundante biomassa de microorganismos.

As camadas de solo que permanecem congeladas ao longo das estações podem variar em espessura e, geralmente, estão associadas a uma estratificação vertical no que diz respeito à idade do congelamento, que pode atingir milhares a milhões de anos. Os estratos mais superficiais (e, portanto, congelados há menos tempo), costumam apresentar uma maior concentração de células microbianas, por estarem mais próximos da camada ativa (i.e., a porção superficial do solo suscetível à ação do tempo). Apesar disso, as camadas mais profundas (e também mais antigas) podem conter grandes quantidades de microorganismos ativos (RIVKINA et al., 2000). Estas comunidades são de grande interesse para estudos sobre sobrevivência e adaptação a condições extremas, além de representarem modelos de vida microbiana em ambientes extraterrestres (GILICHINSKY et al., 2007).

Temperaturas abaixo de zero, pouca disponibilidade de água líquida e baixas taxas de nutrientes constituem condições ideais para a preservação de matéria orgânica, de modo que o *permafrost* representa um enorme repositório de material biológico antigo. Esta biomassa preservada pode incluir grandes quantidades de moléculas de DNA, que podem ser recuperadas, sequenciadas e fornecer informações valiosas sobre as condições ambientais em tempos passados. O sequenciamento de material genético diretamente recuperado do ambiente denomina-se metagenômica, e vem sendo cada vez mais empregada nos estudos de diversidade microbiana. O mesmo termo é comumente empregado para referir-se a duas variações desta metodologia: a análise de todo e qualquer material genômico presente em uma amostra (metagenômica *shotgun*), ou baseada na investigação de um único ou mais genes de

interesse (metagenômica *targeted*) (OULAS et al., 2015). Esta última é comumente aplicada em estudos de diversidade taxonômica, utilizando o gene codificante para o RNA ribossomal 16S. A análise do conteúdo total de um metagenoma, no entanto, fornece uma gama maior de informações sobre a comunidade microbiana, incluindo seu potencial metabólico, funcional e ecológico.

Alguns estudos de metagenômica já foram empregados com o objetivo de acessar a diversidade funcional de comunidades microbianas no *permafrost*, compreendendo as localidades do Ártico canadense (YERGEAU et al., 2010), do Alasca (MACKELPRANG et al., 2011; LIPSON et al., 2013; TAŞ et al., 2014; HULTMAN et al., 2015; JOHNSTON et al., 2016; TRIPATHI et al., 2018), da Suécia (MONDAV et al., 2014b), da Sibéria (KRIVUSHIN et al., 2015), da ilha de Svalbard, na Noruega (TVEIT et al., 2013, 2015) e do Planalto do Tibete (GUAN et al., 2013). Na Antártica, esta abordagem foi aplicada apenas muito recentemente: Bowman e Ducklow (2015) utilizaram dados metagenômicos para descrever a estrutura metabólica das comunidades microbianas do *permafrost* da Península Antártica, porém utilizando a composição taxonômica (via RNAr 16S) para inferência de funcionalidade, e Goordial e colaboradores (2016) compararam os perfis metabólicos de dois diferentes micro-habitats da região dos Vales Secos de McMurdo.

Nos últimos 50 anos, têm-se observado que a camada ativa do *permafrost* vem aumentando em diversas localidades devido a eventos de aquecimento, resultando no descongelamento de extratos cada vez mais profundos e na consequente exposição do material conservado em seu interior (KÖHLER; KNORR; BARD, 2014). A reintrodução de material orgânico, outrora inerte, na atmosfera, pode significar um aumento substancial na emissão de gases do efeito estufa, devido a intensificação da atividade decompositora dos micro-organismos (SCHUUR et al., 2015). Micro-organismos são bastante responsivos a alterações ambientais, além de terem importância fundamental no balanço energético, ciclagem de nutrientes e ciclos biogeoquímicos. O Painel Intergovernamental de Mudanças Climáticas (IPCC) indicou a Península Antártica como uma das três regiões que passam por um rápido aquecimento regional recente (“*RRR warming*”), isto é, um aumento de temperatura superior à média global de aquecimento (VAUGHAN et al., 2003). Sabe-se que a Península Antártica passou por outros dois eventos de aquecimento ao longo do Holoceno: entre 11 e 9,5 mil anos antes do presente (AP), e de 4,5 a 2,8 mil anos AP (BENTLEY et al., 2010). Uma vez que uma grande quantidade de

material orgânico preservado pode ser recuperada a partir do *permafrost*, a análise de metagenomas conservados em estratos de diferentes idades na Península Antártica pode revelar como as comunidades microbianas responderam a estas alterações climáticas no passado.

1.1 OBJETIVOS

1.1.1 Objetivo Geral

Investigar o perfil taxonômico e funcional das comunidades microbianas do *permafrost* da Península Antártica, em busca de indicadores paleoclimáticos.

1.1.2 Objetivos Específicos

a) Analisar comparativamente metagenomas do sedimento marinho moderno e de cinco estratos de *permafrost* da Península Antártica compreendendo diferentes idades dentro do Holoceno;

b) descrever a composição taxonômica dos diferentes estratos;

c) identificar sequências envolvidas em funções metabólicas que forneçam informações sobre características do ambiente;

d) detectar possíveis alterações no perfil metabólico dos micro-organismos entre os diferentes estratos;

e) correlacionar alterações de composição taxonômica e perfil metabólico com dados climáticos e físico-químicos descritos para a localidade em tempos passados.

2 REVISÃO BIBLIOGRÁFICA

2.1 A ANTÁRTICA NO ECOSISTEMA GLOBAL

2.1.1 O Continente dos Extremos

Embora seja popularmente conhecida como o continente mais frio do planeta, a Antártica apresenta uma gama de condições extremas, além da temperatura, que representam um desafio à vida e à condução de estudos na região. É coberta, quase em sua totalidade, por um manto de gelo alimentado continuamente pela deposição de neve e geada, que se mantêm acumuladas devido às baixas temperaturas ao longo de todo o ano. A espessura deste manto varia de região para região, sendo mais fino ao longo da costa, onde avança para o mar mantendo-se unido à borda do continente e dando origem a plataformas de gelo, que constituem 11% da área total da Antártica (TURNER et al., 2009). As regiões livres de gelo se resumem a cerca de 0,34% da área total do continente (~45.000 km²) (CONVEY et al. 2014). Com exceção da porção norte da Península Antártica, a maior parte do continente está localizada ao sul do Círculo Polar Antártico, que enfrenta aproximadamente 24 horas de incidência de luz solar durante o verão austral (nos meses de dezembro a junho) e aproximadamente 24 horas de escuridão contínua durante o inverno austral. É atravessada por uma cordilheira de montanhas rochosas, as Montanhas Transantárticas, que separam o continente entre as regiões Oriental e Ocidental. A Antártica Oriental é majoritariamente localizada numa região de planalto, se estendendo por cerca de mil quilômetros ao redor do polo sul a uma altitude média de 3.000 m acima do nível do mar. A Antártica Ocidental, apesar de muito menos elevada (em média 850 m), é também altamente montanhosa e inclui o ponto mais alto do continente (o Monte Vinson, de 4.892 m de altitude). Estas características conferem à Antártica a maior média de elevação de todos os continentes (cerca de 2.200 m acima do nível do mar). Também se destaca como o continente mais seco do planeta, com médias de precipitação anual de 200 mm ao longo da costa. No interior do continente a precipitação é ainda menor, chegando a extremos de aridez em locais como os Vales Secos de McMurdo, onde não chove há cerca de 2 bilhões de anos. (TURNER et al., 2009).

2.1.2 O Continente dos Micro-organismos

A Antártica é tradicionalmente considerada um ecossistema pobre, sobretudo devido ao seu isolamento geográfico e condições climáticas extremas. Entretanto, as pesquisas realizadas nas últimas décadas indicam que a Antártica apresenta ecossistemas complexos e diversos, moldados justamente pelas condições extremas nos quais se desenvolveram. Os micro-organismos formam a base dos ecossistemas aquáticos e terrestres na Antártica, possuindo um papel fundamental nas cadeias tróficas e nos ciclos biogeoquímicos (WYNN-WILLIAMS, 1996). Sua notável capacidade de adaptação permite com que colonizem praticamente qualquer tipo de ambiente, atuando como pioneiros nos processos de sucessão ecológica e facilitando o estabelecimento de formas de vida mais complexas, embora muitos habitats antárticos ainda sejam colonizados exclusivamente por micro-organismos. São encontrados em veios de geleiras, lagos sub glaciais, solos congelados, lagoas salinas, sítios geotermiais na borda de vulcões ativos e, de forma particularmente abundante, no ambiente marinho. Embora os ambientes mais isolados, geralmente, possuam uma baixa diversidade taxonômica, o repertório metabólico dos organismos que os compõem é bastante vasto, indicando que poucas espécies são capazes de prover a maior parte das funções necessárias para sustentar o ecossistema (WYNN-WILLIAMS, 1996). Dentre as classes ambientais de micro-organismos presentes nos ambientes antárticos, destacam-se, é claro, os psicrófilos (capazes de manter o metabolismo em funcionamento mesmo sob temperaturas abaixo de zero), mas também estão presentes psicotolerantes, termófilos e hipertermófilos, acidófilos, halófitos e as mais diversas classes metabólicas no que diz respeito à obtenção de energia e fontes de carbono. Por sustentarem a vida mesmo diante de tantas condições adversas, os ambientes antárticos e sua microbiota vêm sendo estudados no contexto da Astrobiologia como modelos de ecossistemas extraterrestres (ANDERSEN, 2002; GILICHINSKY et al., 2007).

A Antártica é apontada como o continente mais importante na regulação do clima do planeta e na funcionalidade dos ecossistemas oceânicos. A produtividade primária nos oceanos do hemisfério sul contribui com uma parcela significativa para as cadeias tróficas marinhas de todo o globo, além de participar em larga escala na ciclagem de nutrientes e minerais e no sequestro do excesso de CO₂ de origem antropogênica na atmosfera (WILKINS et al., 2012). Contudo, tratam-se de ecossistemas particularmente sensíveis a alterações

ambientais, e diversos estudos vêm ressaltando a importância de compreender a extensão com que são afetados, bem como as possíveis consequências destes impactos sobre suas funções ecológicas locais e globais (CLARKE et al., 2007; WILKINS et al., 2012; CAVICCHIOLI, 2015a; CHOWN et al., 2015; GOORDIAL et al., 2016; MAKHALANYANE; VAN GOETHEM; COWAN, 2016).

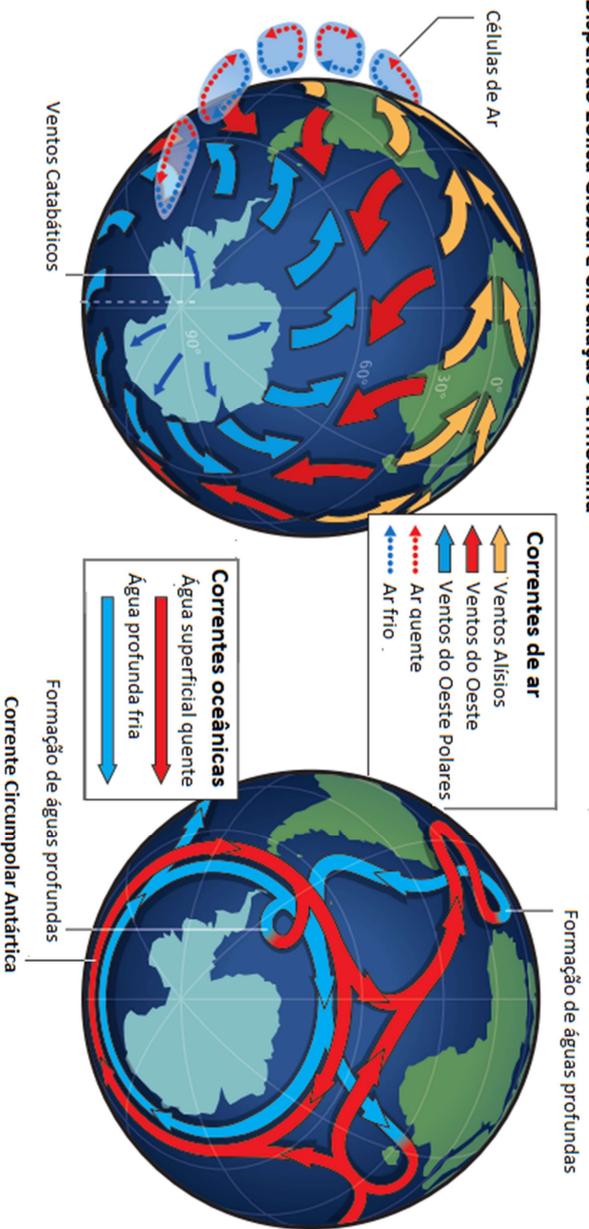
2.1.3 Regulação do Clima Global

(Todas as informações deste capítulo foram consultadas e sistematizadas a partir da revisão do Scientific Committee on Antarctic Research, “Antarctic Climate Change and the Environment: A Contribution to the International Polar Year 2007-2008”, TURNER et al., 2009).

O clima global é majoritariamente influenciado pela quantidade de radiação solar incidente de maneira desigual sobre a superfície terrestre. A diferença de temperatura entre os trópicos e os polos dá origem a um padrão de movimentação de massas de ar conhecido como Célula de Hadley, que atua transportando calor das baixas latitudes para as altas latitudes, com a extensa cobertura de gelo das regiões polares funcionando como centros resfriadores do planeta. O ar resfriado das regiões polares, mais denso e comprimido, gera centros de baixa pressão sobre os polos, que em contraste com as altas pressões originadas pela ascendência de ar quente nas regiões tropicais, provoca o deslocamento das massas de ar no sentido dos polos por gradiente de pressão.

No hemisfério sul, o efeito da força de Coriolis gerada pela rotação da Terra direciona as massas de ar em movimento para o leste, dando origem aos Ventos do Oeste que incidem sobre a costa leste da Antártica. Estes ventos possuem grande influência sobre as correntes marítimas do hemisfério sul, pois deslocam as águas superficiais das regiões polares provocando o ressurgimento de águas profundas, mais frias, densas e ricas em nutrientes. As águas superficiais deslocadas pelos Ventos do Oeste são desviadas para o norte pela força de Coriolis e alimentam a Corrente Circumpolar Antártica, que circunda irrestritamente o globo conectando os oceanos Atlântico, Pacífico e Índico, tornando-a a maior corrente oceânica do planeta (Figura 1).

Dispersão Eólica Global e Circulação Termohalina



A CCA mantém o continente isolado da influência de correntes quentes das latitudes baixas, sendo a principal responsável pela manutenção do manto de gelo da Antártica. O gelo oceânico também possui grande importância na regulação das correntes no hemisfério sul; as águas aquecidas provenientes das regiões tropicais e subtropicais são resfriadas em contato com o mesmo e absorvem o sal não incorporado durante o congelamento, tornando-se mais densas e afundando para dar origem às águas profundas que vêm à tona nos eventos de ressurgência.

O deslocamento de águas superficiais na costa leste da Antártica pelos Ventos do Oeste é determinado pelo Modo Anular do Hemisfério Sul (MAHS), um padrão de oscilação não periódico decorrente de anomalias de pressão atmosférica de sentidos opostos. Um aumento de pressão nas regiões subtropicais, acoplado a uma diminuição de pressão nas regiões polares e subpolares, determinam a fase positiva do fenômeno, e vice-versa. Sua intensidade depende do gradiente de pressão formado entre as altas latitudes (em torno do Círculo Polar Antártico) e as médias latitudes (nas regiões intertropicais). As mudanças de fase do MAHS, bem como sua intensidade, são responsáveis por uma parcela significativa da variabilidade climática de todo o hemisfério sul, sendo também fatores moduladores da circulação termoalina (i.e., padrões de circulação de correntes oceânicas de acordo com a temperatura e a salinidade da água).

2.1.4 Importância no Ciclo do Carbono

O conjunto de propriedades físicas, químicas e biológicas resultantes dos padrões de circulação de massas de ar, correntes marítimas e incidência solar tornam os oceanos do hemisfério sul os mais biologicamente produtivos do planeta (CAVICCHIOLI, 2015). O fenômeno de ressurgência de águas profundas pelos Ventos do Oeste é responsável por grande parte desta característica, trazendo grandes quantidades de nutrientes às águas superficiais que são distribuídas por todo o hemisfério pela forte atuação da CCA. Destacam-se, também, como um enorme sumidouro de carbono, responsável por cerca de 30% da absorção oceânica de CO₂ atmosférico, que ocorre tanto por processos físico-químicos quanto por processos biológicos (SABINE et al., 2004).

Através da fotossíntese, o fitoplâncton contribui massivamente para esta conta, sendo várias ordens de magnitude mais eficiente do que a assimilação físico-química, que se dá pela solubilização do CO₂ na água. Os oceanos do hemisfério sul possuem baixa abundância de

cianobactérias, de modo que os processos fotossintéticos significativos são tipicamente conduzidos por algas eucarióticas. A capacidade de absorção do carbono atmosférico pelo fitoplâncton, no entanto, é bastante limitada pelo inverno austral, em consequência dos longos períodos de ausência de luz. Durante os meses de inverno, os processos físico-químicos passam a dominar a transferência de carbono da atmosfera para os oceanos (a solubilidade do CO₂ é inversamente proporcional à temperatura da água, de modo que águas mais frias possuem maior capacidade absorptiva). A assimilação físico-química, contudo, compete com a transferência de CO₂ no sentido inverso (dos oceanos para a atmosfera), intensificada no inverno devido ao aumento na taxa de ressurgência de águas profundas. Além dos nutrientes, os eventos de ressurgência também trazem à tona grandes quantidades de carbono inorgânico dissolvido (CID), provocando a saturação das águas superficiais e diminuindo sua capacidade de solubilização. Por esta razão, os oceanos do hemisfério sul são considerados como um sumidouro fraco de carbono atmosférico durante o inverno austral (CAVICCHIOLI, 2015). Em contrapartida, dois estudos recentes conduzidos na região da Península Antártica (GRZYMSKI et al., 2012; WILLIAMS et al., 2012) demonstraram que a comunidade microbiana da superfície oceânica aumenta consideravelmente sua capacidade metabólica de fixar CO₂ por processos quimiolitotróficos durante os meses de inverno, representando um sumidouro de carbono até então desconsiderado.

2.2 MUDANÇAS CLIMÁTICAS

2.2.1 Impactos sobre o Continente Antártico

(Todas as informações deste capítulo foram consultadas e sistematizadas a partir da revisão do Scientific Committee on Antarctic Research, “Antarctic Climate Change and the Environment: A Contribution to the International Polar Year 2007-2008”, TURNER et al., 2009).

A Antártica não apenas é um dos continentes com maior influência sobre o clima e a funcionalidade dos oceanos, como também é uma das regiões mais sensíveis a variações ambientais, em grande parte devido aos padrões de circulação de massas de ar e correntes oceânicas descritos anteriormente. Um dos exemplos mais claros de como o continente responde a alterações atmosféricas a nível global foram as anomalias provocadas pelo aumento da emissão de

clorofluorocarbonetos (CFCs) no hemisfério norte a partir da década de 70, acarretando na drástica depleção do ozônio atmosférico que se concentrou sobre o Círculo Polar Antártico. O ozônio atua como um filtro à radiação ultravioleta e mantém a temperatura terrestre através do efeito estufa, mantendo a radiação refletida pela superfície no interior da atmosfera. O buraco na camada de ozônio sobre a Antártica tem aumentado a incidência de UV sobre o continente, ao mesmo tempo em que permite a dissipação dos comprimentos de onda mais elevados, causando uma diminuição de temperatura acima da média dentro do vórtice polar formado durante a primavera. O resfriamento acentuado provoca uma queda de pressão atmosférica anormal sobre a região, aumentando o gradiente de pressão que influencia a intensidade do MAHS.

A partir do final dos anos 70, o MAHS tem apresentado uma tendência cada vez mais positiva, provocando ventos mais fortes sobre a costa da Antártica e intensificando as taxas de ressurgência de águas profundas, o que aumenta a concentração de CID na superfície do oceano. Estudos realizados nas duas últimas décadas demonstraram que as taxas de CID na superfície dos oceanos aumentaram em todos os pontos monitorados a latitudes superiores a 20°S no Oceano Índico, e especificamente em latitudes inferiores a 40°S, aumentaram mais rapidamente do que as taxas de carbono atmosférico desde 1991, sugerindo que os oceanos do hemisfério sul vêm lentamente se tornando um sumidouro menos eficiente de CO₂.

Além da distribuição vertical de CO₂ e nutrientes, os padrões de circulação termoalina também influenciam os processos de formação de gelo oceânico e, conseqüentemente, a temperatura da superfície oceânica. As anomalias provocadas pelo buraco da camada de ozônio e pela intensificação da positividade do MAHS são distribuídas por todo o hemisfério sul pela atuação da CCA, sendo inclusive acentuadas pelo próprio fenômeno por conta do aumento do deslocamento das águas superficiais que compõem a corrente. Tipicamente, a CCA leva em torno de dois anos para propagar anomalias de temperatura da superfície oceânica do sul do Pacífico para o sul do Atlântico, mas o acoplamento entre os hemisférios norte e sul também ocorre, numa escala de cerca de uma década.

Diversas estações de pesquisa e programas de monitoramento ambiental foram estabelecidas a partir do final da década de 50 em consequência do Ano Internacional da Geofísica (1957-1958) da ONU, de modo que a maioria das observações que descrevem a variabilidade climática do continente tiveram início nesta época. Desde então,

variáveis como a temperatura atmosférica, do solo e da superfície dos oceanos, extensão da cobertura de gelo, avanço e retração de geleiras, nível do mar e taxas de precipitação vem sendo acompanhadas periodicamente numa tentativa de compreender os padrões de variabilidade natural e identificar a ocorrência de anomalias e suas consequências. A Tabela 1 descreve algumas das principais tecnologias atualmente empregadas no monitoramento de variáveis ambientais, e a Figura 2 mostra a localização de algumas Estações Meteorológicas Automáticas distribuídas no continente antártico.

Tabela 1. Tecnologias de monitoramento de algumas variáveis ambientais acompanhadas periodicamente no continente antártico. Elaborada pela autora.

TECNOLOGIA	PARÂMETROS MONITORADOS
Estações Meteorológicas Automáticas	Pressão atmosférica, temperatura, umidade, ventos.
Balões Meteorológicos	<p>Radiossondas: Perfil vertical de pressão atmosférica, temperatura, ventos e umidade.</p> <p>Sondas eletroquímicas: Ozônio atmosférico.</p>
Satélites	<p>Parâmetros Atmosféricos: temperatura e composição (concentrações de dióxido de carbono, óxido nítrico, oxigênio, ozônio, vapor de água), inferidos a partir de medições de incidência de radiação</p> <p>Parâmetros Oceanográficos: temperatura da superfície (infravermelho térmico), conteúdo de gelo oceânico, coloração (a partir da qual pode-se inferir a concentração de clorofila).</p>
Testemunhos de gelo	Representações indiretas de temperatura e circulação atmosférica, através de medições de conteúdo geoquímico (isótopos, sal marinho, aerossóis, cálcio, nitrato).
Radares	Acumulação de neve, derretimento de gelo, avanço e recuo de geleiras (estratigrafia e altimetria do manto de gelo).
Sondas Oceânicas (CTDs)	Parâmetros da coluna d'água: condutividade (a partir da qual pode-se inferir a salinidade), temperatura, oxigênio dissolvido, radiação fotossinteticamente ativa, nutrientes, poluentes, carbono total, carbono orgânico e inorgânico dissolvido, alcalinidade, metais traço.
Marégrafos	Variação das marés e nível do mar.

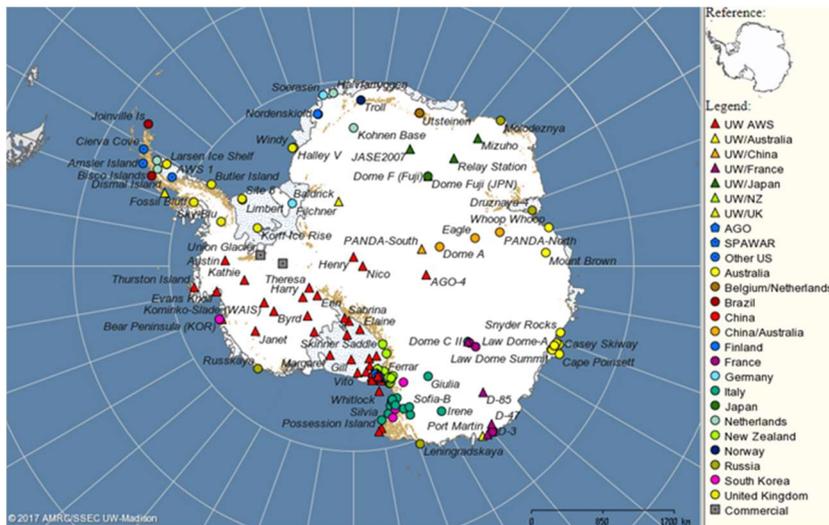


Figura 2. Localização de algumas Estações Meteorológicas Automáticas na Antártica. (Fonte: University of Wisconsin-Madison).

Alterações diversas têm sido observadas em toda a Antártica nas últimas décadas, mas a grande variabilidade natural do clima, aliada à inexistência de dados de períodos anteriores, dificulta com que um prognóstico confiável da extensão do impacto das mudanças climáticas sobre o continente seja concluído. Sabe-se, de forma geral, que algumas regiões da Antártica Marítima, sobretudo a Península Antártica, têm demonstrado uma tendência anormal ao aquecimento, enquanto o interior do continente não tem apresentado alterações conclusivas. Acredita-se que a dinâmica atmosférica sobre o continente, especialmente no que diz respeito ao buraco da camada de ozônio e à formação do vórtice polar, contribuem para mascarar mudanças mais drásticas previstas por modelos computacionais, de modo que as regiões mais afastadas destas influências refletem mais claramente a forma como são afetadas.

2.2.2 A Península Antártica

Diferente das demais regiões do continente, onde as evidências de anomalias não são tão claras e não há consenso quanto às influências antropogênicas no clima, a Península Antártica é uma das regiões que mais vem demonstrando alterações consideradas inconsistentes com a

variabilidade climática natural do continente. O último Painel Intergovernamental sobre Mudanças Climáticas (IPCC) aponta a Península Antártica como uma das três regiões do planeta que passam por um rápido aquecimento regional recente: desde 1951, a média de temperatura anual na Península aumentou em 3°C, em contraste com o aumento de 0.6°C (± 0.2) na média global apontado pelo IPCC.

A Península Antártica é a porção mais ao norte do continente, sendo uma das poucas regiões que experienciam temperaturas superiores a 0°C durante o verão. Banhada pelo mar de Weddel, à leste, e o de Bellingshausen, à oeste, é particularmente suscetível a influências marítimas de diversas fontes, além de possuir diversos arquipélagos que se projetam sobre a Passagem de Drake, uma das zonas com as piores condições meteorológicas para navegação do mundo. Por se tratar de um prolongamento estreito da plataforma continental, rico em córregos de degelo que se formam durante os meses mais quentes, é naturalmente propensa à erosão da cobertura de gelo pela intrusão de águas de temperaturas superiores a 1°C. Apesar da dinâmica de retração e avanço de geleiras ser uma característica natural da Península Antártica, a região vem demonstrando uma tendência à perda da área de plataformas de gelo (mais de 27.000 km² nos últimos 50 anos) e à retração de geleiras (COOK et al., 2005) (Figura 3).

