

UNIVERSIDADE FEDERAL DE MINAS GERAIS
INSTITUTO DE CIÊNCIAS BIOLÓGICAS
DEPARTAMENTO DE BIOLOGIA GERAL
PROGRAMA DE PÓS-GRADUAÇÃO EM BIOINFORMÁTICA



Laura Rabelo Leite

Caracterização taxonômica e funcional de organismos e enzimas de interesse biotecnológico em ambientes de mineração

Tese apresentada ao Colegiado do Programa de Doutorado em Bioinformática do Instituto de Ciências Biológicas da Universidade Federal de Minas Gerais, como requisito parcial para a obtenção do grau de Doutor em Bioinformática.

ORIENTADOR: Guilherme Correa de Oliveira

COORIENTADOR: Gabriel da Rocha Fernandes



Epígrafe

“A tarefa não é tanto ver aquilo que ninguém viu, mas pensar o que ninguém ainda pensou sobre aquilo que todo mundo vê.”

(Arthur Schopenhauer)

Dedicatória

A Élida e Eudes, meus pais

A Natasha e Vitória, minhas irmãs

Ao Bruno, meu amor

À memória de Juraci, minha avó

À memória de Cilon e Samuel, amigos da eternidade

Agradecimentos

Ao Dr. Guilherme Correa de Oliveira e ao Dr. Gabriel da Rocha Fernandes pela paciência, orientação e confiança.

A Ms. Julliane Dutra, minha profunda gratidão e reconhecimento pela dedicação e verdadeiro trabalho em equipe.

A Dra. Sara Cuadros e ao Dr. Victor Pyrlo pelo apoio imprescindível.

Aos professores da banca por aceitarem e contribuírem para o aperfeiçoamento da pesquisa.

Aos amigos e colaboradores do Grupo de Informática de Biosistemas e Genômico e do Centro de Pesquisas René Rachou.

A todos os professores, colegas e funcionários da Pós Graduação em Bioinformática da Universidade Federal de Minas Gerais que colaboraram direta ou indiretamente com a minha formação, e para a realização deste trabalho.

A Vale pelo apoio financeiro.

A minha família, pelo apoio e incentivo. Sem vocês eu não escreveria sequer uma linha da minha história.

A todos os amigos, o meu muito obrigada.

Resumo

Ambientes contaminados com metais pesados apresentam uma comunidade extremamente complexa e bem adaptada. Esses microorganismos possuem enzimas com diversas aplicações biotecnológicas, como a biolixiviação e a biorremediação, além de desempenharem um papel essencial nos ciclos biogeoquímicos. Nesse contexto, a metagenômica permite avaliar o potencial genético de microorganismos cultiváveis e não cultiváveis. Aqui, os sequenciamentos *shotgun* e *amplicon* (16S rDNA) foram utilizados a fim de explorar a diversidade metabólica e funcional dos procariotos presentes em dois ambientes de mineração: uma barragem de rejeitos e uma drenagem ácida de mina. Foram analisadas amostras de água e sedimento nas estações seca (outubro de 2012, julho de 2014) e chuvosa (abril de 2013). O sedimento da barragem de rejeitos e a drenagem ácida de mina foram caracterizados como ambientes mais diversos quando comparados à água da barragem de rejeitos. Todavia, a distribuição de espécies de ambos os grupos se aproximam ao mesmo modelo de sucessão ecológica, no qual o estabelecimento das espécies depende de condições ambientais prévias. A anotação taxonômica das amostras de água da barragem de rejeitos apresentou predominância de sequências dos filos Proteobacteria, Actinobacteria e Cyanobacteria. A drenagem ácida de mina e o sedimento da barragem de rejeitos revelaram maior porcentagem de sequências relacionadas aos filos Firmicutes, Bacteroidetes, Proteobacteria, Verrucomicrobia, Acidobacteria e Planctomycetes. A análise funcional desses ambientes destacou a importância de vias metabólicas relacionadas a fixação de carbono, metabolismo de nitrogênio, oxidação de ferro ferroso e metabolismo de enxofre. Além disso, foram identificados diversos mecanismos adaptativos ligados à resistência a metais pesados, tolerância a solventes orgânicos, adaptação ao baixo pH e à oligotrofia. No presente estudo estabeleceu-se um novo fluxo de trabalho para análise de bibliotecas metagenômicas do tipo *shotgun*, que foi aplicado na construção de um catálogo inédito do potencial biotecnológico dos microorganismos associados a minas brasileiras de cobre, incluindo genes e rotas metabólicas de interesse industrial e ambiental.

Palavras chave: Biomineração, Metagenômica, Diversidade

Abstract

Environments contaminated with heavy metals contain an extremely complex and well adapted community. These microorganisms have enzymes with several biotechnological applications, such as bioremediation and bioleaching, and play an essential role in the biogeochemical cycles. In this context, metagenomics allows to evaluate the genetic potential of cultivable and non-cultivable microorganisms. Here, shotgun and amplicon sequencing (16S rDNA) were used to explore the metabolic and functional diversity of prokaryotes present in two mining environments: a tailings dam and an acid mine drainage. We analyzed water and sediment samples, in the dry (October 2012, July 2014) and wet (April 2013) seasons. The tailings dam and acid mine drainage sediment were characterized as a more diverse environment when compared to other compartments, such as the tailings dam water. However, the species distribution analysis using ecologic succession models demonstrated that in both groups the species establishment depends on previous environmental conditions. The taxonomic annotation in tailings dam water presented predominance of sequences of the phyla Proteobacteria, Actinobacteria and Cyanobacteria. Acid mine drainage and tailings dam sediment showed higher proportion of sequences related to the phyla Firmicutes, Bacteroidetes, Proteobacteria, Verrucomicrobia, Acidobacteria and Planctomycetes. Functional analysis of these environments highlighted the importance of metabolic pathways related to carbon fixation, nitrogen metabolism, ferrous iron oxidation and sulfur metabolism. In addition, were identified several adaptive mechanisms linked to resistance to heavy metals, tolerance to organic solvents, low pH adaptation and oligotrophy. The present study established a new workflow for metagenomic analysis of shotgun libraries, which was used to build an unprecedented catalog of the biotechnological potential of microorganisms in brazilian copper mines, including genes and metabolic pathways with industrial and environmental interest.

Keywords: Biomining, Metagenomics, Diversity

Sumário

Epígrafe.....	2
Dedicatória.....	3
Agradecimentos.....	4
Resumo.....	5
Abstract.....	6
Sumário.....	7
Lista de Tabelas.....	9
Lista de Figuras.....	10
Lista de abreviaturas e símbolos.....	11
1- Revisão da literatura.....	1
1.1 –Mineração.....	1
1.2 – Biodiversidade em ambientes de mineração.....	3
1.2.1 - Drenagem Ácida de Mina.....	3
1.2 .2 - Barragem de Rejeitos.....	5
1.3 – Potencial biotecnológico dos microrganismos em ambientes de mineração.....	7
1.3.1 – Biomineração.....	7
1.3.2 – Biorremediação.....	13
1.4 - Principais abordagens para o estudo de microbiomas.....	17
1.4.1 – Metagenômica.....	17
1.4.2 – Abordagens para o estudo de Genomas Individuais.....	23
2- Justificativa.....	25
3- Objetivos.....	26
3.1 - Objetivo Geral.....	26
3.2 - Objetivos específicos.....	26
4- Metodologia.....	27
4.1 - Áreas de estudo.....	27
4.2 - Coleta.....	27
4.3 - Extração e sequenciamento do DNA.....	28
4.4 – Análise de amplicons (16S rDNA).....	30
4.5 - Análise Shotgun.....	32
4.5 – Análises complementares.....	34
4.6 – Análise comparativa entre abordagens para classificação taxonômica usando bibliotecas <i>shotgun</i>	35
5- Resultados e Discussão.....	38

5.1 – Validação do pipeline para análise metagenômica de bibliotecas <i>shotgun</i> : comparação entre métodos de classificação taxonômica dependentes e independentes de marcadores moleculares	38
5.2 – Influência dos parâmetros ambientais e da sazonalidade na composição da comunidade microbiana da barragem de rejeitos e drenagem ácida de mina.....	41
5.3–Análise de alfa e beta diversidade em ambientes de mineração: buscando unidades taxonômicas operacionais importantes para as comunidades microbianas da barragem de rejeitos e drenagem ácida de mina	49
5.3.1 – Diferenças entre perfis taxonômicos preditos por análise de amplicons 16S rDNA e análise de bibliotecas <i>shotgun</i>	61
5.4–Potencial biotecnológico de genomas isolados a partir de amostras de drenagem ácida de mina.....	64
6- Conclusões e Perspectivas	80
7- Considerações finais	81
9– Anexos	86
8 - Referências.....	92

Lista de Tabelas

Tabela 1 - Resumo da composição taxonômica do metagenoma simulado.	36
Tabela 2 - Comparação entre a abundância relativa de espécies predita por abordagens dependentes e independentes de marcadores moleculares.....	39
Tabela 3 - Características físico-químicas das amostras de água da barragem de rejeitos e drenagem ácida de mina nos períodos de outubro de 2012, abril de 2013 e junho de 2014	44
Tabela 4 - Características físico-químicas das amostras de sedimento da barragem de rejeito e drenagem ácida de mina nos períodos de outubro de 2012, abril de 2013 e junho de 2014	46
Tabela 5 - Principais medidas de alfa diversidade observadas por amostra.....	52

Lista de Figuras

Figura 1- Ciclos do ferro, enxofre e ciclagem de carbono em drenagem ácida de mina.	4
Figura 2 - Modelo de oxidação de ferro caracterizado em <i>At. ferrooxidans</i>	11
Figura 3 - Modelo de oxidação de enxofre em <i>A. ferrooxidans</i> ATCC 23270	12
Figura 4 - Vias assimilatória e dissimilatória de redução de sulfato em procariotos	16
Figura 5 - Representação das áreas de coleta A) Barragem de Rejeitos, outubro de 2012 B) Drenagem ácida de mina, abril de 2013.....	27
Figura 6 - Representação gráfica da análise integrada entre dados de <i>amplicons</i> e <i>shotgun</i>	32
Figura 7 - Comparação entre os métodos dependente e independente de marcadores para classificação de sequências <i>shotgun</i>	40
Figura 8 - Análise de correspondência canônica (CCA) entre a abundância do perfil taxonômico ao nível de gênero e os parâmetros físico-químicos.	49
Figura 9 - Curva de rarefação para as amostras analisadas.	50
Figura 10 - Modelos de distribuição de táxons para amostras de água, baseado nas bibliotecas de 16S rDNA.	54
Figura 11 - Modelos de distribuição de táxons para amostras de sedimento, baseado nas bibliotecas de 16S rDNA.....	55
Figura 12– Comparação entre a abundância de OTUs usando DESeq2. Os círculos plotados representam as OTUs	59
Figura 13 - Gráfico de Venn contendo o nº de filós exclusivos e compartilhados entre todas as predições por análise de <i>amplicons</i> 16S rDNA e análise de bibliotecas <i>shotgun</i>	62
Figura 14 - Recrutamento de sequências <i>shotgun</i> em relação ao genoma de <i>Hydrothalea flava</i>	66
Figura 15 - Visão geral dos atributos funcionais anotados para os genomas A) da <i>Hydrothalea flava</i> B) da ITV01.....	68
Figura 16 - Resumo das vias mais abundantes identificadas nos metagenomas de barragem de rejeitos e drenagem ácida de mina.....	71
Figura 17 - Visão geral dos principais aspectos funcionais encontrados na barragem de rejeitos. A) fixação de carbono, B) oxidação de ferro, C) adaptação ambiental , D) metabolismo de enxofre.....	75
Figura 18 - Visão geral dos principais aspectos funcionais encontrados na drenagem ácida de mina.A) metabolismo de nitrogênio, B) oxidação de ferro, C) adaptação ambiental, D) metabolismo de enxofre	79

Lista de abreviaturas e símbolos

1M.OCT ou BAS_OUT	Água superficial da barragem de rejeitos, coleta de outubro de 2012
15M.OCT ou BAP_OUT	Água de profundidade da barragem de rejeitos, coleta de outubro de 2012
SED.OCT ou BSD_OUT	Sedimento da barragem de rejeitos, coleta de outubro de 2012
1M.APR ou BAS_ABR	Água superficial da barragem de rejeitos, coleta de abril de 2013
15M.APR ou BAP_ABR	Água de profundidade da barragem de rejeitos, coleta de abril de 2013
SED.APR ou BSD_ABR	Sedimento da barragem de rejeitos, coleta de abril de 2013
AMD.APR	Água de drenagem ácida de mina, coleta de abril de 2013
AMD.JUN ou DAM	Água de drenagem ácida de mina, coleta de junho de 2014
AMD.SED.JUN ou DAM_SED	Sedimento de drenagem acida de mina, coleta de junho de 2014
Ag	Prata
As	Arsênio
Au	Ouro
BIF	Banded iron formatios
C	Carbono
Cd	Cádmio
cm	Centímetro
Co	Cobalto
CO2	Dióxido de carbono
CONAMA	Conselho Nacional do Meio Ambiente
Cr	Cromo
Cu	Cobre
Cu	Cobre
DNA	Ácido desoxirribonucleico
Eh	Potencial oxidação-redução
EW	Electrowinning
F	Flúor
FACS	Flow-activated cell sorting
Fe	Ferro
Fe(II)	Ferro ferroso
Fe(III)	Ferro férrico
g	Gramas
Hg	Mercúrio

kDa	Kilodalton
KEGG	Kyoto Encyclopedia of Genes and Genomes
Km	Quilômetros
L	Litro
MDA	Multiple displacement amplification
mg	Miligrama
Mo	Molibdenita
Mt	Toneladas métricas
N	Nitrogênio
NGS	Next-generation sequencing
Ni	Níquel
Ni	Níquel
ORF	Open reading frame
P	Fósforo
pb	Pares de bases
Pb	Chumbo
PCR	Reação em Cadeia da Polimerase
pH	Potencial hidrogeniônico
qPCR	Quantitative real-time PCR
RNA	Ácido ribonucleico
rRNA	RNA ribossômico
SAGS	Isolamento aleatório de células únicas
SCG	Single-cell genomics
SX	Solvent Extraction
t	Tonelada
Th	Tório
U	Urânio
US\$	Dólar dos Estados Unidos
WGA	Whole genome amplification
Y	Ítrio
Zn	Zinco
µg	Micrograma

1- Revisão da literatura

1.1 –Mineração

A indústria de extração e transformação de minerais é um dos setores base da economia Brasileira. Em 2013, a produção mineral brasileira atingiu US\$ 77,5 bilhões o que correspondeu a 4,1% do produto interno bruto (PIB) ¹. Neste setor os estados do Pará e Minas Gerais se destacam pela presença de diversos depósitos minerais, cujos principais produtos são nióbio, ferro, alumínio, amianto e grafita natural¹. Além de ser um dos maiores exportadores mundiais de minério de ferro, o Brasil é também exportador líquido de praticamente todos os metais não ferrosos, com exceção do cobre. O Brasil é o 16º produtor de minério de cobre, com 214 mil toneladas em 2008. Entretanto, um importante volume é importado nas formas de minério concentrado, catodo, fios e cabos. Em 2010, as importações somaram US\$ 952 milhões e, em 2011, US\$ 1.141 milhões, representando 12,3% e 10,1%, respectivamente, do total das importações de bens minerais primários ².

As reservas nacionais de minério de cobre são predominantemente formadas por minerais sulfetados, com ouro e prata associados, e secundariamente por minerais oxidados. Essas reservas estão estimadas em cerca de 21,7 milhões de toneladas, e distribuem-se por nove estados: Alagoas, Bahia, Ceará, Goiás, Mato Grosso, Minas Gerais, Pará, Rio Grande do Sul e São Paulo. Entretanto, é no estado do Pará que se concentram mais de 85% dessas reservas. Destas, cerca de 97% da produção de concentrado de catodo de cobre provêm essencialmente do complexo mineiro de Sossego e Sequeirinho localizado na Província Mineral de Carajás, em Canaã dos Carajás, estado do Pará ³. Esse depósito situa-se perto do contato das rochas vulcânicas do Grupo Grão Pará, que é composto por rochas vulcanossedimentares. Seu minério apresenta anomalias geoquímicas de P e F (apatita), Mo (molibdenita), Y, U, Th, Co e Ni. As reservas desse depósito foram calculadas em 355 Mt de minério com 1,1% de Cu e 0,28 g/t de Au. A 30 km ao norte da Serra dos Carajás, o complexo mineiro de Salobo está

hospedado na sequência vulcanogênica Salobo, e é constituído por quartzitos, anfibolitos, metagrauvas, BIFs estéreis e magnetitos maciços ou foliados aos quais está associada à mineralização cupro-aurífera. Salobo contém reservas da ordem de 789 Mt de minério com 0,96% Cu e 0,52 g/t Au ⁴.

Até o presente momento a extração de cobre no Brasil ocorre por mineração tradicional. O processamento tradicional é composto pelas seguintes etapas: fragmentação primária, granulação ou britagem, moagem e concentração por flotação. O concentrado recolhido na superfície é desidratado, contendo cerca de 30% de cobre. Uma vez concentrado, o cobre composto pode ser convertido a cobre metálico por meio de duas rotas metalúrgicas: a pirometalúrgica, que inclui a fundição e o refinamento eletrolítico, em geral usado em minérios sulfetados; e a hidrometalúrgica, que inclui a lixiviação, extração por solvente (SX) e eletroextração (EW), comumente usada em minérios oxidados. Atualmente, nas minas de Sossego e Salobo, apenas o concentrado de cobre é produzido ⁵.

Se por um lado a demanda por metais é crescente, por outro, as mineradoras estão diante do esgotamento das reservas contendo teores superiores a 1% dos metais de interesse. A energia gasta na britagem e moagem são desperdiçadas com 99% de materiais que não tem valor comercial ⁶. Isso impõe a necessidade de extrair metal a partir de minérios de baixos teores e de rejeitos. Para tanto são necessários processos viáveis que exigem baixos custos de investimento e de operação. Além de um processo oneroso e pouco eficiente, a mineração tradicional ocasiona uma série de alterações ambientais como: poluição das águas subterrâneas e superficiais com rejeitos, poluição do ar com maior quantidade de estéril e poeira em suspensão, poluição sonora com aumento de vibrações e maiores riscos na geração de áreas degradadas. Uma alternativa é a rota biohidrometalúrgica, mais especificamente a biomineração, por meio da qual é possível extrair os metais com baixo custo operacional e menor impacto ambiental ³.

1.2 – Biodiversidade em ambientes de mineração

Bactérias e arqueias desempenham um papel essencial na composição e manutenção da biosfera. Esses microorganismos atuam na ciclagem biogeoquímica, na biotransformação de elementos orgânicos e inorgânicos, decomposição, biointemperismo, e deterioração mineral conduzindo a formação dos solos e sedimentos. Além disso, contribuem ativamente para fenômenos geológicos envolvendo a transformação de minerais, metais e metalóides⁷.

1.2.1 - Drenagem Ácida de Mina

Fluidos ácidos, ricos em metais são formados por intemperismo químico de rochas ricas em metais sulfetados. Estas soluções de drenagem ácida de rochas (ARD) são quentes porque as reações de oxidação dos sulfetos metálicos são altamente exotérmicas. A dissolução de pirita (FeS_2), arsenopirita (FeAsS), calcopirita (CuFeS_2), esfalerita (ZnS) e marcasita (FeS_2) produz soluções ricas em Au, Ag, Cu, Zn e Pb, sendo importantes passivos ambientais de contaminação do solo e recursos hídricos. A atividade mineradora aumenta a área (superfície) de exposição desses minérios ao ar e à água, aumentando assim as taxas de geração de ácido sulfúrico. Regiões onde as rochas têm baixa capacidade de tamponamento geram soluções tóxicas altamente ácidas, que são referidas como drenagem ácida de mina (AMD).

A formação de AMD ocorre naturalmente, mas a atividade microbiana aumenta sua taxa de formação e parece ser responsável pela maior parte da drenagem gerada. Apesar da acidez extrema, calor e altas concentrações de sulfato e metais tóxicos, uma variada gama de microorganismos estão adaptados à AMD. Estes organismos podem formar uma biosfera quimiolitotrófica no subsolo, sustentada por doadores de elétrons derivados de sulfetos, CO_2 , O_2 e N_2 do ar, e derivados de fosfato libertados pela interação água-rocha.

Assim como demonstrado por Baker e Banfield ⁸ a diversidade das comunidades em AMD varia de acordo com pH, temperatura, constituição do minério e concentrações de metais. No entanto, para a maioria dos ambientes já estudados (principalmente na subsuperfície) a comunidade parece realizar oxidação do ferro, oxidação do enxofre, oxidação do carbono orgânico, fixação de carbono, fixação de nitrogênio, produção de iodo polimérico extracelular, bem como redução de ferro e enxofre. A figura 1 ilustra um sistema funcional simplificado e fornece exemplos de organismos que podem controlar o fluxo de ferro, enxofre, nitrogênio, carbono e de energia em AMDs.

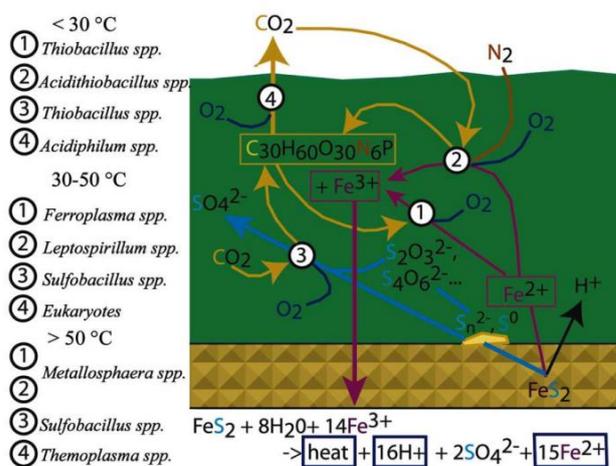


Figura 1 - Ciclos do ferro, enxofre e ciclagem de carbono em drenagem ácida de mina.

Representação baseada nas capacidades metabólicas conhecidas (1, 2, 3 e 4) associadas aos membros da AMD. A pirita (FeS_2) está representada na parte inferior em amarelo e verde é a representação da solução de AMD. O enxofre elementar é apresentado na interface água-pirita como um possível inibidor de dissolução superficial. A oxidação da pirita é mostrada na parte inferior, com Fe^{3+} indicado como o oxidante primário. Compostos de enxofre intermediários são indicados da seguinte forma: $\text{S}_2\text{O}_3^{2-}$ sendo tiosulfato e $\text{S}_4\text{O}_6^{2-}$ como tetrationato. $\text{C}_{30}\text{H}_{60}\text{O}_{30}\text{N}_6\text{P}$ representa os compostos orgânicos de carbono. Fonte: Baker e Banfield ⁸.