Mesmo que estes dados sejam robustos e indiscutíveis, ainda resta o questionamento da origem destas alterações: ainda não é possível concluir se são oscilações naturais de aquecimento e resfriamento, ou se realmente estamos diante de uma anomalia climática provocada por ações antropogênicas. Estudos paleoclimáticos, que buscam a reconstrução da dinâmica das variáveis ambientais em tempos passados, têm sido importantes na busca por respostas para esta questão.

2.3 PALEOCLIMA

(Todas as informações deste capítulo foram consultadas e sistematizadas a partir da revisão do Scientific Committee on Antarctic Research, “Antarctic Climate Change and the Environment: A Contribution to the International Polar Year 2007-2008”, TURNER et al., 2009).

O histórico climático e ambiental da Antártica é uma importante área de estudo no contexto de mudanças climáticas, permitindo compreender a variabilidade natural do continente e determinar quando mudanças podem ser consideradas anômalas.

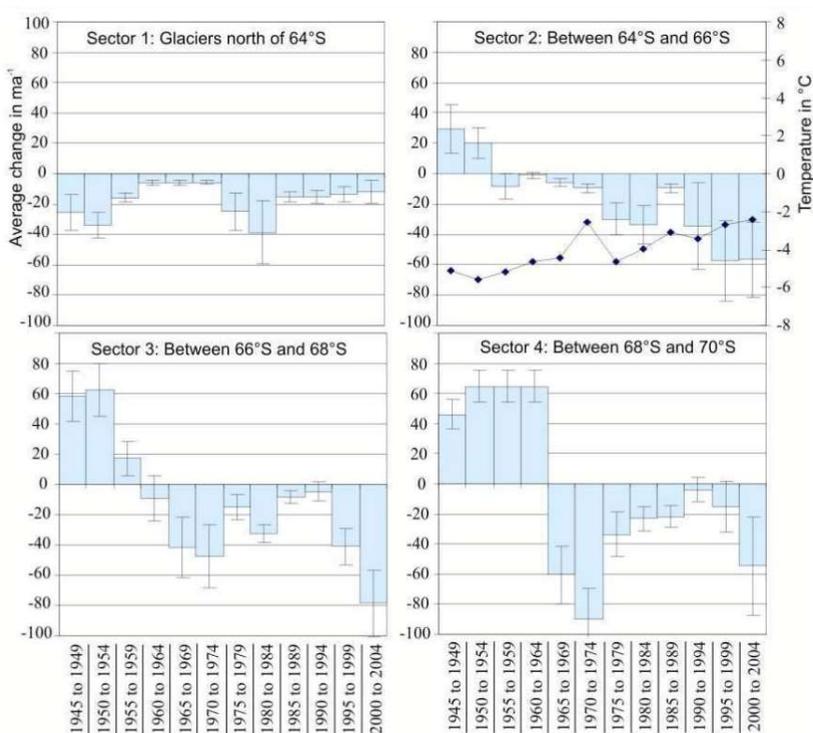


Figura 3. Mudanças observadas na dinâmica de geleiras da Península Antártica ao longo do tempo por latitude, medidas pela proporção entre recuo e avanço (m.a⁻¹, “melting/advance”) (Fonte: Cook et al, 2005).

Projeções do IPCC para os níveis de CO₂ atmosférico preveem concentrações superiores a duas vezes as do período pré-industrial até o fim do século presente. Uma composição atmosférica semelhante só pode ser encontrada em registros geológicos anteriores a 34 milhões de anos, datação aproximada da formação das primeiras coberturas de gelo da Antártica. Nos 150 milhões de anos anteriores, o continente sofreu intensas modificações geológicas, incluindo a deriva continental em decorrência da fragmentação da Gondwana, a abertura da passagem de Drake e o estabelecimento da Corrente Circumpolar Antártica. A intensa atividade tectônica influenciou fortemente a composição da atmosfera do planeta durante todo este intervalo de tempo, sendo referenciado como “greenhouse world” (mundo estufa) devido aos altos níveis de CO₂ de origem vulcânica.

Os estudos do paleoclima neste período se baseiam na datação de isótopos presentes em sítios geológicos e registros fósseis, e indicam que os ciclos glaciais-interglaciais tinham a duração de aproximadamente 40 mil anos. Para tempos mais recentes (até 800 mil anos), testemunhos de gelo (i.e., amostras obtidas por perfurações de grandes profundidades) são capazes de prover informações precisas sobre as condições ambientais passadas usando indicadores indiretos, como o conteúdo de sal marinho, enxofre biogênico, diferentes elementos em partículas de poeira e concentração de gases atmosféricos. Estes registros revelam que no último milhão de anos os ciclos glaciais-interglaciais passaram a ocorrer em intervalos de 100 mil anos, indicando uma intensa mudança na configuração de variáveis que influenciam o clima do planeta.

Esta mudança de frequência coincide com os ciclos de Milankovich, modificações periódicas na órbita da Terra, na escala de milhares de anos, que influenciam a sazonalidade e o posicionamento geográfico da incidência de energia solar. Por este motivo, têm-se especulado que as mudanças climáticas observadas nos últimos tempos, sobretudo no que diz respeito à perda da cobertura de gelo, façam parte da variabilidade natural do clima, induzida por alterações nos padrões de insolação causados pelos movimentos terrestres. Esta hipótese apresenta diversas lacunas que ainda não podem ser totalmente explicadas pelos ciclos de Milankovich, o que ressalta a importância de estudos paleoclimáticos com foco no último milhão de anos.

2.3.1 O Holoceno

A Terra passou por 8 grandes transições de aquecimento-resfriamento nos últimos 800 mil anos, entre os quais os níveis de CO₂ e temperatura média atmosférica oscilaram entre 180 ppm e 10°C nos períodos glaciais, e 300 ppm e 15°C nos períodos interglaciais. Atualmente, vivemos em um período interglacial que teve início há 21 mil anos, e os níveis modernos de CO₂ encontram-se na faixa de 380 ppm, sendo provavelmente os maiores dos últimos 25 milhões de anos. A transição do Último Máximo Glacial para o atual período interglacial, denominada Terminação I, é um dos eventos paleoclimáticos mais bem documentados de que se tem registro, tendo sido reconstruído com base nos dados de testemunhos de gelo de mais de 10 perfurações distribuídas ao longo de todo o continente (Figuras 4 e 5).

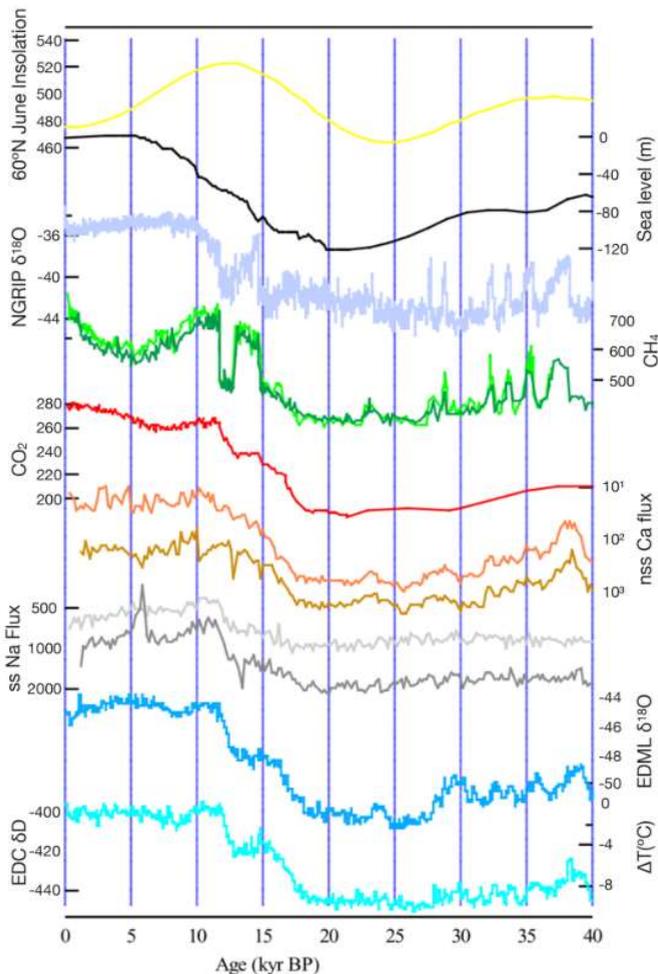


Figura 4. Registros paleoclimáticos da Terminação I, inferidos a partir de dados de testemunhos de gelo: NGRIP – North Grip (Groenlândia), EDC – EPICA Dome C (Antártica), EDML – EPICA Dronning Maud Land (Antártica). Amarelo: Insolação em Junho à latitude de 60° (W/m^2); Preto: mudança global de nível do mar (m); Azul-cinza: porcentagem do isótopo 18 de oxigênio; concentração de gás metano atmosférico (verde claro – NGRIP, verde escuro – EDC); Vermelho – concentração de CO_2 atmosférico (EDC e Vostok); fluxo de cálcio não-marinho (Laranja – EDC, marrom – EDML); Fluxo de cálcio marinho (cinza claro – EDML, cinza escuro – EDC); Azul claro – porcentagem do isótopo 18 de oxigênio (EDML); Azul escuro – porcentagem de isótopo de deutério (EDC). A porcentagem de isótopos de O e D são indicadores de temperatura. (Fonte: *Scientific Committee on Antarctic Research*).

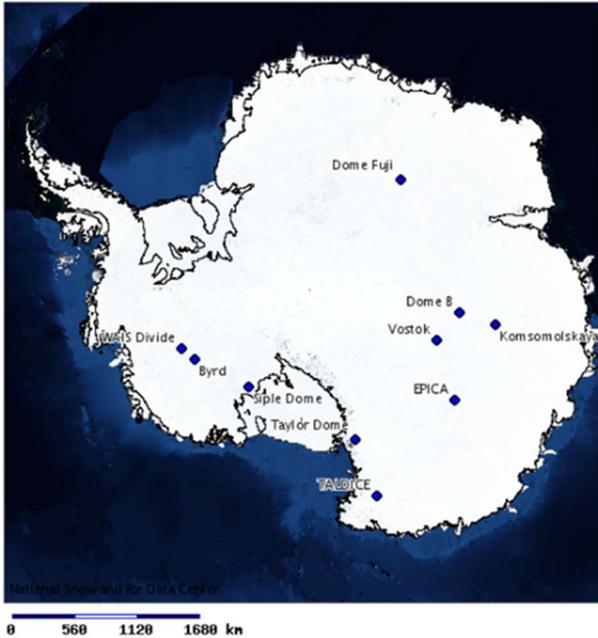


Figura 5. Localização dos testemunhos de gelo a partir dos quais condições paleoclimáticas da Antártica podem ser inferidas. (Fonte: *National Snow and Ice Data Center*).

2.3.2 O Holoceno na Península Antártica

Durante o Último Máximo Glacial (período entre 26,5 e 14,5 mil anos), o manto de gelo da Península Antártica avançou até a borda da plataforma continental, tendo iniciado a retração em torno dos 14 mil anos (BENTLEY et al., 2010). O Holoceno, período interglacial presente, teve início há aproximadamente 11,7 mil anos, e embora seja reconhecido como um período de relativa estabilidade climática, uma análise mais atenciosa dos registros paleoclimáticos revelam que diversas flutuações, ainda pouco compreendidas, marcaram a época de forma bastante pronunciada.

O intervalo entre 11 mil e 9,5 mil anos é conhecido como o Ótimo Climático do Holoceno Inferior, caracterizado por alta produtividade primária documentada nos sedimentos marinhos e de lagos da região da Antártica Oriental. Há 10 mil anos, a deglaciação chegou à parte interna da plataforma na margem pacífica da Península

Antártica, e certas regiões do arquipélago das Shetlands do Sul eram livres de gelo até 9 mil anos. A exposição do sedimento ao longo da Península Fildes (Ilha Rei George) durante este período gerou padrões de estratificação no sedimento de lagos a partir dos quais pode-se inferir a variação do nível do mar na época (Figura 6) (WATCHAM et al., 2011).

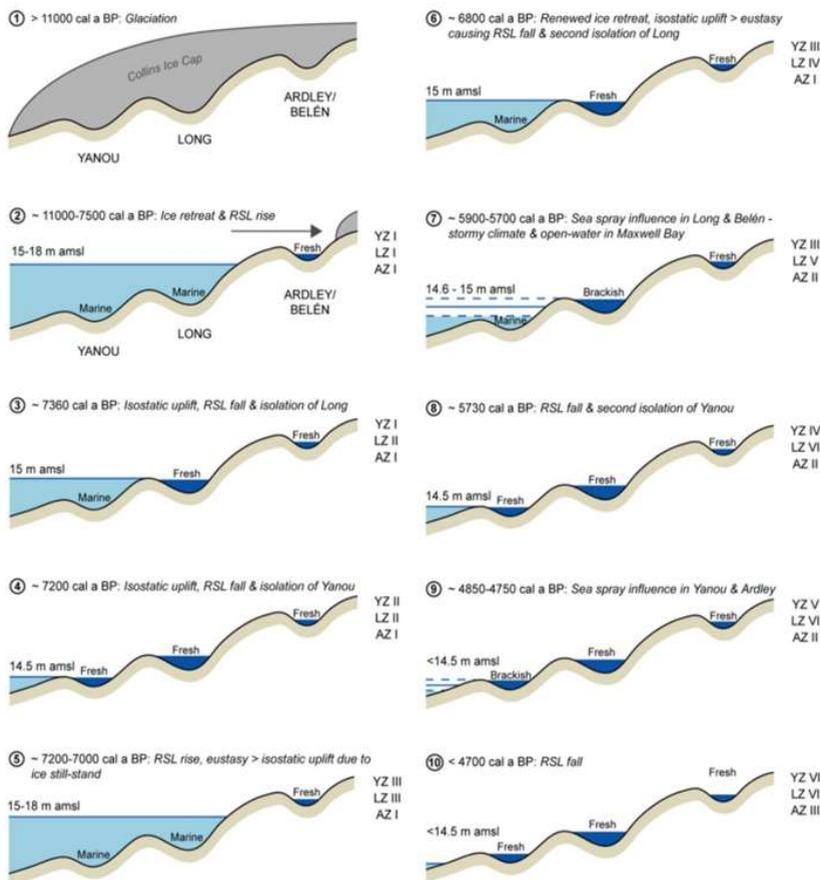


Figura 6. Representação esquemática de mudanças no nível do mar na Ilha Rei George, arquipélago das Shetlands do Sul, com base nos dados de sedimentação de quatro lagos da Península Fildes (Yanou Lake, Long Lake, Ardley Lake e Bélen Lake). (Fonte: (WATCHAM et al., 2011)

Em contrapartida, alguns registros demonstram que a transição glacial-interglacial ocorreu de maneira desigual em diferentes regiões, apresentando uma dinâmica mais gradual ao longo da costa leste.

O período entre 9,5 e 4,5 mil anos demonstra padrões ainda mais complexos de mudanças na Península Antártica. O colapso da plataforma George IV, por exemplo, só ocorreu após o fim do Ótimo Climático do Holoceno Inferior, enquanto a plataforma Larsen B permaneceu intacta. Bentley e colaboradores (2010) sugerem que a deglaciação tenha continuado ocorrendo, a taxas mais lentas e restrita à margem oeste da Península, até pelo menos 7-8 mil anos (idade aproximada da reformação da plataforma George IV). Esta época coincide com registros de intrusão de águas subpolares na plataforma continental, tendo impacto sobre regiões cuja cobertura de gelo era mais fina. A expansão glacial pode ser evidenciada mais significativamente a partir dos 6 mil anos nas regiões costeiras do norte da Península, com registros indicando um clima mais árido e frio e extensas coberturas de gelo oceânico nas baías do entorno do arquipélago das Shetlands do Sul (TURNER et al., 2009).

Um segundo evento de aquecimento significativo ocorreu a partir dos 4.5 mil anos, conhecido como Ótimo Climático do Holoceno Médio. Diferentemente do período anterior, o aumento de temperatura parece ter impactado uma extensão significativamente maior, incluindo flutuações na plataforma Larsen A e o colapso da plataforma do canal Prince Gustav, ambos localizados na porção leste da Península. Inúmeros registros de rápida sedimentação, aumento da produtividade primária e diminuição do gelo oceânico estão presentes deste as Shetlands do Sul até a ilha James Ross. A partir destes registros, Bentley e colaboradores (2010) puderam estimar a datação e a duração deste evento em diversas localidades, compilando-os em um consenso denominado “Mid-Holocene Hypsithermal”, i.e., uma média de temperatura reconstruída a partir de dados paleoclimáticos (testemunhos de gelo, sedimentação de lagos e ambientes marinhos e oscilação de plataformas de gelo) (Figura 7).

Um breve intervalo neoglacial pôde ser observado entre 2.5 e 1.2 mil anos, marcado pelo reavanço de geleiras, gelo marítimo perene e o ressurgimento da plataforma Prince Gustav. No hemisfério norte, há registros de um período mais quente entre 1.2 e 0.6 mil anos, denominado “Período de Aquecimento Medieval”, seguido de um novo resfriamento entre 700 e 150 anos (“Pequena Era do Gelo”). Para a Antártica, as evidências destes três últimos eventos são esparsas, e os poucos registros existentes apresentam datações inconsistentes, ou

mesmo a ausência de datação. Os estudos que apontam para o recente episódio de aquecimento regional na Península enfatizam a necessidade de ampliar o entendimento sobre o histórico climático do Holoceno na região, permitindo que essas informações possam endossar os modelos preditivos de variabilidade climática.

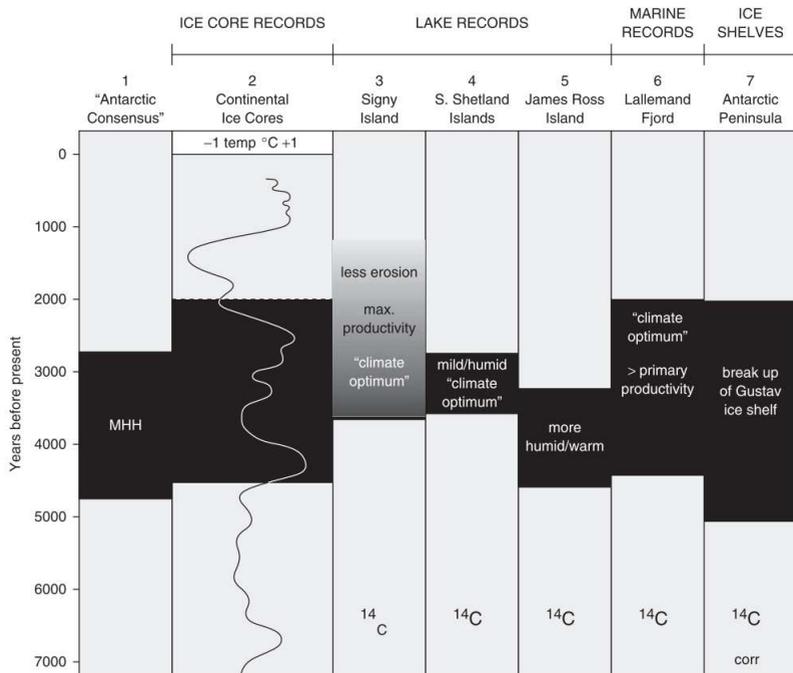


Figura 7. Consenso de temperatura para a Antártica ao longo dos últimos 7 mil anos, compilado a partir de dados de testemunhos de gelo, sedimentação de lagos e ambientes marinhos e oscilação de plataformas de gelo em diversas localidades do continente. MHH – *Mid Holocene Hypsithermal*. (Fonte: BENTLEY et al. 2010).

2.4 PERMAFROST

O *permafrost* (tipo de solo permanentemente congelado por pelo menos dois anos consecutivos) é um importante elemento da criosfera, destacando-se na paisagem antártica por distribuir-se ao longo de praticamente toda a área livre de gelo do continente. A camada mais

superficial deste tipo de solo, denominada camada ativa, compõe a porção do substrato suscetível à ação do tempo, passando por descongelamento periódico nas estações mais quentes. Dependendo da composição mineral, porosidade, quantidade de matéria orgânica e água definidas no momento do congelamento, os estratos mais profundos podem sofrer infiltração a partir da camada ativa, gerando diferentes estruturas que podem estar presentes no *permafrost*. Os processos de modificação dos solos pelo derretimento de gelo e infiltração são denominados crioturbação, e estão representados na Figura 8 (JANSSON; TAŞ, 2014).

Solos que atualmente encontram-se congelados podem ter sido cobertos por vegetação ou água em eras geológicas passadas. A biomassa sedimentada ao longo de períodos mais quentes e soterrada durante eventos de glaciação dá origem a uma estratificação vertical associada à idade de congelamento do solo: camadas mais profundas representam depósitos de materiais cada vez mais antigos. No Ártico, é possível encontrar estratos de até 3 milhões de anos, e na Antártica já foram datadas amostras de solo congelado há 8 milhões de anos. Temperaturas abaixo de 0°C, pouca disponibilidade de água líquida e a baixa atividade metabólica microbiana no interior do substrato constituem condições ideais para preservação de matéria orgânica, representando um grande reservatório de material biológico antigo. O *permafrost* é capaz de preservar por milhares a milhões de anos todo o material biológico sedimentado, incluindo tecidos animais e vegetais, metabólitos, ácidos nucleicos ou mesmo células viáveis. A exemplo disto, Willerslev e colaboradores (2014) conduziram um estudo de reconstrução da diversidade vegetal na região do Círculo Polar Ártico a partir do material preservado no *permafrost*, correlacionando-a com a dieta da megafauna herbívora e dados climáticos dos últimos 50 mil anos – período que inclui a chegada do Último Máximo Glacial (21 mil anos AP) e a transição para o Holoceno (período interglacial presente). Outro exemplo notável da capacidade de preservação de material biológico no *permafrost* foi a recuperação de sementes de *Silene stenophylla* em solos congelados desde o Pleistoceno (30.000 anos) na Sibéria, as quais germinaram com sucesso em laboratório (YASHINA et al., 2012).

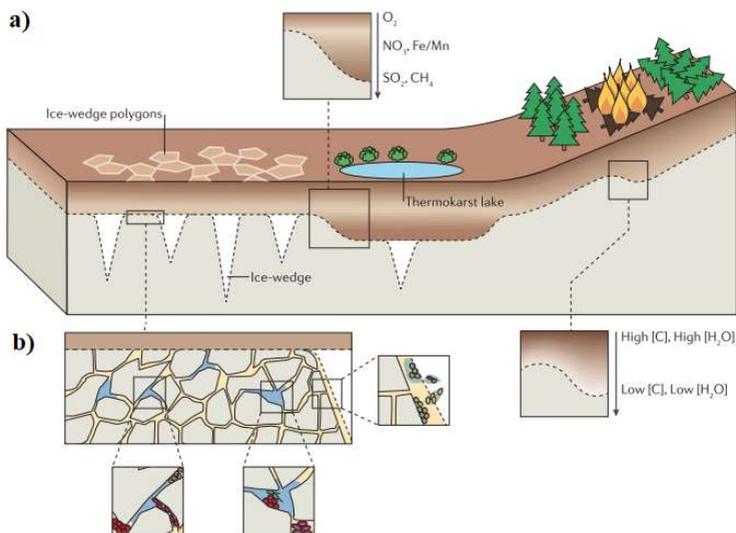


Figura 8. a) Características da *permafrost* geradas por processos de crioturbação: *ice wedges* formadas a partir da infiltração de água de degelo em rachaduras do solo, gerando padrões poligonais na superfície, e *thermokarst lakes*, provenientes do derretimento de *ice wedges*; **b)** poros e veios no substrato contendo água líquida devido à altas concentrações de sais, servindo como hábitat microbiano no *permafrost* (Fonte: JANSSON; TAŞ, 2014);



Figura 8. (continuação) – **c)** e **d)** padrões poligonais na superfície do solo gerados em torno de *ices wedges* nas camadas mais profundas (Fonte: KANEVSKIY et al., 2011; physicalgeography.net); **e)** detalhe de um *ice wedge* exposto devido à erosão do solo (Fonte: WORSLEY, 2014); **f)** *thermokarst lakes* originados a partir do derretimento de *ice wedges* (Fonte: University of Alaska Fairbanks).

2.4.1 Diversidade Microbiana no *Permafrost*

Os estratos mais superficiais do *permafrost* costumam apresentar uma maior concentração de células microbianas, por estarem mais próximos da camada ativa e, conseqüentemente, mais suscetíveis à infiltração. Apesar disso, as camadas mais profundas (e também mais antigas) também podem conter micro-organismos ativos, em zonas com altas concentrações de sais (que diminuem o ponto de congelamento da água) ou em veias e bolhas formadas no gelo (Figura 8.b) (JANSSON; TAŞ, 2014). A capacidade de subsistir neste tipo de ambiente diz muito não apenas sobre os limites, mas também sobre a história evolutiva da vida na Terra – as baixas temperaturas e a dessecação no interior do substrato ocasionam uma drástica queda na atividade metabólica dos micro-organismos ali presentes, levando-os a persistirem no ambiente por tempos muito mais longos do que em outros habitats (WILLERSLEV et al., 2004). Deste modo, muitas linhagens sobreviventes no *permafrost* são bastante antigas e não apresentam evidências significativas de evolução ou adaptação desde a época de congelamento (GILICHINSKY et al., 1992), podendo ser consideradas verdadeiros fósseis vivos. Organismos datados do Neogeno (3.5 milhões de anos), por exemplo, já foram isolados a partir do *permafrost* da Sibéria (SHI et al., 1997; ZHANG et al., 2013).

A maior parte das pesquisas envolvendo micro-organismos do *permafrost* focam na diversidade taxonômica e capacidade de sobrevivência a condições extremas (GILICHINSKY et al., 1992, 2007; RIVKINA et al., 2000; BAKERMANS et al., 2003; STEVEN et al., 2008; TAŞ et al., 2014). A diversidade microbiana no *permafrost* pode ser estudada por diferentes estratégias, que podem ser resumidas entre técnicas dependentes ou independentes de cultivo. A parcela de micro-organismos cultiváveis, no entanto, geralmente representa menos de 1% da diversidade total em qualquer amostra ambiental, de modo que a esmagadora maioria das espécies procariontes permanece oculta quando investigada por técnicas dependentes de cultivo. Steven e colaboradores (STEVEN et al., 2007) constataram, por exemplo, que bactérias formadoras de endósporos do filo Firmicutes predominavam entre os isolados de *permafrost* do Ártico canadense, embora demonstrassem uma baixa representatividade quando avaliados por métodos independentes de cultivo. Por esta razão, técnicas moleculares vêm sendo cada vez mais frequentemente empregadas nos estudos de diversidade, entre elas o sequenciamento de genes codificantes para o RNA ribossomal 16S (RNAr 16S), que permitiu com que a composição

das comunidades microbianas fosse determinada para diversos ambientes outrora inexplorados, incluindo o *permafrost* (Figura 9).

A diversidade de bactérias no *permafrost* geralmente é mais alta do que a dos demais grupos microbianos, embora arqueias e vírus também sejam encontrados em abundância, assim como gêneros ou até mesmo filos ainda não caracterizados (JANSSON; TAŞ, 2014; MONDAV et al., 2014). O grande volume de sequências com baixa porcentagem de similaridade com organismos conhecidos nos bancos de dados públicos ainda é um desafio enfrentado por quaisquer estudos envolvendo sequenciamento direto de amostras ambientais. Também é importante lembrar que baixas temperaturas constituem condições ideais para a conservação de moléculas de DNA (PEDERSEN et al., 2014), o que facilita a recuperação de material genético das amostras, mas ao mesmo tempo impossibilita avaliar se todas as sequências detectadas pertencem a organismos vivos ou ativos. Para isto, as técnicas de isolamento ainda apresentam grande valor (VISHNIVETSKAYA et al., 2000).

2.4.2 O *Permafrost* no Contexto de Mudanças Climáticas

O interesse nas comunidades microbianas do *permafrost* têm aumentado recentemente no contexto de mudanças climáticas. Observações do *permafrost* nos últimos 50 anos têm indicado que a camada ativa vem sofrendo um aumento de espessura em diversas localidades, resultando no descongelamento de estratos cada vez mais profundos e na conseqüente exposição do material conservado em seu interior (KÖHLER; KNORR; BARD, 2014). A reintrodução de material orgânico, outrora inerte, na atmosfera, pode significar um aumento substancial na emissão de gases do efeito estufa, devido a intensificação da atividade decompositora dos micro-organismos (SCHUUR et al., 2015). Diversos estudos têm se proposto a investigar a resposta dos micro-organismos diante de possíveis impactos causados pelo aquecimento global sobre solos congelados (MACKELPRANG et al., 2011; COOLEN; ORSI, 2015; HULTMAN et al., 2015). A produção de metano, por exemplo, demonstrou ser modulada por micro-organismos em resposta ao derretimento do *permafrost* em três diferentes estudos (TVEIT et al., 2013; MCCALLEY et al., 2014; CREVECOEUR et al., 2017).

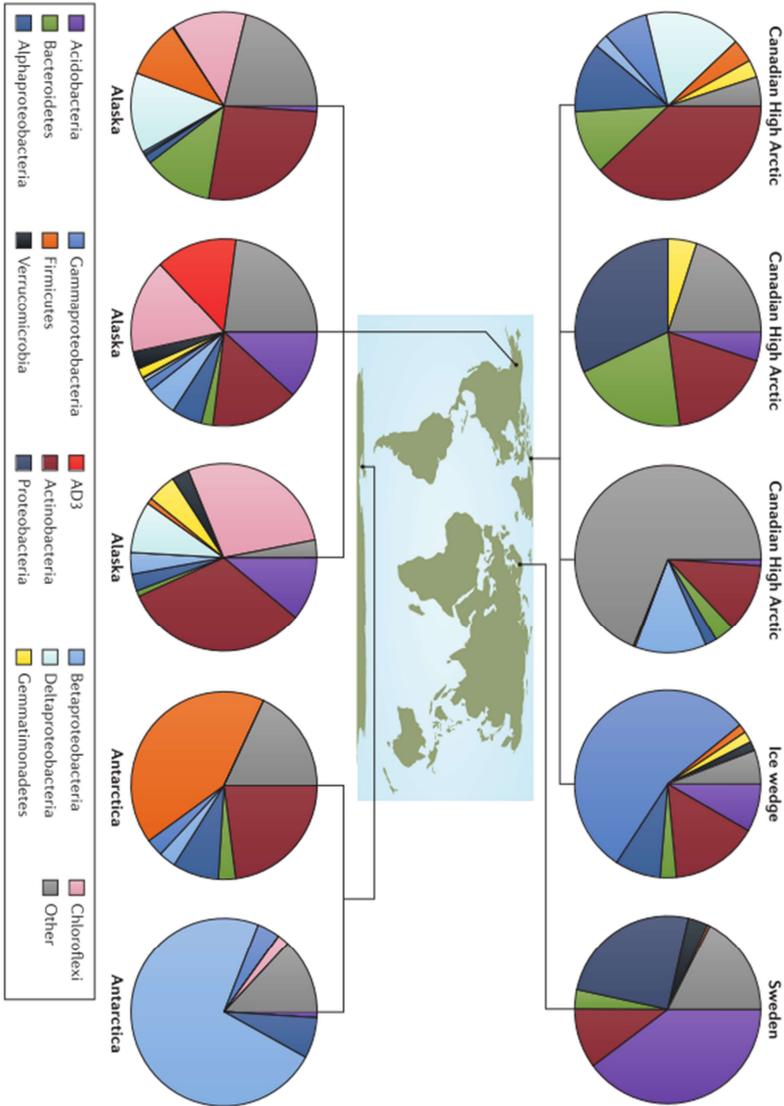


Figura 9. Distribuição taxonômica de filos de bactérias no *permafrost* de diversas localidades, com base no sequenciamento de RNAr 16S. (Fonte: JANSSON; TAŞ, 2014).

As comunidades microbianas são, de fato, extremamente sensíveis a alterações ambientais, podendo responder de diferentes maneiras às pressões exercidas pelo ambiente. Oscilações na abundância de populações específicas, modificações de perfil metabólico e mudanças na concentração de elementos químicos no solo e na atmosfera são características comumente exploradas com o intuito de utilizar micro-organismos como indicadores de mudanças climáticas (BARDGETT; FREEMAN; OSTLE, 2008). Uma vez que o *permafrost* apresenta um rico conteúdo de células microbianas e moléculas de DNA conservadas há milhares de anos, a análise de estratos de diferentes idades pode revelar como as comunidades microbianas responderam a alterações ambientais no passado.

2.5 METAGENÔMICA

A metagenômica é definida como a análise direta do conteúdo genômico de uma amostra ambiental, sendo uma das estratégias que mais vem ganhando força nos estudos de diversidade microbiana. O termo começou a ser utilizado para se referir ao sequenciamento de DNA total de uma amostra, inicialmente realizado a partir da clonagem de fragmentos aleatórios para montagem de bibliotecas contendo DNA de vários organismos (“*shotgun metagenomic libraries*”) (KUNIN et al., 2008). Atualmente, as técnicas de sequenciamento de nova geração permitem a análise de metagenomas a partir de quantidades mínimas de DNA, sem a necessidade de amplificação e clonagem. Contudo, o termo também pode se referir apenas ao estudo de genes de interesse em uma comunidade, utilizando a Reação em Cadeia da Polimerase (PCR) para amplificação de sequências específicas. Esta abordagem é denominada “*marker gene*” ou “*target metagenomics*” de forma geral, mas o termo “*16S metagenomics*” também pode ser encontrado, uma vez que a maior parte destes estudos é realizada sobre genes para o RNA ribossomal 16S. A análise do conteúdo total de metagenomas, por sua vez, é denominada “*full shotgun metagenomics*” ou ainda “*WGS (whole genome sequencing) metagenomics*” (OULAS et al., 2015).

A metagenômica *shotgun* não apenas fornece informações taxonômicas sobre os micro-organismos de uma amostra, mas também permite acessar sua diversidade funcional, isto é, o que a comunidade microbiana é capaz de realizar, em termos metabólicos e ecológicos. Bowman e Ducklow (2015) destacam a importância destas informações:

O gênero de cianobactérias *Trychodesmium*, por exemplo, é associado à baixa concentração de nitrogênio biodisponível nos oceanos tropicais e intertropicais. Porém, são as propriedades metabólicas (p.ex. diazotrofia) do gênero, e não a sua taxonomia, que fornecem uma ligação direta com as condições ambientais. Para compreender o papel das comunidades microbianas nos processos biogeoquímicos, é preferível considerar sua estrutura metabólica, acima de sua estrutura taxonômica.

(BOWMAN; DUCKLOW, 2015)

Embora algumas funções metabólicas e ecológicas possam ser inferidas a partir do RNAr 16S, a metagenômica *shotgun* permite o acesso a informações muito mais completas, precisas e com maior resolução. Cada etapa deste tipo de análise, contudo, é bastante complexa e esbarra em dificuldades intrínsecas aos problemas da Genômica em geral, tais como a cobertura incompleta de genes, o grande volume de dados e o pequeno tamanho das sequências geradas pelas plataformas de sequenciamento atuais (MA; PRINCE; AAGAARD, 2014a). Por ser uma área recente e em constante expansão, as estratégias para a análise de metagenomas são vastas e diversificadas, incluindo desde softwares específicos para cada etapa até *pipelines* completos disponíveis *online*. Algumas possibilidades metodológicas comuns serão abordadas nas próximas sessões, mas é importante lembrar que cada estudo pode aplicar as estratégias e ferramentas que melhor se adequem às suas amostras e aos recursos computacionais disponíveis.

2.5.1 Short Reads x Montagem de Sequências

A análise de metagenomas pode ser realizada diretamente sobre *short reads*, i.e., sequências pequenas provenientes das plataformas de sequenciamento de nova geração mais comuns (*Roche* 454, em torno de 300 pares de base, e *Illumina*, em torno de 150 pares de base). Softwares especializados neste tipo de análise, como o HUMAnN2 (ABUBUCKER et al., 2012) e o MEGAN6 (HUSON et al., 2016), por exemplo, se baseiam no alinhamento de *short reads* contra sequências de referência disponíveis nas bases de dados. Uma vez que sequências curtas são pouco informativas, os algoritmos de busca devem ser robustos e as bases de dados bastante completas para promover um alinhamento de qualidade, o que exige um enorme esforço computacional. Além disso, a maior parte das bases de dados possuem

um grande viés de informações para micro-organismos de ambientes bem estudados, como o microbioma humano, por exemplo.

Por este motivo, a estratégia mais eficaz para metagenomas de ambientes pouco conhecidos é a montagem de *short reads* em seqüências contíguas maiores (*contigs*), orientadas por regiões de sobreposição (Figura 10). Desta forma, as informações dispersas nos milhões de fragmentos gerados pelo sequenciamento são reunidas em segmentos maiores, concisos e mais informativos, reduzindo a dimensão dos dados e maximizando as chances de encontrar seqüências similares nas bases de dados. Este procedimento não é exclusivo para metagenomas – os primeiros algoritmos e softwares, na verdade, foram desenvolvidos para a montagem de genomas completos, provenientes de um único organismo. A performance destas ferramentas sobre *short reads* derivados de metagenomas, contudo, era bastante limitada por problemas intrínsecos à presença de mais de um genoma na mesma amostra. A ocorrência de variações intraespecíficas e grandes similaridades interespecíficas tornavam extremamente prováveis a formação de quimeras, isto é, seqüências montadas com fragmentos provenientes de mais de um organismo. O maior problema, contudo, reside na sub-amostragem de organismos pouco abundantes na comunidade, levando à superestimação da representatividade de organismos mais comuns. Com a popularização da metagenômica, os algoritmos de montagem passaram a ser modificados para atender especificamente a estas dificuldades (OULAS et al., 2015).

Existem duas abordagens possíveis para o processo de montagem: a orientação dos *short reads* utilizando um genoma (ou metagenoma) como referência, ou a montagem *de novo*. O primeiro trata-se de um procedimento relativamente simples e não muito diferente de um alinhamento, mas exige que seqüências de organismos relacionados já tenham sido estudadas e estejam disponíveis nas bases de dados. A segunda abordagem é muito mais complexa e consiste em buscar pequenas regiões de sobreposição contidas nos próprios *reads*. As ferramentas mais comuns utilizam a implementação de gráficos *de-Bruijn* (COMPEAU; PEVZNER; TESLER, 2011), algoritmos derivados da teoria dos grafos que geralmente exigem grandes recursos computacionais. Uma vez que seqüências maiores (ou mesmo genomas inteiros) estejam montados, é possível identificar regiões correspondentes a janelas abertas de leitura (ORFs), seqüências regulatórias, RNAs não codificantes e outros elementos genômicos de interesse. Este procedimento é comumente referenciado como “*gene calling*” ou “*gene prediction*”, embora não seja limitado à busca por

genes propriamente ditos. Estas sequências, por fim, podem ser atribuídas a classificações funcionais, metabólicas ou taxonômicas, processo denominado “anotação”.

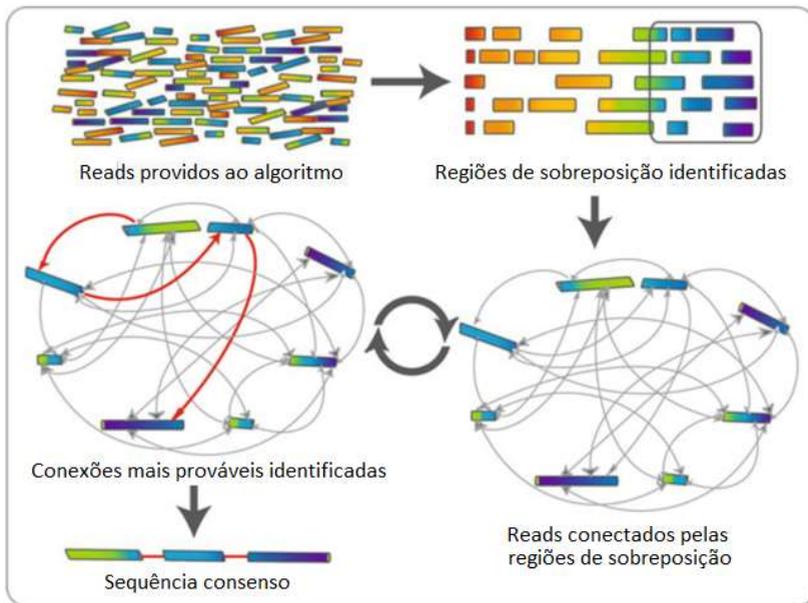


Figura 10. Procedimento de montagem de genomas a partir de *short reads* provenientes das plataformas de sequenciamento. As regiões de sobreposição são identificadas e a orientação mais provável dos *reads* é calculada por algoritmos derivados da teoria dos grafos, denominados gráficos *de-Bruijn* (conjunto de linhas em cinza, que representam as possíveis conexões entre as regiões de sobreposição identificadas). (Adaptado de: COMMINS; TOFT; FARES, 2009).

2.5.2 Anotação de metagenomas

Atribuir funções às sequências preditas é um dos procedimentos mais complexos, custosos e ao mesmo tempo mais cruciais para a análise de metagenomas. O processo de anotação baseia-se em dois pilares principais: a(s) base(s) de dados a partir da(s) qual(is) serão inferidas as informações sobre a sequência, e o método de busca aplicado para acessar estas informações. As diversas ferramentas atualmente disponíveis para anotação de sequências basicamente se resumem à implementação sistematizada das principais metodologias de

busca já consolidadas sobre diferentes bases de dados. Os métodos BLAST (ALTSCHUL et al., 1990), BLAT (KENT, 2002) e DIAMOND (BUCHFINK; XIE; HUSON, 2015), por exemplo, consistem em variações de algoritmos de alinhamento par-a-par baseados na similaridade entre sequências. O método HMMER (EDDY, 1995), por sua vez, utiliza modelos probabilísticos baseados em cadeias de Markov (*profile Hidden Markov Models*) para identificar perfis de distribuição de nucleotídeos ou aminoácidos entre sequências. *Profile* HMMs têm uma extensa aplicação na identificação de sequências remotamente relacionadas (JOHNSON et al., 2006), uma vez que não depende da similaridade direta entre as mesmas.

A escolha da base de dados depende dos objetivos do estudo, mas não é uma tarefa tão trivial quanto parece. Existem bases de dados específicas para informações taxonômicas, funcionais, famílias ou domínios proteicos, rotas metabólicas, bem como consórcios que buscam reunir estas informações de forma integrada e compreensiva (Tabela 2). Embora a maioria das bases de dados estejam disponíveis através de diferentes ferramentas, versões, formatos e tamanhos, o processo de anotação ainda é um dos gargalos da análise de metagenomas, e não apenas pela exigência de recursos computacionais. Os passos subsequentes frequentemente esbarram em dificuldades relacionadas à nomenclatura, formatação de dados e, principalmente, à falta de intercomunicação entre as próprias bases de dados e suas implementações. A construção e a manutenção de uma base de dados exigem uma curadoria criteriosa e a constante revisão e atualização de informações, de modo que a sistematização do enorme volume de dados de sequenciamento, depositados diariamente nas bases de dados públicas, ainda é uma tarefa em aberto sem perspectivas de resolução.

2.5.3 Mapeamento de *reads*

O termo “profundidade da cobertura” refere-se ao número de *reads* sequenciados para uma determinada região do genoma (Figura 11). Trata-se de uma métrica utilizada para avaliar a qualidade de uma montagem, i.e., se a quantidade de *reads* sequenciada foi capaz de cobrir a região de interesse em número suficiente para que a sequência consenso resultante seja considerada confiável. Para a metagenômica, este número pode ser utilizado para estimar a abundância de micro-organismos, genes, funções, ou quaisquer outros elementos genômicos de interesse. Este procedimento é denominado “*read mapping*” (mapeamento de *reads*) e consiste no alinhamento dos *reads* usando as

TAXONOMIA MICROBIANA	ANOTAÇÃO FUNCIONAL	FAMÍLIAS DE PROTEÍNAS	ROTAS METABÓLICAS	CONSORCIOS
Silva (PRUESSSE et al., 2007)	KEGG (Kyoto Encyclopedia of Genes and Genomes) (KANEHISA et al., 2004)	Pfam (FINN et al., 2016)	KEGG (KANEHISA et al., 2000)	NCBI (National Center for Biotechnology Information) (NCBI RESOURCCE COORDINATORS, 2017)
RDB (Ribossomal Database Project) (MADDAK et al., 1996)	SEED (OVERBEEK, 2005)	UniProt (BATEMAN et al., 2017)	MetaCyc (KRIEGER et al., 2004)	EMBL-EBI (European Molecular Biology Laboratory – European Bioinformatics Institute)
GreenGenes (DESANTIS et al., 2006)	eggNOG (HUERTA-CEPAS et al., 2016)	InterPro (HUNTER et al., 2009)	Reactome (CROFT et al., 2014)	Gene Ontology (ASHBURNER et al., 2000)

Tabela 2. Algumas das principais bases de dados utilizadas para anotação de genomas e metagenomas, de acordo com suas funcionalidades. Fonte: elaborado pela autora.

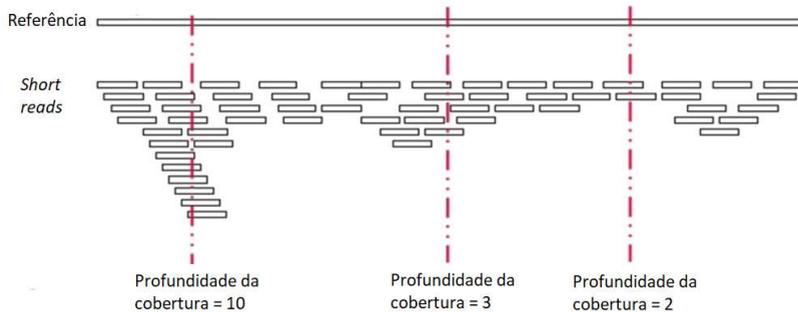


Figura 11. Representação esquemática do conceito de “profundidade da cobertura” (Fonte: *QIAGEN Webinar Series*).

próprias sequências montadas e anotadas como referência. Os resultados obtidos pelo mapeamento devem ser trabalhados por critérios muito atenciosos de normalização, de forma a não subestimar (ou superestimar) a representatividade de uma sequência no metagenoma (MARTIN et al., 2012). O método de normalização mais comum consiste no somatório do total de bases de todos os *reads* mapeados contra uma determinada região, dividido pelo tamanho (número de bases) desta região (YARYGIN et al., 2017).

2.5.4 Pipelines completos

O fluxo de trabalho completo para a análise de metagenomas é denominado *pipeline*. Os procedimentos descritos anteriormente são apenas alguns dos passos mais comuns de um *pipeline* típico. Passos adicionais podem incluir a aplicação de filtros para excluir sequências de baixa qualidade (*trimming*), a remoção de sequências redundantes (*de-replication*) ou de sequências que não interessam ao estudo (*screening*) (provenientes de contaminação, por exemplo), e o agrupamento de sequências que pertencem a um mesmo genoma (*binning*). Oulas e colaboradores (OULAS et al., 2015) sistematizam estes passos e exemplificam algumas ferramentas disponíveis para a realização de cada um deles na Figura 12. De acordo com Ma e colaboradores (2014):

Cada uma destas tarefas é realizada por *softwares* específicos que requerem instalação, configuração e integração. É uma tarefa assustadora até mesmo para especialistas em bioinformática. Muitos grupos de pesquisa constroem *pipelines* de análise escolhendo as ferramentas para cada tarefa de acordo com suas próprias experiências. Para um laboratório sem suporte de bioinformática, pode ser difícil realizar análises significativas com grandes volumes de dados.

(MA; PRINCE; AAGAARD, 2014)

Desta forma, muitas ferramentas estão disponíveis para a análise completa de metagenomas, utilizando servidores remotos e pipelines automáticos. Exemplos incluem o MG-RAST (Metagenomics RAST Server, MEYER et al., 2008), o EBI Metagenomics (MITCHELL et al., 2018) e o IMG/MER (CHEN et al., 2017). Estas ferramentas tornam a análise de metagenomas muito mais acessível para grupos de pesquisa com poucos recursos computacionais ou *expertise* em bioinformática, mas ainda apresentam diversas limitações. As metodologias implementadas apresentam variações de performance para cada tipo de dado e ambiente, e o ajuste de parâmetros nem sempre apresenta opções flexíveis; um *pipeline* unificado está longe de ser aplicável para todo e qualquer metagenoma. Uma análise robusta e a interpretação adequada dos resultados ainda requerem um conhecimento aprofundado dos conceitos envolvidos, de modo que a automatização de análises metagenômicas representa um tópico contraditório.

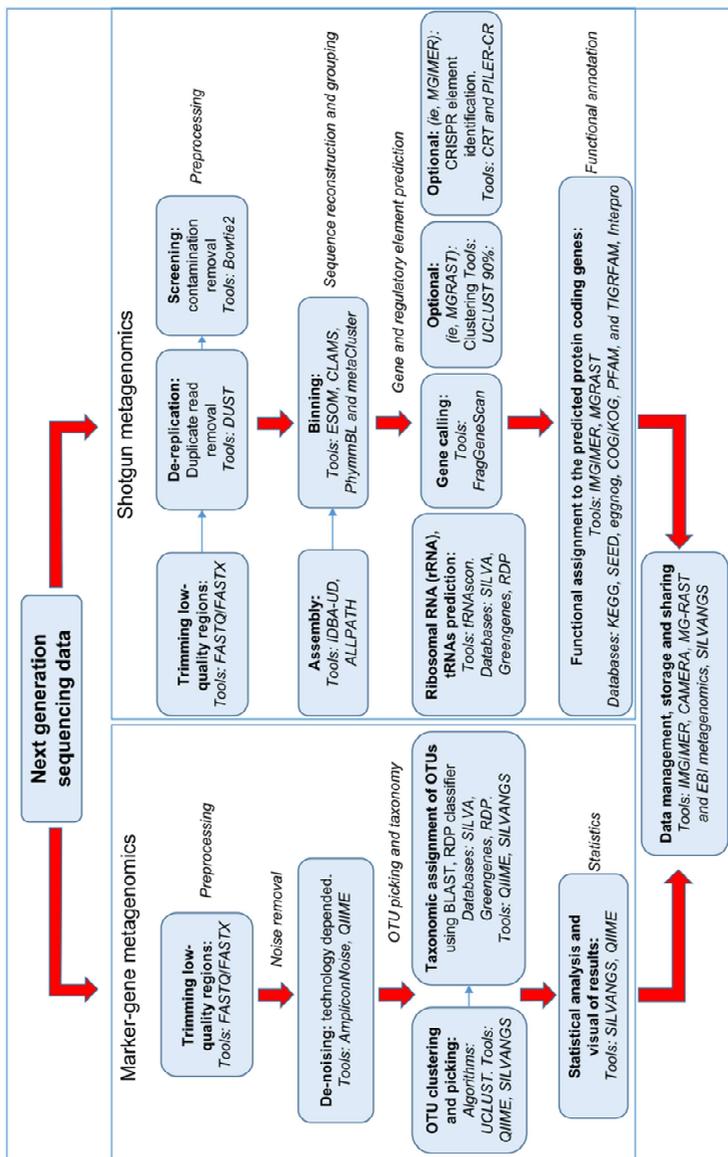


Figura 12. Diagrama de fluxo representando as etapas básicas da análise de dados metagenômicos, com exemplos de ferramentas disponíveis atualmente. (Fonte: OULAS et al., 2015).

3 METODOLOGIA

3.1 COLETA E PROCEDIMENTOS DE BANCADA

A coleta das amostras e os procedimentos de bancada (extração de DNA e sequenciamento) foram realizadas em uma etapa anterior deste trabalho (DUARTE, 2010), cujo foco é a análise *in silico* dos dados gerados.

As amostras de *permafrost* utilizadas neste projeto foram previamente coletadas durante a Operação Antártica Brasileira XXVII e a 54ª Expedição Antártica Russa, que ocorreu no verão austral de 2008-2009. A coleta de amostras fez parte de um projeto conjunto entre o Laboratório de Ecologia Microbiana (Instituto Oceanográfico – USP) e o Laboratório de Criossolos do Instituto de Problemas Físico-químicos e Biológicos do Solo (*Moscow State University*, Rússia). A amostragem foi realizada nas imediações da Estação Antártica Russa Bellingshausen, na Ilha Rei George, Península Antártica (Figura 13) e compreende 5 estratos de profundidades entre 4,6 e 9,2 metros abaixo da superfície do solo, constituindo a perfuração de código A1-09 da 54ª Expedição Antártica Russa. Além destes, uma amostra de superfície do sedimento marinho da Baía Maxwell, denominada B2, foi coletada no verão 2011/2012 por pesquisadores do Laboratório de Criossolos, de modo a representar o sedimento marinho moderno para comparações entre as comunidades microbianas do presente e do passado (Figura 14). A datação das amostras de *permafrost* foi realizada através da análise de radioisótopos em espectrômetro de massa com acelerador (*Accelerator Mass Spectrometry* - AMS) no *Swiss Federal Institute of Technology Zurich* (ETHZ) sob o código UZ-5776/ETH-39470 (ABRAMOV et al. 2011).

O *permafrost* a exatos 9,0 m de profundidade indicou uma idade de congelamento de 7485 ± 40 anos antes do presente (ABRAMOV et al. 2011) e dados geológicos indicam que a taxa de sedimentação foi razoavelmente constante ao longo do tempo. Dessa forma, as datações das demais amostras de *permafrost* foi estimada com base na amostra de 7485 anos (aproximado para o valor de 7500 anos). As datações indicam a idade aproximada em que os sedimentos foram congelados: 2250 anos antes do presente na amostra mais superficial (P13) até 7500 anos antes do presente na amostra mais profunda (P17) (Tabela 3).