A biodiversidade obtida a partir de pilhas de rejeito da empresa Myanmar estava associada a prevalência de arqueias heterotróficas inicialmente isoladas e classificadas como *Ferroplasma cupricumulans* ⁹ e, posteriormente, reclassificadas como *Acidiplasma cupricumulans* ¹⁰. Já na Agnes Gold Mine em Barberton, África do Sul ¹¹ mixo- termofílicas moderadas e heterotróficas do gênero *Sulfobacillus* foram isoladas e identificadas como a população dominante. Em rejeitos sólidos da mina de cobre La Andina no Chile notou-se, por análise de sequências de 16S rDNA, maior presença de quimiolitotróficos

oxidadores de ferro do gênero *Leptospirillum*¹². Drenagens ácidas de mina são os locais ideais para o estudo da biodiversidade de microorganismos biomineradores. Isso porque os mesmos consórcios microbianos são os responsáveis pela geração da lixívia e da drenagem ácida de mina.

1.2 .2 Barragem de Rejeitos

A oxidação de minerais sulfetados e a formação de AMD pode ser limitada, desde que o rejeito permaneça saturado em água, evitando um acesso rápido ao oxigênio. Barragens de Rejeitos são estruturas que têm por finalidade o depósito e retenção de resíduos sólidos e água oriunda do processo de beneficiamento de minério. Os rejeitos apresentam na sua composição partículas dissolvidas e em suspensão, além de metais pesados e reagentes químicos. A Vale iniciou a operação da mina Sossego/Sequeirinho em 2004, e criou uma barragem de rejeitos a partir da água do Rio Parauapebas. Estima-se que esta barragem tenha recebido, desde o início da operação até o momento, mais de 90 milhões de toneladas de detritos com um teor de 0,07% de cobre. Atualmente 99% da água empregada no processo de mineração (900 mil m³/ano) é reutilizada a partir da barragem, promovendo um aporte cíclico de material exógeno no local. Em longos períodos de estiagem a água reutilizada tem se caracterizado por concentrações elevadas de sais dissolvidos como sulfatos, cloretos e íons metálicos como cálcio, potássio e magnésio¹³. Esses valores também apresentam diferenças sazonais em função da estação do ano. Se todo o cobre presente nesses rejeitos fosse recuperado, a Vale teria uma receita bruta de US\$ 1,4 bilhão, valor superior ao US\$ 1,2 bilhão investido pela companhia entre 1997 e 2004 para colocar a mina em operação¹⁴.

A diversidade microbiana em barragem de rejeitos de minas compreende espécies aeróbias e anaeróbias que podem ser autotróficas (fixação de CO²) ou heterotróficas (material orgânico como fonte de carbono), bem como litotróficas (compostos inorgânicos como fonte de energia) ou organotróficas (material orgânico como fonte de energia)¹⁵. A mineralização do carbono em ecossistemas de água doce com baixa concentração de sulfato é

tipicamente controlada por metanogênese. No entanto, a redução de sulfato pode ser importante em muitos sistemas de água doce e uma quantidade significativa de carbono pode ser mineralizada através desta via. Em ecossistemas de água doce com alta concentração de sulfato, procariotos redutores de sulfato (SRP) podem desempenhar um papel importante no ciclo do carbono e competir com as metanogênicas pelos doadores de elétrons. Nesses ambientes, Alphaproteobacteria, Deltaproteobacteria, e Firmicutes foram relacionados a variações nas concentrações de ferro férrico e ferroso. Bactérias quimiolitotróficas tais como as oxidadoras de amoníaco, redutoras e oxidadoras de ferro e metanogênicas foram descritas nas amostras de sedimento ¹⁶.

A contaminação dos solos com metais pesados tem sido relatada como fator que afeta negativamente a diversidade e uniformidade de microrganismos do solo e os seus processos metabólicos ¹⁷. Muitas espécies microbianas em barragem de rejeitos também são tolerantes a elevadas concentrações de metais através de uma variedade de mecanismos. Esses mecanismos incluem transformações tipo redox, síntese de proteínas e peptídeos que se ligam aos metais, precipitação orgânica ou inorgânica, transporte ativo, efluxo e compartimentalização intracelular ou a adsorção dos metais à superfície celular (biosorção), através da parede celular e outros componentes estruturais ¹⁸.

Com o rápido desenvolvimento da mineração de cobre, os impactos das barragens de rejeitos têm se expandido em todo o mundo ¹⁹. Os rejeitos de minas também são fonte de contaminação por metais pesados e, portanto, a descontaminação dos rejeitos é uma necessidade ²⁰. O principal desafio tem sido a forma de impedir a acidificação dos rejeitos de minas ao longo da operação e lidar com os rejeitos já acidificados em minas abandonadas ²¹. Os microrganismos, especialmente as bactérias que metabolizam enxofre, desempenham um papel na biorremediação desses metais e metalóides ⁷.

1.3 – Potencial biotecnológico dos microrganismos em ambientes de mineração

1.3.1 – Biomineração

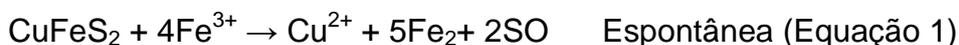
A biomineração é um método de extração de minerais e metais a partir de minérios complexos utilizando processos biológicos. A prática não exige nenhum dos procedimentos ambientalmente nocivos encontrados em métodos convencionais de refino, mas depende inteiramente da interação natural dos organismos biológicos. Um dos maiores sítios de biolixiviação na Kennecott Bingham Canyon Mine, em Utah, EUA tem produzido até 200 mil toneladas de cobre por dia, sendo responsável por aproximadamente 25% da produção de cobre do mundo.

No geral, empresas de mineração utilizam a biolixiviação como uma forma de explorar minérios de baixo teor e recursos minerais localizados em áreas de difícil viabilização técnica ou financeira com a tecnologia tradicional. Além disso, existem várias vantagens evidentes para o uso de biolixiviação, tais como (i) a utilização de componentes que ocorrem naturalmente, (ii) a utilização de pressão atmosférica e temperatura ambiente, e (iii) a prevenção da geração de poeira e SO_2 ²². O processo também tem provado ser vantajoso porque durante a ação microbiana, a produção de ácido sulfúrico é gradual, e a redução no pH do meio ocorre de forma lenta, permitindo uma dissolução diferencial de distintos metais e sua separação seletiva a partir da lixívia²³.

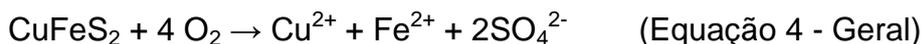
Para que ocorra o processo de biolixiviação são necessários: um doador de elétrons (elementos do minério como Fe^{2+} e S e suas formas reduzidas), uma fonte de carbono (CO_2), um acceptor final de elétrons (O_2) e um meio rico em microorganismos (como os pulverizados com água)²⁴. Existem dois mecanismos biolixiviadores dominantes. No primeiro mecanismo, a ação catalítica dos microrganismos é responsável pela dissolução do mineral, e os microrganismos interagem com a superfície do mineral diretamente, aumentando a taxa de dissolução. No mecanismo direto, a oxidação bacteriana converte partículas de sulfeto metálico em sulfatos solúveis. Na dissolução do cobre, por exemplo, o sulfeto mineral é oxidado e o sulfato de cobre é liberado

em solução. A mesma reação pode ocorrer por ação de seus produtos de metabolismo (principalmente sulfato férrico e ácido sulfúrico) conforme demonstrado nas equações abaixo:

Biolixiviação da calcopirita ²³:



O ferro ferroso (Fe²⁺) gerado nessa reação (equação 1) é rapidamente oxidado por procariotos oxidadores de ferro (equação 2), e mais Fe³⁺ é liberado no meio para que a oxidação espontânea da calcopirita ocorra (equação 1).



Já os procariotos oxidadores de enxofre realizam a redução do SO₂ (equação 3) que facilitará a produção do ácido sulfúrico.

No segundo mecanismo, referido como indireto, a ação microbiana ocasiona a oxidação dos íons ferrosos para íons férricos, e os íons férricos em seguida oxidam quimicamente os minerais sulfetados ²⁵. A principal abordagem utilizada para extração de cobre a partir de calcocita (Cu₂S) ou covelita (CuS) envolve o particionamento dos minérios, acidificação do meio com ácido sulfúrico e aglomeração em tambor rotativo. As pilhas são irrigadas com uma solução contendo ferro que possibilita o crescimento de bactérias sobre a superfície do minério para que seja produzida a solução catalisadora ²⁵.

De um modo geral, os microorganismos capazes de promover a lixiviação de minérios contendo ferro e enxofre (ex. pirita e calcopirita) são acidófilos, quimiotróficos (obtem energia a partir da oxidação de compostos inorgânicos), autotróficos (executam a biossíntese de todos os constituintes celulares utilizando o dióxido de carbono, CO₂, como única fonte de carbono) e são classificados de acordo com a temperatura em que se desenvolvem, distinguindo-se em: mesófilos (até ~40° C), termófilos moderados (~40 - ~55°

C) e termófilos extremos (~55 - ~80° C) ²⁶. A maioria deles pertence aos domínios Bacteria e Archaea, porém eucariotos unicelulares já foram reportados ²⁷. As bactérias mesófilas mais frequentemente isoladas em ambientes de lixiviação são linhagens das espécies *Acidithiobacillus ferrooxidans*, *Acidithiobacillus thiooxidans* e *Leptospirillum ferrooxidans* ²⁸. Em geral, devido às similaridades fisiológicas e ambientais, essas bactérias coexistem, intensificando a solubilização dos metais constituintes dos sulfetos minerais. Uma variedade de outros consórcios está bem descrita na literatura: *Acidithiobacillus caldus*, *Acidimicrobium ferrooxidans*, *Leptospirillum ferriphilum*, *Ferrimicrobium acidophilum*, *Acidiphilium acidophilum*, *Thiomonas cuprina*, *Sulfobacillus sp.*, *Alicyclobacillus sp.*, *Sulfolobus sp.*, *Ferroplasma sp.* e *Acidiplasma sp.* ²⁹.

Além disso, as interações entre os membros da comunidade (consórcios microbianos) são críticas na otimização dessas atividades. Por exemplo, existe uma simbiose importante entre as espécies autotróficas e heterotróficas: autotróficos podem depender de heterotróficos coexistentes para remover compostos orgânicos que são tóxicos para eles. Acidófilas heterotróficas são capazes de utilizar materiais orgânicos produzidos por acidófilas autotróficas ³⁰. Oliveira and Duarte ³¹ conduziram experimentos com cultura mista contendo *Acidimicrobium ferrooxidans*, *A. ferrooxidans*, *S. thermosulfidooxidans* e *S. acidophilus*. O filtrado da cultura dos autotróficos *Acidithiobacillus ferrooxidans* continha matéria orgânica suficiente para suportar o crescimento heterotrófico de *S. thermosulfidooxidans* (estirpe TH1) ³². Também foi relatada oxidação de ferro mais eficiente por culturas mistas de *Sulfobacillus spp.* e *Acidithiobacillus ferrooxidans* do que por culturas puras. Eles atribuíram essa maior lixiviação para o crescimento mixotrófico do *Sulfobacillus*, removendo subprodutos orgânicos de *A. ferrooxidans*. Isto pode explicar a correlação estreita em populações de *A. ferrooxidans* / *T. thiooxidans* e *Acidiphilium* em ambientes estudados por Peccia et al. ³³. Recentemente além do destaque na presença de membros *Acidithiobacillus* e *Sulfobacillus*, foi relatada uma prevalência média de 13% de membros *Anoxybacillus spp.* em modelos controlados de biolixiviação ³⁴. Dopson e Lindstrom ³⁵ examinaram o efeito de culturas mistas de *A. caldus* e *S. thermosulfidooxidans* (oxidadoras de Fe e S) na dissolução

de arsenopirita. A espectroscopia Raman mostrou formas abundantes de enxofre elementar (S^0) nas superfícies de arsenopirita e pirite durante a dissolução oxidativa ³⁶. A remoção microbiana de S^0 , e outros compostos extremamente hidrofóbicos, é crítica para o acesso de outros grupos microbianos as áreas de oxidação de ferro. *A. caldus* cresce na superfície de compostos de enxofre e produz agentes tensoativos para mobilizar S^0 , aumentando taxa de dissolução da arsenopirita. Alguns estudos têm demonstrado que os transportadores relacionados às proteínas da substância polimérica extracelular (EPS) são capazes de acumular enxofre coloidal e melhorar a biolixiviação ³⁷. A melhor adesão dos microorganismos à superfície do minério também pode ser mediada pela formação de biofilmes e produção de EPS por lixivadores como *A. thiooxidans* ³⁸.

Ainda não foi estabelecido um consórcio microbiano igualmente eficiente em diferentes ambientes. Uma das hipóteses seria que a introdução de microorganismos exógenos ocasiona um desequilíbrio ecológico incompatível com a atividade lixivadora ³⁹. Neste contexto, a seleção de um consórcio local pode apresentar vantagens em relação a utilização de consórcios importados, sendo de suma importância para a aplicação da biolixiviação no Brasil.

As vias metabólicas relacionadas à oxidação de ferro nos microorganismos acidófilos e neutrófilos permanecem, até o momento, pouco esclarecidas. Em vez de ter uma organização horizontal, como nas proteínas redox localizadas ao longo da membrana citoplasmática na maioria das cadeias respiratórias, a topografia dos transportadores de elétrons nos organismos oxidadores de ferro é vertical, permitindo a formação de um nanofio que conecta o meio exterior ao citoplasma. Essa topografia só é possível devido ao elevado número de aceptores de elétrons através da membrana celular. Esta organização é vantajosa, pois permite: (i) manter o ferro do lado de fora da célula, para evitar a precipitação de Fe (III) no citoplasma a um pH neutro, (ii) fugir do stress oxidativo pela reação entre Fe(II) e O_2 pois a oxidação de Fe(II) ocorre em um compartimento subcelular diferente do local da redução de O_2 e, (iii) manter um pH neutro no citoplasma por consumir os prótons que entram na matriz celular devido ao gradiente de

prótons entre o exterior e o interior da célula, por redução de O_2 em H_2O . Esses mecanismos foram recentemente revisados por Bonnefoy and Holmes ⁴⁰.

No modelo da bactéria acidófila oxidadora de ferro mais bem caracterizada (a *At. Ferrooxidans*) sugere-se que há a transferência de elétrons do ferro ferroso via citocromo c e rusticianina A através do periplasma para a citocromo aa₃-oxidase na membrana celular interna. Assim, os elétrons são usados para a redução do oxigênio para água ⁴¹ como ilustrado na figura 2. Em outras estirpes de *At. ferrooxidans* e na espécie *At. ferrivorans* existem outras duas vias de oxidação de Fe (II): uma através do modelo rusticianina rusA e outra através do complexo da proteína HiPIP e, eventualmente, uma isoenzima da rusticianina RusB ⁴².

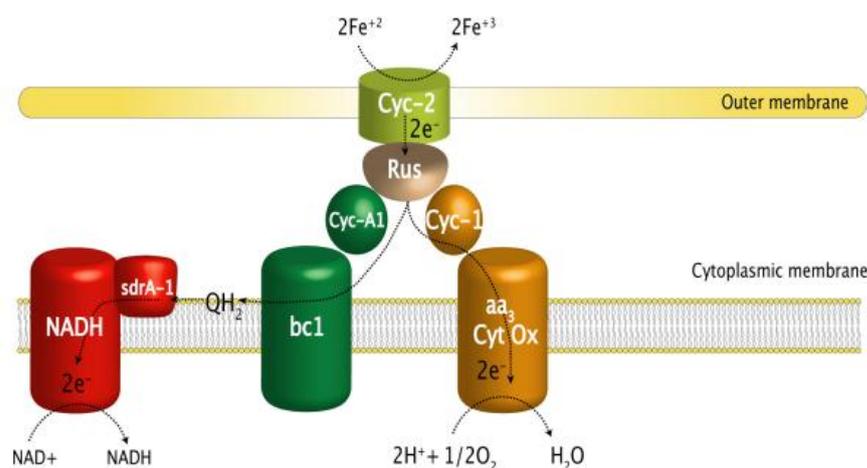


Figura 2 - Modelo de oxidação de ferro caracterizado em *At. ferrooxidans*.

Esta via envolve componentes das membranas externa, periplasmática e interna, que constituem um super-complexo conduzindo os elétrons a partir de minerais sulfetados para o aceptor final de elétrons localizado no citoplasma. Este nanofio é constituído pela membrana externa do citocromo c Cyc2 que contém um domínio voltado para o ambiente externo, no qual a oxidação do Fe (II) ocorre. A cadeia continua pela proteína rusticianina no periplasma, passando para o citocromo c cyc1 e para proteína integral de membrana aa₃ citocromo-oxidase, que catalisa a redução de O_2 . A partir da rusticianina os elétrons também podem ter outra via, impulsionada energeticamente pela força motriz de prótons através do citocromo c CycA1, pela proteína integral bc1 e ubiquinonas associadas ao periplasma, passando para o complexo NADH-1, que catalisa a redução do NAD^+ , no lado interior da membrana citoplasmática. Fonte: Quatrini, et.al. ⁴³.

A transformação do enxofre elementar por *A. ferrooxidans* com vários estados de oxidação consiste em vias metabólicas mais complexas quando comparadas ao metabolismo de ferro. Enxofre e compostos inorgânicos reduzidos de enxofre (RISCs) podem acumular-se nos minérios e outros tipos de minerais sulfetados que foram oxidados a Fe (III). Assim, os substratos energeticamente favoráveis podem ser utilizados por bactérias acidófilas oxidadoras de enxofre. Recentemente, há uma série de mecanismos enzimáticos conhecidos que estão relacionados com o metabolismo do enxofre, incluindo sulfeto quinona redutase (SQR), tiosulfato quinona oxidoredutase (TQO) e tetrionato hidrolase (teth) conforme demonstrado na figura 3⁴².

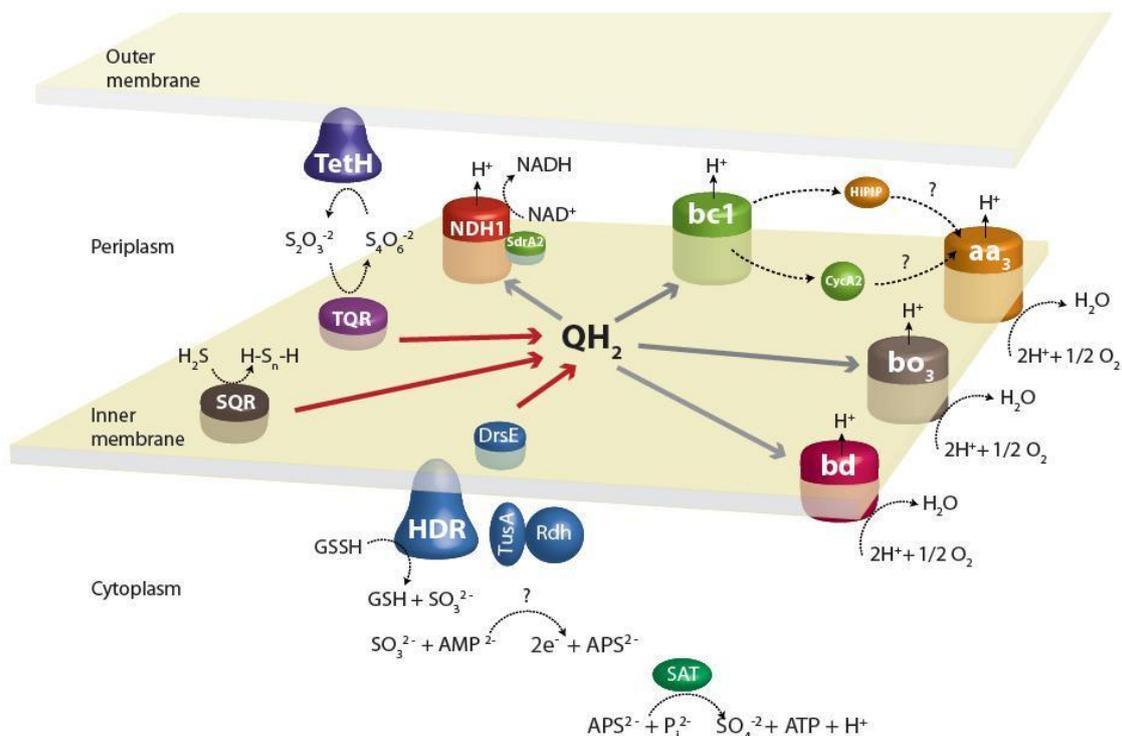


Figura 3 - Modelo de oxidação de enxofre em *A. ferrooxidans* ATCC 23270

A via de oxidação de compostos inorgânicos de enxofre reduzido (RISC) é composta por uma série de complexos enzimáticos e transportadores de elétrons situados em diferentes compartimentos celulares: na membrana externa de em contato com o periplasma (tetrionato redutase, TetH), no periplasma (potencial proteína de ferro-enxofre, HiPIP), ligada à membrana citoplasmática no lado periplasmático (citocromo c, CycA2), na membrana citoplasmática (sulfeto quinona redutase (SQR), tiosulfato quinona redutase (TQR), complexo bc 1, complexo NADH I, e oxidases terminais BD e BO3) e no citoplasma (heterodisulfeto redutase (HDR), e ATP sulfurilase (SAT)). O enxofre insolúvel é primeiro convertido em sulfano sulfato (GSSH), que é então transferido para a heterodisulfeto redutase (HDR) através de uma cascata de enxofre

transferases (DsrE, TusA e Rdh). Os elétrons provenientes do sulfeto (H_2S), tiosulfato ($\text{S}_2\text{O}_3^{2-}$) ou sulfano sulfato (GSSH) são transferidos pelo quinol (QH_2) por três maneiras: (1) diretamente ao terminal bd ou bo3 oxidase, (2) indiretamente, passando pelo complexo bc 1 e um citocromo c (CycA_2) ou uma proteína HiPIP, provavelmente, para a aa3 oxidase onde ocorre a redução de O_2 , (3) para o complexo de NADH I para gerar uma redução de energia. Fonte: Váldez et.al.⁴¹.

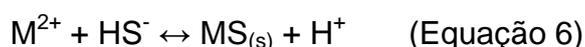
Além disso, uma variedade de mecanismos parece contribuir para a resistência das espécies microbianas a elevadas concentrações de metais⁴⁴. Dois estudos têm descrito as respostas proteômicas globais de *Acidithiobacillus ferrooxidans* ATCC 23270 ao estresse causado por altas concentrações de cobre (>100 mM de sulfato de cobre)^{45; 46}. As adaptações moleculares incluem uma regulação positiva de sistemas Cus do tipo RND, um aumento nas proteínas relacionadas com a biossíntese de cisteína e histidina, bem como dissulfeto isomerases putativas, provavelmente envolvidas em mecanismos de reparação das ligações dissulfeto danificadas após a exposição ao cobre. A regulação negativa da proteína principal da membrana externa (OMP40) e alguns transportadores iônicos sugerem a ocorrência de uma redução no influxo de metais e outros cátions⁴⁵. Curiosamente, as proteínas codificadas para o operon rus estavam aumentadas, sugerindo a sua participação também em um mecanismo de resistência a cobre, e apoiando o papel da rusticianina como uma chaperona de cobre⁴⁶. Em biofilmes formados sob pirita houve aumento nas funções relacionadas ao transporte de membrana, proteínas envolvidas na biossíntese de EPS, bombas de efluxo, lipoproteínas, transportadores ABC, e proteínas relacionadas à resposta ao estresse oxidativo e osmótico⁴⁷.