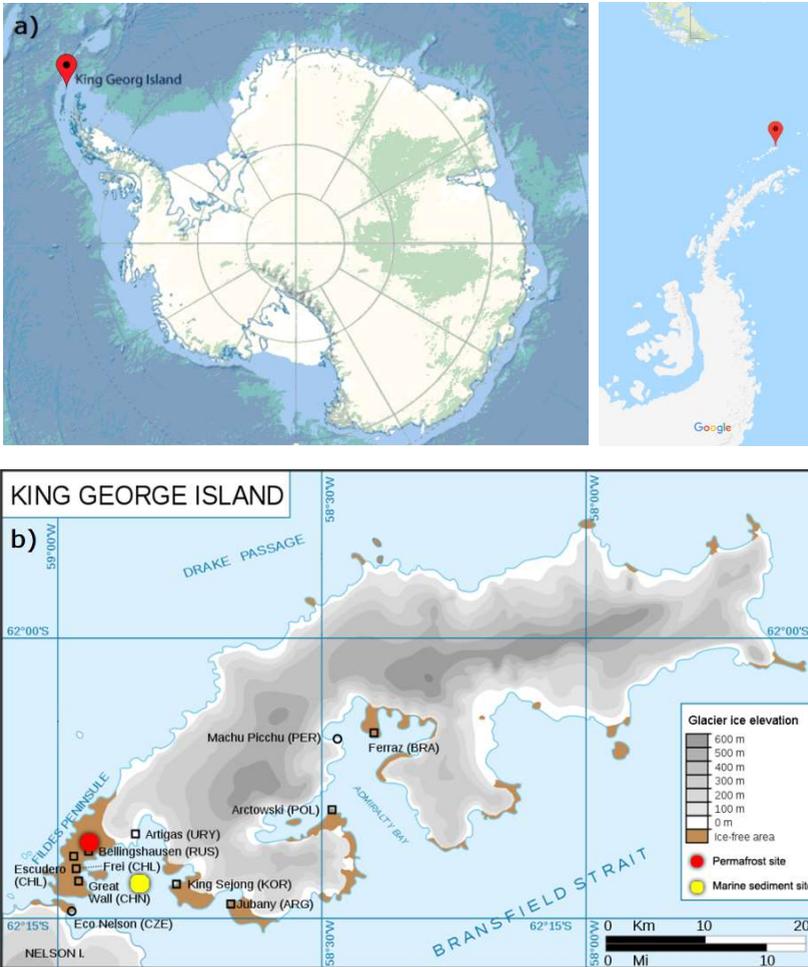


Figura 13. a) Mapa da Antártica indicando a localização da Ilha Rei George (King George Island); **b)** Mapa da Ilha Rei George indicando o ponto de coleta das amostras de *permafrost* P13-P17 (círculo vermelho) e do sedimento marinho moderno B2 (círculo amarelo). Fonte: Alexrk2, *Wikimedia Commons* CC BY-SA 3.0; Giovanni Fattori, *Wikipedia*, CC BY-SA 3.0.



Figura 14. a) Fotografia da área de coleta próxima à estação russa Bellingsausen, Ilha Rei George, Antártica; b) Equipamento de perfuração à seco utilizado na coleta das amostras de *permafrost*; c) Detalhe do testemunhador utilizado na perfuração; d) Exemplos de algumas amostras de *permafrost* coletadas; e) Procedimento de descontaminação da superfície do *permafrost*. (Fonte: Acervo pessoal do Prof. Rubens Duarte).

Tabela 3. Amostras de sedimento marinho e *permafrost* utilizadas neste trabalho.

Amostra	Coordenadas	Profundidade (m)	Idade estimada (Anos antes do presente)
B2	62°20'16"S; 58°93'87"O	0 m	Sedimento atual
P13	62°19'68"S; 58°96'62"O	4,6 m	2.250 anos
P14	62°19'68"S; 58°96'62"O	5,8 m	3.100 anos
P15	62°19'68"S; 58°96'62"O	6,9 m	4.300 anos
P16	62°19'68"S; 58°96'62"O	8,1 m	5.600 anos
P17	62°19'68"S; 58°96'62"O	9,2 m	7.500 anos

Após a coleta, as amostras foram descontaminadas conforme descrito em Duarte (2010), armazenadas em freezer a -20°C e transportadas ao laboratório de Ecologia Microbiana da Universidade de São Paulo, onde permaneceram sob congelamento até a realização da extração de DNA e sequenciamento. A extração do DNA foi realizada em triplicatas de 10g cada com o kit *PowerMax Soil*™ DNA (Mo Bio, USA), e em seguida o DNA foi purificado com o kit DNA *Clean & Concentrator*™ (Zymo Research, USA).

O sequenciamento de genoma completo (WGS) das amostras foi realizado pela plataforma Illumina HiSeq 2000, no Laboratório Central de Tecnologias de Alto Desempenho em Ciências da Vida (LaCTAD – Unicamp, SP), através de terceirização de serviço. Os kits Illumina utilizados para preparação de bibliotecas de DNA geraram fragmentos *paired-end* com *reads* de até 2x 100 pb.

3.2 ANÁLISES *IN SILICO*

3.2.1 Hardware

Três computadores foram utilizados nas análises de dados:

1) Dell Inspiron, processador Intel Core i7-3537U 2.00 GHz, 8Gb de RAM e HD de 1 Tb, sistema operacional Bio-Linux 8, *desktop* do Laboratório de Ecologia Molecular e Extremófilos, Departamento de Microbiologia, Imunologia e Parasitologia (MIP) do Centro de Ciências Biológicas da UFSC;

2) Sony VAIO modelo SVF15A1C5E, processador Intel Core i73537U 2 GHz, 8Gb de RAM e HD de 500Gb, sistema operacional *dual boot* com Bio-Linux 8 e Windows 10 Home, *laptop* pessoal da autora;

3) Dell PowerEdge R910 com 4 Processadores Intel Ten-Core Xeon E7-4870 2.40GHz, 30M cache, 6.4 GT/s QPI, Turbo, HT, 1066MHz MaxMem 512GB de Memória (32 x 16GB), 1066 MHz, Quad Ranked RDIMMs Low Voltage, 4 Discos Rígidos de 600GB 10K RPM SAS 6Gbps 2.5 in Hotplug, acoplado a um Dell PowerVault MD1200 com 12 discos rígidos de 2Tb cada, modelo Near Line SAS 6Gbps 7.2K 3.5 polegadas. O servidor Dell PowerEdge faz parte do Centro de Facilidades de Apoio à Pesquisa (CEFAP) da Universidade de São Paulo (USP), sendo acessado remotamente via protocolo *ssh*.

3.2.2 Pré-processamento das sequências

Os *reads* gerados para cada amostra foram pré-processados para remoção de sequências de baixa qualidade com base em *score Phred* através do *software* Mothur (SCHLOSS et al., 2009), utilizando um *cut-off* de média inferior a 20 (i.e., probabilidade de erro de 1:100). De modo a maximizar as chances de encontrar sequências similares nas bases de dados atuais, decidiu-se por realizar a montagem das sequências (*assembly*) em *contigs* de maior tamanho, utilizando algoritmos de montagem *de novo* por gráficos *de-Bruijn* através do *software* IDBA-UD v.1.1.1 (PENG et al., 2012). Cada *contig* resultante do processo de montagem foi submetido a um processo de predição de genes pelo *software* FragGeneScan (RHO; TANG; YE, 2010). Os comandos utilizados em cada programa, seus respectivos parâmetros e os passos intermediários necessários para estes procedimentos encontram-se detalhados no Apêndice A.

3.2.3 Anotação das sequências do metagenoma

A anotação das sequências de cada metagenoma foi realizada utilizando duas abordagens: um *pipeline* online automático, e um *pipeline* próprio aplicado localmente.

A primeira abordagem consistiu em submeter os *reads* brutos (sem passar pelo pré-processamento) diretamente ao *pipeline* do MGRAST. Este método tem a vantagem de utilizar um conjunto de algoritmos especificamente desenhados para a anotação de *reads* curtos (~50 pb), como é o caso do sequenciamento na plataforma Illumina HiSeq, além de quantificar automaticamente a abundância de cada gene identificado no metagenoma. Entretanto, o processamento dos dados não é imediato devido ao grande número de usuários, obrigando o usuário aguardar o início das análises por uma média de 2 a 6 meses.

A segunda abordagem consistiu no processamento local, isto é, o pré-processamento das sequências conforme descrito no item 3.2.2, e a anotação das sequências geradas utilizando diferentes opções de ferramentas disponíveis. Esta abordagem tem a vantagem de permitir flexibilidade no uso de algoritmos e parâmetros de anotação, bem como na escolha de bancos de dados mais adequados para interpretação dos resultados. O princípio motivador da busca por diferentes metodologias foi determinar um método de anotação e bases de dados capazes de classificar o maior número de sequências possível sem exigir um esforço computacional excessivo; por esse motivo, o método BLAST e

demais ferramentas dependentes do seu algoritmo foram descartadas. A ferramenta online eggNOG-*mapper* (HUERTA-CEPAS et al., 2017) foi utilizada para anotar sequências aminoacídicas resultantes do processo de predição de genes com o FragGeneScan através do método de mapeamento rápido DIAMOND, uma vez que o modo HMMER só está disponível para um número máximo de 5.000 sequências. As mesmas sequências foram anotadas localmente através dos *softwares* Prokka (SEEMANN, 2014) e InterProScan (QUEVILLON et al., 2005), neste último utilizando as bases de dados Pfam, InterPro, Gene Ontology, KEGG Pathways, MetaCyc e Reactome. O MEGAN6 foi testado apenas para a visualização de dados, sobre os resultados obtidos pelo eggNOG-*mapper* e InterProScan.

3.2.4 Mapeamento de *reads*

De modo a recuperar os dados quantitativos, perdidos durante o processo de montagem devido à própria natureza do algoritmo, os *reads* foram mapeados contra as sequências nucleotídicas geradas pelo FragGeneScan. Para isso, utilizou-se o *software* Bowtie2 (LANGMEAD; SALZBERG, 2012) em conjunto com o pacote de algoritmos SAMTools para edição de arquivos em formato SAM (Sequence Alignment/Map) (LI et al., 2009). Os comandos e parâmetros utilizados encontram-se detalhados no Apêndice A.

3.2.5 Pós-processamento

Os procedimentos descritos nas próximas sessões dizem respeito às anotações geradas pelo InterProScan utilizando a base de dados Pfam. Os critérios de escolha da ferramenta estão descritos na sessão 5.1 (Discussão Metodológica).

A contagem de *reads* atribuída a cada sequência pelo Bowtie2 foi acoplada ao arquivo proveniente do processo de anotação utilizando comandos em linguagem *Bash* (detalhados no Apêndice A). *Reads* mapeados contra diferentes sequências com a mesma anotação foram somados para representar a contagem total do produto gênico (família/domínio proteico) no metagenoma. Nos casos em que uma mesma sequência recebeu mais de uma anotação (por exemplo, quando a sequência completa corresponde a uma família de proteínas e também possui um domínio proteico anotado), o número de *reads* foi contabilizado para cada diferente anotação. O termo OFU (*Operational Functional Unit*) passará a ser empregado para se referir, de forma

generalizada, às anotações realizadas de acordo com a base de dados Pfam em níveis funcionais distintos (famílias ou domínios proteicos). O termo proposto é uma alusão à sigla OTU (*Operational Taxonomic Unit*) utilizada para se referir a um conjunto de sequências gênicas correspondentes ao RNAr 16S, agrupadas por similaridade quando o nível taxonômico não pode ser definido.

3.2.6 Seleção de dados

A contagem absoluta de *reads* atribuída a cada OFU (Identificador Pfam) pelo processo de mapeamento foi transformada em abundância relativa utilizando a normalização proposta por Yarygin et al. (2017), que pondera a representatividade dos *reads* no metagenoma de acordo com o tamanho da sequência ao qual foram mapeados:

$$\text{Abundância Relativa} = \frac{\sum \text{Reads Mapeados} \times \text{Taman do read}}{\text{Taman da sequência} \times \sum \text{Total de reads}}$$

Na fórmula acima, o somatório do total de *reads* incluiu mesmo aqueles que não foram mapeados contra sequências anotadas.

Os valores obtidos para cada OFU foram normalizados pela soma da abundância relativa de todas as OFUs do metagenoma, para que pudessem ser comparados entre metagenomas de diferentes tamanhos:

$$\text{A.R. Normalizada} = \frac{\text{Abundância Relativa} \times 100}{\sum \text{Abundância relativa de todas as OFUs}}$$

A partir dos valores normalizados, foram calculadas proporções de distribuição de cada OFU entre as amostras, dividindo o valor de abundância de cada OFU em cada amostra pelo somatório dos valores de abundância daquela OFU entre todas as amostras:

$$\text{Distribuição} = \frac{\text{A.R.N. na amostra}}{\sum \text{A.R.N. em todas as amostras}}$$

OFUs com até 90% de distribuição em uma única amostra foram consideradas exclusivas daquela amostra e analisadas separadamente. Para as demais, foram calculados a variância e o desvio padrão, e a partir da ordenação dos valores de desvio, foram considerados dois subgrupos para análise: 25% das OFUs mais diferencialmente distribuídas entre as amostras (i.e., valores acima do terceiro quartil do

desvio padrão), e 25% das OFUs mais homoganeamente distribuídas entre as amostras (valores abaixo do primeiro quartil do desvio padrão).

Para ambos os subgrupos, os identificadores relacionados a cada OFU anotada foram manualmente consultados *online* na base de dados Pfam, e as informações disponíveis foram utilizadas para atribuir, a cada sequência, uma nomenclatura referente às 27 categorias de primeiro nível do SEED Subsystems. A abundância de cada categoria foi calculada somando-se os valores de abundância relativa normalizada de todas as OFUs classificadas dentro de um mesmo grupo funcional. Os resultados referentes aos 25% das OFUs mais diferencialmente distribuídas foram utilizados para construir uma tabela de abundância diferencial, a partir da qual foi realizada a visualização dos dados e a análise de estatística multivariada.

3.2.7 Estatística Multivariada

A diversidade funcional das amostras de *permafrost* e sua correlação com os parâmetros físico-químicos foi avaliada através de estatística multivariada no software Canoco v4.5 (LEPS; SMILAUER, 2003). Para isso foi realizada uma análise de componentes principais (PCA) utilizando a abundância de OFUs em cada categoria funcional como dados biológicos e os resultados das análises físico-químicas como variáveis explanatórias. Os dados biológicos foram transformados pelo cálculo da raiz quadrada dos valores e os dados físico-químicos foram convertidos em módulo. Nesta análise a amostra de sedimento marinho moderno B2 foi excluída por não possuir dados físico-químicos avaliados.

4 RESULTADOS

4.1 PRÉ-PROCESSAMENTO

As amostras submetidas ao *pipeline* automático do MG-RAST não apresentaram resultados satisfatórios com relação ao controle de qualidade (Tabela 4). A ferramenta utiliza o método DRISSEE (KEEGAN et al., 2012), que embora seja mais eficiente para detecção de erros em sequências metagenômicas, demonstrou-se demasiado rigoroso para as sequências do *permafrost*. Uma vez que as amostras do *permafrost* datam de milhares de anos, era esperado que uma grande parcela das sequências obtidas não encontrasse correspondentes nas bases de dados atuais, dificultando o processo de anotação. Como a identidade precisa das sequências não é de importância fundamental para este trabalho, considerou-se mais adequada a aplicação de um filtro de qualidade mais flexível, que mantivesse um número maior de sequências mesmo sob o ônus de incluir aquelas consideradas de baixa qualidade. Deste modo, decidiu-se trabalhar com os resultados da abordagem local de pré-processamento e anotação. Os resultados do *pipeline* próprio aplicado localmente encontram-se reunidos na Tabela 5.

Tabela 4. Resultados de controle de qualidade e anotação das sequências obtidos pelo *pipeline* automático do MG-RAST.

Amostra	Total de reads	Failed quality control	Unknown	Predicted feature
B2	34.935.245	3.033.754 (8,68%)	2.463.436 (7,05%)	29.438.055 (84,26%)
P13	30.648.173	26.631.331 (79,83%)	981.831 (2,94%)	5.747.058 (17,23%)
P14	33.360.220	17.791.547 (58,05%)	2.284.084 (7,45%)	10.572.542 (34,5%)
P15	107.688.314	71.056.129 (65,98%)	5.705.690 (5,3%)	30.926.495 (28,72%)
P16	19.592.317	10.443.574 (53,3%)	1.409.725 (7,2%)	7.739.018 (39,5%)
P17	24.171.152	7.573.780 (31,33%)	2.472.220 (10,23%)	14.125.152 (58,44%)

Tabela 5. Resultados do *pipeline* de pré-processamento aplicado localmente.

Amostra	Total de reads	Reads pós controle de qualidade	Contigs	Genes preditos
B2	2x 17.645.005	2x 15.909.276	82.910	138.611
P13	2x 18.709.144	2x 16.327.610	22.903	38.696
P14	2x 15.641.744	2x 14.646.462	35.195	54.685
P15	2x 55.714.600	2x 52.216.929	88.233	138.092
P16	2x 9.912.802	2x 9.280.734	20.753	31.464
P17	2x 12.220.391	2x 10.948.081	42.662	67.348

4.2 ANOTAÇÃO

Os resultados de anotação realizados pelas outras duas ferramentas testadas (*InterProScan* e *eggNOG-mapper*) encontram-se ilustrados na Tabela 6. Embora o *eggNOG-mapper* tenha apresentado uma melhor performance com relação ao número de sequências anotadas, houve dificuldade na visualização dos resultados devido ao sistema de nomenclatura utilizado pela ferramenta (discutido na Seção 5: Discussão Metodológica). Deste modo, a visualização e interpretação dos dados foi realizada sobre os resultados do *InterProScan*, utilizando as anotações provenientes das bases de dados Pfam e KEGG Pathways (Tabela 7).

Tabela 6. Resultados do processo de anotação de sequências utilizando as ferramentas *InterProScan* e *eggNOG-mapper*.

Amostra	Sequências preditas	Sequências anotadas (InterProScan)	Sequências anotadas (eggNOG-mapper)
B2	138.611	82.308 (59,38%)	108.053 (77,95%)
P13	38.696	15.724 (40,63%)	19.255 (49,75%)
P14	54.685	28.277 (51,71%)	31.211 (57,07%)
P15	138.092	74.826 (54,18%)	81.174 (58,78%)
P16	31.464	18.130 (57,62%)	19.387 (61,61%)
P17	67.348	37.527 (55,72%)	40.636 (60,33%)

Tabela 7. Resultados do processo de anotação utilizando a ferramenta InterProScan, demonstrando o número de sequências anotadas de acordo com as bases de dados Pfam e KEGG Pathways, e o número de OFUs (Pfam Ids e KEGG Pathways) não redundantes identificados em cada amostra.

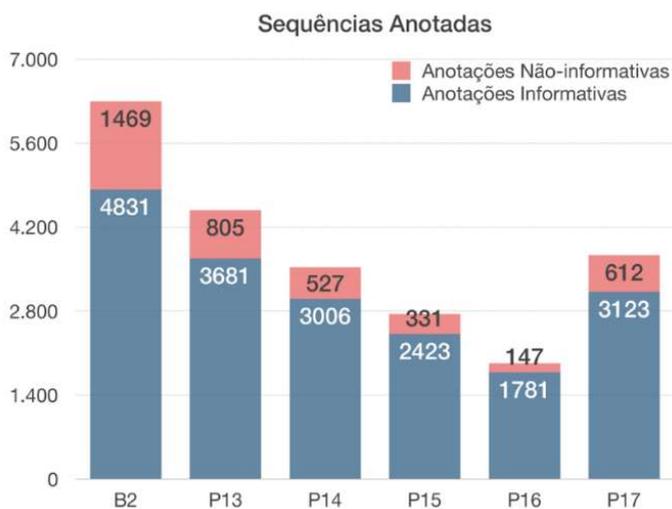
Amostra	Sequências preditas	Sequências anotadas (Pfam)	OFUs Identificadas	Sequências anotadas (KEGG Pathways)	Rotas metabólicas identificadas
B2	138.611	82.308 (59,38%)	6.300	8.143 (5,87%)	119
P13	38.696	15.724 (40,63%)	4.486	1.240 (3,20%)	110
P14	54.685	28.277 (51,71%)	3.533	3.041 (5,56%)	102
P15	138.092	74.826 (54,18%)	2.754	7.730 (5,59%)	96
P16	31.464	18.130 (57,62%)	1.928	2.035 (6,46%)	93
P17	67.348	37.527 (55,72%)	3.735	3.954 (5,87%)	102
Total de OFUs não redundantes identificados:		6.406	Total de rotas metabólicas não redundantes identificados:		119

A Tabela 8 demonstra o número de OFUs compartilhadas entre cada par de amostras, a partir do total de 6.406 sequências anotadas de acordo com a base de dados Pfam. Com exceção da amostra P17, cada amostra mais recente compartilhou todas as OFUs da amostra anterior; isto é: a amostra B2 conteve todas as OFUs das amostras P13, P14, P15 e P16; a amostra P13 conteve todas as OFUs das amostras P14, P15 e P16, e assim sucessivamente.

Do total de 6.406 sequências não redundantes anotadas pelo InterProScan, 1.495 (aproximadamente 23% das OFUs identificadas) foram atribuídas a domínios ou famílias de função desconhecida ou hipotéticos. A representatividade das OFUs de função desconhecida ou hipotética em cada amostra está representada na Figura 15.

Tabela 8. Número de OFUs compartilhadas entre cada par de amostras.

	B2	P13	P14	P15	P16	P17
B2	6.300	4.486	3.533	2.754	1.928	3.629
P13	4.486	4.486	3.533	2.754	1.928	3.452
P14	3.533	3.533	3.533	2.754	1.928	3.125
P15	2.754	2.754	2.754	2.754	1.928	2.630
P16	1.928	1.928	1.928	1.928	1.928	1.905
P17	3.629	3.452	3.125	2.630	1.905	3.735

**Figura 15.** Representatividade das sequências atribuídas a famílias ou domínios proteicos de função desconhecida ou hipotética em cada metagenoma, utilizando a ferramenta de anotação InterProScan e a base de dados Pfam.

A partir do procedimento descrito no item 3.2.5, foram identificadas 1.915 OFUs (29,89% do total de OFUs não redundantes entre as amostras) encontradas apenas na amostra B2. Este resultado era esperado, uma vez que a amostra B2 representa o sedimento marinho atual, sendo melhor representada nas bases de dados atuais. Com o objetivo de diminuir este viés de representatividade, as OFUs exclusivas da amostra B2 foram removidas para que pudessem ser comparadas com as amostras do *permafrost* (Apêndice B). No entanto, para obter uma visualização menos enviesada da abundância diferencial entre as OFUs do *permafrost*, a amostra B2 foi removida por completo e o procedimento repetido considerando apenas as amostras de *permafrost*. Após removidas as sequências correspondentes a produtos gênicos desconhecidos ou hipotéticos, foram selecionadas 3.872 OFUs para a análise, das quais: 607 estiveram presentes apenas na amostra P13, 4 apenas na amostra P14, e 203 apenas na amostra P17. As amostras P15 e P16 não apresentaram nenhuma OFU exclusiva.

As OFUs homogeneamente distribuídas entre as amostras (desvio padrão $< 0,098$, excetuando as amostras exclusivas) totalizaram 764 (11,92% do total de sequências anotadas, e 19,73% das sequências anotadas excetuando os produtos gênicos desconhecidos e OFUs exclusivas da B2). Entre as OFUs consideradas mais desigualmente distribuídas entre as amostras (desvio padrão $> 0,244$, excetuando as amostras exclusivas), foram selecionadas 766 para a análise (11,95% de todas as sequências anotadas, e 19,78% excetuando os produtos gênicos desconhecidos e OFUs exclusivas da B2). A Figura 16 ilustra a representatividade de cada subgrupo de OFUs definido pela seleção de dados, em termos de porcentagem de *reads* mapeados, em cada metagenoma: sequências de função desconhecida ou hipotéticas, de acordo com as anotações na base de dados Pfam (verde escuro), sequências consideradas exclusivas de cada amostra (distribuição superior a 90% em uma única amostra – cinza), sequências homogeneamente distribuídas (valores de distribuição inferiores ao primeiro quartil do desvio padrão – rosa), sequências mais diferencialmente distribuídas (valores de distribuição superiores ao terceiro quartil do desvio padrão – verde claro) e o restante das sequências (valores entre o primeiro e o terceiro quartil – azul).

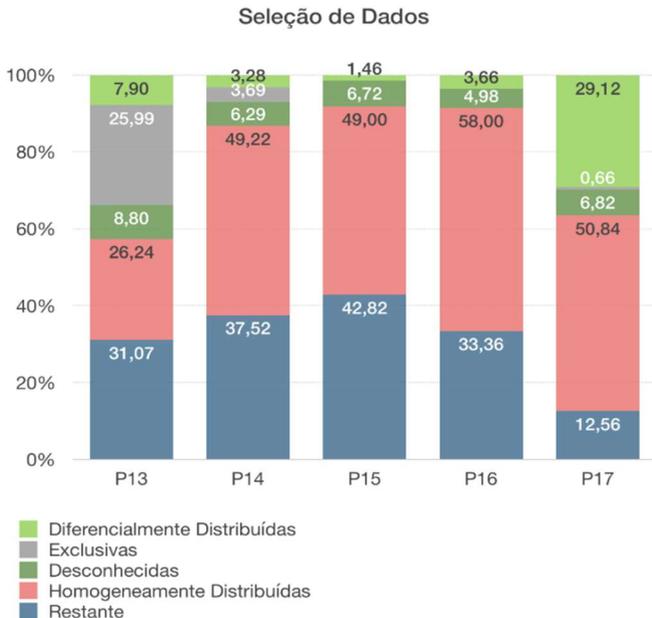


Figura 16. Representatividade de cada subgrupo de OFUs definidos pela seleção de dados a partir dos valores de distribuição entre as amostras.

4.3 ABUNDÂNCIA DIFERENCIAL

A Tabela 9 apresenta os resultados de abundância relativa das 25 categorias funcionais de primeiro nível do SEED *Subsystems*, considerando as 766 OFUs mais diferencialmente distribuídas entre as amostras, mais as OFUs exclusivas de cada amostra.

A representatividade de cada categoria encontra-se ilustrada nas Figuras 17 (por amostra) e 18 (entre as amostras). A partir destes resultados, algumas categorias foram selecionadas para uma análise mais detalhada, sendo reagrupadas e/ou subcategorizadas de acordo com o que se considerou mais pertinente avaliar em conjunto. Estes novos subgrupos são apresentados nas sessões seguintes, Figuras 19 a 46, acompanhados dos resultados de abundância relativa de rotas metabólicas, de acordo com a base de dados KEGG Pathways.

Tabela 9. Resultados de abundância relativa para as 25 categorias funcionais do SEED *Subsystems* entre as amostras de *permafrost*.