1.3.2 – Biorremediação

O processamento e/ou beneficiamento de minérios, independentemente do método de extração (químico ou biológico), gera grandes quantidades de rejeito mineral disposto em barragens ou pilhas de estéril que são os formadores de AMD discutidos até o momento. A remediação desses

ambientes inclui estratégias de transporte ativo, precipitação, quelação, neutralização, volatilização, oxidação e redução de possíveis contaminantes ⁴⁸. A neutralização química do pH de águas residuais utilizando calcário pode formar uma camada impermeável de hidróxido de metal que diminua eficiência do processo de precipitação de metais em hidróxidos e carbonatos. O sulfato é a forma mais oxidada de enxofre, e não pode ficar disponível no ambiente por ser um composto tóxico, de forte odor e corrosivo ⁴⁹. Neste sentido, a utilização de processos de biorremediação oferece várias vantagens potenciais como: i) menor custo quando comparada as tecnologias convencionais, ii) os contaminantes são geralmente convertidos em produtos inócuos, iii) os contaminantes são destruídos, e não simplesmente transferidos para outros meios, iv) é um método com relativa facilidade de implementação, v) permite a precipitação e conseqüente recuperação diferencial dos metais sulfatados ⁵⁰.

Um importante grupo de microrganismos que atuam como biorremediadores são as Bactérias Redutoras de Sulfato (SRB). A SRBs são um grupo anaeróbio, que suprem sua demanda energética a partir de reações redox, em que o sulfato serve como acceptor final de elétrons. A reação consiste na redução do sulfato vinculada à oxidação de matéria orgânica. Por esse motivo, SRBs são ubíquos em ambientes anóxicos ricos em sulfato, como esgotos e sedimentos de lagos ^{51; 52}, onde possuem um papel importante nos ciclos de carbono e enxofre. Substratos definidos, como lactato, etanol, metanol e gás rico em hidrogênio são preferivelmente utilizados como doadores de elétrons para a redução do sulfato ^{18; 53}. A neutralização do meio ácido se deve ao consumo de prótons com produção de sulfetos de hidrogênio (equação 5). Deste modo, os metais solúveis sulfatados são reduzidos a sulfetos metálicos, insolúveis, podendo ser então precipitados (equação 6) e removidos utilizando outros métodos, como a fitoacumulação.



*M= metal; MS= sulfeto metálico

As bactérias representam a grande maioria dos gêneros de redutores de sulfato descritos até o momento, entretanto já foram descobertos representantes no domínio *Archaea*. O maior grupo de procariontes redutores de sulfato (SRP) é composto por famílias da classe Deltaproteobactéria, como *Desulfobacterales*, *Desulfovibrionales* e *Syntrophobacterales*. O segundo maior grupo é representado pelo filo Firmicutes, incluindo os gêneros *Desulfotomaculum*, *Desulfosporomusa* e *Desulfosporosinus*. No filo Nitrospira é possível citar a espécie *Thermodesulfovibrio*. Outro grupo inclui as bactérias termófilas como *Thermodesulfobacteria* e *Thermodesulfobium*. Dentre os gêneros de arqueias conhecidas destacam-se *Archaeoglobus*, *Thermocladium* e *Caldivirga*⁵⁴. SRPs constituem um grupo diversificado que contribuem para uma variedade de funções essenciais em muitos ambientes, incluindo participação nos ciclos de carbono, enxofre e mercúrio, e biodegradação de poluentes orgânicos, tais como cresol e fenilbenzeno⁵⁵.

Os SRP podem ser autotróficos ou heterotróficos tendo como fonte de carbono o CO₂ e compostos orgânicos. Nos heterotróficos, a necessidade de matéria orgânica como fonte de carbono é uma característica importante, que influencia o processo de biorremediação. Lactato, etanol, acetato, glicose, glicerol, metanol e substratos orgânicos complexos (como esterco animal, lodo ativado e serragem) são exemplos de fontes de carbono, e consequentemente doadores de elétrons utilizados na redução microbiana do sulfato²⁷.

As equações químicas que representam a redução do sulfato acoplado à oxidação da matéria orgânica são:



* CH₂O representando fonte orgânica simples de carbono

A redução do sulfato pode ocorrer tanto com consumo de energia pela via assimilatória, quanto para produção de energia pela via dissimilatória (figura 4). Na via de assimilação, encontrada em uma ampla gama de organismos, o sulfato é reduzido em pequenas quantidades de sulfeto, e este é transformado em cisteína, a partir da qual outras moléculas biológicas contendo enxofre são

derivadas. Na via dissimilatória, que é restrita a arqueias e linhagens bacterianas anaeróbias obrigatórias ou microaerofílicas, o sulfato é o aceptor final de elétrons da cadeia respiratória, gerando grandes quantidades de sulfeto inorgânico ⁵⁶.

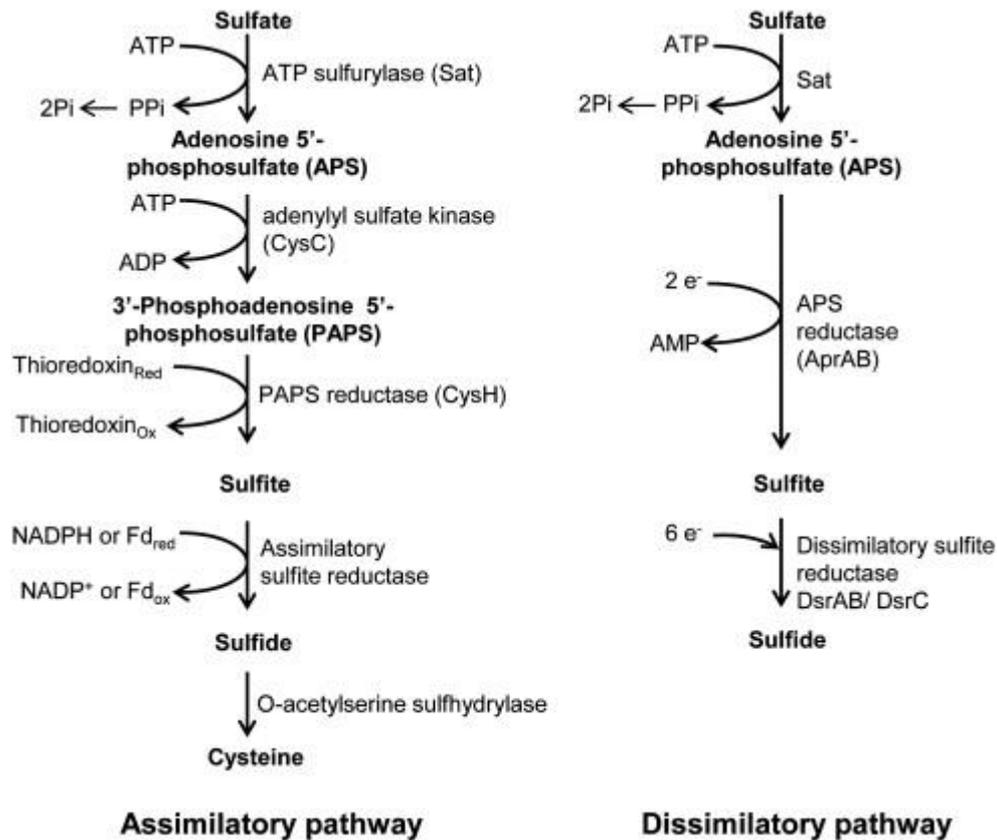


Figura 4 - Vias assimilatória e dissimilatória de redução de sulfato em procariontes

Ambas as vias se iniciam com a ativação de sulfato por reação com ATP para formar 5'-adenosina-fosfosulfato (APS) um passo catalisado pela ATP sulfurilase (SAT). Na via de assimilação, APS é convertido em 3'-fosfoadenosina-5'-fosfosulfato (PAPS) pela adenilil sulfato quinase (CysC) e, em seguida, reduzido a sulfito pela PAPS redutase dependente de tioredoxina (CysH). O sulfito é ainda mais reduzido para sulfeto pela sulfito redutase dependente de NADPH (CysIJ), ou uma enzima dependente de ferredoxina monomérica. Na via dissimilatória APS é diretamente reduzida a sulfito pela APS redutase (AprBA), uma flavoenzima heterodimérica de ferro-enxofre, e o sulfito é novamente reduzido à sulfeto pela sulfito redutase DsrAB, com o envolvimento da proteína dsrC. Outra proteína pequena dsrM, que muitas vezes é codificada a jusante do dsrAB, também pode estar envolvida na redução de sulfito, possivelmente em um papel regulador, mas a sua função exata ainda é desconhecida. O gene dsrM é fortemente regulado negativamente na presença de concentrações elevadas de

sulfeto. DsrAB, dsrC, e o complexo dsrMKJOP também estão presentes em organismos que reduzem sulfito, tiosulfato ou compostos organosulfonados. Fonte: Grein et. al. ⁵⁶

1.4 - Principais abordagens para o estudo de microbiomas

1.4.1 – Metagenômica

Microorganismos são fontes potenciais de enzimas de interesse industrial, além de vários outros produtos, como os metabólitos secundários, que possuem grande impacto no delineamento de estratégias de interesse biotecnológico ⁵⁷. No entanto, estima-se que apenas uma pequena fração (1% a 10%) da diversidade microbiana existente seja cultivável com técnicas disponíveis até o momento. Neste contexto, a metagenômica representa um importante campo de estudo, no qual se busca caracterizar comunidades complexas e descrever a diversidade biológica de amostras ambientais sem a necessidade de isolamento ou cultivo microbiano ⁵⁸. Usando dados metagenômicos é possível acessar a informação genética sobre: potenciais novos biocatalisadores ou enzimas, ligações genômicas entre função e filogenia de organismos não cultiváveis, perfis evolutivos baseados na função e estrutura da comunidade. Esses estudos também podem ser complementados com abordagens como a metatranscriptômica ou metaproteômica para descrever quais atividades potenciais estão expressas em determinada condição ⁵⁹.

A estratégia metagenômica tradicional inclui a construção de bibliotecas com dezenas de milhares de clones, tipicamente em vetores de elevada capacidade, que retratem de forma aproximada a composição biológica de determinada comunidade. As tecnologias de sequenciamento de DNA de nova geração (NGS) possibilitaram a aplicação de duas outras estratégias em metagenômica: a) Estudo de diversidade utilizando sequências de genes amplificados (*amplicons*). Nesse tipo de estudo amplifica-se por PCR a sequência de interesse (ex. 16s, 18s, ITS) e se compara o resultado contra um banco de dados de bactérias conhecidas. A análise de genes que codificam para o gene 16S rRNA tornou-se uma das abordagens mais importantes em projetos voltados ao estudo da diversidade microbiana ⁶⁰; b) Estudo sem a

seleção de alvos (*shotgun*). Todo o DNA extraído da amostra é fragmentado e sequenciado. A análise consiste em avaliar e montar fragmentos metagenômicos a fim de identificar, além da diversidade de genomas, novos genes.

Atualmente os novos kits comerciais de extração de DNA e as plataformas NGS permitem a construção de bibliotecas com alta cobertura, proporcionando uma maior representatividade taxonômica das amostras ^{61; 62}. As plataformas NGS, como 454 GS FLX Sequencing (Roche), MiSeq e HiSeq (Illumina), e Ion Torrent (Life Technologies), compreendem equipamentos com capacidade de sequenciamento da ordem de milhões ou bilhões de bases em algumas horas de corrida, porém sob o custo de uma redução significativa no tamanho das leituras geradas. A tecnologia Ion Torrent trouxe ao mercado um chip semicondutor capaz de traduzir diretamente sinais químicos em informação digital com alta acurácia (>99,5%) e de aproximadamente 400 pb, o que é extremamente útil na montagem de leituras do tipo *shotgun* ⁶³. Já a plataforma MiSeq (Illumina) possui uma capacidade de produção de 15 GB com leituras de 300 pares de base (pb) de comprimento. Para análise de mais de uma região variável do 16S rDNA, por exemplo, os fragmentos agrupados podem ser sequenciados a partir de ambas as extremidades do 16S rDNA. Esses pares de sequências (*paired-end*) podem ser fundidos de modo que o fragmento final tenha 600 pb ⁶⁴.

1.4.1.1 – Abordagens para análise metagenômica

Embora o volume de dados metagenômicos depositados em repositórios públicos esteja aumentando exponencialmente ⁶⁵ ainda não existem padrões para métodos experimentais e computacionais necessários para analisar tais conjuntos de dados ⁶⁶, tornando-se difícil comparar os resultados destes estudos.

O pré processamento da sequências, anterior aos passos de montagem, predição gênica e anotação, é um aspecto crítico, pois evita *contigs* quiméricos e frameshifts (deslocamento do quadro de leitura na identificação dos genes

por inserções e deleções). O objetivo dessa etapa é remover possíveis erros de sequenciamento, vetores, adaptadores, primers e organismos não relevantes para o estudo (contaminantes). Um dos índices de qualidade amplamente utilizados é o Phred, que corresponde a uma estimativa de probabilidade de erro por base de modo que: $q = -10 \times \log_{10}(p)$, onde a pontuação da qualidade atribuída a uma base (q) está relacionada com a probabilidade estimada de uma atribuição errada (p). Assim, uma qualidade Phred desejável para análise de metagenomas seria superior a 20 em todas as bases, que corresponde a uma probabilidade de erro de 1% por base analisada ⁶⁷. Os principais programas que fornecem os pré processamentos descritos são PRINSEQ ⁶⁸, NGS QC Toolkit ⁶⁹ e Trimmomatic ⁷⁰.

Uma primeira classe de abordagens computacionais atribui às sequências metagenômicas uma categoria taxonômica utilizando buscas por similaridade contra um banco de dados de genes ou proteínas conhecido. O programa MEGAN ⁷¹ por exemplo, usa alinhamento local em suas atribuições e calcula o menor ancestral comum entre todos os organismos que tiveram *hit* com cada sequência. Alternativamente, a análise filogenética de genes marcadores, incluindo o gene 16S rRNA, também tem sido aplicada para a determinação de distribuição taxonômica (MLTreeMap ⁷² usa TreePuzzle ⁷³, AMPHORA ⁷⁴ usa PHYML ⁷⁵). Essa atribuição requer a disponibilidade de dados de referência filogeneticamente próximos nos bancos de dados, para classificações mais informativas, como ao nível de gênero e espécie ⁶⁷.

A segunda classe de algoritmos tenta resolver o problema de *binning*, que consiste no agrupamento das sequências em diferentes grupos (*bins*) de organismos estreitamente relacionados. A maioria das ferramentas computacionais existentes para *binning* utilizam a composição do DNA, como conteúdo GC, frequência de nucleotídeos e preferência de códons, para esse agrupamento (TETRA ⁷⁶, MetaClust ⁷⁷, MaxBin ⁷⁸, TACOA ⁷⁹). Para a correta utilização de qualquer algoritmo *binning* é importante verificar o tipo de dado de entrada e a existência de um completo conjunto de dados de treinamento ou genomas de referência suficientes para classificar os organismos presentes no metagenoma. *Binning* baseado em composição não é adequado para o sequências muito curtas (por exemplo uma sequência de 100 pb contém

menos de metade de todos os 256 tetrâmeros possíveis), pois elas não representam informações suficientes para uma acurada classificação ⁶⁷.

Com base em bibliotecas 16S rDNA do Projeto Microbioma Humano (HMP) ⁸⁰, o programa PICRUSt ⁸¹ (*phylogenetic investigation of communities by reconstruction of unobserved states*) cria uma predição da abundância de famílias gênicas a partir de marcadores filogenéticos. Ele pré contabiliza o número de cópias do 16S rDNA de acordo com os táxons descritos no HMP e ajusta a abundância taxonômica presente nas amostras analisadas de acordo com esse catálogo pré estabelecido. A partir das classificações, PICRUSt retorna apenas as funções previamente anotadas para os correspondentes genomas completos que estão depositados em bancos de dados públicos. MetaPhlAn ⁸² (*metagenomic phylogenetic analysis*) é um programa para caracterizar os perfis taxonômicos usando bibliotecas *shotgun*. Ele faz a inferência da abundância relativa de táxons a partir da identificação de marcadores clado específicos e como PICRUSt só identifica táxons para marcadores pré computados. O programa GSMer ⁸³, usa uma metodologia baseada em kmers para identificar os marcadores específicos dos genomas sequenciados pelo Projeto Microbioma Humano. Apesar de GSMer flexibilizar os marcadores a serem usados de acordo com a amostra de entrada, não possibilita a identificação de espécies de grupos taxonômicos não representados no HMP.

Uma vez encontrados os *taxa* de interesse, é possível reconstruir genomas microbianos a partir do conjunto de sequências metagenômicas. O mapeamento contra genomas de referência pode ser usado para as espécies esperadas na amostra. No entanto, esse mapeamento não elucida os genomas de espécies novas ou cepas muito variáveis de espécies já sequenciadas. Nesse caso, a montagem *de novo* é o a abordagem de escolha, e geralmente é realizada utilizando montadores específicos tais como Metavelvet ⁸⁴, Genovo ⁸⁵, Meta-IDBA ⁸⁶. Montadores baseados em grafos *deBruijn* foram desenvolvidos para as sequências curtas. Um grafo *deBruijn* é uma estrutura de dados direcionada que representa compactamente uma sobreposição de sequências curtas, de modo que cada k-mer (palavra de comprimento k criada a partir da decomposição das sequências originais) é atribuído a um nó, e

assim, o tamanho do grafo torna-se independente do tamanho da entrada. Os *contigs* resultantes são formados pelos caminhos que percorrem o maior número de k-mers no grafo. Os dois principais desafios encontrados por montadores que utilizam sequências curtas como entrada são: a abundância variável dos organismos na amostra e a baixa sensibilidade dos montadores. A variação de abundância entre os organismos leva a uma variação de amostragem que confunde montadores que utilizam "a cobertura esperada" como um parâmetro para diminuição de erros derivados da montagem de regiões repetitivas. Algumas soluções foram propostas como cálculo de cobertura local ou particionamento dos grafos de montagem, entretanto uma abordagem híbrida de montagem pode ser mais efetiva ⁸⁷. A sensibilidade está relacionada ao dimensionamento do estudo. É necessária uma amostragem profunda para montar robustamente membros raros. Em amostras metagenômicas ribotipadas (estudos de 16s rRNA) é possível detectar espécies raras, que possuem uma abundância de 1 para 100.000 espécies dominantes ou inferior. Estas espécies podem ser inobserváveis com cultivo, todavia temos ciência de que estas podem ser membros ecologicamente significativos (a hipótese de "biosfera rara" ⁸⁸).

A predição gênica é utilizada para a identificação de proteínas e RNAs presentes na amostra. É possível mesclar predições realizadas nos contigs e nas sequências do metagenoma não montado. Existem duas abordagens para predição de genes: a) "baseada em evidências" que usa busca por homologia a partir de comparações simples (ex. BLAST ⁸⁹) contra bancos de dados de proteínas (ex. Crítica ⁹⁰, Orpheus ⁹¹) e b) "ab initio" que depende de características intrínsecas da sequência de DNA para discriminar entre regiões codificadoras e não codificadoras, permitindo a identificação de genes que não possuem homólogos nas bases de dados disponíveis. Nesse caso, o uso de conjuntos de treinamento, estratégia usada por FgenesB ⁹² e Genemark ⁹³, pode melhorar a qualidade dos genes preditos. Em procariotos, os genes possuem sinais específicos e relativamente bem compreendidos como promotores, fatores de transcrição, ligação a *Pribnow Box* (sequência TATAAT, parte essencial dos promotores). Além disso, sequências codificadoras de proteína ocorrem como um formato de leitura aberta (ORF), com estatísticas

próprias de tamanho da janela até a identificação do *stop* códon. Genes codificadores de RNAs (tRNA e rRNA) são preditos utilizando ferramentas como tRNAscan para tRNAs e buscas de similaridade para rRNAs. Outros tipos de RNA não codificadores (ncRNA) podem ser detectados por comparação com os modelos de covariância ⁹⁴ e *sequence-structure motifs* ⁹⁵. A predição é geralmente seguida de uma anotação funcional; comparações entre genes preditos e sequências previamente anotadas (similaridade ou ortologia). Sequências de proteínas podem ser comparadas com perfis de alinhamento de sequências de proteínas das famílias TIGRFAM ⁹⁶ e Pfam ⁹⁷ usando, por exemplo, RPS-BLAST ⁹⁸ ou DIAMOND ⁹⁹. RPS-BLAST usa o algoritmo “*reverse position-specific BLAST*”, no qual a matriz de alinhamento utilizada uma pontuação de scores modificada de acordo com a posição específica dos aminoácidos, penalizando substituições em sítios de ligação dos domínios conservado. Já DIAMOND, assim como o BLASTX e BLASTP ¹⁰⁰, é baseado no alinhamento local de k-tuplas, estruturas de dados ordenadas, mas é mais rápido pela dupla indexação da referência e da *query*.

Métodos para a anotação de contexto genômico de cada gene incluem *genomic neighborhood* ¹⁰¹, *gene fusion* ¹⁰², perfis filogenéticos ¹⁰³ e co-expressão ¹⁰⁴. KEGG orthology (KO) ¹⁰⁵ possibilita uma análise de enzimas e vias metabólicas, de modo que cada leitura corresponda a um número de acesso KO (PathwayTools ¹⁰⁶). MG-RAST ¹⁰⁷ e CloVR ¹⁰⁸ são exemplos de *pipelines* que integram anotações funcional e metabólica. As informações obtidas das várias espécies podem ser agrupadas em vias biológicas representativas da atividade bioquímica da comunidade microbiana ^{58; 59}. MOCAT2 (*metagenomic analysis toolkit*) também pode gerar perfis taxonômicos a partir de conjuntos de referência de genes marcadores, além de perfis funcionais, bem como e fazer a predição gênica das sequências montadas. No entanto, nenhuma das ferramentas supracitadas é capaz de quantificar as abundâncias, pois retornam uma simples contagem do mapeamento de genes ou espécies conhecidos. Além disso, não levam em consideração o comprimento do gene, das leituras ou o tamanho do genoma de referência. Destes *pipelines* apenas MOCAT2 ¹⁰⁹ é capaz de realizar a montagem automática das sequências, sendo, portanto, os demais, incapazes

de identificar operons e genes com múltiplos domínios em metagenomas de baixa ou média complexidade. *SmashCommunity*¹¹⁰ (Simple Metagenomics Analysis Shell for microbial communities) é capaz de suprir essas demandas, e realizar análises quantitativas e comparativas, entretanto foi desenvolvido apenas para conjuntos de dados menores como leituras de Sanger e 454.

A metagenômica comparativa é importante para estudos funcionais e evolutivos em diferentes amostras, revelando variações na composição microbiana, composição das sequências, tamanho dos genomas, *taxas* de evolução, enriquecimentos específicos de vias metabólicas, grupos funcionais ou grupos de ortologia que caracterizam a funcionalidade bioquímica de uma comunidade¹¹¹.