Subsystems	P13	P14	P15	P16	P17
Amino Acids and Derivatives	0,386	0,671	0,854	1,314	1,355
Carbohydrates	1,682	18,602	9,893	7,142	11,548
Cell Division and Cell Cycle	0,373	0,034	0,000	0,588	0,942
Cell Wall and Capsule	1,308	0,767	4,252	1,608	2,645
Cofactors, Vitamins, Prosthetic Groups, Pigments	1,863	10,762	2,657	6,490	3,820
DNA Metabolism	0,599	2,701	1,453	4,111	4,526
Dormancy and Sporulation	0,070	0,329	0,166	0,882	1,035
Fatty Acids, Lipids, Isoprenoids	0,273	0,170	0,106	0,780	0,410
Iron Acquisition and Metabolism	0,093	0,014	0,000	0,028	0,029
Membrane Transport	0,885	1,799	1,391	3,518	3,331
Metabolism of Aromatic Compounds	0,032	0,433	0,399	0,704	0,708
Miscellaneous	2,492	4,957	4,000	8,194	8,596
Motility and Chemotaxis	0,064	0,046	0,000	1,717	0,610
Nitrogen Metabolism	0,681	0,117	0,000	0,401	0,340
Nucleotides and Nucleosides	0,215	3,457	0,989	8,174	3,411
Phages, Prophages, Transposable Elements, Plasmids	47,932	8,012	0,594	3,422	3,190
Photosynthesis	0,002	0,000	0,000	1,033	0,370
Protein Metabolism	0,859	28,006	25,732	11,379	9,947
Regulation and Cell Signaling	16,477	16,582	46,476	26,628	34,975
Respiration	1,265	0,922	0,283	3,556	1,682
RNA Metabolism	0,102	0,087	0,053	2,092	1,135
Secondary Metabolism	0,056	0,046	0,000	0,448	0,074
Stress Response	3,324	0,813	0,347	2,194	2,796
Sulfur Metabolism	0,003	0,054	0,052	0,117	0,091
Virulence, Disease and Defense	18,964	0,453	0,304	3,275	2,378

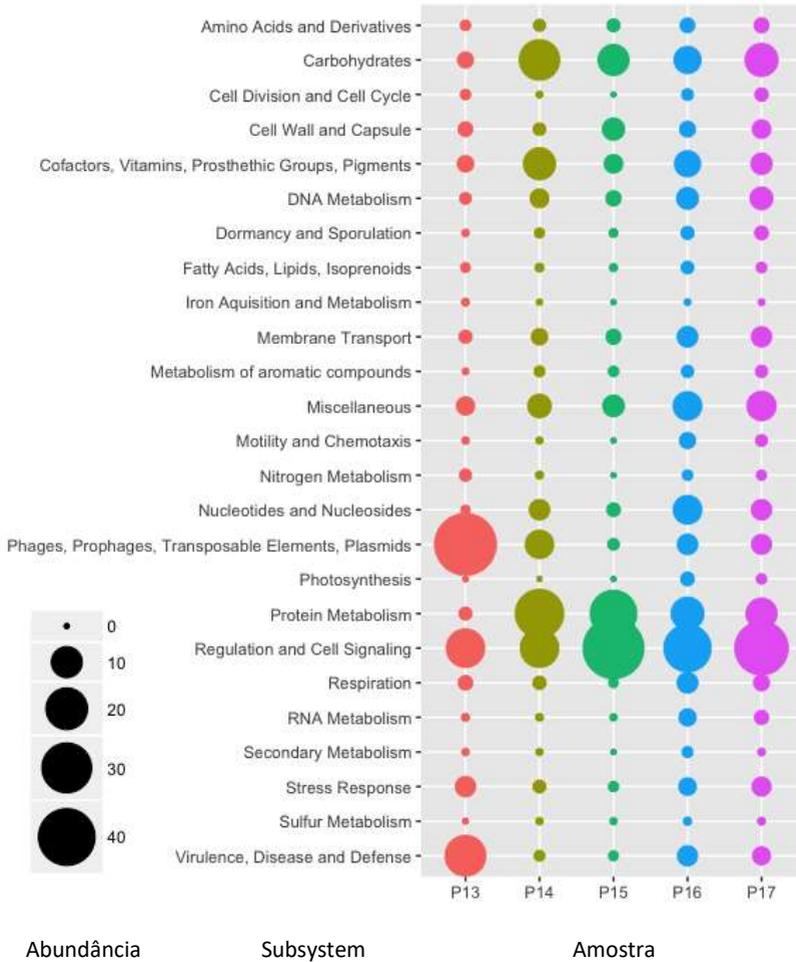


Figura 17. Representatividade (em termos percentuais) das 25 categorias funcionais do SEED *Subsystems* em cada amostra do *permafrost*.

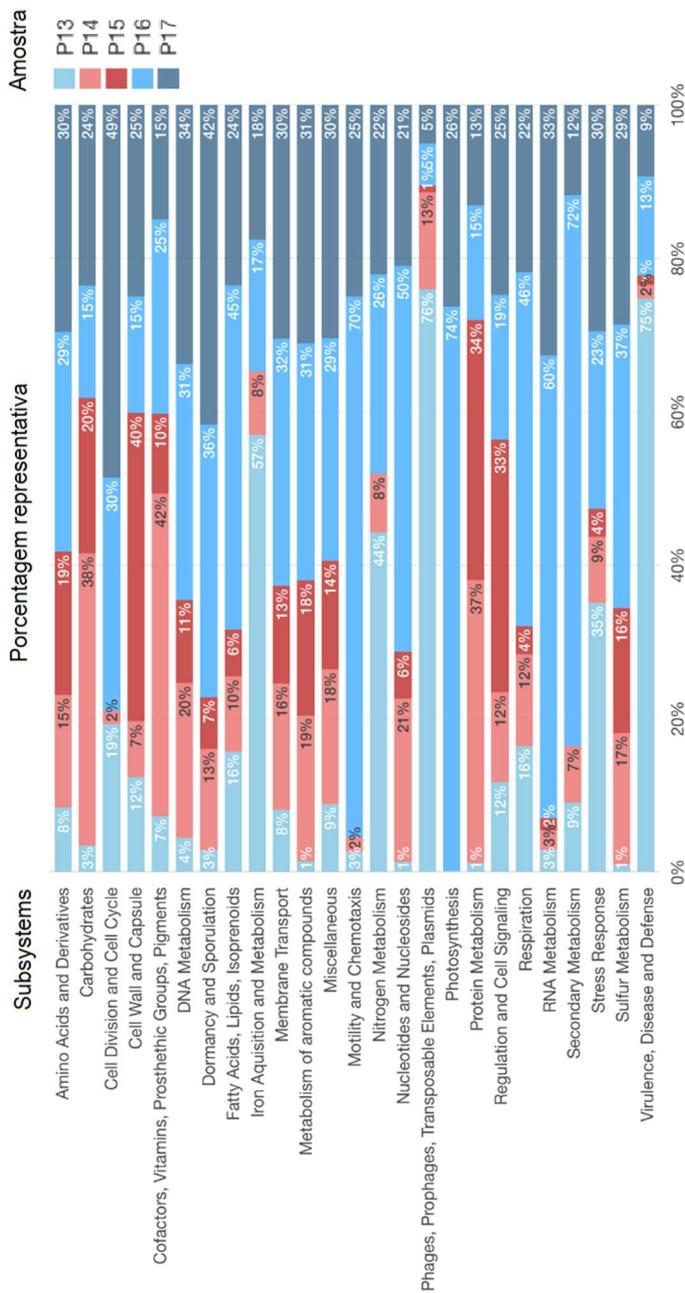


Figura 18. Representatividade (em termos percentuais) das 25 categorias funcionais do SEED Subsystems entre as amostras de *permafrost*. Em tons de vermelho estão representadas as amostras correspondentes ao período do Ótimo Climático do Holoceno Médio (P14 e P15). As demais (P13, P16 e P17) estão representadas em tons de azul.

4.3.1 Metabolismo de biomoléculas

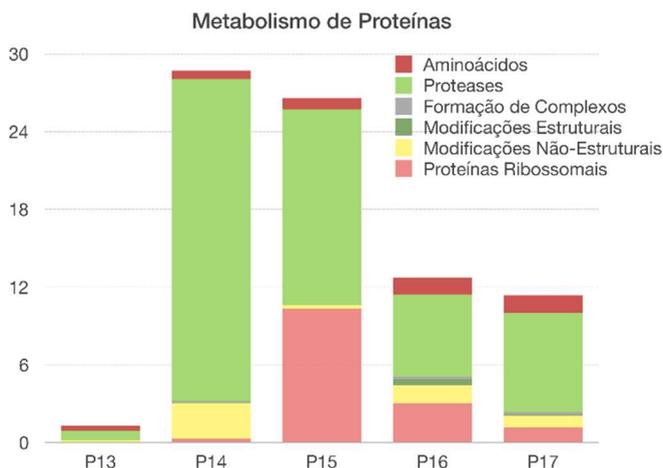


Figura 19. Abundância relativa das diferentes subcategorias relacionadas ao metabolismo de aminoácidos e proteínas, derivadas dos *Subsystems* “*Amino Acids and Derivatives*” e “*Protein Metabolism*”.

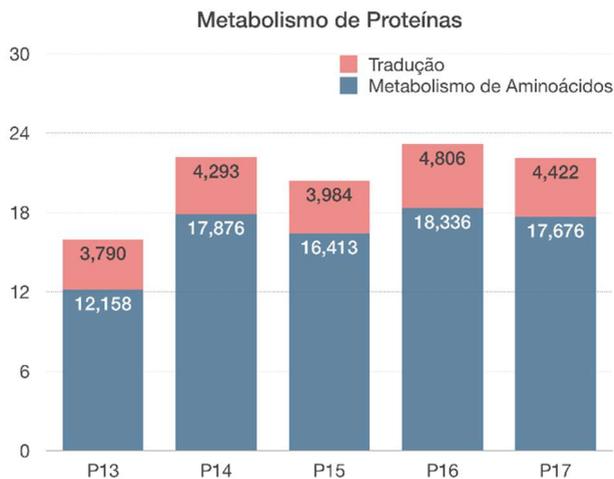


Figura 20. Abundância relativa de rotas metabólicas relacionadas ao metabolismo de proteínas, derivadas dos KEGG *Pathways* “*Amino Acid Metabolism*”, “*Metabolism of Other Amino Acids*” e “*Translation*”.

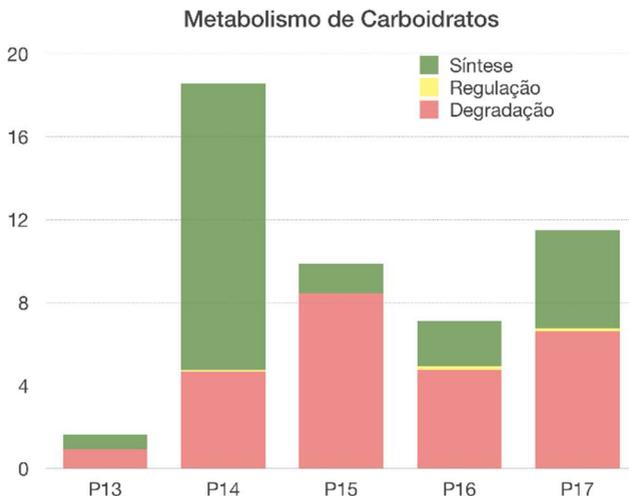


Figura 21. Abundância relativa das diferentes subcategorias relacionadas ao metabolismo de carboidratos, derivadas do *Subsystem* “*Carbohydrates*”.

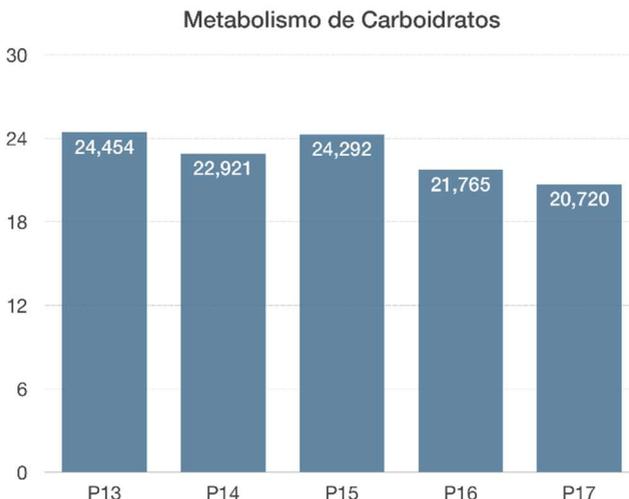


Figura 22. Abundância relativa de rotas metabólicas relacionadas ao metabolismo de carboidratos, derivadas do KEGG Pathway “*Carbohydrate Metabolism*”.

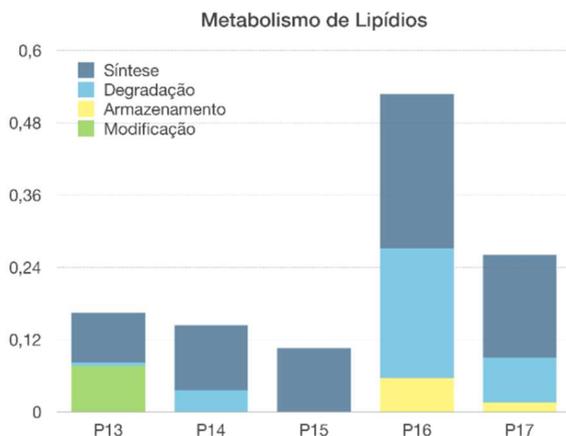


Figura 23. Abundância relativa das diferentes subcategorias relacionadas ao metabolismo de lipídios, derivadas do *Subsystem* “*Fatty Acids, Lipids, Isoprenoids*”. As OFUs referentes ao metabolismo de fosfolipídios não estão incluídas nesta visualização, sendo consideradas no gráfico “Estrutura de Membrana”.

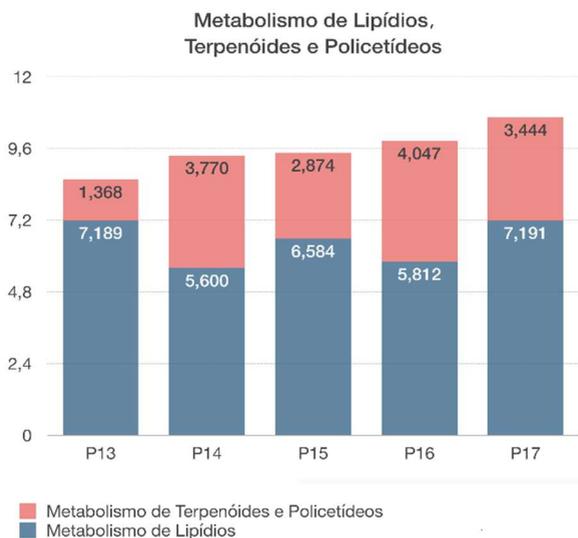


Figura 24. Abundância relativa de rotas metabólicas relacionadas ao metabolismo de lipídios, derivadas dos KEGG Pathways “*Lipid Metabolism*” e “*Metabolism of Terpenoids and Polyketides*”

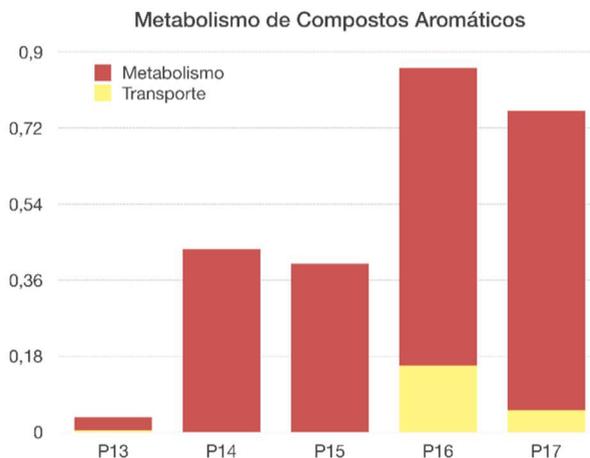


Figura 25. Abundância relativa das OFUs relacionadas ao metabolismo de compostos aromáticos, derivadas do *Subsystem* “*Metabolism of Aromatic Compounds*”. A subcategoria “Transporte” inclui OFUs referentes ao transporte dos compostos tolueno, benzeno e organoclorados, identificadas no *Subsystem* “*Membrane Transport*”.

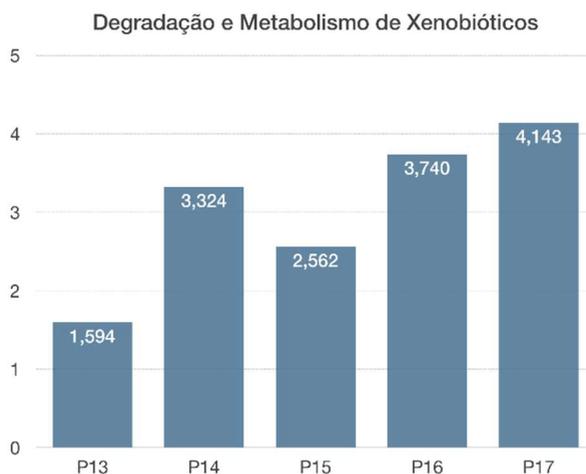


Figura 26. Abundância relativa de rotas metabólicas relacionadas à degradação de compostos xenobióticos, derivadas do *KEGG Pathway* “*Xenobiotics Biodegradation and Metabolism*”.

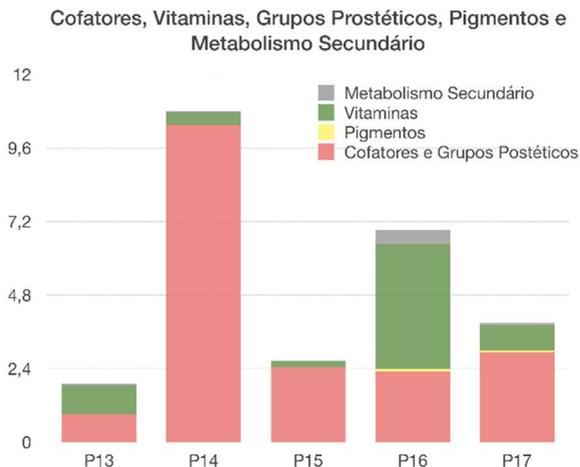


Figura 27. Abundância relativa das diferentes subcategorias relacionadas ao metabolismo de cofatores, vitaminas, grupos prostéticos e pigmentos (derivadas do *Subsystem* “*Cofactors, Vitamins, Prosthetic Groups, Pigments*”), e compostos do metabolismo secundário, derivados do *Subsystem* “*Secondary Metabolism*”.

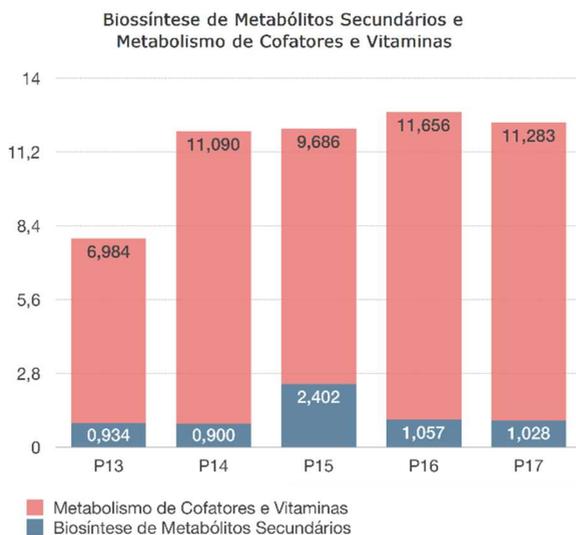


Figura 28. Abundância relativa de rotas metabólicas relacionadas ao metabolismo secundário e metabolismo de cofatores e vitaminas, derivadas dos *KEGG Pathways* “*Secondary Metabolism*” e “*Metabolism of Cofactors and Vitamins*”.

4.3.2 Membrana Plasmática

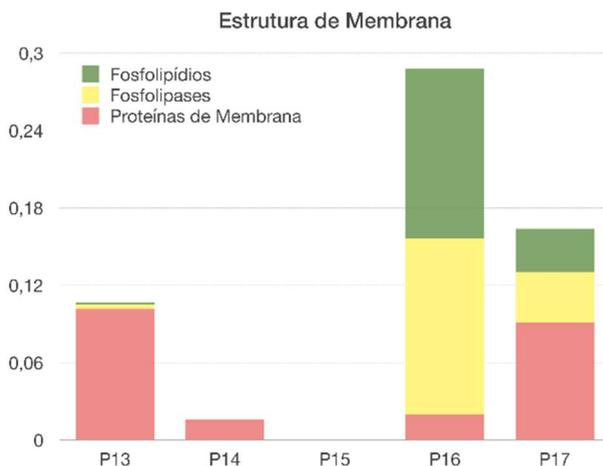


Figura 29. Abundância relativa de OFUs relacionadas ao metabolismo de fosfolipídios e fosfolipases (identificadas a partir do *Subsystem* “*Fatty Acids, Lipids, Isoprenoids*”) e proteínas de membrana (originalmente classificadas no *Subsystem* “*Protein Metabolism*”).

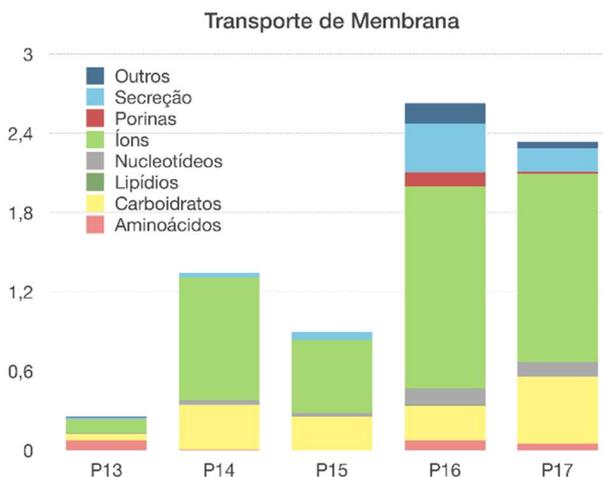


Figura 30. Abundância relativa das diferentes subcategorias de transporte de membrana, derivadas do *Subsystem* “*Membrane Transport*”. A subcategoria “Outros” inclui transportadores de membrana não identificados.

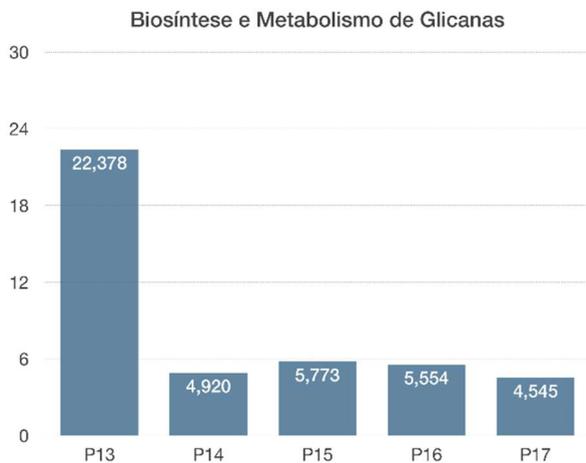


Figura 31. Abundância relativa de rotas metabólicas relacionadas à biossíntese e metabolismo de glicanas, derivadas do KEGG Pathway “*Glycan Biosynthesis and Metabolism*”.

4.3.3 Metabolismo Energético

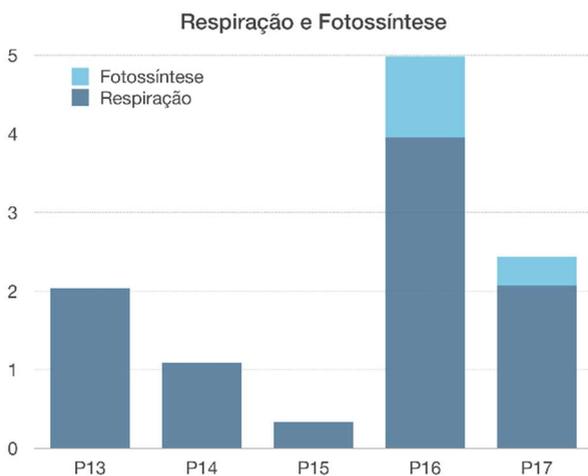


Figura 32. Abundância relativa de OFUs relacionadas à fotossíntese e respiração. A subcategoria “Respiração” inclui os *Subsystems* “*Respiration*”, “*Iron Acquisition and Metabolism*”, “*Sulfur Metabolism*” e as OFUs relacionadas à respiração de nitrato e nitrito do *Subsystem* “*Nitrogen Metabolism*”.

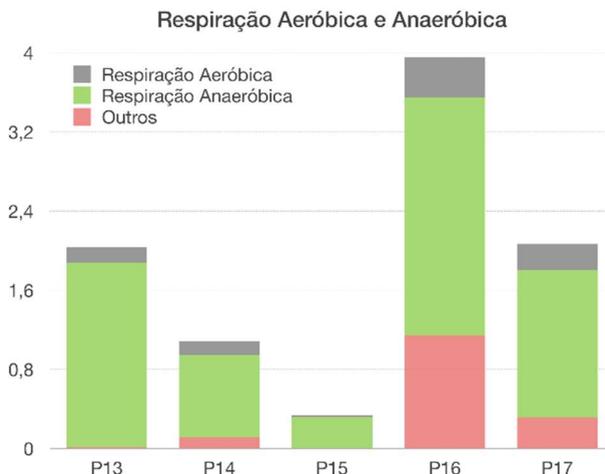


Figura 33. Abundância relativa de OFUs relacionadas à respiração aeróbica e anaeróbica. A subcategoria “Respiração Anaeróbica” inclui os Subsystems “*Iron Acquisition and Metabolism*”, “*Sulfur Metabolism*”, as OFUs relacionadas à respiração de nitrato e nitrito do Subsystem “*Nitrogen Metabolism*” e as OFUs relacionadas à anaerobiose do Subsystem “*Respiration*”. A subcategoria “Outros” inclui OFUs relacionadas aos processos respiratórios de maneira geral.

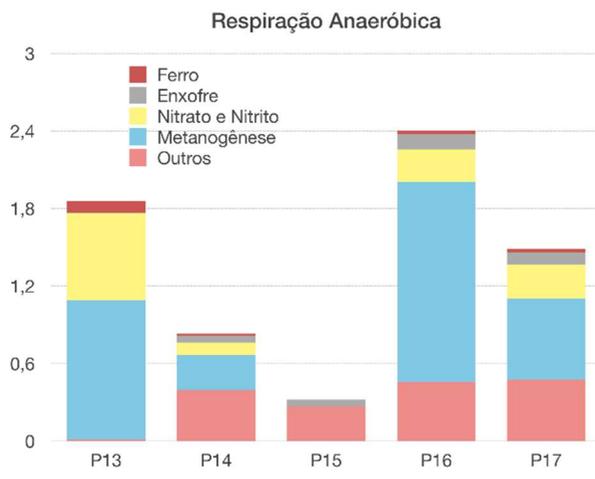


Figura 34. Abundância relativa de classes respiratórias identificadas na subcategoria “Respiração Anaeróbica”. A subcategoria “Outros” inclui OFUs relacionadas à anaerobiose de maneira geral.

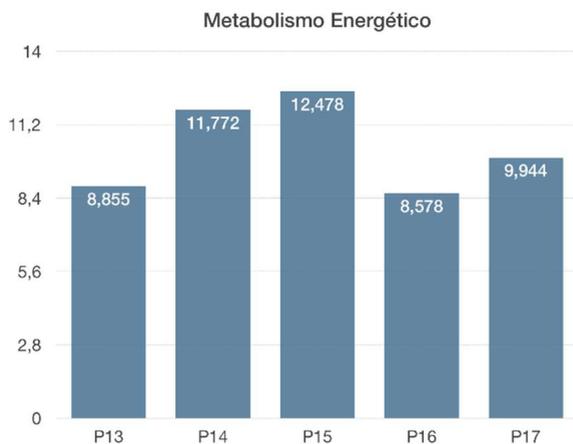


Figura 35. Abundância relativa de rotas metabólicas relacionadas ao metabolismo energético, derivadas do KEGG Pathway “Energy Metabolism”.

4.3.4 Metabolismo de Ácidos Nucleicos

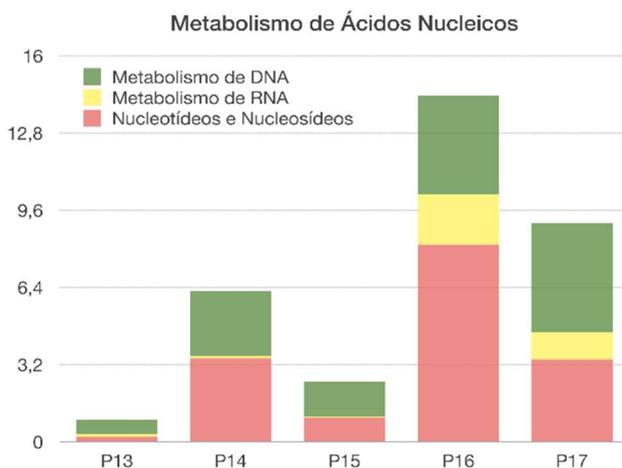


Figura 36. Abundância relativa das subcategorias relacionadas ao metabolismo de ácidos nucleicos, derivadas dos Subsystems “DNA Metabolism”, “RNA Metabolism” e “Nucleotides and Nucleosides”.

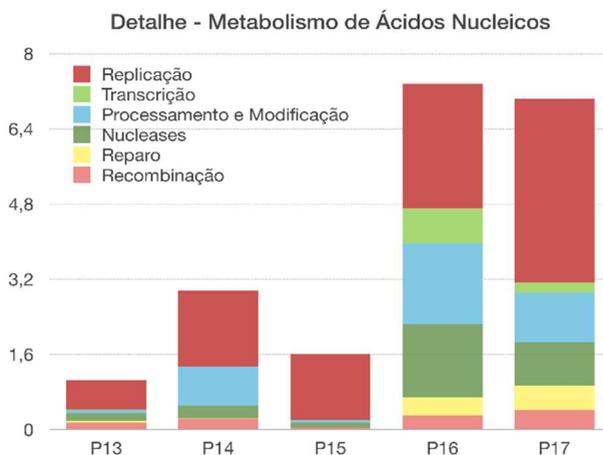


Figura 37. Detalhamento das subcategorias relacionadas ao metabolismo de ácidos nucleicos. A subcategoria “Replicação” inclui algumas OFUs do *Subsystem* “*Cell Division and Cell Cycle*”; a subcategoria “Processamento e Modificação” inclui helicases, topoisomerases, proteínas relacionadas à adição de grupos às cadeias nucleotídicas (metilações, glicosilações, fosforilações, etc) e *capping* de RNAm; a categoria “Nucleases” inclui tanto DNA nucleases quanto ribonucleases; a subcategoria “Transcrição” não inclui fatores de transcrição, considerados no gráfico “Regulação da Transcrição”.

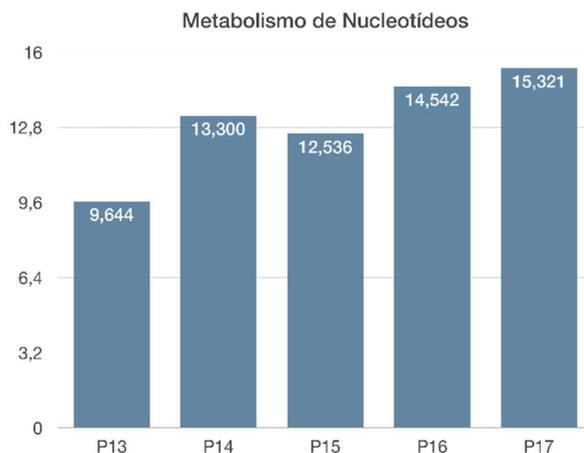


Figura 38. Abundância relativa de rotas metabólicas relacionadas ao metabolismo de ácidos nucleicos, derivadas do KEGG *Pathway* “*Nucleotide Metabolism*”.

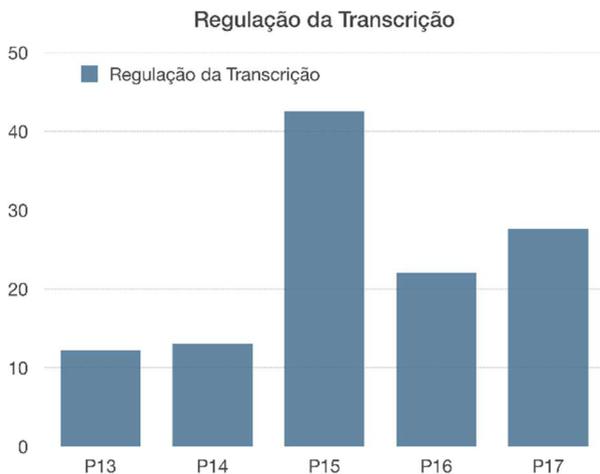


Figura 39. Abundância relativa de OFUs relacionadas à regulação da transcrição, pertencentes ao *Subsystem* “*Regulation and Cell Signaling*”.

4.3.5 Regulação e Sinalização Celular

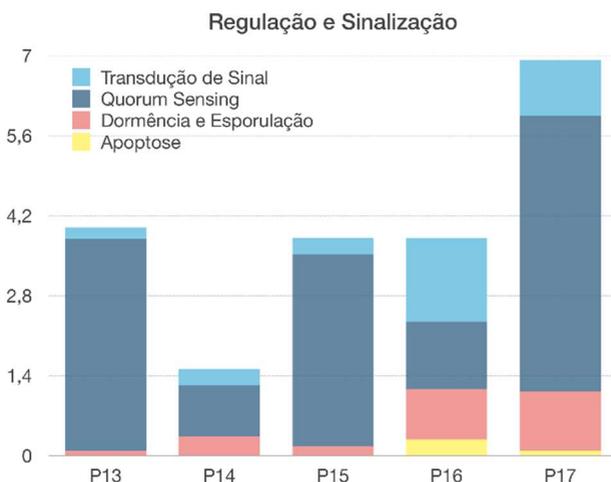


Figura 40. Abundância relativa de OFUs relacionadas à regulação e sinalização celular, identificadas a partir do *Subsystem* “*Regulation and Cell Signaling*”. As OFUs referentes ao *Subsystem* “*Dormancy and Sporulation*” foram incluídas nesta visualização por se tratarem de proteínas regulatórias destes processos.

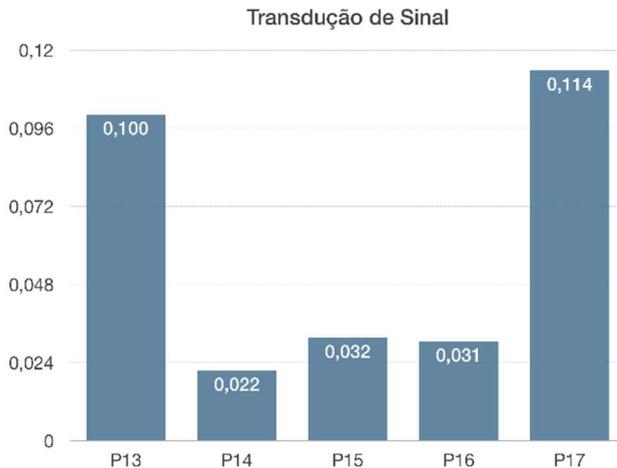


Figura 41. Abundância relativa de rotas metabólicas relacionadas à transdução de sinal, derivadas do KEGG Pathway “*Signal Transduction*”.

4.3.6 Promotores de Diversidade Genética

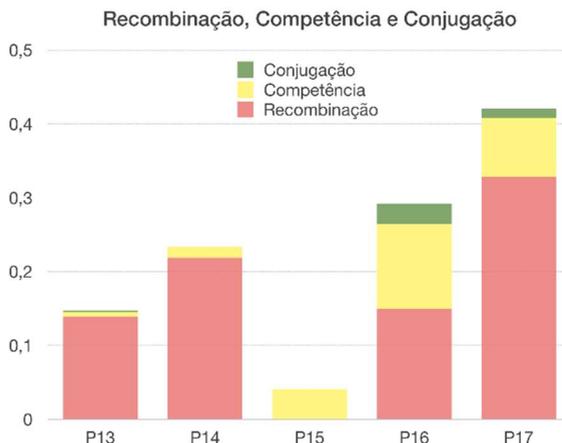


Figura 42. Abundância relativa de OFUs relacionadas aos processos de recombinação, competência e conjugação. Estas categorias reúnem OFUs identificadas nos *Subsystems* “*DNA Metabolism*”, “*Regulation and Cell Signaling*” e “*Phages, Prophages, Transposable Elements, Plasmids*”.

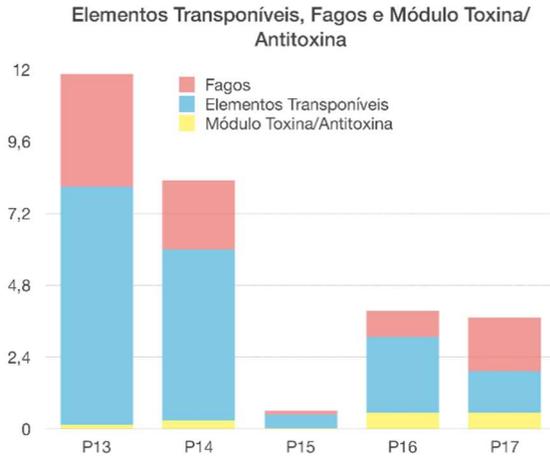


Figura 43. Diversidade de OFUs relacionadas ao processamento de elementos transponíveis e DNA de fagos (derivadas do *Subsystem* “*Phages, Prophages, Transposable Elements, Plasmids*”). A categoria “Módulo Toxina-Antitoxina” foi inicialmente classificada no *Subsystem* “*Regulation and Cell Signaling*”, mas foi incluída nesta visualização uma vez que estes sistemas são adquiridos através da transferência de plasmídios.

4.3.7 Resistência, Stress e Defesa

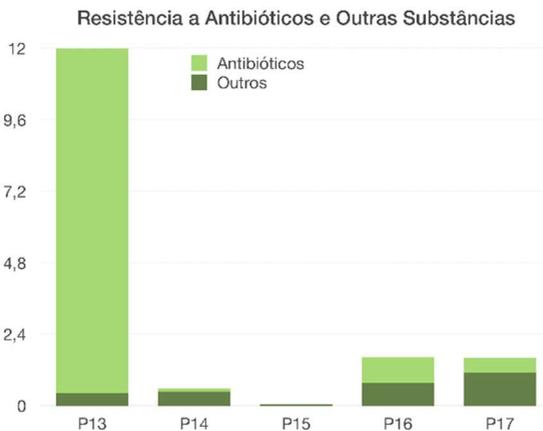


Figura 44. Abundância relativa de OFUs relacionadas à resistência a antibióticos e outras substâncias (identificadas a partir dos *Subsystems* “*Virulence, Disease and Defense*” e “*Stress Response*”).

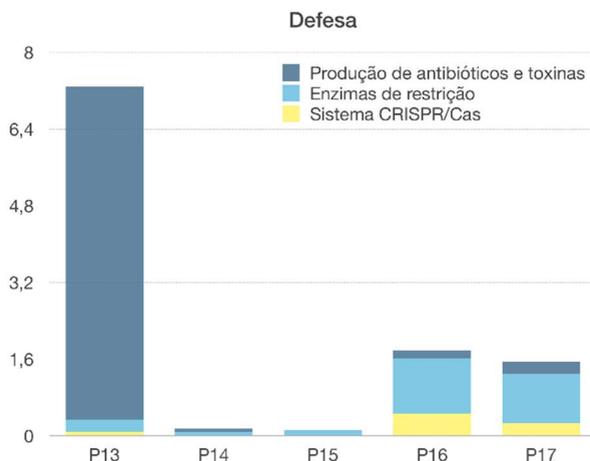


Figura 45. Abundância relativa de OFUs relacionadas à produção de antibióticos e toxinas e sistemas de defesa (identificadas a partir do *Subsystem* “*Virulence, Disease and Defense*”).

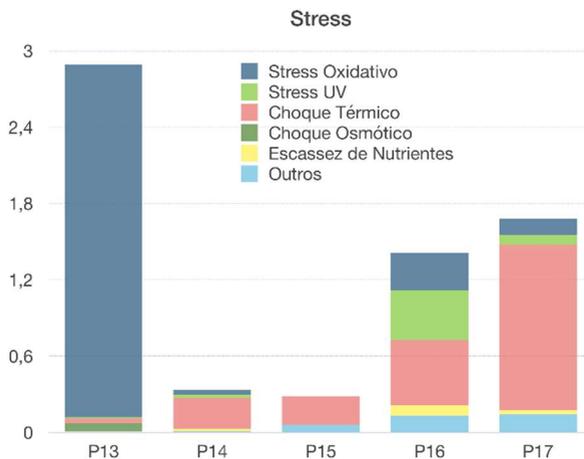


Figura 46. Abundância relativa de OFUs relacionadas à resposta ao stress (identificadas a partir do *Subsystem* “*Stress Response*”).

4.4 ESTATÍSTICA MULTIVARIADA

A análise de componentes principais (PCA) foi aplicada sobre os valores de abundância relativa referentes às categorias do SEED *Subsystems*, a fim de ordenar as amostras com base em sua diversidade funcional. Além disso, os parâmetros físico-químicos do *permafrost* (Tabela 4.4.1) foram utilizados como variáveis explanatórias, isto é, a influência de cada parâmetro foi calculada e correlacionada com as variações de abundância obtidas para cada amostra (Figura 47). A componente principal 1 (PC1, eixo X) contribuiu com 53,0% da variação encontrada entre as amostras, enquanto a componente principal 2 (PC2, eixo Y) contribuiu com 23,8%, totalizando 76,8% de explanação da diversidade. A PCA indicou que a diversidade funcional dos *permafrosts* foi agrupada em três partes: a) amostra P13; b) amostras P14 e P15; c) amostras P16 e P17. Este agrupamento foi observado apenas com os dados de abundância relativa de categorias funcionais, sendo a influência dos parâmetros físico-químicos calculada independentemente e seus vetores sobrepostos no mesmo gráfico.

Tabela 10. Dados físico-químicos (metadados) obtidos para os estratos de *permafrost*. (Fonte: ABRAMOV et al. 2011).

Amostra	P13	P14	P15	P16	P17
Profundidade	4,6 m	5,8 m	6,9 m	8,1 m	9,2 m
Idade	2,250	3,100	4,300	5,600	7,500
Cálcio	0,1	0,1	0,075	0,2	0,2
Cloro	0,175	0,175	0,175	0,175	0,175
Magnésio	0,2	0,2	0,2	0,5	0,5
Metano	1,386	3,757	5,675	4,356	1,742
Células/g	295.594.618	257.830.584	96.037.212	107.490.046	196.995.109
pH	8,95	9,56	9,82	9,39	9,48
Potássio	0,218	0,1	0,1	0,146	0,1345
Sódio	0,695	1,76	1,956	3,326	3,293
Sulfato	1,146	1,898	2,018	4,962	4,449
Carbono total	0,288	0,383	0,337	0,512	0,64
Carbono orgânico	0,111	0,131	0,163	0,266	0,263
Água	6,274	31,368	20,473	24,157	27,157

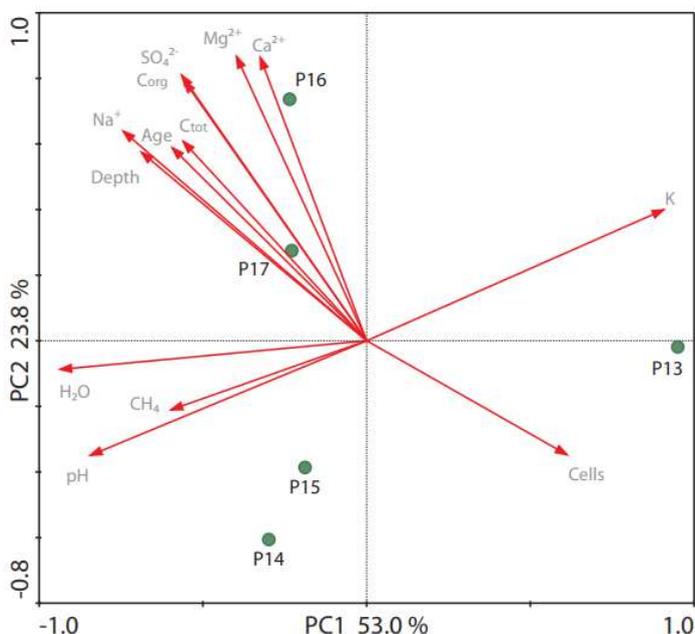


Figura 47. Análise de Componentes Principais (PCA) entre os dados de abundância relativa de categorias funcionais do SEED Subsystems e os parâmetros físico-químicos das amostras de *permafrost*. As duas componentes de maior explanação (PC1 e PC2) foram utilizadas para construir a figura. As amostras de *permafrost* estão representadas por círculos verdes, enquanto as variáveis físico-químicas estão apresentadas como vetores (Age = idade de congelamento; Depth = profundidade do *permafrost*; Cells = número total de células; H₂O = porcentagem de água congelada; CH₄ = gás metano; pH = potencial hidrogeniônico; Ctot = carbono total; Corg = carbono orgânico; SO₄²⁻ = íon sulfato; Cl⁻ = íon cloro; K⁺ = íon potássio; Ca²⁺ = íon cálcio; Mg²⁺ = íon magnésio; Na⁺ = íon sódio).

Os parâmetros que mais influenciaram a separação das amostras nestes três grupos foram o total de células e íon K⁺ para a amostra P13, gás CH₄ e pH para as amostras P14 e P15, e as demais variáveis correlacionadas às amostras P16 e P17. Estes resultados indicam que as amostras de *permafrost* diferenciam-se na passagem do período frio para o período quente na Antártica com base na diversidade funcional, sendo os parâmetros gás CH₄ e pH as variáveis que mais influenciaram esta transição. Cabe ressaltar que a análise de PCA não indica se o aumento da quantidade de metano ou da alcalinidade do meio é uma causa ou uma consequência da transição de período frio para quente.

4.5 TAXONOMIA

Os resultados obtidos a partir do pipeline automático do MG-RAST foram utilizados para avaliar a distribuição taxonômica dos filios de Bacteria (Figura 48) e Archaea (Figuras 49) entre as amostras de sedimento marinho (B2) e *permafrost* (P13 a P17). As classificações foram realizadas utilizando a base de dados RefSeq e o algoritmo “contigLCA”, que busca identidades taxonômicas para todas as seqüências funcionais presentes no metagenoma.

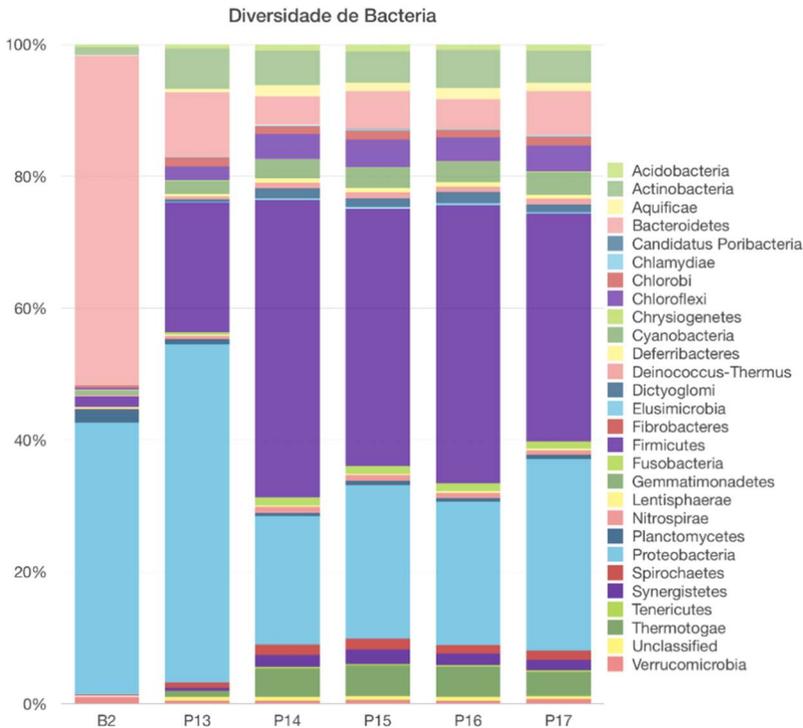


Figura 48. Distribuição taxonômica de filios de Bacteria entre as amostras, de acordo com os resultados obtidos a partir do pipeline automático do MG-RAST.

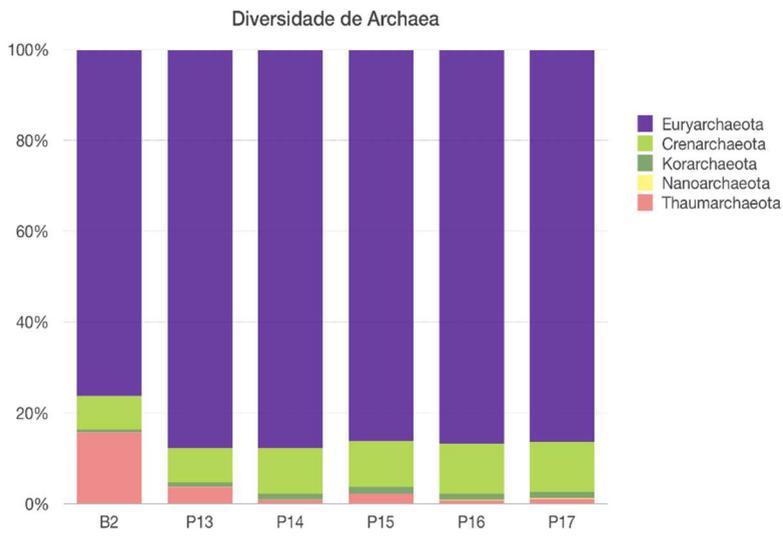


Figura 49. Distribuição taxonômica de filós de Archaea entre as amostras, de acordo com os resultados obtidos a partir do pipeline automático do MG-RAST.

5 DISCUSSÃO

5.1 DISCUSSÃO METODOLÓGICA

Embora diversas ferramentas estejam atualmente disponíveis para a análise de metagenomas, trabalhar com as sequências do *permafrost* demonstrou-se uma tarefa particularmente desafiadora. Os micro-organismos de ambientes extremos ainda são pouquíssimo representados nas bases de dados disponíveis, de modo que as ferramentas de anotação possuem um grande viés para ambientes mais bem conhecidos, como o microbioma humano, por exemplo. A classificação taxonômica das sequências só foi possível através do *pipeline* automático do MG-RAST, que apresentou resultados insatisfatórios com relação ao número de sequências mantidas pelo filtro de qualidade, limitando a análise. A classificação das sequências em categorias funcionais também apresentou diversas dificuldades na visualização e interpretação dos dados, permanecendo limitada à ferramenta que forneceu um maior número de anotações, não necessariamente estruturadas de uma forma elucidativa.

A base de dados SEED *Subsystems* possui um bom sistema de categorização, mas sua utilização é limitada ao *pipeline* automático do MG-RAST, que não apresentou uma boa performance sobre as sequências do *permafrost*; a submissão de *contigs* já montados maximizou bastante o número de sequências anotadas, mas as informações quantitativas (*read mapping*) não puderam ser associadas às anotações. O software MEGAN6 foi testado apenas para a visualização dos dados, uma vez que o processo de classificação depende de arquivos de alinhamento BLAST. O classificador *online* eggNOG-*mapper* foi capaz de classificar boa parte das sequências submetidas, mas o sistema de identificadores próprio da base de dados impossibilitou a visualização dos dados através do MEGAN6. As anotações obtidas pelo InterProScan tampouco puderam ser visualizadas, pois o reconhecimento de identificadores da base de dados Pfam só está disponível na versão paga do *software*. Boa parte das sequências anotadas também foram associadas a identificadores do InterPro, mas o MEGAN6 reconhece apenas aqueles referentes a famílias, de modo que as sequências anotadas a nível de domínio, por exemplo, não foram incluídas na visualização. Outras bases de dados que fornecem anotações a nível metabólico (KEGG *Pathways*, Metacyc, Reactome, *Gene Ontology*) produziram pouquíssimos *hits*, tanto através do InterProScan quanto do eggNOG-*mapper*.

O processo de classificação manual das sequências em grupos funcionais limitou muito a capacidade de análise das amostras como um todo. Um método de seleção das porções mais informativas dos metagenomas teve de ser definido para tornar este estudo possível, de modo que uma quantidade significativa das amostras não pôde ser avaliada. O quanto estas sequências não selecionadas poderiam modificar os resultados obtidos é uma informação que não pode ser quantificada, a menos que todo o metagenoma fosse manualmente classificado. No entanto, é possível obter uma boa estimativa a partir do número de *reads* associados às sequências anotadas (Figura 15): aproximadamente metade dos *reads* de todos os metagenomas (exceto da P13) foram mapeados contra sequências que apresentaram pouca variação entre as amostras. Se estas sequências fossem incluídas, provavelmente mudariam muito pouco as proporções diferenciais observadas.

5.2 DISCUSSÃO BIOLÓGICA

O Ótimo Climático do Holoceno Médio na Península Antártica ocorreu entre 4.5 e 2.8 mil anos antes do presente, e as amostras P14 (3.1 mil anos) e P15 (4.3 mil anos) situam-se exatamente dentro deste período. Para auxiliar a discussão, a Figura 55 demonstra o posicionamento temporal das amostras de *permafrost* utilizadas neste trabalho com relação a alguns dos principais eventos dos últimos 21 mil anos na Península Antártica.

De modo geral, as amostras de *permafrost* apresentaram um perfil taxonômico bastante semelhante entre filões de Bacteria e Archaea, diferindo bastante da amostra de sedimento marinho atual (B2). As amostras P14 a P17 demonstraram a predominância de bactérias do filo Firmicutes, enquanto a amostra B2 foi praticamente dominada pelos filões Proteobacteria e Bacteroidetes. As proteobactérias são amplamente distribuídas e abundantes em praticamente todos os ambientes, de modo que sua predominância em todas as amostras era esperada. O filo Firmicutes possui uma grande diversidade de bactérias formadoras de esporos, e sua grande abundância nas amostras de *permafrost* mais antigas pode estar relacionada a uma maior capacidade de preservação de moléculas de DNA em células em estado de dormência. De fato, OFUs relacionadas aos processos de dormência e esporulação apresentaram-se mais abundantes nas amostras mais antigas (Figura 40).