1.4.2 – Abordagens para o estudo de Genomas Individuais

Com o objetivo de prever o potencial metabólico de comunidades microbianas utilizamos a metagenômica, porém essa análise não permite o estudo da variabilidade genética intra-específica. Ainda sem a necessidade de cultivo prévio, é possível reconstruir pequenos genomas ou conjuntos de linhagens estreitamente relacionadas a partir de bibliotecas *shotgun* de uma comunidade complexa. Esse *binning* metagenômico possui dois componentes principais: (i) clusterização e (ii) representação de dados¹¹². A clusterização envolve o agrupamento (ex. agrupamento hierárquico e redes neurais) dos *contigs*, *scaffolds* ou genes com base em suas características genéticas, como por exemplo, a frequência de oligonucleotídeos (2T-binning¹¹³, Databionic ESOM tools¹¹⁴, tetramerFreqs¹¹⁵) e a cobertura dos *contigs* (Multi-metagenome¹¹⁶, MGS Canopy algorithm¹¹⁷). Existem ainda programas que combinam as duas abordagens, criando uma matriz de distância composta da combinação dessas duas análises (MaxBin⁷⁸, GroopM¹¹⁸, MetaBAT¹¹⁹).

Atualmente sabemos que a diversidade de espécies unicelulares dentro de um ecossistema é muito maior do que se pode medir com precisão através do estudo de grupos mistos. Estudos de células únicas permitem *insights* sobre adaptação ecológica, interações tróficas e versatilidade metabólica pontual,

incluindo a análise de polimorfismos de organismos geneticamente adaptados ao ambiente sem necessidade de cultivo prévio ¹²⁰. Métodos para sequenciamento genômico de células únicas (*Single Cell Genomics*, SCG) consistem em uma série dos processos integrados: separação física (*sorting*), lise, amplificação do genoma inteiro das células individuais e sequenciamento. A separação das células únicas pode ser realizada por fluorescência (FACS) ¹²¹ ou por micromanipulação ¹²². Uma única célula procariótica contém de 10 a 15 fentogramas de DNA, aproximadamente 10^9 vezes menos DNA do que o necessário para construção de uma biblioteca NGS que permita a montagem de um genoma procariótico. A amplificação do genoma inteiro (*Whole Genome Amplification* - WGA) se torna possível através da técnica de amplificação por deslocamento múltiplo (*Multiple Displacement Amplification* - MDA), que envolve a utilização de uma DNA polimerase (ϕ 29) ¹²³.

As técnicas de sequenciamento, filtragem das sequências e anotação dos *scaffolds* são as mesmas descritas para a metagenômica, incluindo um passo anterior de isolamento e cultivo microbiano. Todavia, o viés de amplificação da técnica de MDA resulta em uma grande cobertura de algumas regiões do genoma, em detrimento da ausência de cobertura em outras regiões. Uma montagem de novo usando algoritmos convencionais, que pressupõe uma profundidade uniforme de sequenciamento, gera *contigs* quiméricos ligando fragmentos genômicos não contíguos. Uma solução proposta por Chitsaz *et al.* foi implementada em Velvet-SC ¹²⁴, e adaptada em outras ferramentas de montagem como SPADES ¹²⁵ e IDBA-UD ¹²⁶. O algoritmo baseado em grafos *deBruijn* usa um limiar variável de corte para as regiões de baixa cobertura, que começa em 1 e aumenta gradualmente. Desse modo, alguns contigs podem ser conectados mesclando regiões de alta e baixa cobertura a fim de formar um *scaffold* longo.

A abordagem SCG permitiu um novo olhar sob alguns aspectos estruturais do genoma, como a conservação gênica, conservação de operons, duplicação gênica, ilhas genômicas, transferência horizontal de genes e vias metabólicas pouco elucidadas ¹²⁷. No entanto, para estudos focados em genomas cultiváveis (como para desenvolvimento de consórcios microbianos),

de difícil separação por SCG ou de alta complexidade, a genômica ainda é o método de escolha.

2- Justificativa

A indústria de mineração é um dos motores econômicos do mundo, com grandes investimentos em pesquisas nos setores de prospecção e implantação de novas minas ¹. Apesar disso, ainda não existe informação disponível suficiente sobre a diversidade de bactérias e arqueias de minas brasileiras de cobre, bem como poucas pesquisas estão focadas na biorecuperação desses metais ²⁹. O conhecimento da diversidade de comunidades microbianas procarióticas associadas a minérios complexos contendo cobre e a reconstrução de genes e rotas metabólicas envolvidos nas interações entre esses microrganismos é pré-requisito para que se possa compreender como os microrganismos influenciam na biooxidação e bio-redução desses metais. Esta visão ampla pode ser melhor estabelecida através de métodos como a metagenômica e genômica, que identificam porções gênicas de microrganismos cultiváveis e não cultiváveis. Os resultados do presente trabalho constituem um catálogo genético inédito e de grande valor para condução de estudos mais focados em aplicações biotecnológicas como: (1) *screening* funcional ¹²⁸, (2) isolamento e caracterização de novas espécies ¹²⁹, (3) estudos de microcosmos ¹³⁰, (4) ampliação do biomonitoramento ambiental com o desenvolvimento de novos biomarcadores relacionados à contaminação com metais pesados ¹³¹, (5) criação de consórcios microbianos comerciais para biolixiviação de biorremediação ¹³² e (6) desenvolvimento de subprodutos para biomineração por biologia sintética ²⁷. Além disso, a correlação entre esses achados e atividades microbianas em íons metálicos pode ser a chave para o desenvolvimento de novas tecnologias visando, principalmente: aumentar a extração de cobre (de forma mais econômica), remediar ambientes contaminados sem a necessidade de processos químicos (tecnologia biologicamente correta), possibilitar a recuperação de metais a partir de depósitos de muito baixo teor (até o momento, economicamente inviáveis).

3- Objetivos

3.1 - Objetivo Geral

Caracterizar taxonômica e funcionalmente comunidades microbianas de ambientes de mineração de cobre utilizando abordagens dependentes e independentes de marcadores moleculares.

3.2 - Objetivos específicos

1. Determinar e comparar a diversidade taxonômica e funcional de amostras provenientes de diferentes ambientes de minas de cobre;
2. Reconstruir genes, rotas metabólicas e genomas de interesse para a biolixiviação e bioxidação de cobre;
3. Comparar as abordagens dependentes e independentes de marcadores moleculares, avaliando suas vantagens e limitações;
4. Associar as mudanças ambientais e parâmetros físico químicos com os perfis encontrados no item 1.

4- Metodologia

4.1 - Áreas de estudo

A mina de Sossego (06°23'14.9"S - 50°03'19.5"W, 238 metros de altitude) é uma mina de cobre da Vale localizada na província mineral de Carajás, no estado do Pará, uma área que faz parte da Amazônia legal brasileira, compreendendo o bioma cerrado e floresta tropical. Neste macroambiente, a barragem de contenção de rejeitos recebe e armazena os rejeitos produzidos em grande escala em uma das etapas do beneficiamento da calcopirita. Essa barragem de rejeitos foi escolhida como área de estudo por ser um circuito fechado com alto teor de cobre, sulfato e outros compostos que favorecem o ciclo de vida quimiolitotrófico, principalmente de organismos biorremediadores. Neste ambiente coletamos amostras de água de superfície, água de profundidade a 15 metros e sedimento (figura 5 A). Representando ambientes impactados, locais com formação de drenagem ácida de mina foram selecionados para coletas de água e sedimento, por serem amostras alvo para o desenvolvimento de bactérias biolixiviadoras (figura 5 B).



Figura 5 - Representação das áreas de coleta A) Barragem de Rejeitos, outubro de 2012. Coordenadas geográficas 06°27'83" S - 050°024' W ; B) Drenagem ácida de mina, abril de 2013. Coordenadas geográficas 06°25'62" S - 050°03°068' W.

4.2 - Coleta

As coletas ocorreram em período chuvoso (outubro de 2012 para as amostras de barragem de rejeitos, e junho de 2014 para as amostras de drenagem ácida de mina) e período seco (abril de 2013 para as amostras de

barragem de rejeitos). Foram coletadas triplicatas da água de superfície e profundidade da barragem de rejeitos. O sedimento da barragem de rejeitos, por se tratar de uma amostra formada por camadas complexas, possui hexaduplicatas. Já a drenagem ácida de mina, até o presente momento está representada com duplicatas para a água e amostra única para o sedimento. Amostras de AMD são de difícil coleta, pois são formadas em locais de difícil acesso, onde é preciso uma autorização prévia da equipe de segurança do trabalho e acompanhamento de profissional treinado para dirigir em áreas de lavra. Uma parte do material coletado foi utilizada para a extração do DNA metagenômico, outra parte foi destinada às análises físico-químicas realizadas pela SGS Geosol Laboratórios Ltda, uma terceira porção foi destinada ao cultivo de microrganismos em meios específicos por uma parceria entre o Instituto Tecnológico Vale e a Universidade de Bangor. Os pontos escolhidos para a coleta levaram em consideração a autorização pela empresa que mantém a concessão da área, além da facilidade de acesso aos ambientes de interesse, para possível localização de microrganismos autotróficos. As amostras foram armazenadas a -20°C até o momento da extração.

4.3 - Extração e sequenciamento do DNA

Todas as sequências utilizadas nas análises bioinformáticas foram coletadas para o projeto Vale (Grant 1786221), extraídas por Julliane Dutra Medeiros doutoranda do programa de Genética da UFMG e sequenciadas pela Plataforma de Sequenciamento do Centro de Pesquisas René Rachou.

Para amostras de sedimento (1 g), foi usado o PowerSoil® DNA Isolation Kit (MoBio), conforme instruções do fabricante. Para amostras de água (membrana de 0,22 µm saturada com biomassa microbiana, por filtração de cerca de 3 a 5 litros de água) foi utilizado o PowerWater® DNA Isolation Kit (MoBio). A qualidade e quantidade do DNA foram avaliadas por eletroforese em gel e Qubit®.

Utilizamos duas abordagens para o estudo taxonômico das comunidades microbianas em ambientes de mineração. A primeira é a

amplificação de duas regiões hipervariáveis do gene 16S rDNA procariótico, denominadas V3 e V4, a partir dos iniciadores universais 515F e 806R. Estes oligonucleotídeos estão indicados na literatura como referências para os estudos metagenômicos ¹³³ e foram anteriormente padronizados em experimentos no laboratório Centro de Pesquisas René Rachou/Fiocruz-Minas e do projeto Brazilian Microbiome Project ¹³⁴. A PCR foi feita em um volume final de 25 µL, as condições usadas estão descritas na Tabela Suplementar 1. Conforme esperado, a amplificação do gene 16S rRNA resultou em um tamanho de fragmento de aproximadamente 550pb. A preparação das bibliotecas do tipo paired-end de DNA genômico foram construída a partir de 50 ng de DNA. Em seguida, as amostras foram submetidas a uma reação de fragmentação aleatória na qual o DNA foi simultaneamente fragmentado e ligado a adaptadores específicos utilizando o kit Nextera® DNA Sample Preparation (Illumina) conforme instrução do fabricante. Em seguida, o DNA genômico foi purificado e submetido a uma reação de amplificação utilizando iniciadores complementares aos adaptadores. Os produtos foram quantificados através do qPCR utilizando o Kit Sybr Fast qPCR kit (Kapa). As bibliotecas de amplicons foram diluídas em uma solução de Tris-HCl e Tween 0,1%, depositadas em uma flowchip e submetidas a 600 ciclos de sequenciamento do tipo paired-end (gerando 300 pb no sentido *forward* e 300 pb no sentido *reverse* para cada molécula de DNA sequenciada) utilizando o MiSeq Reagent Kit v3 (Illumina). As imagens obtidas foram processadas e analisadas pelo programa “*Imaging and Base Calling*” fornecido pelo fabricante.

Como segunda abordagem nenhum alvo foi selecionado e foram construídas bibliotecas do tipo *shotgun*. Todo DNA extraído da amostra foi fragmentado e sequenciado na plataforma Ion Torrent, gerando aproximadamente 400 pb para cada molécula seqüenciada. Para a construção das bibliotecas foram utilizados os reagentes Ion Xpress™ Plus Fragment Library Kit, Ion PGM Template OT2 200 Kit, Ion PGM Sequencing 200 Kit v2, Ion Sphere™ Quality Control Kit, Ion Library Quantification Kit e Ion 318 Chip Kit , todos fornecidos pela ThermoFisher. Cerca de 500 ng de DNA total foram fragmentados enzimaticamente, usando o Ion Shear Plus Reagent Kit. Em seguida, as extremidades do DNA foram reparadas e ligadas a adaptadores

específicos fornecidos nos kits de sequenciamento. Com o uso de barcodes, foram sequenciadas até três amostras em um único chip, mantendo uma profundidade elevada de sequenciamento (equivalente a 1.000 vezes o tamanho de um único amplicon ou 4 vezes o tamanho médio de 100 genomas bacterianos). Os fragmentos de DNA de cada biblioteca foram ligados a esferas magnéticas e em seguida amplificados através de PCR em emulsão. As Ion Sphere™ particles foram então aplicadas nos poros do Ion chip por meio de centrifugação. As bibliotecas foram posteriormente enriquecidas e preparadas para deposição no Chip 318 a fim de serem então processadas pelo PGM Ion Torrent™.

4.4 – Análise de amplicons (16S rDNA)

Os dados gerados foram submetidos a um pré-processamento usando Trimmomatic v0.32 ⁷⁰. Os adaptadores Illumina foram identificados e removidos. A filtragem por qualidade foi aplicada percorrendo uma janela de 4 bases ao longo de toda sequência, a partir da extremidade 5'. Foram eliminadas regiões com pontuação média da janela inferior a Phred 15. Desse modo, todas as janelas subsequentes à janela filtrada, extremidade 3' da sequência, também foram removidas. Os pares de sequências (*forward* and *reverse* de cada produto amplificado) foram montados a partir de suas regiões de sobreposição utilizando o programa Flash v1.2.11 ¹³⁵. Todos os pares montados foram então truncados em 400 pares de base usando USEARCH v.7 ¹³⁶, a fim de garantir um melhor alinhamento múltiplo possível nas etapas posteriores. O mesmo programa foi aplicado nas etapas de derreplicação, remoção de quimeras e agrupamento. Derreplicação é a identificação de sequências 100% idênticas, de modo que apenas uma cópia de cada sequência seja reportada. Isso diminui o esforço computacional nas etapas de agrupamento e classificação taxonômica.

As sequências filtradas foram agrupadas em unidades taxonômicas operacionais (OTUs) com uma identidade mínima de 97%, que corresponderia à classificação ao nível de espécie. Quimeras (sequências errôneas geradas pela amplificação por PCR de mais de um molde pai) foram removidas com

base no banco rdp_gold.fa disponível no pacote QIIME v1.8.0 ¹³⁷. A classificação taxonômica dos grupos foi atribuída usando QIIME v1.8.0 e a base de dados Greengenes v13_08 como referência. Os arquivos de saída (seqs_otu.txt e tax_assignments.txt) foram utilizados para construir uma tabela de OTUs no formato BIOM. As análises de diversidade foram realizadas através da execução do fluxo de trabalho core_diversity_analyses.py.

A riqueza da população e os índices de alfa e beta diversidade foram calculados usando QIIME com rarefação de 28.000 sequências e scripts R que utilizam rotinas do pacote Vegan. Para comparações de beta diversidade, conjuntos de dados de sequência rarefeitos foram utilizados para calcular a distância filogenética Unifrac. As relações de dominância e modelos AIC (akaike information criterion) foram geradas com base em uma matriz formada pelo logaritmo da abundância de espécies em relação as unidades taxonômicas encontradas para todas as amostras. Essa matriz serviu como entrada para um script R que utiliza a função radfit do pacote vegan. Uma visão geral dos passos para a análise integrada entre dados de *amplicons* e *shotgun* está representada na figura 6.

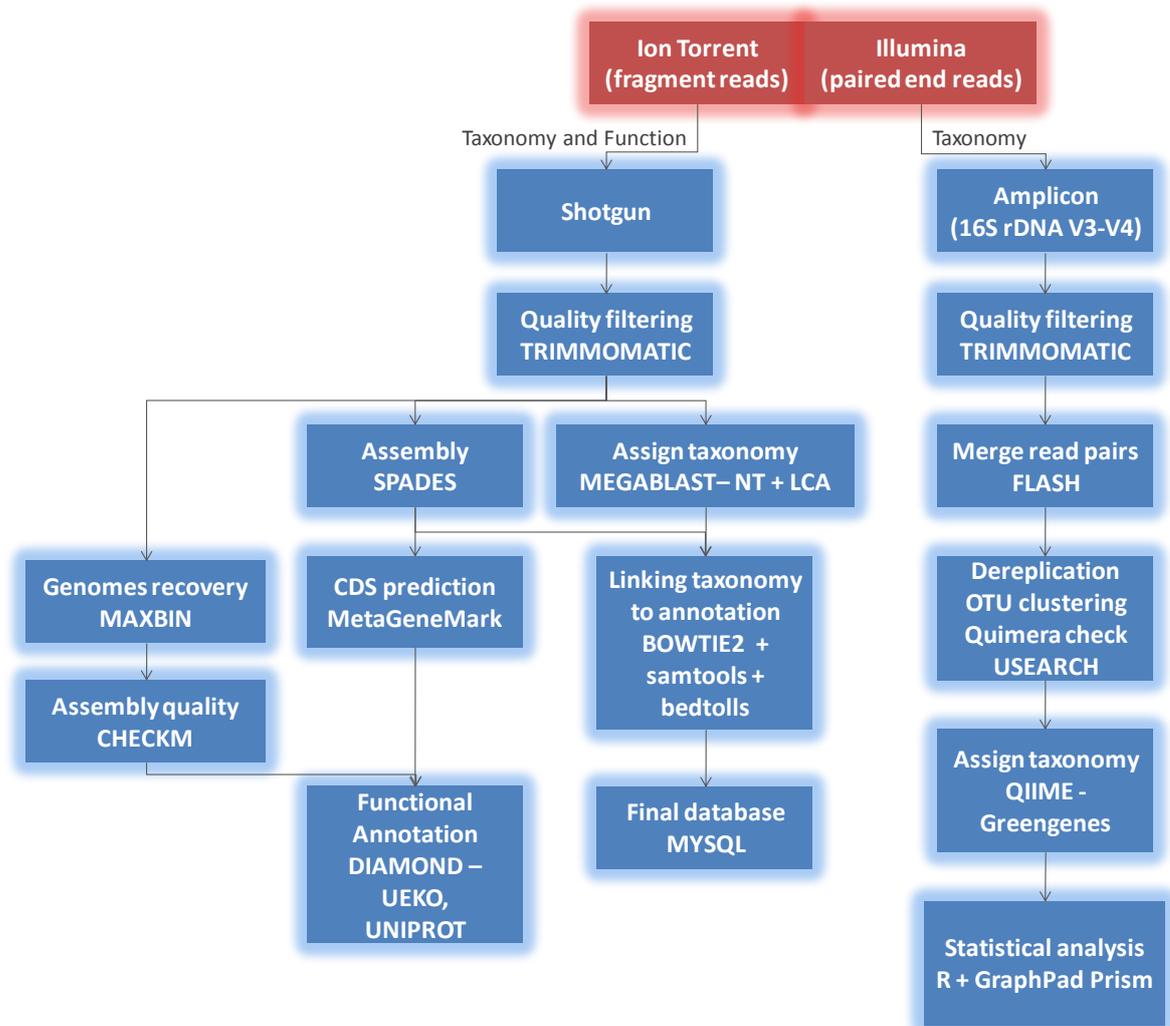


Figura 6 - Representação gráfica da análise integrada entre dados de *amplicons* e *shotgun*.

4.5 - Análise Shotgun

Uma vez concluído o sequenciamento os dados foram automaticamente transferidos para o servidor Torrent, onde foram submetidos a um processamento inicial de sinal e algoritmos de *base calling* que geram seqüências de DNA associadas a reads individuais. Os resultados foram então baixados nos formatos padrão SFF e FASTQ. A separação dos barcodes foi realizada através de plugins instalados no próprio sequenciador (SFFCreator e FasqCreator).

As análises dos dados shotgun foram divididas em três etapas: 1ª) Análise da diversidade taxonômica de todas as seqüências filtradas; 2ª) Análise

da diversidade funcional dos dados montados; 3º) Reconstrução parcial de genomas a partir dos dados metagenômicos.

Sequências de alta qualidade foram obtidas usando Trimmomatic-v.0.32⁷⁰, com qualidade média correspondente a Phred 15 em janelas deslizantes de 4 bases conforme descrito para os dados de 16S rDNA. Um filtro adicional foi implementado para remover sequências menores que 100 pares de base. O perfil taxonômico foi obtido por comparação de todas as sequências com a base de dados referência do NT (10_04), utilizando a ferramenta blastn⁸⁹. Os melhores resultados foram selecionados de acordo com os critérios: cobertura maior que 50% da referência, e-value menor que 0.001, identidade maior que 90%. O menor ancestral comum de todos os *hits* filtrados foi obtido utilizando o módulo lca do programa MEGAN v5¹³⁸ em linha de comando.

O mesmo conjunto e sequências pré-processamento foi utilizado para a realização de uma montagem de novo, baseada em grafos *deBruijn*, usando SPADESv3.6.0¹²⁵. Para esta etapa, os parâmetros foram selecionados de acordo com as recomendações dos autores. Uma variedade de valores de kmer foram testados, e apenas o melhor (K=127) foi selecionado para as etapas subsequentes. A fim de reduzir o número de inconsistências relacionadas a curtas inserções e deleções a opção *careful* também foi utilizada nas montagens. As regiões codificadoras foram preditas utilizando o programa MetaGeneMark-v1¹³⁹. As proteínas foram extraídas utilizando o script Perl *aa_from_gff.pl* do MetaGeneMark. A anotação funcional dos genes preditos se baseou em comparações com sequências previamente anotadas (ortologia). As sequências proteicas traduzidas a partir dos genes preditos puderam ser classificadas usando a ferramenta DIAMOND v0.7.10.59⁹⁹ contra a base de dados completa do UniProt¹⁴⁰. Identificadores Uniprot foram mapeados a outras classificações de famílias protéicas do TIGRFAM⁹⁶, Pfam⁹⁷, COG¹⁴¹ e Gene Ontology¹⁴². Os genes preditos também associados aos seus produtos enzimáticos e suas respectivas vias metabólicas originárias usando DIAMOND v0.7.10.59⁹⁹ contra a base de dados UEKO¹⁴³ (integração da base de dados KO¹⁰⁵ com o UniProtKB). O melhor *hit* do DIAMOND para cada proteína predita, em cada banco de dados, foi selecionado de acordo com os critérios:

cobertura maior que 50% da referência, *e-value* menor que 0.001, identidade maior que 30%, melhor valor de *bitscore*.

A ligação entre os achados taxonômicos (classificação das sequências filtradas) e funcionais (classificação das proteínas preditas) foi possível mapeando as sequências filtradas nos contigs montados usando Bowtie2 v2.2.3¹⁴⁴. O arquivo gff resultante da predição gênica foi usado para encontrar a posição inicial e final de cada gene. Para encontrar o valor de cobertura de sequências em cada base do gene, o gff foi combinado ao arquivo “.bam” do mapeamento usando as ferramentas samtools v0.1.19¹⁴⁵ e bedtools v2.17.0¹⁴⁶. A integração dos dados taxonômicos e funcionais foi realizada incorporando as tabelas de cada um dos resultados em um banco mysql, utilizando como chave primária os identificadores das sequências ou dos genes ou os KOs (figura suplementar 3).