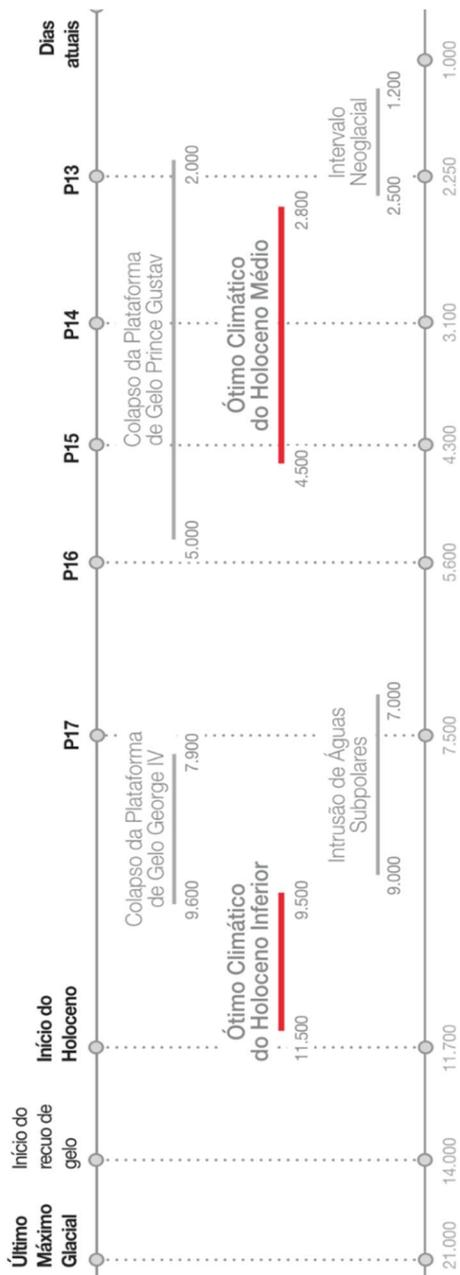


Figura 55. Principais eventos paleoclimáticos dos últimos 21 mil anos na Península Antártica; as linhas vermelhas representam o posicionamento temporal e a duração dos dois períodos de aquecimento registrados na história climática do Holoceno; as linhas em cinza indicam o posicionamento temporal e a duração de alguns eventos relacionados às alterações climáticas observadas na região da Península; os círculos em cinza indicam eventos pontuais e o posicionamento temporal de cada amostra de *permafrost* utilizada neste trabalho com relação aos eventos indicados. Para a construção da linha do tempo, foi utilizada uma escala logarítmica composta para facilitar a visualização dos períodos e eventos em escalas de tempos diferentes.

As amostras P14 e P15 demonstraram menor abundância de quase todas as categorias funcionais avaliadas entre as OFUs selecionadas como diferencialmente abundantes. Embora a camada mais profunda (P17) tenha apresentado uma variabilidade maior de OFUs com relação às amostras P16, P15 e P14, a maior parte dos *reads* do metagenoma P17 encontram-se distribuídos entre poucas sequências, que acabaram por se destacar entre as OFUs mais diferencialmente abundantes. As amostras P14 e P15, por sua vez, apresentaram uma menor variabilidade de OFUs, mas a maior parte dos *reads* parecem estar distribuídos entre sequências igualmente abundantes em todas as amostras, não tendo sido incluídas pela metodologia utilizada. Isto não quer dizer que as amostras P14 e P15, como um todo, possuam menor abundância das funções metabólicas avaliadas, mas sim que as sequências consideradas diferencialmente abundantes representam uma porcentagem pequena de seus metagenomas. Isto demonstra, de forma generalizada, que as amostras referentes ao período mais quente possuem um metabolismo mais bem distribuído e balanceado, sem a predominância de alguma função metabólica muito específica.

Diferenças na composição das comunidades microbianas já foram associadas à estratificação vertical do *permafrost*, demonstrando que as camadas mais superficiais comumente apresentam maior diversidade e abundância de micro-organismos devido à maior disponibilidade de nutrientes (YERGEAU et al., 2010; EILERS et al., 2012). Tripathi e colaboradores (2018), por exemplo, demonstraram que a abundância de grupos copiotróficos (i.e., adaptados a altas taxas de matéria orgânica) diminui com o aumento da profundidade do *permafrost*. Em contrapartida, os estratos intermediários de *permafrost* analisados no presente estudo (P14 e P15) demonstraram um perfil metabólico mais diversificado e relacionado à copiotrofia do que o estrato mais superficial (P13), contrariando os resultados dos estudos mencionados. Nenhum destes estudos, no entanto, associou o perfil das comunidades à idade de congelamento dos estratos, o que representa uma grande diferença para a interpretação de resultados no presente estudo.

Gittel e colaboradores (2014) observaram que micro-organismos degradadores de matéria orgânica estão mais associados às camadas imediatamente abaixo da superfície nos solos suscetíveis à crioturbação. Estes estratos apresentam uma grande concentração de matéria orgânica devido à infiltração proveniente do derretimento da superfície, representando um nicho bastante favorável para decompositores anaeróbicos facultativos. As amostras utilizadas no presente trabalho são referentes a solos congelados há, no mínimo, 2.225 anos, não

apresentando, portanto, processos de crioturbação atualmente. Porém, estes resultados sugerem que uma condição semelhante pode ter ocorrido durante o Ótimo Climático do Holoceno Médio, quando as amostras de profundidade intermediária (P14 e P15) estavam mais próximas da superfície. O aumento de temperatura teria intensificado o derretimento do conteúdo de gelo dos solos, aumentando a espessura da camada ativa e dando origem a um estrato altamente rico em matéria orgânica logo abaixo da superfície, representado pela amostra P15. Este ambiente teria favorecido populações de estratégia metabólica mais generalista e impactado negativamente aquelas altamente adaptadas à sobrevivência nos solos congelados.

Lauro e colaboradores (LAURO et al., 2009) demonstraram que as estratégias tróficas dos micro-organismos marinhos podem ser inferidas a partir de diversos elementos genômicos relacionados a funções distintas. Organismos copiotróficos, por exemplo, apresentam maiores quantidades de genes relacionados à regulação da transcrição, transporte e utilização de açúcares, mecanismos de defesa, motilidade celular e transdução de sinal. Os oligotróficos, por sua vez, apresentam maiores proporções de genes relacionados ao metabolismo de lipídios, compostos aromáticos e metabólitos secundários diversos. Alguns destes resultados foram consistentes com as diferenças de abundância observadas entre os metagenomas do *permafrost*. Embora os estudos não possam ser diretamente comparados (por se tratarem de ambientes muito distintos), a ideia de que é possível delinear diferenças de estratégia metabólica a partir de elementos genômicos foi massivamente aplicada na interpretação dos resultados deste trabalho.

Entre as poucas categorias funcionais que se destacaram nas amostras P14 e P15, foi observada uma maior abundância de OFUs relacionadas ao metabolismo de proteínas, principalmente proteases (Figura 19). As proteínas são uma fonte primária de nitrogênio orgânico para os micro-organismos, de modo que o aumento da capacidade de metabolizar estes compostos pode ter sido favorecido na comunidade como forma de utilizar os nutrientes disponíveis em abundância no ambiente. O aumento substancial de proteínas ribossomais na amostra P15 pode indicar um favorecimento da síntese proteica a partir da reciclagem de aminoácidos disponíveis no meio. O metabolismo de aminoácidos apresentou-se levemente mais abundante nas amostras P16 e P17, podendo significar uma maior necessidade de síntese *de novo* em um ambiente oligotrófico. O metabolismo de lipídios e compostos aromáticos (Figuras 23, 24 e 25) também é mais abundante nas amostras de clima mais frio, indicando uma tendência à utilização e

armazenamento de moléculas altamente energéticas, em detrimento de manter um repertório metabólico diversificado. Na amostra P13, a subcategoria “Modificação” corresponde a uma desaturase de ácidos graxos, responsável por adicionar ligações duplas a estes compostos. Micro-organismos adaptados ao frio apresentam um maior teor de ácidos graxos de cadeia curta e insaturados, que auxiliam na manutenção da fluidez de membrana uma vez que não tendem a endurecer sob baixas temperaturas. De fato, as amostras correspondentes ao período mais frio também apresentaram maior abundância de metabolismo de fosfolipídios (Figura 29), o que pode indicar uma maior necessidade de alterar a composição lipídica da membrana para se adequar às baixas temperaturas.

As OFUs relacionadas ao transporte de membrana (Figura 30) se destacaram nas amostras P16 e P17, tanto em termos de abundância relativa quanto em termos de diversidade. Resultados semelhantes foram observados por Sowell e colaboradores (2009) em um estudo conduzido no Mar dos Sargaços durante um mês de alta depleção de nutrientes. A prevalência de proteínas ligantes de substrato no metaproteoma da região indica que, em ambientes altamente oligotróficos, os micro-organismos tendem a otimizar sua capacidade de absorção de nutrientes, aumentando a produção de transportadores de membrana. A osmolaridade também representa um problema em ambientes com temperaturas abaixo de 0°C: devido à pouca disponibilidade de água líquida, as células microbianas tendem a perder água para o ambiente. Micro-organismos psicrófilos e psicrotolerantes apresentam diversas estratégias para manter o equilíbrio positivo de água, como o bombeamento de íons inorgânicos para o interior da célula e o acúmulo de solutos compatíveis (i.e., substâncias que podem ser acumuladas sem alterar significativamente os processos celulares). Além da pouca disponibilidade de nutrientes, a maior abundância de proteínas relacionadas ao transporte de membrana (principalmente íons), nas amostras P16 e P17 pode indicar um ambiente de pouca atividade de água ou alta concentração de sais. De fato, uma grande área das Sheltands do Sul estava coberta por água marinha até 7.5 mil anos antes do presente, devido à deglaciação durante o Ótimo Climático do Holoceno Inferior (Watcham et al, 2011). A partir dos 4.7 mil anos o nível do mar baixou permanentemente até os dias atuais, isolando o sedimento da influência de água marinha. Contudo, a região sofreu diversas alterações de nível do mar que influenciaram fortemente a salinidade dos lagos da região até pelo menos 4.750 anos (Figura 6). Estas alterações podem estar relacionadas a uma maior necessidade de

controle osmótico, refletida na comunidade microbiana pela diversidade e abundância de transportadores de membrana.

Embora uma maior disponibilidade de nutrientes seja geralmente associada a maiores taxas de crescimento e divisão celular, as amostras P16 e P17 demonstraram uma abundância significativamente maior de OFUs relacionadas aos processos gerais de metabolismo de ácidos nucleicos, tanto DNA quanto RNA (Figuras 36 e 37). É importante observar, no entanto, que as amostras P14 e P15 não necessariamente possuem uma menor quantidade de sequências relacionadas a estes processos, apenas que, em termos de abundância relativa, estas sequências não se destacam em seus metagenomas. Estes resultados podem indicar que as comunidades microbianas presentes nas amostras P16 e P17 tendem a dedicar uma porção maior de seus genomas aos processos de divisão celular, provavelmente como uma adaptação para manter o crescimento a temperaturas mais baixas. Em contrapartida, as OFUs referentes aos processos de regulação da transcrição estão muito mais presentes na amostra P15 (Figura 39), indicando que o aumento de temperatura a partir dos 4.5 mil anos pode ter favorecido microorganismos com maior capacidade regulatória para se adaptar às alterações do ambiente. Uma vez que a disponibilidade de nutrientes não é mais um problema, as comunidades microbianas das amostras P14 e P15 puderam maximizar seu repertório metabólico, mantendo em seus genomas mais sequências relacionadas ao metabolismo e reciclagem de diferentes biomoléculas.

No que diz respeito ao metabolismo energético, as amostras P16 e P17 demonstram maior abundância de OFUs relacionadas à fotossíntese (ausentes na porção observada das amostras P14 e P15) e à respiração aeróbica (pouquíssimo abundantes nas amostras P14 e P15) (Figuras 32 e 33). Mesmo entre as OFUs relacionadas à respiração anaeróbica, dominantes em todas as porções dos metagenomas avaliadas, a comunidade microbiana das amostras P16 e P17 apresentou maior abundância de OFUs relacionadas à metanogênese, à redução de nitrato e nitrito, ao metabolismo do enxofre e ao metabolismo do ferro (Figura 34). Novamente, cabe ressaltar que as OFUs selecionadas como diferencialmente abundantes entre as amostras representam apenas 1,5% dos *reads* da P15, contra 29% da P17, por exemplo, (Figura 16). Isto indica que a maior representatividade de diferentes classes energéticas nas amostras P16 e P17 não significa que estas funções estão ausentes na P15, mas podem encontrar-se distribuídas entre OFUs igualmente abundantes entre as todas as amostras. As categorias aparentemente mais abundantes nas amostras de clima frio na verdade representam

funções que *se destacam* na comunidade microbiana com relação às demais, em termos relativos, e não absolutos. Esta *aparente* diversidade parece ser mantida na comunidade por um equilíbrio bastante frágil entre as diferentes classes energéticas. O aumento da disponibilidade de matéria orgânica no ambiente pode ter deslocado este equilíbrio e favorecido micro-organismos de estratégia mais generalista, capazes de ocupar vários nichos ecológicos de acordo com a disponibilidade de nutrientes. Um possível aumento destas populações pode ter impactando negativamente as populações de metabolismo altamente específico, reduzindo sua *representatividade* na comunidade de forma que estas funções não *se destacam* no metagenoma da amostra P15. Uma possível explicação para que estes tipos de metabolismo não tenham se destacado entre as funções diferencialmente distribuídas é que, na amostra P15, parece haver uma maior *diversidade* destas populações, enquanto nas amostras P16 e P17, os poucos micro-organismos que realizam estas funções acabam se destacando em termos de *abundância relativa*.

A abundância de OFUs relacionadas ao stress oxidativo (Figura 46) chama atenção na amostra P13, embora também tenha aparecido, em menor proporção, nas amostras P16 e P17. Estes resultados, novamente, corroboram a hipótese de um ambiente rico em matéria orgânica ocasionado pelo aumento de temperatura durante o Ótimo Climático, provavelmente gerando um ambiente altamente anóxico e favorecendo processos quimiolitotróficos. Uma grande intensificação no metabolismo de carboidratos, principalmente relacionada à síntese (Figura 21), demonstram que uma grande disponibilidade de açúcares pode ter favorecido, novamente, os processos respiratórios aeróbicos. Com mais oxigênio na atmosfera, é possível que as comunidades anaeróbicas estabelecidas e dominantes nas amostras P14 e P15 tenham sido desfavorecidas, e que uma pressão seletiva tenha atuado no sentido da diversificação de resposta ao stress oxidativo.

Por fim, as correlações observadas entre as abundâncias nas categorias “Defesa”, “Resistência a Antibióticos e Outras Substâncias” e “Elementos Transponíveis e Fagos” podem ser interpretadas como uma extrapolação grosseira da história evolutiva dos micro-organismos do *permafrost*. A maior abundância de processos relacionados à defesa celular nas amostras P16 e P17 (Figura 45) pode indicar que seus genomas eram mais protegidos contra processos promotores de variabilidade, reflexo de um ambiente estável e taxas metabólicas lentas devido às baixas temperaturas. As perturbações ambientais decorrentes do aumento de temperatura ao longo dos 1.700 anos seguintes podem ter impulsionado processos de evolução rápida nas comunidades

microbianas; os reflexos desta transição podem ser observados nos grandes saltos nas categorias de “Resistência” (Figura 44), e “Produção” (Figura 45) de antibióticos e toxinas na amostra P13, sugerindo um ambiente onde a competição por recursos é bastante acirrada. É possível que o grande aumento e diversificação populacional durante o Ótimo Climático do Holoceno Médio não pôde mais se sustentar à medida que as condições climáticas foram se encaminhando para o breve intervalo neoglacial subsequente. A abundância de elementos transponíveis e DNA de fagos (Figura 43) é capaz de ilustrar a mesma história. Embora não tenha sido incluída na visualização, a amostra P13 apresentou 36,19% dos *reads* mapeados contra uma proteína FLP responsável por catalisar a recombinação entre repetições invertidas em plasmídeos.

De acordo com Madigan e colaboradores (2016):

Bactérias que estão em rápida mudança evolutiva contêm números relativamente elevados de elementos móveis, especialmente sequências de inserção, elementos transponíveis cujos genes codificam exclusivamente para transposição. A recombinação entre elementos idênticos gera rearranjos cromossômicos como deleções, inversões ou translocações. Acredita-se que tal processo represente uma fonte de diversidade genômica sobre a qual a seleção pode atuar. Assim, rearranjos cromossômicos que se acumulam em bactérias durante o crescimento em condições de estresse são frequentemente flanqueados por repetições ou sequências de inserção. Contrariamente, uma vez que uma espécie se estabeleça em um nicho evolutivo estável, a maioria dos elementos móveis é aparentemente perdida.

Uma vez que a amostra P13 corresponde à transição do Ótimo Climático do Holoceno Médio para um clima levemente mais frio, é possível que as comunidades microbianas representadas em seu metagenoma ainda estejam em processo de reconfiguração e restabelecimento após as intensas perturbações sofridas no período anterior.

6 CONCLUSÕES

O conteúdo dos metagenomas do *permafrost* utilizados neste trabalho demonstraram diversas alterações de perfil funcional entre amostras de diferentes idades, a partir das quais foi possível inferir como a comunidade microbiana foi impactada pelo Ótimo Climático do Holoceno Médio.

Os resultados obtidos sugerem que o aumento de temperatura ao longo do Ótimo Climático teria ocasionado o derretimento do conteúdo de gelo dos solos, intensificando os processos de crioturbação e gerando um ambiente altamente rico em matéria orgânica nos estratos mais próximos da camada ativa. Este processo teria favorecido populações de estratégia mais generalista, representadas nas amostras P14 e P15 por uma maior abundância de grupos funcionais relacionados à regulação do metabolismo. Categorias referentes ao metabolismo de proteínas e carboidratos também sustentam a hipótese do aumento de populações de repertório metabólico flexível. Estas teriam sido favorecidas pela abundância de nutrientes, impactando negativamente as comunidades altamente adaptadas à sobrevivência nos solos congelados, que demonstraram uma maior representatividade nos metagenomas P16 e P17.

As amostras referentes ao período anterior ao Ótimo Climático demonstraram um perfil funcional menos diverso e mais concentrado em poucas categorias metabólicas, que podem estar relacionadas a adaptações a um modo de vida tipicamente psicrófilico e/ou psicrotolerante. A maior abundância nas categorias referentes ao metabolismo de lipídios e compostos aromáticos e transporte de membrana nas amostras P16 e P17 indicam populações especializadas na aquisição e armazenamento de nutrientes altamente energéticos em um ambiente oligotrófico. A grande representatividade de produtos gênicos relacionados ao metabolismo de ácidos nucleicos demonstra que estes organismos são especializados à manutenção do crescimento e divisão celular em baixas temperaturas.

A maior abundância nas categorias relacionadas à resistência e produção de antibióticos, elementos transponíveis e fagos na amostra P13 indica um ambiente altamente competitivo, sugerindo que o aumento e diversificação populacional do período anterior não pôde mais ser sustentado à medida que as condições climáticas foram retornando a um ambiente polar típico após o Ótimo Climático do Holoceno Médio.

Estes resultados demonstram que a resposta das comunidades microbianas a alterações ambientais fica registrada em seu conteúdo metagenômico, podendo ser consultada e servir como indicativo de mudanças paleoclimáticas.

7 REFERÊNCIAS

- ABRAMOV, A. et al. Geocryological conditions of Antarctica. **Earth's Cryosphere**, v. 15, n. 3, p. 3-19, 2011.
- ABUBUCKER, S. et al. Metabolic reconstruction for metagenomic data and its application to the human microbiome. **PLoS Computational Biology**, v. 8, n. 6, p. e1002358, 2012.
- ALTSCHUL, S. F. et al. Basic local alignment search tool. **Journal of Molecular Biology**, v. 215, n. 3, p. 403–410, 1990.
- ANDERSEN, D. T.; Cold springs in permafrost on Earth and Mars. **Journal of Geophysical Research**, v. 107, n. E3, p. 1–7, 2002.
- ASHBURNER, M. et al. Gene Ontology: tool for the unification of biology. **Nature Genetics**, v. 25, n. 1, p. 25–29, 2000.
- BAKERMANS, C. et al. Reproduction and metabolism at - 10 degrees C of bacteria isolated from Siberian permafrost. **Environmental Microbiology**, v. 5, n. 4, p. 321–326, 2003.
- BARDGETT, R. D.; FREEMAN, C.; OSTLE, N. J. Microbial contributions to climate change through carbon cycle feedbacks. **The ISME journal**, v. 2, n. 8, p. 805–814, 2008.
- BATEMAN, A. et al. UniProt: the universal protein knowledgebase. **Nucleic Acids Research**, v. 45, p. D158-D169, 2017.
- BENTLEY, M. J. et al. Mechanisms of Holocene palaeoenvironmental change in the Antarctic Peninsula region. **The Holocene**, v. 19, n. 1, p. 51–69, 2010.
- BOCKHEIM, J.G; MUNROE, J.S. Organic carbon pools and genesis of alpine soils with permafrost: a review. **Arctic, Antarctic, and Alpine Research**, Vol. 46, No. 4, pp. 987–1006, 2014.
- BOWMAN, J. S.; DUCKLOW, H. W. Microbial communities can be described by metabolic structure: a general framework and application to a seasonally variable, depth-stratified microbial community from the coastal west Antarctic Peninsula. **PLoS ONE**, v. 10, n. 8, p. 1–18, 2015.
- BUCHFINK, B.; XIE, C.; HUSON, D. H. Fast and sensitive protein alignment using DIAMOND. **Nature Methods**, v. 12, n. 1, p. 59–60, 2015.

CAVICCHIOLI, R. Microbial ecology of Antarctic aquatic systems. **Nature Reviews Microbiology**, v. 13, n. 11, p. 691–706, 2015.

CHEN, I.-M. A. et al. IMG/M: integrated genome and metagenome comparative data analysis system. **Nucleic Acids Research**, v. 45, p. D507–D516, 2017.

CHOWN, S. L. Antarctic marine biodiversity and deep-sea hydrothermal vents. **PLoS Biology**, v. 10, n. 1, p. E1001232, 2012.

CLARKE, A. et al. Climate change and the marine ecosystem of the western Antarctic Peninsula. **Philosophical Transactions of the Royal Society B**, v.362, p. 149–166, 2007.

COMMINS, J.; TOFT, C.; FARES, M. A. Computational biology methods and their application to the comparative genomics of endocellular symbiotic bacteria of insects. **Biological Procedures Online**, v. 11, n. 1, 2009.

COMPEAU, P. E. C.; PEVZNER, P. A.; TESLER, G. How to apply de-Bruijn graphs to genome assembly. **Nature Biotechnology**, v. 29, n. 11, p. 987–991, 2011.

CONVEY, P. et al. The spatial structure of Antarctic biodiversity. **Ecological Monographs**, v. 84, p. 203–244, 2014.

COOK, A. et al. Retreating glacier fronts on the Antarctic Peninsula over the past half-century. **Science**, v. 308, n. 5721, p. 541–544, 2005.

COOLEN, M. J. L.; ORSI, W. D. The transcriptional response of microbial communities in thawing Alaskan permafrost soils. **Frontiers in Microbiology**, v. 6, n. 197, 2015.

CREVECOEUR, S. et al. Diversity and potential activity of methanotrophs in high methane-emitting permafrost thaw ponds. **PLoS ONE**, v. 12, n. 11, p. e0188223, 2017.

CROFT, D. et al. The Reactome pathway knowledgebase. **Nucleic Acids Research**, v. 42, p. D472–D477, 2014.

DeSANTIS, T. Z. et al. Greengenes, a chimera-checked 16s rRNA gene database and workbench compatible with ARB. *Applied and Environmental Microbiology*, v. 72, n. 7, p. 5069–72, 2006.

- DUARTE, R. T. D. **Micro-organismos em ambientes criogênicos: gelo glacial, solos expostos por recuo de geleiras e permafrost polares**. 2010. 200 f. Tese de doutorado, Programa de Pós-Graduação Interunidades em Biotecnologia, Instituto Butantan, Universidade de São Paulo, São Paulo, SP, 2010.
- EDDY, S. R. Multiple alignment using hidden Markov models. **ISMB-95 Proceedings**, v. 3, p. 114-120, 1995.
- EILERS, K. G. et al. Digging deeper to find unique microbial communities: The strong effect of depth on the structure of bacterial and archaeal communities in soil. **Soil Biology and Biochemistry**, v. 50, p. 58–65, 2012.
- FINN, R. D. et al. The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Research*, v. 44, p. D279–D285, 2016.
- GILICHINSKY, D. A. et al. Microbial populations in antarctic permafrost: biodiversity, state, age, and implication for astrobiology. **Astrobiology**, v. 7, n. 2, p. 275–311, 2007.
- GILICHINSKY, D. A. et al. Long-term preservation of microbial ecosystems in permafrost. **Advances in space research**, v. 12, n. 4, p. 255–263, 1992.
- GITTEL, A. et al. Distinct microbial communities associated with buried soils in the Siberian tundra. **The ISME Journal**, v. 8, n. 4, p. 841–853, 2014.
- GOORDIAL, J. et al. Nearing the cold-arid limits of microbial life in permafrost of an upper dry valley, Antarctica. **The ISME Journal**, v. 10, n. 7, p. 1613–1624, 2016.
- GRZYMSKI, J. J. et al. A metagenomic assessment of winter and summer bacterioplankton from Antarctica Peninsula coastal surface waters. **The ISME Journal**, v. 6, n. 10, 2012.
- GUAN, X. et al. Soil bacterial communities shaped by geochemical factors and land use in a less-explored area, Tibetan Plateau. **BMC genomics**, v. 14, n. 1, 2013.
- HAFT, D. H.; SELENGUT, J. D.; WHITE, O. The TIGRFAMs database of protein families. *Nucleic Acids Research*, v. 31, n. 1, p. 371–3, 2003.

HUERTA-CEPAS, J. et al. Fast genome-wide functional annotation through orthology assignment by eggNOG-Mapper. **Molecular Biology and Evolution**, v. 34, n. 8, p. 2115–2122, 2017.

HUERTA-CEPAS, J. et al. eggNOG 4.5: a hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences. **Nucleic Acids Research**, v. 44, p. D286–D293, 2016.

HULTMAN, J. et al. Multi-omics of permafrost, active layer and thermokarst bog soil microbiomes. **Nature**, v. 521, n. 7551, p. 208–212, 2015.

HUNTER, S. et al. InterPro: the integrative protein signature database. **Nucleic Acids Research**, v. 37, p. D211-5, 2009.

HUSON, D. H. et al. MEGAN Community Edition – Interactive Exploration and Analysis of Large-Scale Microbiome Sequencing Data. **PLOS Computational Biology**, v. 12, n. 6, p. e1004957, 2016.