Alguns genomas puderam ser parcialmente reconstruídos a partir do metagenoma, usando o programa MaxBin v2.1¹⁴⁷. Esse programa agrupa os *contigs* montados a partir de suas características genômicas (frequência de tetranucleotídeos e porcentagem de cobertura) e classifica-os a partir de um conjunto modelo composto por genes marcadores com cópia única. O algoritmo expectation-maximization é então utilizado para calcular a probabilidade do *contig* pertencer ao dado genoma. A avaliação da qualidade dos genomas recuperados foi realizada com o programa checkm v.1.0¹⁴⁸. Essa avaliação se baseia em um conjunto de genes mínimo para cada táxon em uma árvore de referência. Modelos Ocultos de Markov dos genes de referência são alinhados aos contigs agrupados. Existe um modelo para a classificação taxonômica e outro para a predição da integridade do genoma final (*completeness coverage*).

4.5 – Análises complementares

Para identificar táxons e funções significativamente super ou sub-representadas em cada metagenoma, foi realizada uma comparação do tipo “todos contra todos” usando ANOVA e o método de correção de Tukey (P

<0,05) com o software GraphPad Prism-v6.05¹⁴⁹. A fim de evitar a inserção de um viés técnico, anteriormente à todas as análises estatísticas, as abundâncias (nº de sequências) atribuídas a cada gene predito foram normalizadas de acordo com o tamanho do gene.

A fim de encontrar variações ambientais que influenciasses a composição microbiana, foi aplicada a análise de variância multivariada (função *anova.cca* no pacote *vegan* do R v3.2.2; teste de permutação) de modo que fossem filtradas apenas associações significativas ($p < 0.01$).

Todas as amostras coletadas foram agrupadas de acordo com o local de coleta (drenagem e barragem). A partir desses dois grupos, também foram realizadas comparações taxonômicas e funcionais, utilizando a distribuição binomial negativa presente no pacote *DESeq2*¹⁵⁰ do R v3.2.2. Nesse caso, o teste t do R também foi utilizado a fim de averiguar quais táxons ou funções estavam diferencialmente abundantes entre os dois grupos.

4.6–Análise comparativa entre abordagens para classificação taxonômica usando bibliotecas *shotgun*

A metodologia utilizada para a classificação taxonômica baseada em 16S rDNA já foi amplamente avaliada e discutida por Pyrlo e colaboradores^{134; 151; 152}. Como validação da metodologia proposta no presente estudo, para classificação usando dados do tipo *shotgun*, o *pipeline* apresentado foi comparado com as seguintes ferramentas: mapeamento simples usando *Bowtie2*¹⁴⁴, classificação baseada em marcadores moleculares *MetaPhlAn2*⁸², classificação baseada em marcadores preditos por frequência de k-mers *GSMer*¹⁴⁴. Existem muitas ferramentas disponíveis para analisar metagenomas. Para a presente análise, os seguintes critérios foram utilizados para escolher as ferramentas: (a) apenas ferramentas livremente disponíveis para *download* com versão estável e utilização local; (b) ferramentas com documentação clara e adequada; (c) ferramentas com potencial de ser amplamente adotadas pela comunidade científica.

A fim de validar as anotações taxonômicas preditas usando o pipeline aqui proposto, e compará-las a ferramentas conhecidas (previamente citadas), foi utilizado um conjunto simulado de sequências metagenômicas a partir de dados disponibilizados pelo Human Microbiome Project (NCBI SRR172903). A biblioteca simulada é do tipo escalonada, foi convertida para o formato FASTA usando a ferramenta sra toolkit do NCBI, e sua composição taxonômica está mostrada a tabela 1.

Tabela 1 - Resumo da composição taxonômica do metagenoma simulado.

Acesso	Nome da espécie	Nº de sequências de 16S rDNA por conjunto	Nº de reads mapeadas por genoma
SRR172903	<i>Acinetobacter baumannii</i> (ATCC 17978)	10000	35168
	<i>Actinomyces odontolyticus</i> (ATCC 17982)	1000	785
	<i>Bacillus cereus</i> (ATCC 10987)	100000	24502
	<i>Bacteroides vulgatus</i> (ATCC 8482)	1000	1857
	<i>Candida albicans</i> (SC5314)	1000	0
	<i>Clostridium beijerinckii</i> (ATCC 51743)	100000	97503
	<i>Deinococcus radiodurans</i> (DSM 20539)	1000	22252
	<i>Enterococcus faecalis</i> (ATCC 47077)	1000	898
	<i>Escherichia coli</i> (ATCC 70096)	1000000	0
	<i>Helicobacter pylori</i> (ATCC 700392)	10000	0
	<i>Lactobacillus gasserii</i> (DSM 20243)	10000	580
	<i>Listeria monocytogenes</i> (ATCC BAA679)	10000	6283
	<i>Methanobrevibacter smithii</i> (ATCC 35061)	1000000	307410
	<i>Neisseria meningitidis</i> (ATCC BAA335)	10000	0
	<i>Propionibacterium acnes</i> (DSM 16379)	10000	0
	<i>Pseudomonas aeruginosa</i> (ATCC 47085)	100000	200832
	<i>Rhodobacter sphaeroides</i> (ATCC 17023)	1000000	2069570
	<i>Staphylococcus aureus</i> (ATCC BAA1718)	100000	1838371
	<i>Staphylococcus epidermidis</i> (ATCC 12228)	1000000	1732753
	<i>Streptococcus agalactiae</i> (ATCC BAA611)	100000	38387
<i>Streptococcus mutans</i> (ATCC 700610)	1000000	625899	
<i>Streptococcus pneumoniae</i> (ATCC BAA334)	1000	0	

Os valores absolutos preditos por cada programa foram usados para medir, independentemente, a divergência total nos níveis taxonômicos gênero e espécie. Para calcular sensibilidade (SEM), especificidade (SPEC), valor preditivo positivo (VPP) e valor preditivo negativo (VPN), foram definidos os verdadeiros positivos (TP), falsos positivos (FP), verdadeiros negativos (TN) e falsos negativos (FN). Uma vez que sabemos a proveniência de cada

sequência, podemos obter FP como o número de sequências que foram preditas para táxons não existentes na amostra e FN como o número de sequências que não foram atribuídas aos táxons existentes na amostra. Como verdadeiros negativos foram inseridas 1.000 sequências modificadas que serão contabilizadas se permanecerem sem classificação. Verdadeiro positivo corresponde às sequências corretamente classificadas. As demais métricas de qualidade foram calculadas conforme as fórmulas:

$$\text{SEM} = \text{TP} / (\text{TP} + \text{FN})$$

$$\text{SPEC} = \text{TN} / (\text{TN} + \text{FP})$$

$$\text{PPV} = \text{TP} / (\text{TP} + \text{FP})$$

$$\text{NPV} = \text{TN} / (\text{TN} + \text{FN})$$

5- Resultados e Discussão

5.1 – Validação do pipeline para análise metagenômica de bibliotecas *shotgun*: comparação entre métodos de classificação taxonômica dependentes e independentes de marcadores moleculares

Para cada ferramenta, a abundância taxonômica predita por gênero foi comparada com a abundância conhecida da comunidade simulada: a) pelo seu número de sequências de 16S fornecida pelo *Human Microbiome Project* e calculada por RT-PCR, b) pelo número de sequências mapeadas em cada genoma de referência correspondente à cepa indicada. Uma medida importante relacionada à acurácia é a ocorrência de falsos positivos, ou seja, a frequência com que uma ferramenta prediz a presença de taxons que não foram incluídos nos conjuntos de dados. Em geral, as ferramentas são consistentes, pois conforme demonstrado na tabela 2, os resultados ao nível de gênero/espécie não diferiram significativamente entre si (1 way anova, $p < 0,05$). Vale lembrar que mesmo não sendo diferenças significativas, nenhum dos programas conseguiu prever corretamente todas as abundâncias. Curiosamente, as abundâncias referentes à proporção conhecida não foram compostas por organismos raros, mas pelos pertencentes a filos abundantes em bancos de dados como acidobacteria e proteobacteria. No caso de organismos menos representados como *Methanobrevibacter* as predições corresponderam a menos da metade da abundância original da amostra. Segundo Tsai e colaboradores¹⁵³ a classificação taxonômica baseada em bibliotecas *shotgun* é dependente da relação profundidade de sequenciamento / complexidade da amostra e da composição da comunidade (número de subespécies ou espécies com genomas semelhantes). Conseqüentemente, isso pode resultar em reconstruções fragmentadas e incompletas.

Tabela 2 - Comparação entre a abundância relativa de espécies predita por abordagens dependentes e independentes de marcadores moleculares

Gênero	% relativa de sequências de 16S rDNA	% relativa de sequências mapeadas em seus respectivos genomas de referência	% relativa de sequências identificadas usando o fluxo de trabalho proposto nesse estudo (similarity)	% relativa de sequências identificadas usando Metaplan (marker genes)	% relativa de sequências identificadas usando GSM (marker - kmer composition)
Acinetobacter	0,18	0,50	4,11	0,00	0,23
Actinomyces	0,02	0,01	0,00	0,00	5,10
Bacillus	1,80	0,35	2,88	8,07	0,04
Bacteroides	0,02	0,03	0,00	0,00	0,00
Candida	0,02	0,00	0,00	0,00	0,00
Clostridium	1,80	1,39	2,55	0,02	0,00
Deinococcus	0,02	0,32	0,00	0,00	0,04
Enterococcus	0,02	0,01	0,49	0,73	0,61
Escherichia	17,96	0,00	20,57	1,30	0,00
Helicobacter	0,18	0,00	0,00	0,00	0,00
Lactobacillus	0,18	0,01	0,80	0,19	0,15
Listeria	0,18	0,09	0,00	0,00	0,00
Methanobrevibacter	17,96	4,39	5,44	0,74	0,00
Neisseria	0,18	0,00	0,00	0,00	0,00
Propionibacterium	0,18	0,00	0,00	0,00	0,00
Pseudomonas	1,80	2,87	2,24	0,40	0,04
Rhodobacter	17,96	29,55	8,16	13,71	4,22
Staphylococcus	19,77	50,99	31,03	6,16	15,03
Streptococcus	19,77	9,49	4,01	4,02	74,50
Unclassified	0,01	-	0,02	0,05	0,02
Others	-	-	17,70	64,61	0,03

A identificação confiável dos fragmentos de cada genoma dentro do metagenoma também pode ser confundida pela natureza conservada e alta similaridade de algumas regiões como genes 16S rRNA, genes *housekeeping* e eventos de duplicação gênica ou transferência horizontal de genes. Esse ruído genômico pode ocasionar altas taxas de resultados falsos positivos¹⁵⁴. As regiões intergênicas compreendem aproximadamente 15% dos genomas bacterianos¹⁵⁵ e são normalmente descartadas da análise dos dados metagenômicos após o passo de predição gênica. O interesse atual em regiões intergênicas está focado na exploração de novas unidades funcionais, tais como pequenos RNAs, pequenas ORFs, pseudogenes, transposons, locais de ligação de integrase e elementos repetitivos¹⁵⁶. Essas regiões também estavam presentes na comunidade simulada e pareceram contribuir para a identificação taxonômica, sugerindo seus papéis na identificação de espécies microbianas. O resultado também corrobora a idéia de que amostras que foram submetidas a diferentes fluxos de trabalho de bioinformática são comparáveis

¹⁵⁷ e reforça a necessidade de uma padronização dessas análises em metagenômica ¹⁵⁸. Além disso, não foram incluídas sequências de eucariotos para essa avaliação, sendo essas observações relativas somente a procaritos.

O desempenho de cada método, dadas as métricas matemáticas valor preditivo positivo (PPV), valor preditivo negativo (NPV), SEN (sensibilidade) e SPEC (especificidade) é ilustrado na figura 7, em que valores próximos de 1 são as melhores. Em geral, os métodos apresentaram bom desempenho. O método baseado em k-mer apresentou maior especificidade, enquanto o método baseado em similaridade apresentou maior sensibilidade e precisão (PPV).

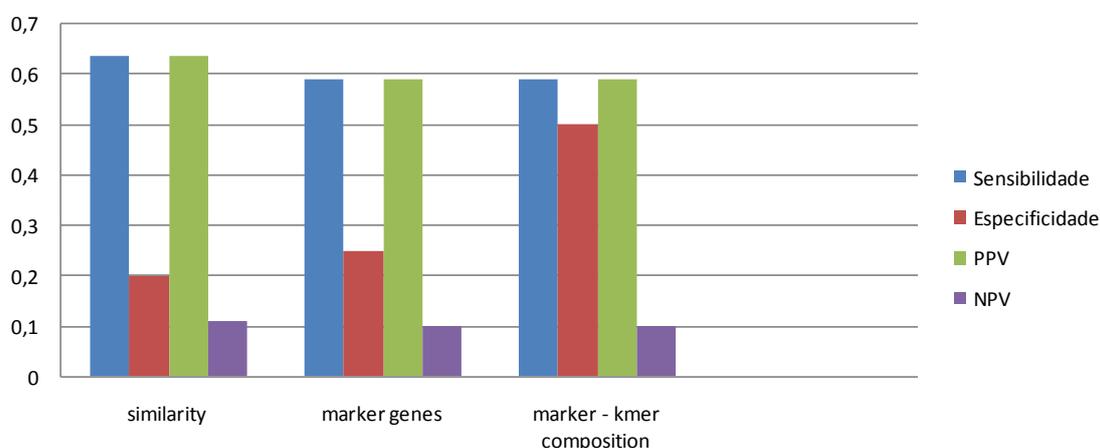


Figura 7 - Comparação entre os métodos dependente e independente de marcadores para classificação de sequências *shotgun*. As equações são descritas respectivamente por: $SEM = TP / (TP + FN)$; $SPEC = TN / (TN + FP)$; $PPV = TP / (TN+FN)$; $NPV = TN / (TN +FN)$. De modo que TP corresponde aos verdadeiros positivos, FP aos falsos positivos, TN aos verdadeiros negativos e FN aos falsos negativos.

A definição clara de cepas e espécies microbianas para cada método de classificação também é um assunto amplamente debatido. O método de classificação do fluxo de trabalho aqui proposto define estirpes e espécies com base no sistema de classificação do NCBI, em que a parte binomial da nomenclatura define uma espécie e a identificação seguinte à nomenclatura binomial define uma cepa. Em MetaPhlAn a atribuição taxonômica é realizada mapeando todas as leituras contra um conjunto de marcadores usando Bowtie,

de modo que a atribuição das nomeclaturas depende da classificação taxonômica usada durante o depósito do genoma mapeado. Em GSM (*genome specific marker*), em relação ao nó mais geral de sua árvore pré-determinada, o programa possui um conjunto de cerca de 1 milhão de marcadores. Esse conjunto é redimensionado em marcadores específicos para o clado predito como o mais próximo da amostra analisada, em que um clado pode ser tão amplo como um filo. Deste modo, as sequências de marcadores têm de ser fortemente conservadas dentro desse clado sem ser localmente semelhante às sequências fora do clado. Por esse motivo, já não era esperado o mapeamento de todas as *reads*, diminuindo seu cálculo de verdadeiros positivos e impactando em sua sensibilidade. Além disso, essa não seria a melhor abordagem para análise de ambientes extremos ou pouco estudados. Isso porque grande parte dos organismos constitutivos desses ambientes não está representada nos bancos de dados genômicos e nos marcadores utilizados por esse programa.

5.2 – Influência dos parâmetros ambientais e da sazonalidade na composição da comunidade microbiana da barragem de rejeitos e drenagem ácida de mina

A cinética dos processos geoquímicos envolvidos na geração da drenagem ácida pode ser afetada pela concentração de oxigênio na superfície da partícula de sulfeto, granulometria das partículas, superfície específica do mineral sujeito à oxidação, temperatura, concentração do íon férrico, pH, entre outros fatores ¹⁵⁹. As características físico-químicas das amostras de água e sedimento da barragem de rejeito e da drenagem ácida de mina estão representados nas Tabelas 3 e 4.

Fósforo (P) e nitrogênio (N) são os principais fatores limitantes e responsáveis pela eutrofização. Robertson (2011) ¹⁶⁰ cita que a presença de P e N nos ambientes de água doce, em concentrações superiores a 0,01 e 0,03 mg/l, respectivamente, determinam proliferações algais, enquanto que valores entre 0,005 e 0,006 mg/l (= 5 e 6 µg/l de P) são limítrofes entre o estado de Mesotrofia e Eutrofia para lagos tropicais. Neste contexto, a barragem de

rejeitos e a drenagem ácida de mina são ambientes tipicamente oligotróficos ($P < 0,1$ mg/l e $N < 3,8$ mg/l).

As concentrações de metais dissolvidos na água da barragem não excedem os valores máximos estabelecidos nas diretrizes brasileiras para enquadramento de corpos d'água dentro de padrões de qualidade (CONAMA, resoluções de 2005 ¹⁶¹). Em contraste, a água da drenagem ácida de mina apresenta diversos parâmetros físico-químicos alterados. O pH de 5,02 maior que os observados em drenagens ácidas típicas, que podem chegar a pH de 1 ¹⁶². Valores de pH em torno de 5 ou inferiores não são naturais para o ambiente em estudo, e a correção química do pH, por parte da empresa mineradora, é feita por meio da utilização de calcário. A presença de calcário nesse ambiente explica os elevados níveis de cálcio solúvel detectado das amostras de drenagem, que atingiram 200 mg/L. A elevação do pH de drenagens ácidas é uma medida remediadora comumente adotada, e visa à precipitação de metais, como Cu, Fe e Al, os quais permaneceram imobilizados no sedimento dessa drenagem, não contaminando cursos d'água da região. Outra medida adotada pela empresa é a fitorremediação, com a colocação de aguapés em toda a área.

Segundo Abu Al-Rub ¹⁶³ é bem conhecido que o pH do meio aquoso é um parâmetro importante que interfere em muitos processos metabólicos de microorganismos, como, por exemplo, o controle a captação de muitas espécies metálicas pela biomassa. Por outro lado, quando os valores de pH se elevam, a superfície dos bioabsorventes torna mais negativamente carregada com a subsequente atração dos íons metálicos com carga positiva e a bioabsorção é favorecida até o máximo a um pH em torno de 5 ¹⁶⁴.

Em virtude de seu menor coeficiente de difusão em água, a concentração máxima de O_2 dissolvido em ambiente aquático natural é cerca de 25.000 vezes menor do que a concentração de O_2 na atmosfera. Quando esse oxigênio disponível na água é consumido, inclusive na oxidação dos sulfetos, a oxidação diminui significativamente, pois sua reposição na água é relativamente lenta ¹⁶⁵. No período chuvoso (abril) a concentração de oxigênio dissolvido na água da barragem de rejeitos chega a ser duas vezes

maior que em todas as outras amostras. Apesar da diferença não apresentar resultado significativo ($p=0,8$), o aumento da disponibilidade de O_2 é uma medida simples que retrata condições mais favoráveis à oxidação de sulfetos nesse período (1M_ABR=12,6 ; 15M_ABR= 11,7).

O potencial de oxi-redução é uma referência para as taxas de oxidação das espécies iônicas de ferro, ao passo que representa a tendência da solução de ser oxidada ou reduzida, ou seja, a sua capacidade de capturar ou liberar elétrons. Nos sistemas de lixiviação, o potencial de oxi-redução da solução é determinado pela relação das concentrações de Fe^{+2}/Fe^{+3} . Essa proporção foi maior no período seco (1M_OUT= 0,19/0,07; 15M_OUT=0,38/0,06), nas águas de superfície e profundidade. A oxidação dos íons ferrosos, indicada pela elevação do potencial de oxi-redução, é uma indicação da dissolução dos sulfetos minerais e consequentemente, da solubilização do metal de interesse ¹⁶⁶.

As altas concentrações de cobre (74,7 mg/L) e sulfato (1.170 mg/L) observadas nas amostras de água de drenagem ácida de mina eram esperadas, já que a formação desse tipo de drenagem envolve a solubilização de metais e sulfatos em decorrência da oxidação de ferro e enxofre em minerais sulfetados, por microorganismos quimiolitotróficos ¹⁶⁷.

Tabela 3 - Características físico-químicas das amostras de água da barragem de rejeitos e drenagem ácida de mina nos períodos de outubro de 2012, abril de 2013 e junho de 2014. O caracter - representa medidas que não foram realizadas. Em negrito estão os valores acima dos previstos nas resoluções do CONAMA (para implementação da Política Nacional do Meio Ambiente e do Sistema Nacional do Meio Ambiente). O caracter * corresponde a parâmetros que apresentaram diferença significativa ($p < 0.01$) entre as amostras da barragem de rejeitos (superfície e profundidade de outubro : 1M_OUT,15M_OUT; superfície e profundidade de abril : 1M_ABR,15_ABR) e da drenagem ácida de mina (água superficial de junho: AMD_JUN).

Parâmetros	Unidades	1M_OUT	15M_OUT	1M_ABR	15M_ABR	AMD_JUN	CONAMA
Temperatura	°C	-	-	31,00	31,50	-	>40
O2 dissolvido	mg/mL	5,60	4,00	12,62	11,70	4,10	-
Turbidez	NTU	12,80	177,00	2,60	2,40	42,80	-
pH	NOUNIT	8,50	8,23	7,36	7,10	5,02	5 a 9
Potencial Redox	mV	-	-	310	256	-	-
Alcalinidade Total	mg CaCO ₃ /L	36,50	45,40	41,10	38,10	3,03	-
Alumínio Total	mg Al/L	0,21	0,28	0,12	0,14	0,09	-
Arsênio Total	mg As/L	<0,01	<0,01	0,01	<0,01	<0,01	0,50
Cádmio Total	mg Cd/L	<0,001	<0,001	<0,001	<0,001	<0,001	0,20
Calcio Total	mg Ca/L	58,10	61,50	68,70	66,80	219,00	-
Chumbo Total	mg Pb/L	<0,01	<0,01	<0,01	<0,01	<0,01	0,50
Cloreto	mg Cl-/L	126,00	120,00	97,80	96,90	1,94	-
Cobalto Total	mg Co/L	<0,01	<0,01	<0,01	<0,01	2,78	1,00
Cobre Total	mg Cu/L	<0,009	0,01	0,01	0,01	74,7	1,00
Cromo Total	mg Cr/L	<0,01	<0,01	<0,01	<0,01	<0,01	15,00
Ferro Ferrico	mg Fe ₃ /L	0,07	0,06	0,10	0,12	0,12	-
Ferro Ferroso	mg Fe ₂ /L	0,19	0,38	<0,05	<0,05	<0,05	-
Fluoreto	mg F-/L	0,80	0,79	0,56	0,56	<0,1	-
Fósforo Total	mg P/L	0,02	0,10	0,02	0,01	0,02	1,00
Magnésio Total	mg Mg/L	2,71	3,06	3,81	3,55	136,00	2,00
Manganês Total	mg Mn/L	<0,05	<0,05	<0,05	<0,05	18,40	-
Níquel Total	mg Ni/L	<0,02	<0,02	<0,02	<0,02	<0,02	-
Nitrogênio Amoniacal	mg N_NH ₃ /L	0,20	0,04	3,80	3,60	-	-
Potássio Total	mg K/L	36,50	36,00	33,80	33,50	9,69	-
Sódio Total	mg Na/L	160,00	156,00	158,00	176,00	65,40	-
* Sulfato	mg SO ₄ /L	275	263	160	154	1.170	-
Sulfeto	mg S ₂ -/L	<0,002	0,003	0,002	0,002	<0,002	-
Sulfito	mg SO ₃ /L	<2,00	<2,00	<2,00	<2,00	<2,00	-
Zinco Total	mg Zn/L	<0,10	<0,10	<0,10	<0,10	0,38	5,00

- representa medidas que não foram realizadas. * Diferença significativa ($p < 0.001$) entre barragem e drenagem.