JANSSON, J. K.; TAŞ, N. The microbial ecology of permafrost. **Nature Reviews Microbiology**, v. 12, n. 6, p. 414–425, 2014.

JOHNSON, S. **Remote protein homology detection using Hidden Markov Models**. 2006. 92 f. PhD Dissertation, Program in Molecular Genetics, Division of Biology and Biomedical Sciences, Washington University, Saint Louis, Missouri, 2006.

JOHNSTON, E. R. et al. Metagenomics Reveals Pervasive Bacterial Populations and Reduced Community Diversity across the Alaska Tundra Ecosystem. **Frontiers in Microbiology**, v. 7, p. 1–16, 2016.

KANEHISA, M.; GOTO, S.; KAWASHIMA, S.; OKUNO, Y.; HATTORI, M. The KEGG resource for deciphering the genome. **Nucleic Acids Research**, v. 32, p. D277-80, 2004.

KANEVSKIY, M. Z. et al. Proceedings of the Twenty-first International Offshore and Polar Engineering Conference, Maui, Hawaii, USA, 2011. **Permafrost of Northern Alaska**. International Society of Offshore and Polar Engineers (ISOPE), p. 1179–1186.

KENT, W. J. BLAT – the BLAST-like alignment tool. **Genome research**, v. 12, n. 4, p. 656–64, abr. 2002.

- KÖHLER, P.; KNORR, G.; BARD, E. Permafrost thawing as a possible source of abrupt carbon release at the onset of the Bølling/Allerød. **Nature Communications**, v. 5, n. 5520, 2014.
- KRIEGER, C. J. et al. MetaCyc: a multiorganism database of metabolic pathways and enzymes. **Nucleic Acids Research**, v. 32, p. D438-42, 2004.
- KRIVUSHIN, K. et al. Two Metagenomes from Late Pleistocene Northeast Siberian. **Genome Announcements**, v. 3, n. 1, p. 4–5, 2015.
- KUNIN, V. et al. A bioinformatician's guide to metagenomics. **Microbiology and Molecular Biology Reviews**, v. 72, n. 4, p. 557–578, 2008.
- LANGMEAD, B.; SALZBERG, S. L. Fast gapped-read alignment with Bowtie 2. **Nature Methods**, v. 9, n. 4, p. 357–9, 2012.
- LAURO, F. M. et al. The genomic basis of trophic strategy in marine bacteria. **PNAS**, v. 106, n. 37, p. 15527–33, 2009.
- LEPS, J.; SMILAUER, P. **Multivariate Analysis of Ecological Data Using CANOCO 5**. 2nd Edition. New York: Cambridge University Press, 2003. 269 p.
- LI, H. et al. The Sequence Alignment/Map format and SAMtools. **Bioinformatics**, v. 25, n. 16, p. 2078–79, 2009.
- LIPSON, D. A. et al. Metagenomic Insights into Anaerobic Metabolism along an Arctic Peat Soil Profile. **PLoS ONE**, v. 8, n. 5, p. e64659, 2013.
- MA, J.; PRINCE, A.; AAGAARD, K. M. Use of whole genome shotgun metagenomics: A practical guide for the microbiome-minded physician scientist. **Seminars in Reproductive Medicine**, v. 32, p. 5–13, 2014.
- MACKELPRANG, R. et al. Metagenomic analysis of a permafrost microbial community reveals a rapid response to thaw. **Nature**, v. 480, n. 7377, p. 368–371, 2011.
- MADIGAN, M. T. **Microbiologia de Brock**, 14ª Edição. Porto Alegre: Artmed, 2016. 1160 p.
- MAIDAK, B. et al. The Ribosomal Database Project (RDP). **Nucleic Acids Research**, v. 24, n. 1, p. 82–85, 1996.

MAKHALANYANE, T. P.; VAN GOETHEM, M. W.; COWAN, D. A. Microbial diversity and functional capacity in polar soils. **Current Opinion in Biotechnology**, v. 38, p. 159–166, 2016.

MARTIN, J. et al. Optimizing read mapping to reference genomes to determine composition and species prevalence in microbial communities. **PLoS ONE**, v. 7, n. 6, 2012.

MCCALLEY, C. K. et al. Methane dynamics regulated by microbial community response to permafrost thaw. **Nature**, v. 514, n. 7523, p. 478–481, 2014.

MEYER, F. et al. The metagenomics RAST server – a public resource for the automatic phylogenetic and functional analysis of metagenomes. **BMC Bioinformatics**, v. 9, n. 1, p. 386, 2008.

MITCHELL, A. L. et al. EBI Metagenomics in 2017: enriching the analysis of microbial communities, from sequence reads to assemblies. **Nucleic Acids Research**, v. 46, p. D726–D735, 2018.

MONDAV, R. et al. Discovery of a novel methanogen prevalent in thawing permafrost. **Nature communications**, v. 5, n. 3212, 2014.

NCBI Resource Coordinators. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Research*, v. 41, p. D8–D10.

OULAS, A. et al. Metagenomics: tools and insights for analyzing next-generation sequencing data derived from biodiversity studies. **Bioinformatics and Biology Insights**, v. 9, p. 75–88, 2015.

OVERBEEK, R. et al. The Subsystems Approach to Genome Annotation and its Use in the Project to Annotate 1000 Genomes. **Nucleic Acids Research**, v. 33, n. 17, p. 5691–5702, 2005.

PEDERSEN, M. W. et al. Ancient and modern environmental DNA. **Philosophical Transactions B**, v. 370, p. 1–11, 2014.

PENG, Y. et al. IDBA-UD: A *de novo* assembler for single-cell and metagenomic sequencing data with highly uneven depth. **Bioinformatics**, v. 28, n. 11, p. 1420–1428, 2012.

- PRUESSE, E. et al. SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. **Nucleic Acids Research**, v. 35, n. 21, p. 7188–96, 2007.
- QUEVILLON, E. et al. InterProScan: protein domains identifier. **Nucleic Acids Research**, v. 33, p. W116-20, 2005.
- RHO, M.; TANG, H.; YE, Y. FragGeneScan: predicting genes in short and error-prone reads. **Nucleic Acids Research**, v. 38, n. 20, p. e191, 2010.
- RIVKINA, E. M.; FRIEDMANN, E. I.; MCKAY, C. P.; GILICHINSKY, D. A. Metabolic activity of permafrost bacteria below the freezing point. **Applied and Environmental Microbiology**, v. 66, n. 8, p. 3230–3, 2000.
- SABINE, C. L. et al. The oceanic sink for anthropogenic CO₂. **Science**, v. 305, n. 5682, p. 367–371, 2004.
- SCHLOSS, P. D. et al. Introducing mothur: Open-source, platform-independent, community-supported software for describing and comparing microbial communities. **Applied and Environmental Microbiology**, v. 75, n. 23, p. 7537–41, 2009.
- SCHUUR, E. A. G. et al. Climate change and the permafrost carbon feedback. **Nature**, v. 520, n. 7546, p. 171–179, 2015.
- SEEMANN, T. Prokka: rapid prokaryotic genome annotation. **Bioinformatics**, v. 30, n. 14, p. 2068–2069, 15 jul. 2014.
- SHI, T. et al. Characterization of viable bacteria from Siberian permafrost by 16s rDNA sequencing. **Microbial Ecology**, v. 33, n. 3, p. 169–179, 1 abr. 1997.
- SOWELL, S. M. et al. Transport functions dominate the SAR11 metaproteome at low-nutrient extremes in the Sargasso Sea. **ISME Journal**, v. 3, n. 1, p. 93-105, 2009.
- STEVEN, B. et al. Characterization of the microbial diversity in a permafrost sample from the Canadian high Arctic using culture-dependent and culture-independent methods. **FEMS Microbial Ecology**, v. 59, n. 2, p. 513–523, 2007.
- STEVEN, B.; POLLARD, W. H.; GREER, C. W.; WHYTE, L. G. Microbial diversity and activity through a permafrost/ground ice core

profile from the Canadian high Arctic. **Environmental Microbiology**, v. 10, n. 12, p. 3388–3403, 2008.

TAŞ, N. et al. Impact of fire on active layer and permafrost microbial communities and metagenomes in an upland Alaskan boreal forest. **The ISME Journal**, v. 8, n. 9, p. 1904–1919, 2014.

TRIPATHI, B. M. et al. Variations in bacterial and archaeal communities along depth profiles of Alaskan soil cores. **Scientific reports**, v. 8, n. 1, p. 504, 2018.

TURNER, J. et al. (Org). **Antarctic Climate Change and the Environment**. Scientific Committee on Antarctic Research, 2009. 555 p.

TVEIT, A. et al. Organic carbon transformations in high-Arctic peat soils: key functions and microorganisms. **The ISME Journal**, v. 7, n. 2, p. 299–311, 2013.

TVEIT, A. T. et al. Metabolic and trophic interactions modulate methane production by Arctic peat microbiota in response to warming. **PNAS**, v. 112, n. 19, p. E2507-16, 2015.

VAUGHAN, D. G. et al. Recent rapid regional climatewarming on the Antarctic Peninsula. **Climate Change**, v. 60, n. 3, p. 243–274, 2003.

VISHNIVETSKAYA, T.; et al. Low-temperature recovery strategies for the isolation of bacteria from ancient permafrost sediments. **Extremophiles**, v. 4, p. 165–173, 2000.

WATCHAM, E. P. et al. A new Holocene relative sea level curve for the South Shetland Islands, Antarctica. **Quaternary Science Reviews**, v. 30, n. 21–22, p. 3152–3170, 2011.

WILKINS, D. et al. Key microbial drivers in Antarctic aquatic environments. **FEMS Microbiology Reviews**, v. 37, p. 303–335, 2012.

WILLERSLEV, E. et al. Fifty thousand years of Arctic vegetation and megafaunal diet. **Nature**, v. 506, n. 7486, p. 47–51, 2014.

WILLERSLEV, E. et al. Long-term persistence of bacterial DNA. **Current Biology**, v. 14, n. 1, p. 13–14, 2004.

WILLIAMS, T. J. et al. A metaproteomic assessment of winter and summer bacterioplankton from Antarctic Peninsula coastal surface waters. **The ISME Journal**, v. 6, n. 10, p. 1883–1900, 2012.

WILKE, A. et al. **MG-RAST Manual for version 4, revision 1**. 129 f. Argonne National Library, University of Chicago. 2016.

WORSLEY, P. Ice-wedge growth and casting in a Late Pleistocene, periglacial, fluvial succession at Baston, Lincolnshire. **Mercian Geologist**, v. 18, n. 3, p. 159–170, 2014.

WYNN-WILLIAMS, D. D. Antarctic microbial diversity: the basis of polar ecosystem processes. **Biodiversity and Conservation**, v. 5, p. 1271–1293, 1996.

YARYGIN, K.; et al. Abundance profiling of specific gene groups using precomputed gut metagenomes yields novel biological hypotheses. **PLoS ONE**, p. 1–23, 2017.

YASHINA, S. et al. Regeneration of whole fertile plants from 30,000-y-old fruit tissue buried in Siberian permafrost. **PNAS**, v. 109, n. 10, p. 4008–4013, 2012.

YERGEAU, E. et al. The functional potential of high Arctic permafrost revealed by metagenomic sequencing, qPCR and microarray analyses. **The ISME Journal**, v. 4, n. 9, p. 1206–1214, 2010.

ZHANG, D. C. et al. Isolation and characterization of bacteria from ancient Siberian permafrost sediment. **Biology**, v. 2, n. 1, p. 85–106, 2013

8 APÊNDICES

8.1 APÊNDICE A – Programas, Comandos e Parâmetros

1) *Quality Trimming*

Aplicação de um filtro de qualidade baseado em score Phred sobre os arquivos de *paired end short reads*, provenientes da plataforma Illumina HiSeq. Cada amostra corresponde a dois arquivos, *forward* e *reverse*, e os passos descritos foram aplicados para ambos os arquivos.

Software: Mothur (Versão 1.35.1)

```
$ fastq.info(fastq=file.fastq)
```

Input: arquivos contendo *short reads* e os respectivos *scores* de qualidade (.fastq).

Output: arquivos contendo *short reads* (.fasta) e *quality scores* (.qual) separados.

```
$ trim.seqs(fasta=file.fasta, qfile=file.qual,
qaverage=20, maxambig=0, maxhomop=10)
```

Input: arquivos contendo *short reads* (.fasta) e *quality scores* (.qual)

Output: arquivos contendo *short reads* mantidos (.trim.fasta) e removidos (.scrap.fasta) de acordo com os parâmetros de qualidade definidos.

```
$ list.seqs(fasta=file.scrap.fasta)
```

Input: arquivos contendo *short reads* removidos pelo filtro de qualidade (.scrap.fasta)

Output: arquivos contendo uma lista de identificadores referentes aos *short reads* removidos pelo filtro de qualidade (scrap.accnos).

Software: Unix Shell (Bash Script)

```
$ cut -d \ -f1 input.accnos > output.accnos
```

Input: lista de identificadores dos *short reads* removidos (scrap.accnos)

Output: lista de identificadores dos *short reads* removidos sem os caracteres adicionais inseridos pelo comando anterior.

Software: Mothur (Versão 1.35.1)

```
$ remove.seqs (accnos=file1.scrap.accnos,
fasta=file2.trim.fasta)
```

```
$ remove.seqs (accnos=file2.scrap.accnos,
fasta=file1.trim.fasta)
```

Input: lista de identificadores dos *short reads* removidos do arquivo *forward* (scrap.accnos), arquivo contendo os *reads* mantidos no arquivo *reverse* (.trim.fasta), e vice-versa.

Output: arquivos contendo os *reads* mantidos no arquivo *reverse* sem os *reads* removidos no arquivo *forward* e vice-versa (.trim.pick.fasta)

```
$ summary.seqs (file1.trim.pick.fasta)
```

```
$ summary.seqs (file2.trim.pick.fasta)
```

Input: arquivos contendo os *reads* mantidos no arquivo *reverse* sem os *reads* removidos no arquivo *forward* e vice-versa (.trim.pick.fasta)

Output: informações sobre as sequências contidas nos arquivos. O número de sequências nos dois arquivos gerados deve ser a mesma.

2) Assembly**Software: IDBA_UD (Versão 1.1.1)**

```
$ fq2fa --merge --filter file1.trim.pick.fasta
file2.trim.pick.fasta output.fa
```

Input: arquivos *forward* e *reverse* referentes aos *reads* mantidos pelo filtro de qualidade (.trim.pick.fasta)

Output: arquivos *forward* e *reverse* reunidos em um único arquivo

```
$ idba_ud -r outputfq2fa.fa -step=5
--pre_correction --num_threads=x -o output.fa
```

Input: arquivos *forward* e *reverse* reunidos em um único arquivo

Output: arquivo contendo *contigs* montados a partir dos *short reads* (.fa)

Obs: parâmetro `--num_threads` opcional, ajustado para o número de processadores (x) disponíveis no momento da execução.

3) Gene Calling

Software: FragGeneScan (Versão 1.30)

```
$ ./run_FragGeneScan.pl -genome=file.fa
-out=./path/output -complete=0 -train=illumina10
-thread=x
```

Input: arquivo contendo *contigs* montados a partir dos *short reads* (.fa)

Output: arquivos contendo ORFs identificadas nos *contigs* montados

- file.out – coordenadas das ORFs (*strand, frame, start, end, size*)
- file.ffn – sequências de nucleotídeos das ORFs preditas
- file.faa – sequências de aminoácidos das ORFs preditas

4) Anotação

Software: InterProScan (Versão 5.29-68.0)

```
$ bash interproscan.sh -appl Pfam -iprlookup
-goterms -pa -i file.faa -f tsv
```

Input: sequências de aminoácidos das ORFs preditas (.faa)

Output: arquivo separado por tabulações (.tsv) contendo identificadores correspondentes às anotações nas bases de dados selecionadas para cada sequência do arquivo de *input*.

5) Read Mapping

Software: Bowtie2 (Versão 2.3.3)

```
$ bowtie2-build -f file.ffn file_index
> file_index.out.txt 2> file_index.err.txt
```

Input: sequências de nucleotídeos das ORFs preditas (.ffn)

Output: arquivo de índice das sequências de referência (ORFs preditas)

```
$ bowtie2 --very-sensitive-local -p x -f -x
file_index -1 file1 -2 file2 -S file_index.sam
2> file.bowtie2.err.txt
```

Obs: parâmetro *-p* opcional, ajustado para o número de processadores (x) disponíveis no momento da execução.

Input: arquivos *forward* e *reverse* referentes aos *reads* filtrados (.fasta) e arquivo de índice do genoma de referência (ORFs preditas)

Output: arquivo de mapeamento dos *reads* contra as sequências de referência (.sam)

```
$ samtools view -Su file_index.sam > file_index.bam
```

Input: arquivo de mapeamento dos *reads* contra as sequências de referência (.sam)

Output: arquivo binário correspondente ao mapeamento dos *reads* contra as sequências de referência (.bam)

```
$ samtools sort file_index.bam > file_index.sorted
```

Input: arquivo binário correspondente ao mapeamento dos *reads* contra as sequências de referência (.bam)

Output: arquivo binário ordenado (.sorted)

```
$ samtools index file_index.sorted
> file_index.sorted.bai
```

Input: arquivo binário ordenado (.sorted)

Output: arquivo de índice do arquivo binário ordenado (.sorted.bai)

```
$ samtools idxstats file_index.sorted | grep -v "*"
| cut -f 1 > file_names.txt
```

Input: arquivo binário ordenado (.sorted)

Output: arquivo correspondente aos identificadores das sequências de referência (.txt)

```
$ samtools idxstats file_index.sorted | grep -v "*"
| cut -f 3 > file_counts.txt
```

Input: arquivo binário ordenado (.sorted)

Output: arquivo correspondente ao número de *reads* mapeados contra cada sequência de referência (.txt)

Software: Unix Shell (Bash Script)

```
$ paste -d "\t " file_names.txt file_counts.txt
> read_mapping.txt
```

Input: arquivos correspondentes aos identificadores das sequências de referência e ao número de *reads* mapeados contra cada sequência de referência (.txt)

Output: arquivo de *read mapping* indicando o número de *reads* mapeados contra cada sequência de referência.

5) Abundância De OFUs

Software: Unix Shell (Bash Script)

```
$ cut -f 1,x file.tsv > file.txt
```

Input: arquivo correspondente às anotações (output do InterProScan)

Output: arquivo separado por tabulações contendo os identificadores correspondentes à anotação na base de dados desejada para cada sequência anotada (1 = coluna de nomes das sequências, x = coluna correspondente aos identificadores na base de dados desejada)

```
$ join <(sort file1.txt) <(sort read_mapping.txt)
> read_mapping_annotado.txt
```

Input: arquivo correspondente às anotações (output do InterProScan) e ao *read mapping* contra cada sequência de referência

Output: arquivo correspondente ao número de *reads* associado às sequências de referência anotadas de acordo com os identificadores na base de dados desejada.

```
$ awk '!a[$1$2]++' read_mapping_annotado.txt
> read_mapping_annotado_nr.txt
```

Input: *output* do passo anterior

Output: arquivo de *input* removendo as sequências de referência duplicadas que possuem a mesma anotação

```
$ awk '{seen[$2]+= $3}END{for (i in seen) print i,
seen[i]}' read_mapping_annotado_nr
> soma_read_mapping_annotado_nr
```

Input: *output* do passo anterior

Output: arquivo de *input* com a soma do número de *reads* mapeados contra sequências com a mesma anotação

6) Normalização

Software: Unix Shell (Bash Script)

```
$ cat file.ffn | awk '$0 ~ ">" {print c; c=0;printf
substr($0,2,100) "\t"; } $0 !~ ">" {c+=length($0);}
END { print c; }' > file_seq_length
```

Input: arquivo de output do FragGeneScan correspondente às sequências de nucleotídeos

Output: arquivo indicando o tamanho das sequências de referência (número de bases)

```
$ join <(sort soma_read_mapping_annotado_nr) <(sort
file_seq_length) > file_counts_length
```

Input: arquivo correspondente ao número de *reads* associado às sequências de referência anotadas e arquivo indicando o tamanho das sequências de referência

Output: arquivos de *input* acoplados em um único arquivo separado por tabulações

Software: Microsoft Excel 2013

Aplicação das fórmulas de normalização e renormalização sobre as células correspondentes no arquivo de *output* do passo anterior.

O arquivo resultante deve conter uma lista de identificadores referentes às sequências anotadas (Coluna 1) acompanhadas dos valores de abundância relativa normalizada de cada sequência (Coluna 2). Este arquivo deve ser salvo em formato .csv ou .txt, para que seja unido aos demais em uma única tabela de abundância no passo subsequente.

IMPORTANTE: *TODOS* os arquivos devem conter a *EXATA* mesma formatação de identificadores (sem caracteres especiais), separação de colunas (separados por vírgula OU tabulação) e numeração (decimais com ponto OU vírgula).

7) Tabelas De Abundância

Software: Unix Shell (Bash Script)

```
$ join -a 1 -a 2 -o auto
<(sort B2_relative_abundance)
<(sort P13_relative_abundance)
> B2_P13_relative_abundance

$ join -a 1 -a 2 -o auto
<(sort B2_P13_relative_abundance)
<(sort P14_relative_abundance)
> B2_P13_P14_relative_abundance
```

E assim sucessivamente, até que os arquivos referentes a todas as sequências estejam unidos em um mesmo arquivo.

8.2 APÊNDICE B – Resultados de Abundância Diferencial incluindo a amostra B2.

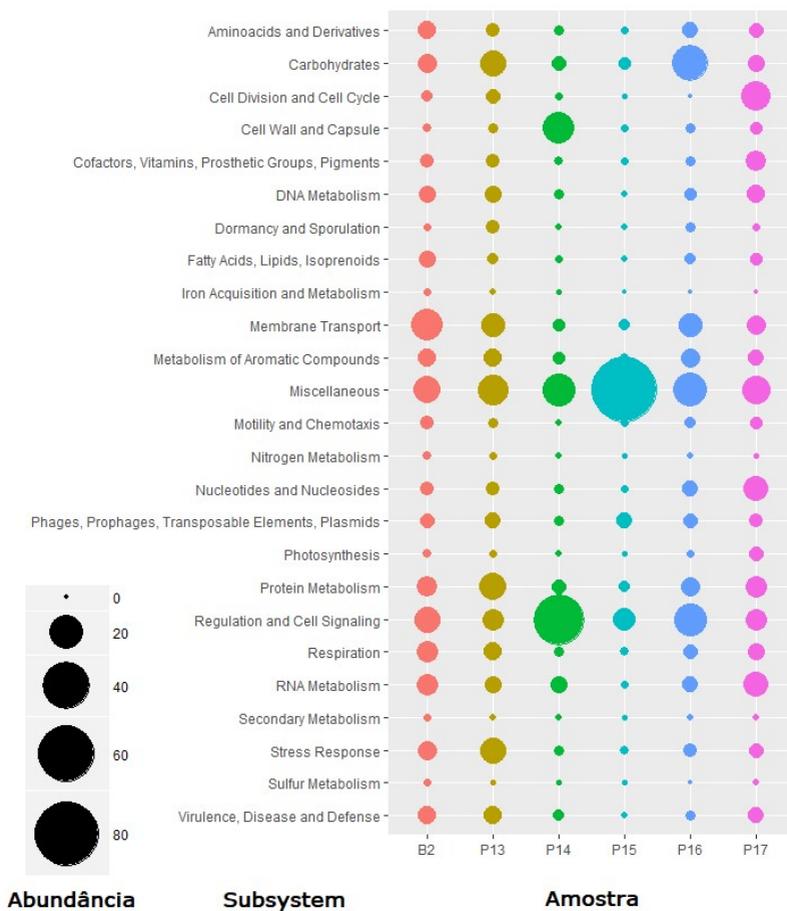


Figura B.1. Representatividade (em termos percentuais) das 25 categorias funcionais do SEED *Subsystems* em cada amostra do *permafrost*.

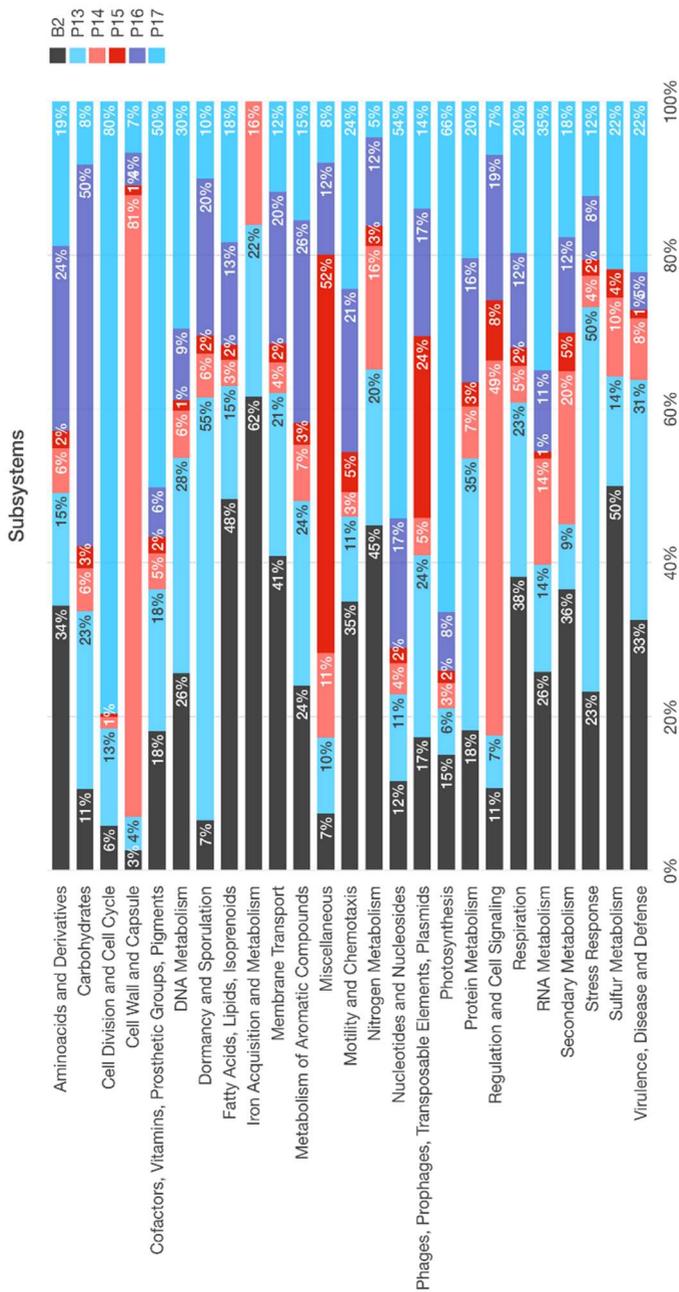


Figura B.2. Representatividade (em termos percentuais) das 25 categorias funcionais do SEED Subsystems entre as amostras de *permafrost*. Em tons de vermelho estão representadas as amostras correspondentes ao período do Ótimo Climático do Holoceno Médio (P14 e P15). As demais (P13, P16 e P17) estão representadas em tons de azul. A amostra B2 (sedimento marinho atual) está