Com relação aos metadados avaliados para as amostras de sedimento da barragem de rejeitos e da drenagem ácida de mina, destacam-se as altas concentrações de metais (tabela 4) quando comparadas às amostras de água.

Isso pode ser explicado pela precipitação ou adsorção dos metais que são depositados no sedimento tornando baixas as concentrações dos metais dissolvidos na água. Além disso, os sedimentos possuem papel importante no transporte e armazenamento de contaminantes, fazendo com que a sua qualidade determine a saúde ecológica do ambiente ¹⁶⁸. No sedimento da barragem os valores de cobre atingiram 1.256 ppm e na drenagem ácida estão acima do nível de detecção do método usada (valores maiores que 10.000 ppm). A alta concentração de cobre observada no sedimento da drenagem também era esperada, pois, como ressaltado anteriormente, essas drenagens foram corrigidas com calcário o que ocasionou a precipitação desse metal, antes solúvel (sulfato de cobre), formando óxidos de coloração azulada nessas drenagens. Com a correção em longo prazo, quando a concentração de sulfato excede 2.000 mg/L, ocorre a precipitação de CaSO_4 . Assim, sulfato ferroso pode ser removido com cal, sendo que o hidróxido ferroso solúvel pode ser oxidado a hidróxido férrico insolúvel pelo oxigênio dissolvido ¹⁶⁸.

Tabela 4–Características físico-químicas das amostras de sedimento da barragem de rejeito e drenagem ácida de mina nos períodos de outubro de 2012, abril de 2013 e junho de 2014. O caracter - representa medidas que não foram realizadas. Em negrito estão os valores acima dos previstos nas resoluções do CONAMA (para implementação da Política Nacional do Meio Ambiente e do Sistema Nacional do Meio Ambiente). O caracter * corresponde a parâmetros que apresentaram diferença significativa ($p < 0.01$) entre as amostras da barragem de rejeitos (sedimentos de outubro e abril: SED_OUT, SED_ABR) e da drenagem ácida de mina (sedimento de junho: AMD_JUN_SED).

Parâmetros	SED_OUT	SED_ABR	AMD_JUN_SED	CONAMA
C	0,51	0,45	0,91	-
Ag	<0,01	0,21	2,85	-
Al	2,88	2,62	3,28	-
As	2	<1	1	5,9
Au	<0,1	-	-	-
B	<10	<10	<10	-
Ba	65	93	98	-
Be	0,7	1	2,4	-
Bi	0,08	0,08	0,08	-
Ca	3,3	3,59	0,54	-
Cd	0,04	0,05	0,12	0,6
Ce	131,57	228,02	139,49	-
Co	53	67,6	57,9	-
Cr	65	49	62	37.3
Cs	0,52	0,62	0,55	-
*Cu	1256,1	1112	>10000	35.7
Fe	6,55	7,04	7	-
Ga	19,2	22,5	17,5	-
Ge	0,1	<0,1	0,2	-
Hf	0,21	<0,05	<0,05	-
Hg	<0,01	0,01	<0,01	-
In	0,04	0,05	0,05	-
K	0,48	0,53	0,28	-
La	79,5	124,7	84,6	-
Li	12	14	17	-
Lu	0,23	0,28	0,5	-
Mg	2,26	2,19	1,65	-
Mn	0,04	0,04	337	-
Mo	3,81	4,25	6,11	-
Na	0,13	0,2	0,04	-
Nb	0,53	0,2	1,32	-
Ni	399,1	395,9	232,1	18
*P	5217	5084	1233	2000
Pb	5,8	5,5	8,3	35
Rb	36,2	51,9	29,2	-
Re	<0,1	<0,1	<0,1	-
S	0,2	0,16	0,6	-
Sb	0,19	0,26	<0,05	-
Sc	9,7	12,7	13,4	-
Se	<1	2	2	-
Sn	3,8	4,5	4,4	-
Sr	42	43,2	19,7	-
Ta	<0,05	<0,05	<0,05	-
Tb	0,65	3,53	1,28	-
Te	0,34	0,47	0,19	-
Th	11,8	13,4	17,3	-
Ti	0,16	0,15	0,11	-
Tl	0,18	0,08	0,11	-
U	9,98	9,91	7,52	-
V	135	151	116	-
W	0,7	0,4	0,7	-
Y	18,76	23,79	41,91	-
Yb	1,5	2,2	3,6	-
Zn	23	24	53	123
Zr	3,8	4,4	9,2	-

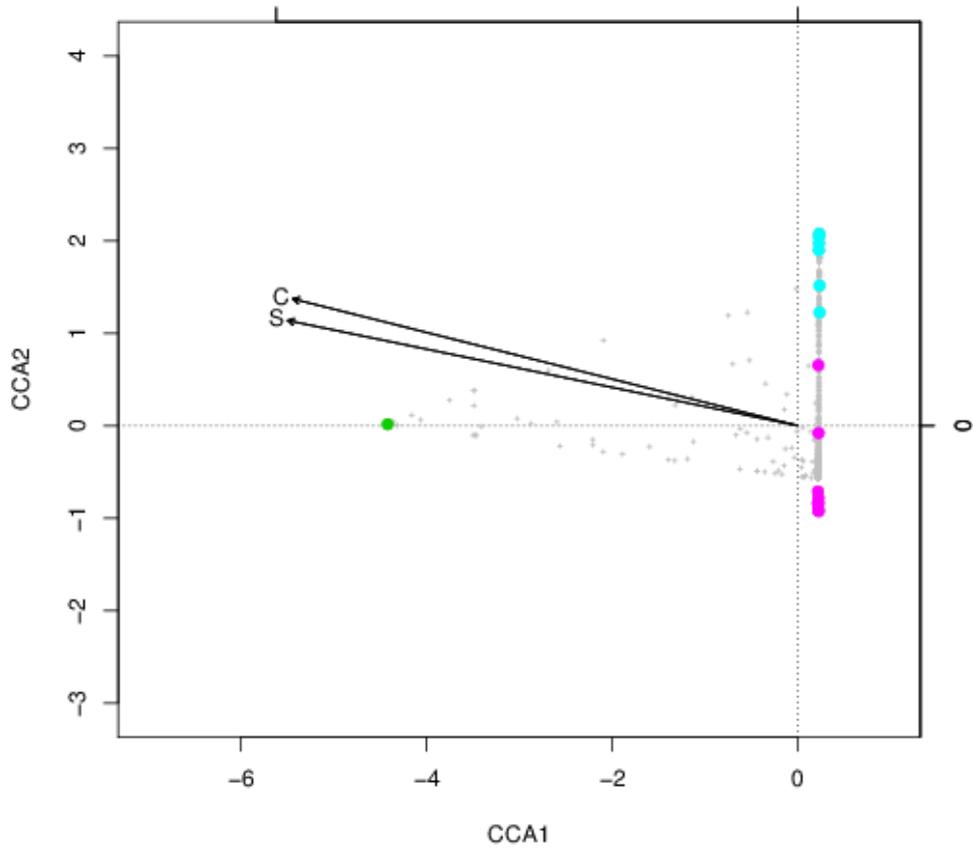
O perfil taxonômico, que será discutido no próximo tópico, foi correlacionado com os parâmetros físico-químicos por análise de correspondência canônica (CCA), agrupando as amostras em quatro habitats distintos: (i) amostras de água coletadas na estação chuvosa, (ii) água coletada na estação seca, (iii) sedimentos do período chuvoso, e (iii) de sedimentos da estação seca, com exceção de uma amostra de profundidade da água da barragem coletada na estação seca (iv) amostras de drenagem ácida de mina. Alcalinidade, turbidez, pH, O₂ dissolvido, concentrações de cálcio e alumínio foram as principais variáveis ambientais que influenciam a composição da comunidade microbiana da água, principalmente nas amostras da barragem de rejeitos (Figura 8 B). Enquanto as concentrações de carbono e enxofre foram os principais metabólitos correlacionados ao sedimento (Figura 8 A). Os valores de pH verificados podem estar relacionados a adição de surfactantes, como ditiofosfatos e xantatos usados no processo de flotação.

Um teor mais baixo de O₂ foi observado no período chuvoso. Segundo Méndez ¹⁶⁹ a concentração de O₂ cai conforme aumenta a distância da interface ar-água, enquanto a interface água-sedimento pode sofrer perdas substanciais de oxigênio, em razão de atividade bacteriana ¹⁷⁰. Os valores de turbidez mais altos no período chuvoso eram esperados. Há um aumento considerável da concentração de material em suspensão na água visto que este é o período que normalmente apresenta um aumento no escoamento superficial após intensas precipitações, e conseqüentemente um aumento de material drenado para dentro dos cursos d'água. O aumento dos valores de turbidez pode ainda influenciar em uma diminuição considerável da atividade fotossintética.

Além disso, no sedimento, a matéria orgânica poderia complexar traços metálicos e outros contaminantes. Sua degradação, seja aeróbia ou anaeróbia, influencia nos processos de oxirredução do ambiente. Em ambientes anóxicos, por exemplo, bactérias redutoras de sulfato entram em ação diminuindo a carga de matéria orgânica, o que pode promover o sequestro e imobilização de diversos metais (Zn, Cd, Pb, Ni e Cu) na forma de sulfetos secundários. Assim, é possível afirmar que sedimentos que apresentam abundância de sulfato e escassez de matéria orgânica, apresentarão uma baixa concentração de

sulfetos, já que ocorrerá pouca sulfidização pela falta de matéria orgânica a ser oxidada ¹⁷¹.

A



B

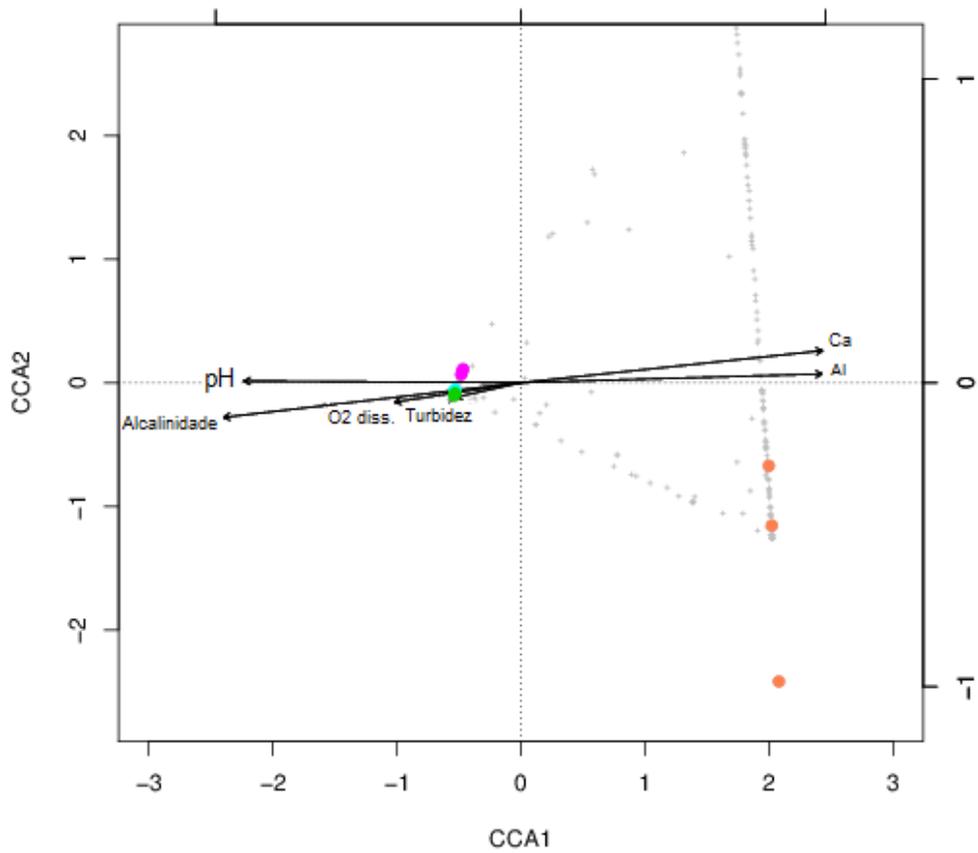


Figura 8 - Análise de correspondência canônica (CCA) entre a abundância do perfil taxonômico ao nível de gênero e os parâmetros físico-químicos. As OTUs estão representadas com o símbolo +. As amostras estão representadas com círculos, coloridos de acordo com o ambiente a qual pertencem. Os parâmetros físico-químicos, com maior correspondência aos perfis de cada grupo de amostras, estão representados com os vetores. A) Amostras de sedimento: OUT (azul), ABR (rosa), AMD (verde); B) Amostras de água: 1M_OUT (rosa), 15M_OUT (azul), 1M_ABR e 15M_ABR (verde), AMD (alaranjado).

5.3–Análise de alfa e beta diversidade em ambientes de mineração: buscando unidades taxonômicas operacionais importantes para as comunidades microbianas da barragem de rejeitos e drenagem ácida de mina

O número de sequências obtidas, após o filtro de qualidade, foi superior a 300 mil por amostra (tabela suplementar 2). A curva de rarefação é uma ferramenta que permite avaliar a profundidade da amostragem realizada. Para a maioria das amostras analisadas, a curva de rarefação resultante atingiu um platô (figura 9), indicando que a amostragem e o sequenciamento foram suficientes para representar os principais grupos taxonômicos do ambiente. É interessante considerar que as amostras de sedimento da barragem representam as curvas da parte superior da figura, o que indica uma maior diversidade desse ambiente em relação aos demais ambientes representados. Existe uma tendência dos solos mais argilosos, como os Amazônicos, em apresentar maiores quantidades de biomassa microbiana, por serem capazes de promover maior retenção de umidade por formar complexos organo-minerais e servindo de tampão às mudanças de pH ¹⁷². Inesperadamente a amostra proveniente de drenagem ácida também apresentou maior número de unidades taxonômicas identificadas em comparação com outras drenagens ácidas descritas com pH próximo a 4¹⁷³. Essa maior riqueza pode estar relacionada a inúmeros fatores como o período chuvoso em que a amostra foi coletada, à correção de pH adotada pela empresa nas proximidades ao local de coleta ou ao tempo de formação dessa drenagem.

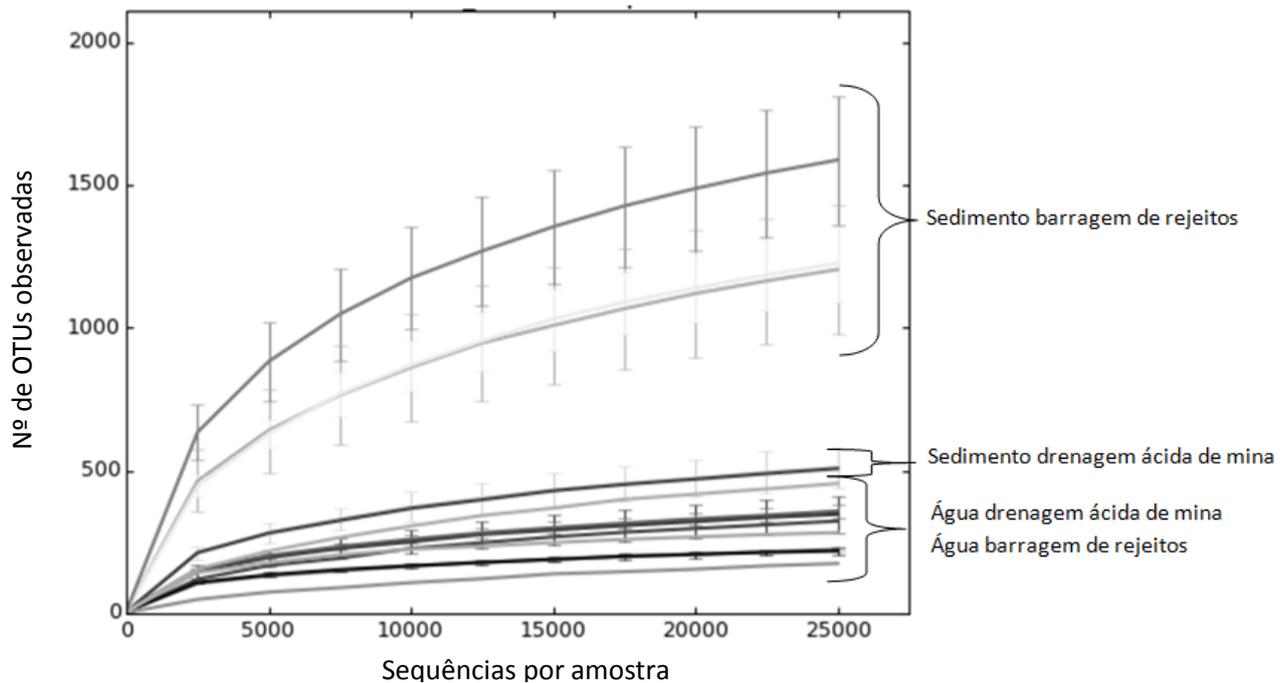


Figura 9 - Curva de rarefação para as amostras analisadas. O eixo X apresenta o número de sequências por subset. O eixo Y apresenta o número de OTUs observadas a cada subset. Cada amostra é representada por uma curva no gráfico. As descrições das amostras estão apresentadas ao lado das respectivas curvas.

Alguns índices de alfa e beta diversidade foram calculados a fim de verificar diferenças quantitativas entre as comunidades (tabela 5). O índice de Shannon fornece a idéia do grau de incerteza em prever qual seria a espécie pertencente a um indivíduo da população, de modo que quanto maior o valor de seu H' , maior a diversidade da área em estudo. Esse índice considera peso igual para as espécies raras e abundantes, e em ecossistemas reais varia entre 1.5 e 3.5, raramente ultrapassando 4¹⁷⁴. Devido ao elemento logarítmico da função, não é possível afirmar se pequenas diferenças numéricas são traduzidas em uma grande ou pequena diversidade real. Isso porque o valor de H' pode ser influenciado tanto por sua riqueza (número de espécies identificadas) quanto por sua equitabilidade (proporção de indivíduos atribuídos a cada espécie em relação ao total de indivíduos na comunidade). Ainda sim,

esse índice é rotineiramente usado em estudos de ecologia e particularmente útil para comparação de ecossistemas semelhantes ¹⁷⁵.

O índice de Simpson é derivado da teoria de probabilidades. A partir dele é possível averiguar, por exemplo, se, ao coletar aleatoriamente 2 indivíduos da comunidade, estes obrigatoriamente pertencerão a 2 espécies diferentes. Neste contexto, valores mais próximos de 1 indicam menor diversidade amostral, de modo que uma comunidade de espécies com maior diversidade deve possuir uma menor dominância. Os resultados para as amostras de água sugerem que este ambiente é ligeiramente menos diverso (diversidade moderada) que o sedimento. Vale lembrar que o índice de Shannon dá maior peso a riqueza de espécies do que o índice de Simpson ¹⁷⁶. A drenagem ácida apresenta um maior número de espécies que o sedimento da barragem, porém menores índices de Shannon e Simpson, pelo maior grau de incerteza estatística na atribuição das unidades taxonômicas por uma maior diversidade amostral (grupos raros que podem não influenciar na conservação e estrutura da comunidade). Além disso, foi possível observar uma diminuição da equitabilidade ao longo do gradiente longitudinal (menor proporção de indivíduos dentro de cada espécie em relação ao total de indivíduos da superfície para o sedimento).

Tabela 5 - Principais medidas de alfa diversidade observadas por amostra. Não foi encontrada nenhuma diferença significativa entre as amostras para os parâmetros aqui apresentados.

	Nº de sequências filtradas	Nº de OTUs	Dominância	Simpson	Shannon	Evenness	Equitabilidade	Chao-1
1M.OCT.1	862.116	107,000	0,015	0,985	4,411	0,769	0,944	107,000
1M.OCT.2	751.365	89,000	0,016	0,984	4,290	0,820	0,956	89,000
1M.OCT.3	697.272	79,000	0,018	0,983	4,194	0,839	0,960	79,000
15M.OCT.1	586.789	45,000	0,029	0,971	3,679	0,880	0,966	45,000
15M.OCT.2	811.531	73,000	0,021	0,979	4,073	0,805	0,949	73,000
15M.OCT.3	717.446	42,000	0,032	0,968	3,583	0,856	0,959	42,000
1M.APR.1	472.386	54,000	0,032	0,969	3,702	0,750	0,928	54,000
1M.APR.2	451.205	58,000	0,029	0,971	3,792	0,765	0,934	58,000
1M.APR.3	696.408	91,000	0,020	0,980	4,184	0,721	0,927	91,000
15M.APR.1	945.350	52,000	0,028	0,972	3,741	0,811	0,947	52,000
15M.APR.2	642.310	74,000	0,069	0,931	3,039	0,282	0,706	74,000
15M.APR.3	1.051.774	127,000	0,015	0,985	4,477	0,693	0,924	127,000
SED.OCT.1	433.093	166,000	0,008	0,992	4,927	0,831	0,964	166,000
SED.OCT.2	631.325	120,000	0,011	0,989	4,622	0,848	0,966	120,000
SED.OCT.3	656.756	113,000	0,011	0,989	4,624	0,902	0,978	113,000
SED.OCT.4	573.519	189,000	0,008	0,992	4,995	0,782	0,953	189,000
SED.OCT.5	539.001	129,000	0,009	0,991	4,752	0,898	0,978	129,000
SED.OCT.6	673.523	191,000	0,009	0,991	4,896	0,700	0,932	191,000
SED.APR.1	544.771	131,000	0,011	0,989	4,640	0,791	0,952	131,000
SED.APR.2	630.150	157,000	0,009	0,991	4,878	0,837	0,965	157,000
SED.APR.3	359.331	104,000	0,017	0,983	4,160	0,616	0,896	104,000
SED.APR.4	322.311	159,000	0,011	0,989	4,673	0,673	0,922	159,000
SED.APR.5	375.613	151,000	0,009	0,991	4,860	0,855	0,969	151,000
SED.APR.6	296.627	242,000	0,008	0,992	4,990	0,607	0,909	242,000
SED.APR.7	335.913	75,000	0,018	0,982	4,115	0,816	0,953	75,000
SED.APR.8	338.896	150,000	0,009	0,991	4,812	0,820	0,960	150,000
SED.APR.9	335.591	205,000	0,008	0,992	4,982	0,711	0,936	205,000
AMD.APR.1	862.116	127,000	0,014	0,986	4,484	0,697	0,926	127,000
AMD.APR.2	751.365	77,000	0,032	0,968	3,619	0,484	0,833	77,000
AMD.JUN 1	171.802	76,000	0,020	0,980	4,116	0,807	0,950	76,000
AMD.JUN 2	86.986	61,000	0,021	0,979	3,992	0,888	0,971	61,000
AMD.SED.JU	458.764	185,000	0,009	0,991	4,907	0,731	0,940	185,000
N								

Dominância: $1 - \text{Simpson}$. $D = \sum ((n_i / n)^2)$, onde n_i é o número de indivíduos do taxon i . Varia de 0, onde todos os táxons são igualmente presentes, a 1, onde um taxon domina a comunidade completamente.
Simpson: $1 - D$. Mede a "uniformidade" da comunidade, variando de 0 a 1. Observe a confusão na literatura, muitos confundem dominância e índice de Simpson.
Shannon: $H = -\sum ((n_i/n) \ln(n_i/n))$. Índice de diversidade que leva em consideração o número de indivíduos, bem como número de taxons. Varia de 0 para comunidades com apenas um único taxon para valores máximos para as comunidades com muitos taxons, cada um com alguns indivíduos.
Evenness: (D/S) . Representa a uniformidade das amostras.
Equitability: H/S , mede a regularidade com que os indivíduos são divididos entre os táxons presente.
Chao-1: $S + N_{-1}(N_{-1}-1)/(2*(N_{-2}+1))$, onde S é o número de espécies observadas, N_{-1} são os *singletons* e N_{-2} os *doubletons*. Estima a riqueza total a partir do número de espécies representadas por apenas um indivíduo (*singletons*) e por dois indivíduos (*doubletons*) nas amostras.

Vale destacar que nenhuma comunidade apresentou uma diversidade máxima, de modo que todas as unidades fossem igualmente abundantes. Para o número total de grupos, a maior parte dos seus constituintes é rara (por conter poucos indivíduos representados), enquanto poucos são dominantes (comuns; o que pode significar altos valores de importância). Com base nessa observação, foram testados diferentes modelos de abundância de espécies, que utilizam o conjunto total dos valores de importância de cada espécie, plotando tais valores contra um *ranking* de abundância (as espécies são

plotadas das mais comuns para as mais raras). A linha de referência que melhor se ajusta aos pontos plotados (menor valor de AIC – Aikaike Information Criterion) corresponde ao modelo que explica a diversidade encontrada. Quatro modelos principais foram avaliados neste estudo: a série geométrica (preemption), a série logarítmica, a distribuição logaritmo-normal, e o modelo Zipf. Conforme apresentado nas Figuras 10 e 11, as amostras se ajustaram melhor, na maioria dos casos à série Mandelbrot. Os modelos Mandelbrot e Zipf são interpretados como o processo de sucessão no qual espécies colonizadoras tardias têm as maiores necessidades de um nicho especializado e por isso são mais raras que as espécies colonizadoras iniciais. Esse seria um indício de que este ambiente se encontra em constante transformação, já que a mina está em atividade há apenas 12 anos e o aporte de rejeitos minerais (tipo e teor do minério) e as condições climáticas mudam ao longo do ano ¹⁷⁷. Isso significa que a presença de uma espécie depende de condições físicas e ecológicas prévias. As espécies pioneiras possuem baixo custo ecológico, sendo pouco exigentes (como exemplo os microorganismos autotróficos), enquanto as espécies tardias têm o custo ecológico elevado de energia, tempo e organização do ecossistema antes de se estabelecerem (como exemplo os organismos heterotróficos) ¹⁷⁸.

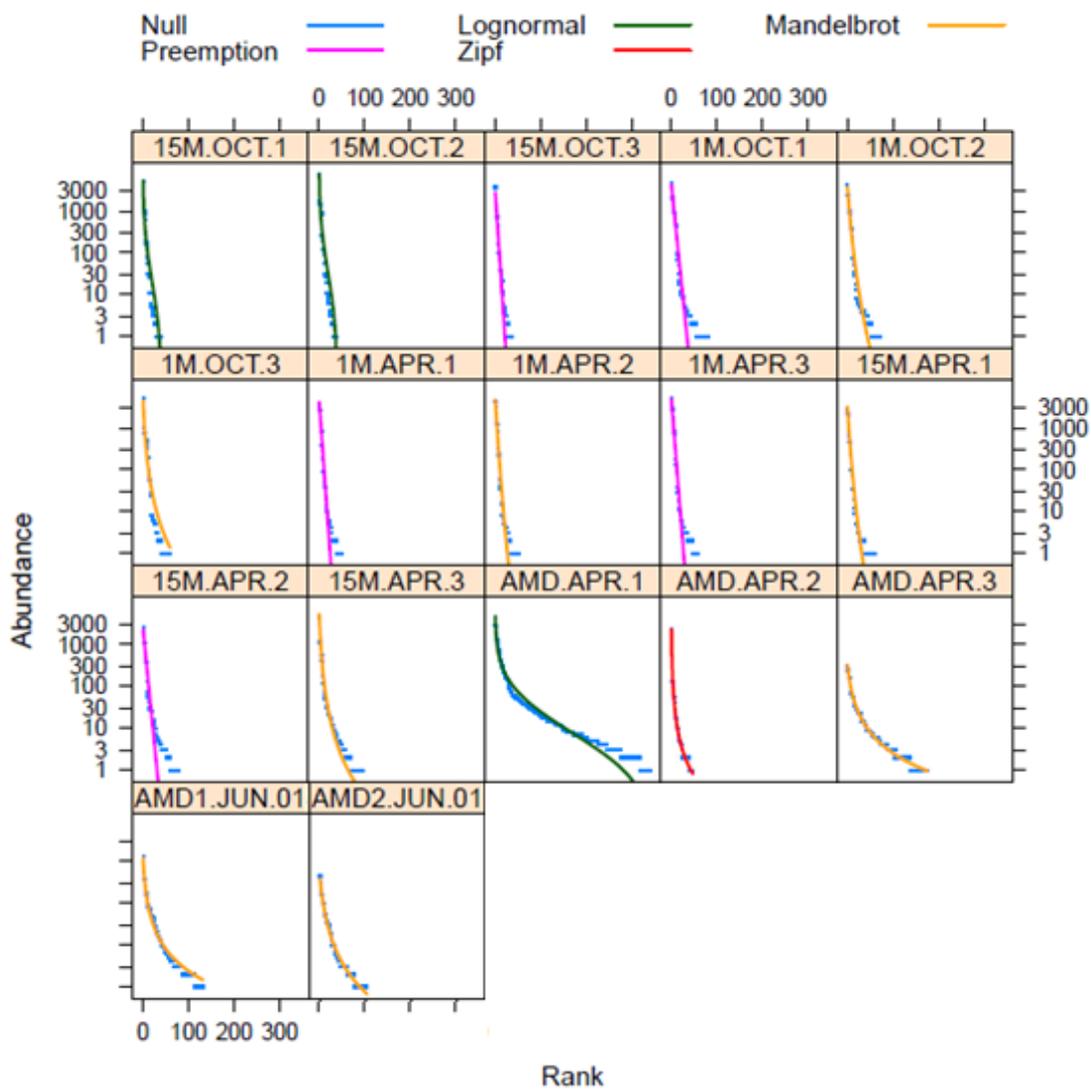


Figura 10 - Modelos de distribuição de táxons para amostras de água, baseado nas bibliotecas de 16S rDNA.

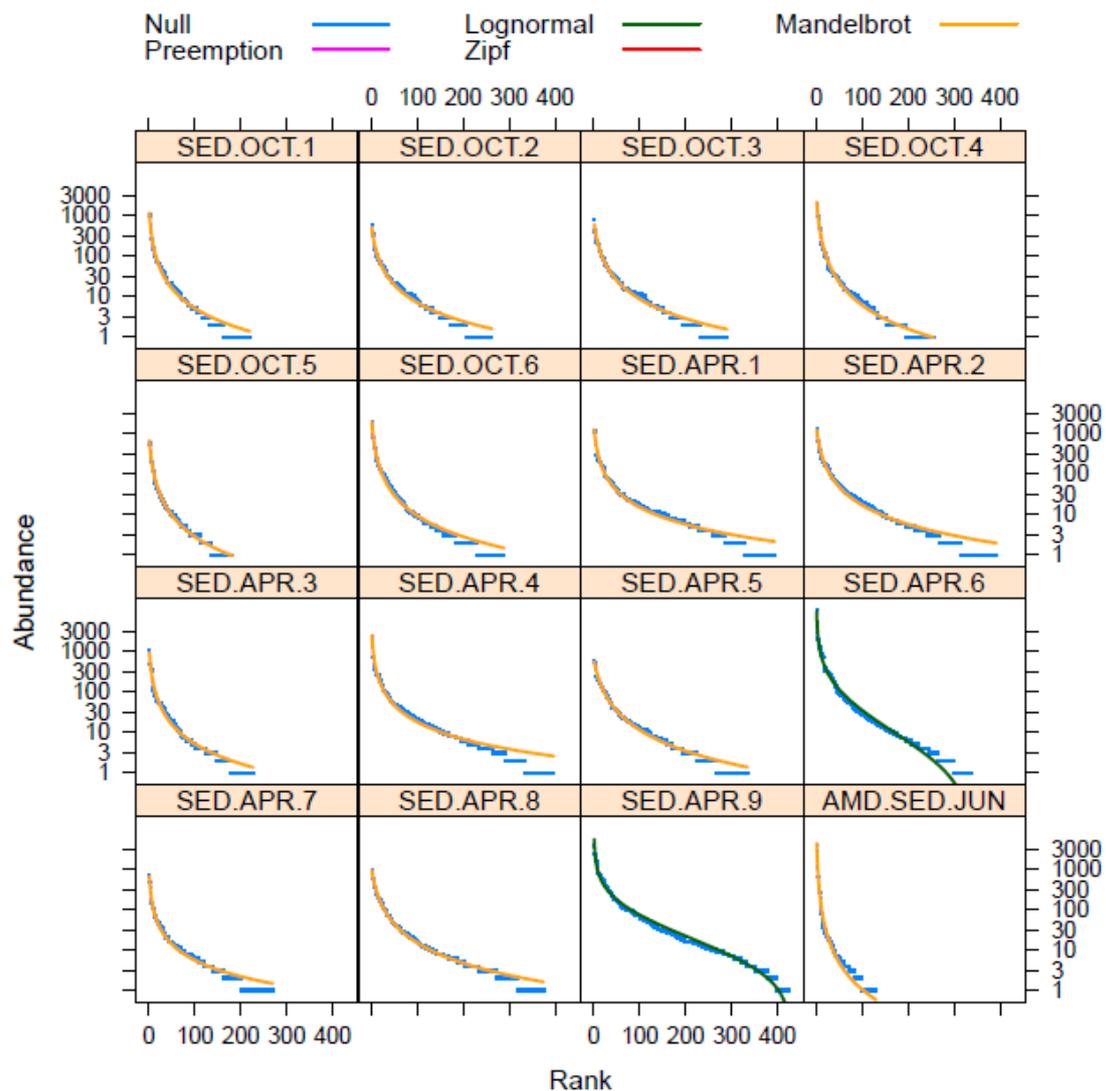


Figura 11 - Modelos de distribuição de táxons para amostras de sedimento, baseado nas bibliotecas de 16S rDNA.

Duas das três amostras de profundidade da água da barragem de rejeitos de outubro (período seco) se adequaram ao modelo Log-normal. Esse modelo indica padrões mais complexos de sobreposição de nicho e uma uniformidade intermediária. Nesse modelo a maioria das espécies possui abundâncias intermediárias, não sendo nem raras, nem comuns. Deste modo, a distribuição logaritmo-normal provavelmente deriva de espécies coexistindo em condições de competição parcial, em vez de direta, com adaptações

promovendo uma diferenciação de nicho sem uma exclusão competitiva no habitat.

O número de sequências que puderam ser classificadas nas diferentes amostras variou de 80 a 95%. Considerando as sequências que tiveram algum hit com sequências nos bancos de dados usados, houve um predomínio significativo do domínio Bacteria em todos os ambientes de barragem de rejeitos, com exceção de uma amostra de água de profundidade, em que o domínio Archaea representou cerca de 27%. Nas amostras AMD_ABR_1 e AMD_ABR_2 de drenagem ácida de mina também foi possível observar uma alta representatividade de Crenarchaeota, Euryarchaeota e Parvarchaeota. Essa diferença pode ser explicada pelos parâmetros de acidez e temperaturas supracitados, pelos quais a drenagem ácida se caracterizou como um ambiente favorável à sobrevivência de organismos extremófilos, como os da família Halobacteriaceae, os da ordem WCHD3-30 e da classe Thermoplasmata E2, esses dois últimos descritos em ambiente hipersalino. Além disso, no *sorting* celular, a ser descrito posteriormente, foram separadas células com baixa concentração de DNA, possivelmente células ARMAN-like do filo Parvarchaeota, descritas anteriormente em AMD^{179; 180}.

A comunidade microbiana presente nos diferentes ambientes e épocas do ano é diferente já ao nível de filo (figura suplementar 1). Essa diferença também está evidenciada na análise de agrupamento das amostras. Através dos resultados observados na figura suplementar 2, pode-se perceber que houve a formação de três grupos distintos sendo eles compostos por amostras de sedimento de barragem, água de barragem e AMD.

A Figura suplementar 1 apresenta um panorama geral ao nível de filo, e algumas classes de Proteobacteria, presentes na comunidade microbiana das diferentes amostras estudadas. Com relação à comunidade microbiana da água da barragem de rejeito, os grupos taxonômicos predominantes foram Actinobacteria seguido de Betaproteobacteria e Bacteroidetes, com exceção da amostra da superfície de abril de 2013, em que Alphaproteobacteria foi mais prevalente que Bacteroidetes. É interessante considerar que em abril (período chuvoso) houve um aumento na incidência do Filo Verrucomicrobia.

Actinobacteria e Proteobacteria são filos numericamente dominantes em uma variedade de sistemas de aquáticos, tais como rios, mares, água salobra ¹⁸¹. Proteobacterias são os principais membros que contribuem para o metabolismo de ferro e enxofre em comunidades de drenagem ácida de mina. Em geral Bacteroides são importantes decompositores de matéria orgânica e estão diretamente envolvidos no ciclo de carbono. Além disso, foi possível notar um aumento de Alphaproteobacteria, intimamente relacionado com um aumento do gênero *Acidiphilium*, na água na estação chuvosa. Estas são bactérias redutoras de ferro, e esse aumento pode ter relação com a maior solubilidade do ferro durante este período. A maior abundância de Betaproteobacteria em outubro, período seco, já havia sido descrita em lagos por Pernthaler e colaboradores ¹⁸². No ambiente de drenagem ácida os microorganismos oxidadores de ferro e nitrito do filo Nitrospirae (ex. *Nitriliruptorales*, *Nitrosomonadales*, *Nitrospirales*) estavam diferencialmente representados. A oxidação anaeróbica da amônia com nitrito levando a N₂ (ANAMMOX) é uma reação mediada por microorganismos que apresenta grande potencial para remoção de nitrogênio de águas residuárias¹⁸³.

As amostras de sedimento de barragem e drenagem ácida aparentam ser ambientes mais diversos. Na barragem provavelmente devido aos microorganismos e matéria orgânica depositada a partir da camada superior de água, fornecendo uma matriz mais complexa de nutrientes e regiões anóxicas, e dando suporte a diferentes tipos de metabolismo para o crescimento microbiano ¹⁸⁴. Dentro desse grupo, apesar das amostras apresentarem o mesmo perfil taxonômico geral, foram encontradas diferenças nas abundâncias relativas dentro de cada táxon considerando as duas épocas do ano. Em outubro, o filo mais representado foi Firmicutes seguido de Bacteroidetes, Betaproteobacteria e Delta/Epsilonproteobacteria. Os dois filos pouco caracterizados WS5 e Elusimicrobia também estavam mais representados na estação seca. Em abril, houve um aumento dos grupos Verrucomicrobia, Gammaproteobacteria, Betaproteobacteria, Acidobacteria e Planctomycetes. Verrucomicrobia é um grupo ubíquo e abundante em ecossistemas terrestres ¹⁸⁵, incluindo na floresta amazônica ¹⁸⁶, bioma natural da região de Carajás. A ecologia deste grupo não é clara, mas eles têm sido descritos recentemente

com potenciais papéis na oxidação de metano ¹⁸⁷. Aumentos de matéria orgânica (nitrogênio total e carbono total), mais disponíveis na estação chuvosa, pela mistura da coluna de água, foram significativamente associados com a incidência de Verrucomicrobia e Acidobactéria ¹⁸⁸. Cianobactérias fotoautotróficas são comuns em ambientes aquáticos com pH neutro ou ligeiramente alcalino e estão frequentemente associadas à presença de partículas sólidas. Planctomycetes foram detectados em ambientes terrestres, tais como solos e sedimentos. Firmicutes aparecem como a maior parte de todas as bactérias descobertos em sedimentos de lagos contaminados por metais ^{189; 190}.

No sedimento da barragem e da drenagem ácida foram encontradas importantes famílias de bactérias redutoras de sulfato a enxofre como *Peptococcaceae*, *Thermodesulfovibrionaceae*, *Desulfobacteriaceae*, *Desulfobulbaceae*, *Nitrospinaceae*. A redução dissimilativa de sulfato está diretamente relacionada ao aumento da alcalinidade e a formação de sulfetos, que são aspectos de grande interesse nos processos de biorremediação ¹⁹¹. Dentro da família Peptococcaceae diferentes OTUs foram identificadas, essas OTUs estavam representadas por membros dos gêneros *Desulfurispora*, *Desulfotomaculum* e *Desulfitobacter*. Esta última mais representada nas amostras do sedimento de abril, enquanto Desulfuromonadales são mais representados no sedimento da barragem de outubro.

A fim de averiguar se a alta abundância das OTUs mostradas na Figura suplementar 3 também correspondiam às OTUs significativamente super representadas em cada um dos grupos, foi realizado o teste DESeq com base em um modelo de distribuição binomial negativa. Essa análise comparou a abundância de OTUs em todas as amostras de água da barragem de rejeitos (quadrante superior direito da Figura 12 A) contra todas as amostras de água da drenagem ácida de mina (quadrante inferior direito da mesma figura). Em outra análise, foram comparadas todas as amostras de sedimento da barragem (quadrante superior direito da Figura 12 B) contra todas as amostras de sedimento da drenagem ácida de mina (quadrante inferior direito da mesma figura).

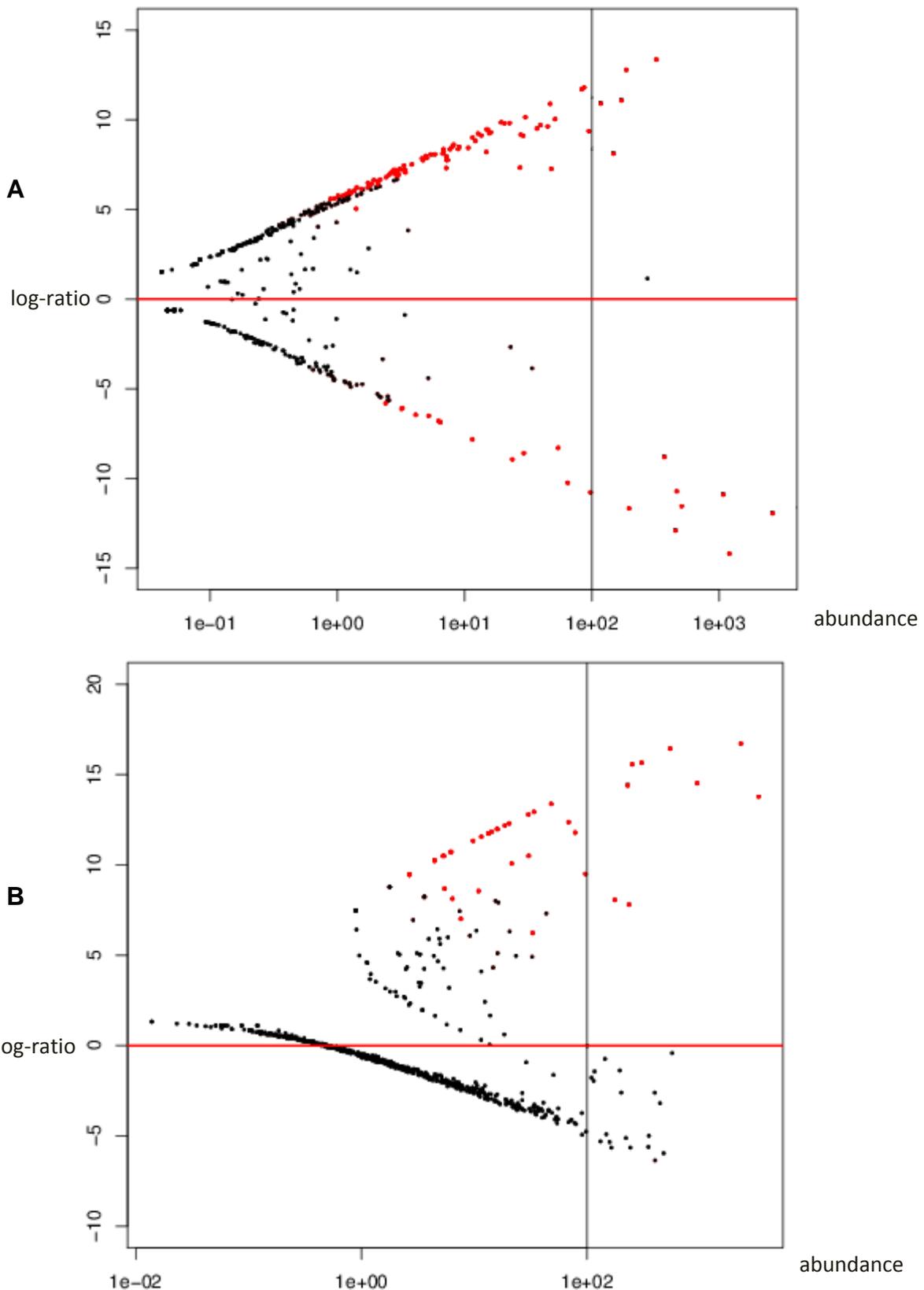


Figura 12- Comparação entre a abundância de OTUs usando DESeq2. Os círculos plotados representam as OTUs. O eixo x mostra a abundância de cada OTU plotada. O eixo y representa os valores de log-ratio (medida na escala logarítmica relacionada a quantas vezes as OTUs plotadas são mais abundantes em um grupo quando comparado ao outro). A linha vermelha do eixo X divide os conjuntos de amostras da seguinte forma: A) Entre amostras de água da barragem de rejeiros (quadrante superior) e de água de drenagem ácida de mina (quadrante inferior). B) Entre amostras de sedimento da barragem de rejeitos (quadrante superior) e de sedimento de drenagem ácida de mina (quadrante inferior). Em vermelho estão indicadas as OTUs diferencialmente abundantes. A direita da linha preta (quadrantes superior e inferior) do eixo X estão mostradas OTUs com abundância superior a 100 seqüências de 16S rDNA.

Na água da barragem de rejeitos estavam diferencialmente abundantes as seguintes OTUs: 01 OTU da ordem Actinomycetales, da família ACK-M1; 01 OTU de arquea da família Methanomassiliicoccaceae; 01 OTU de Crenarchaeota da ordem NRP-J; 01 OTU da classe Betaproteobacteria; 01 OTU Unassigned. Estudos recentes têm levantado a importância de membros da família *Methanomassiliicoccaceae* envolvidos na digestão anaeróbia de resíduos orgânicos¹⁹². Foi identificada uma interação entre espécies pouco descritas, chamada "sintrofia elétrica", que parece acelerar a metanogênese por uma rápida transferência de elétrons entre os organismos, através de correntes elétricas formadas por minerais semicondutores de óxido de ferro¹⁹³. Outro grupo importante e igualmente pouco caracterizado são as actinobactérias ACK-M1, que parecem apresentar quimiotaxia para regiões ricas em amônio¹⁹⁴.

Para as amostras de sedimento só houve diferença significativa na abundância de OTUs pouco descritas da barragem de rejeitos: 02 OTUs do gênero *Kaistobacter*; 02 OTUs Unassigned; 01 OTU da ordem Methylophilales; 01 OTU da família Acidobacteriaceae; 01 OTU de Actinobacteria da ordem 0319-7L14; 01 OTU do filo TM6, classe SJA-4; 01 OTU da classe Betaproteobacteria; Na água da drenagem ácida de mina estavam super representados: 02 OTUs da ordem Acidimicrobiales, família C111; 01 OTU Unassigned; 01 OTU da ordem Phycisphaerales (filo Planctomycetes); 01 OTU da ordem Methylophilales, família LD19 (Verrucomicrobia); 01 OTU do gênero *Synechococcus* (Cyanobacteria); 01 OTU da ordem Actinomycetales (Actinobacteria), família ACK-M1; 01 OTU da família Chitinophagaceae. Muitas das OTUs acima estão relacionadas apenas a membros não cultiváveis e pouco descritos. Em estudo filogenético baseado em 16S, os membros da família C111 aparecem próximos às oxidadoras de ferro da espécie *Acidimicrobium ferrooxidans*¹⁹⁵. A família Chitinophagaceae apresenta uma abundância maior nas amostras de sedimento da barragem de rejeitos (SED OCT 18.784.327 e SED APR 27.195.908 sequências de 16S rDNA) e drenagem ácida de mina (22.764.081 classificadas) quando comparadas às amostras de água de superfície da barragem de rejeitos (5.908.459 sequências).

Os resultados apresentados nos capítulos anteriores estão apresentados em formato de artigo nos materiais anexos 1 e 2:

- **L.,R.L.**, M.,J.D., P.,V.S., A.,F., S.,A.C., V.,A., O.,S.C., O.,G., F.,G.R. Linking biodiversity and potential functional role in a copper mine tailings dam (material ainda não publicado).
- M., J.D. ,**L.,R.L.** , O.,S.C. , O.,G. Taxonomic and functional diversity of microbial community from a mining environment. BMC Bioinformatics, 16(Suppl 8):A3. Resumo apresentado no Latin American Student Council Symposium 2014 (LA-SCS 2014). 2015.

5.3.1 – Diferenças entre perfis taxonômicos preditos por análise de amplicons 16S rDNA e análise de bibliotecas shotgun

Por causa das curtas sequências NGS, atribuições taxonômicas precisas, com base em alinhamentos de sequência, permanecem um desafio computacional tanto para 16S como para bibliotecas *shotgun*. Além disso, a falta de organismos referência específicos de ambientes pouco estudados, como os de mineração, complicam ainda mais a correta classificação taxonômica das sequências metagenômicas.

No presente estudo, em diferentes amostras, ao nível de filo, já houve diferença entre as predições taxonômicas usando 16S rDNA e *shotgun* (Figura 13). Os filios mais abundantes foram preditos por ambas metodologias, como Armatimonadetes, Cyanobacteria, Fusobacteria, Crenarchaeota, Gemmatimonadetes, Proteobacteria, Spirochaetes, Tenericutes, Elusimicrobia, Euryarchaeota, Nitrospirae, Parvarchaeota, Actinobacteria, Planctomycetes, Firmicutes, Chloroflexi.

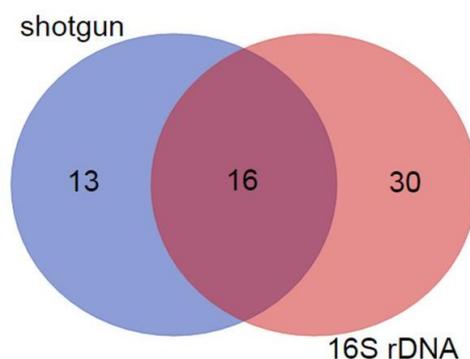


Figura 13 - Gráfico de Venn contendo o nº de filós exclusivos e compartilhados entre todas as predições por análise de *amplicons* 16S rDNA e análise de bibliotecas *shotgun*

Entre os filós exclusivamente preditos pela abordagem *shotgun* estão Chrysiogenetes, Korarchaeota, Deferribacteres, Nitrospinae, Synergistetes, Thermodesulfobacteria, Deinococcus-Thermus, Dictyoglomi, Thermotogae, Aquificae, Nanoarchaeota, Thaumarchaeota, Caldiserica. Uma das possíveis explicações para que eles não tenham sido identificados por 16S rDNA é que o par de *primers* utilizado nesse estudo, para amplificação das regiões variáveis V3 e V4, não é capaz de identificar esses organismos. Buscando uma comprovação para essa afirmação fizemos uma amplificação *in silico* do par de *primers* utilizado nesse estudo usando a ferramenta probematch do *Ribosomal Database Project*¹⁹⁶, testando uma base de dados diferente da Greengenes. Cada um dos filós supracitados possuía pelo menos 10 sequências representadas nos respectivos bancos de dados. Mesmo permitindo 3 *mismatches* entre as regiões amplificadas e os *primers*, os filós Chrysiogenetes, Korarchaeota, Nitrospinae, Dictyoglomi e Nanoarchaeota não foram identificados. Os outros filós tinham poucas sequências identificáveis: Deferribacteres (apenas 25 sequências de 16S rDNA tiveram hit em 1.462 presentes no banco de dados), Synergistetes (119 em 2.996), Thermodesulfobacteria (2 em 249), Deinococcus-Thermus (485 em 5.035), Thermotogae (160 em 1.199), Aquificae (111 em 2.085), Thaumarchaeota (22 em 38.702), Caldiserica (53 em 361). Isso significa que muitas espécies, representadas como *no hit*, não seriam passíveis de identificação em nossa

amostra. Possivelmente a classificação usando dados *shotgun* ocorreu porque todos esses filós possuem pelo menos um representante com genoma completo depositado no NCBI, possibilitando a identificação por outras regiões genômicas.

Filos novos ou pouco descritos na literatura, e conseqüentemente com poucos representados nos bancos de dados, foram preditos exclusivamente pela análise do 16S rRNA, que são os descritos a seguir: 30SBR1093, TM6, GN04, NKB19, Fibrobacteres, WS6, Lentisphaerae, WS3, OP3, AC1, NC10, GAL15, Thermi, GN02, Bacteroidetes, OP1, GOUTA4, AD3, Chlamydiae, BHI80-139, PAUC34f, Acidobacteria, OP11, OC31, BRC1, SR1, FCPU426, Chlorobi.

Até o momento, nenhuma abordagem isolada é capaz de oferecer uma visão completa da comunidade microbiana. A escolha de qual abordagem usar é geralmente ditada pela natureza dos estudos a serem realizados. Por exemplo, 16S rDNA é bem adequado para a análise de um grande número de amostras, como estudos longitudinais, mas oferece uma resolução limitada (geralmente a gênero e poucas espécies) ¹⁹⁷. O uso de iniciadores para diferentes regiões do gene 16S rRNA pode conduzir a resultados discordantes. Isso ocorre porque existem distintas afinidades de ligação para as diferentes regiões flangeadoras conservadas. Além disso, para cada região variável existem diferentes níveis de resolução entre os táxons ¹⁹⁸. A metagenômica *shotgun*, por outro lado, é mais custosa financeira e computacionalmente, mas oferece uma maior resolução, permitindo uma classificação taxonômica e funcional mais específica para cada sequência ou *contig*, bem como a descoberta de novos genes e genomas bacterianos. Além disso, dependendo do método de coleta e objetivo do projeto, também permite o estudo simultâneo de arqueias, vírus, fagos e eucariotos ¹⁹⁹.

5.4–Potencial biotecnológico de genomas isolados a partir de amostras de drenagem ácida de mina

Com o intuito de melhor descrever a diversidade não cultivável presente na drenagem ácida de mina, em Junho de 2014, foi coletada uma amostra de água para o isolamento aleatório de células únicas (SAGs). A técnica de *Single Cell Genomics* visa proporcionar novas perspectivas para a compreensão da genética, permitindo uma maior resolução de diferenças ao nível celular, e uma melhor compreensão da função da célula individual no microambiente. O isolamento das células individuais e seus respectivos sequenciamentos, do 16S rDNA e do genoma completo, foram realizados no *Single Cell Genomics Center Bigelow* (CSGS). A partir da análise filogenética do 16S rDNA fornecida pelo laboratório Bigelow, onde as análises foram realizadas, foram identificadas quatro SAGs de arqueias, duas classificadas como *Methanosalsum*, e uma como *Methanomethylovorans* (cada táxon com apenas um genoma seqüenciado até o momento), e seis SAGs de bactérias pertencentes à família Chitinophagaceae (filogeneticamente próximas a *Hydrotalea acidophilamas* sem genoma sequenciado). As arqueias supracitadas apresentaram alta representatividade nos metagenomas *shotgun* de drenagem ácida, compreendendo 52.898, 66.450 e 36.647 pares de seqüências mapeados exatamente uma vez nos respectivos genomas montados. Uma análise detalhada das SAGs de Chitinophagaceae será parte da tese de doutorado de Julliane Dutra Medeiros para o programa de pós graduação em genética. A tabela suplementar 3 exibe uma descrição geral desses genomas.

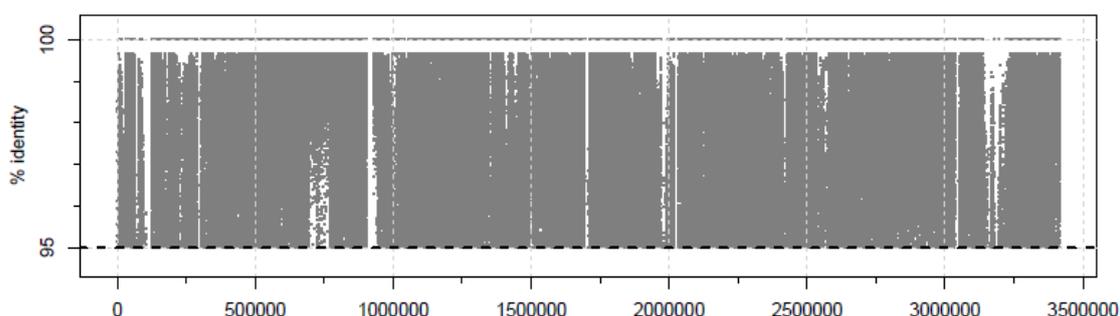
O sulfeto de dimetilo (DMS) representa um fator de impacto no aquecimento global, na precipitação ácida e no ciclo global do enxofre devido a formação de seus produtos de oxidação, como ácido metanossulfônico e SO₂. Haloarqueias foram avaliadas para a biorremediação em ambientes com contaminantes orgânicos, tratamento anaeróbico de lamas, resíduos agrícolas, municipais e no tratamento de águas residuais da indústria têxtil. Por causa da produção de biotensoativos elas podem desempenhar um papel importante na reparação rápida de ambientes salinos poluídos por petróleo²⁰⁰. *Methanosalsum* é um grupo recentemente descrito, caracterizado pela elevada

tolerância ao sal (até 3,5 M Na⁺), crescimento em ambiente alcalino (pH ótimo de 9,5) e metilotrofia, utilizando metanol, metilamina e sulfeto de dimetilo (a baixas concentrações) como substratos metanogênicos.

Methanomethylovorans, isolada inicialmente de sedimento de lago, é caracterizada por também ser metanogênica. Além de ser metilotrófica, não halofílica e codificar várias proteases extracelulares²⁰¹. Em todos os genomas foi encontrado o gene *rhIAB* que codifica uma ramnosiltransferase envolvida na síntese do biossurfactante ramnolípido. A propriedade emulsificante desse biossurfactante iônico vem sendo estudada como meio de potencializar o aumento da degradação de hidrocarbonetos em ambientes contaminados com petróleo²⁰².

Devido a baixa cobertura dos *scaffolds* montados das SAGs de Chitinophagaceae, foi preciso validar as anotações utilizando um genoma referência. Para tal sequenciamos, montamos e anotamos o genoma da espécie-tipo *Hydrotalea. flava*. A alta representatividade de genes dessa referência foi evidenciada pelo recrutamento do genoma montado contra todas as sequências da biblioteca *shotgun* (Figura 14), sendo possível afirmar a presença de organismos proximamente relacionados à *H. flava*, com identidade acima de 95%, na drenagem ácida de mina.

A



B

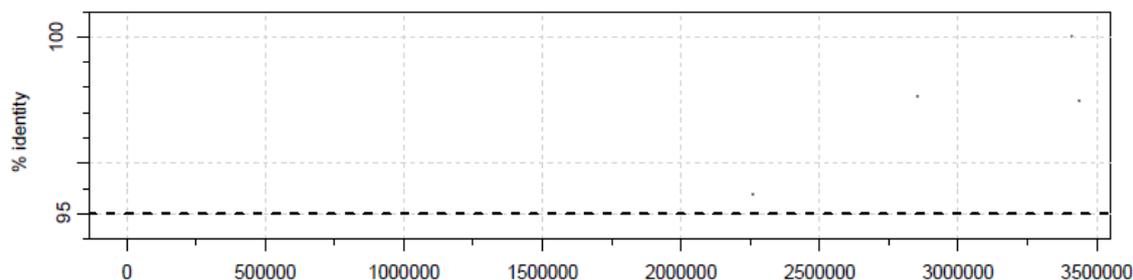


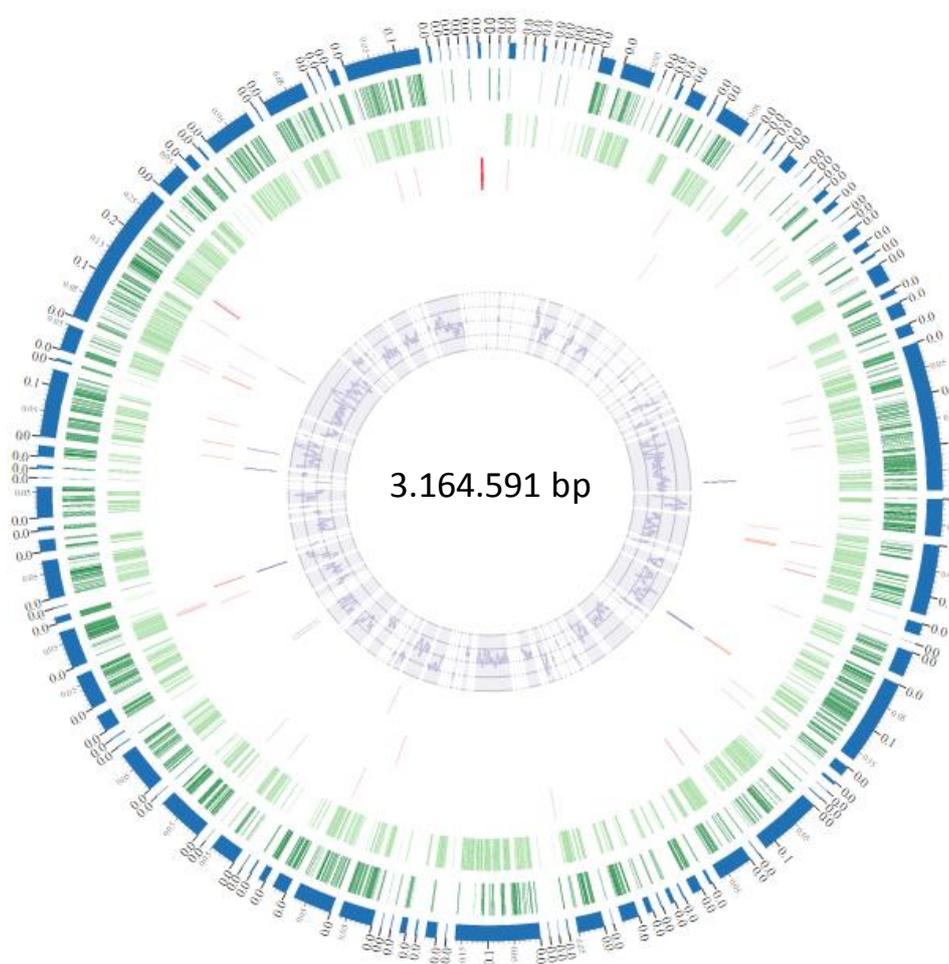
Figura 14 Recrutamento de sequências *shotgun* em relação ao genoma de *Hydrotalea flava*. A) sequências *shotgun* de água de drenagem ácida de mina B) sequências *shotgun* de barragem de rejeitos

A *Hydrotalea flava* CCUG 51397T é a espécie tipo do gênero *Hydrotalea*, e seu DNA foi obtido da coleção de cultura da universidade de Gotemburgo (Suécia). Essa bactéria foi originalmente isolada a partir de amostras de água no sul da Suécia, sendo caracterizada como gram-negativa, aeróbia, bastonete e não flagelada. Em uma análise filogenética baseada no gene 16S rDNA essa espécie mostrou menos do que 93% de semelhança com todas as espécies descritas pertencentes à família Chitinophagaceae. A análise do conteúdo genético de *H. flava* revelou a presença de genes que codificam para algumas características bem conhecidas da família Chitinophagaceae, como genes que codificam quitinases (endoglicosil-hidrolases que clivam as ligações glicosídicas da quitina), genes que codificam proteínas da parede celular relacionados a *rod cells* (mreBCD, PBP2, RodA), e genes de motilidade/deslizamento (gldBCDFGHJN). Também foi possível identificar a presença de genes envolvidos em processos de adaptação, bem como relacionados à resistência a antibióticos, compostos tóxicos e resposta ao estresse (Figura 15 A).

Outro genoma completo foi isolado em Dezembro de 2013 a partir de drenagem neutra formada sob pilhas de rejeito de calcopirita de baixo grau. Nomeado de ITV01, o isolado é filogeneticamente próximo a *Alicyclobacillus ferrooxidans*, sendo igualmente heterotrófica e acidofílica. Ele possui um repertório completo de genes necessários para uma vida livre, estilo de vida

quimiolitotrófico, incluindo de fixação de CO₂, biossíntese de nucleotídeos e cofatores, além de conter importantes genes envolvidos na oxidação de ferro e enxofre, tais como uma *DYP-like* peroxidase (EC 1.11.1.7), ferritina-like (EC:1.16.3.1), sulfato adenylyltransferase (2.7.7.4) (Figura 14 B).

A



B

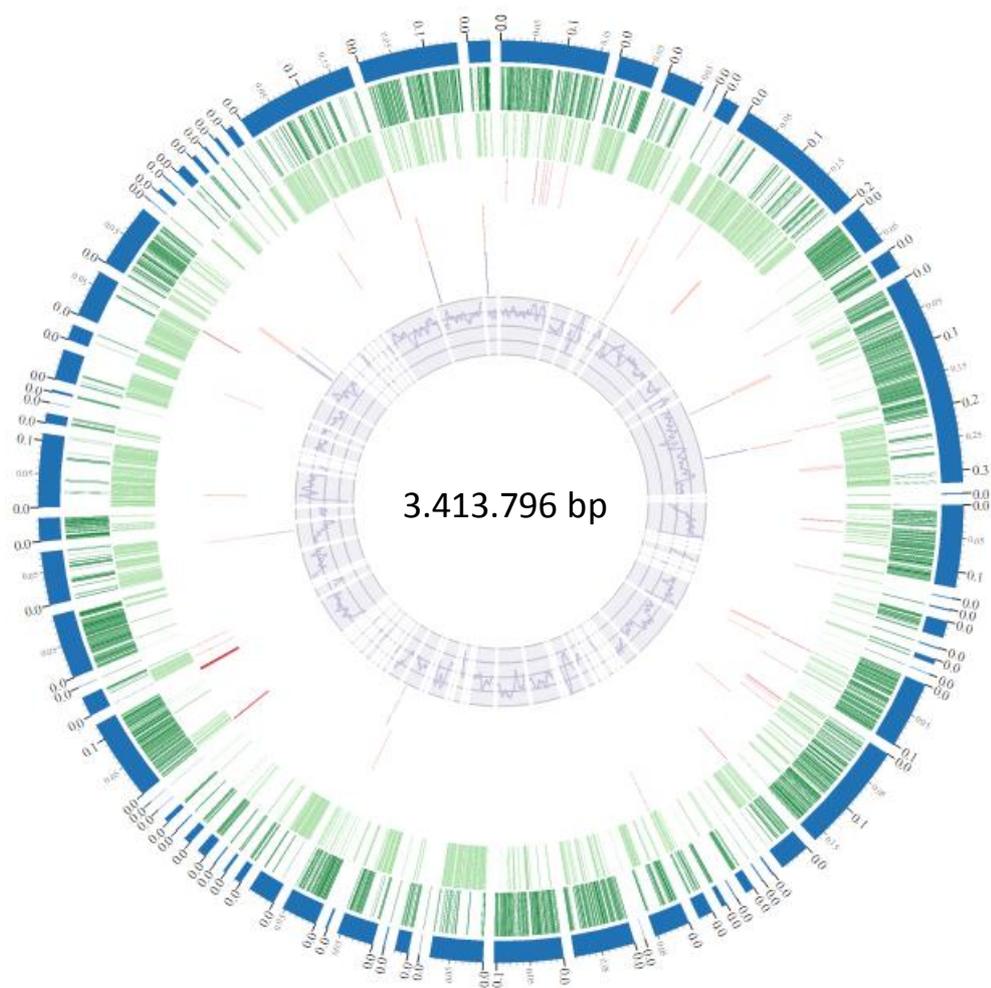


Figura 15 - Visão geral dos atributos funcionais anotados para os genomas A) da *Hydrotalea flava* B) da ITV01. Lista de faixas, de fora para dentro: 1) Cromossomos / plasmídeos / scaffolds, 2) CDS, *forward strand*, 3) CDS, *reverse strand*, 4) RNAs, 5) CDS relacionados ao metabolismo de enxofre (*both strands*), 6) CDS relacionadas ao metabolismo de ferro (*both strands*), 7) conteúdo GC.

Alguns dos resultados apresentados nesse tópico também estão descritos nos seguintes artigos, que se encontram anexos, no fim da sessão de materiais suplementares:

- **L.,R.L.**, M.,D.J., F.,G.R., A.,F., S.,A.C., V.,A., O.,G., O.,S.C. Draft Genome Sequence of *Hydrotalea flava* Strain CCUG 51397 T. *Genome Announc.* May/June 2016 vol. 4 no. 3 e00527-16.
- D.,H., Ñ.,I., J.,D.B., O.,R., **L.,R.L.**, P.,V.S., H.,R., G.,B., C. N, N.,G.L., T.,G., F.,G.R., D.,J., O.,S.C., O.G. 2016. Draft Genome Sequence of “*Acidibacillus ferrooxidans*” ITV01, a Novel Acidophilic Firmicute Isolated from a Chalcopyrite Mine Drainage Site in Brazil. *Genome Announc.* 2016 Mar-Apr; 4(2): e01748-15

5.5–Perfil funcional dos ambientes de mineração: a flexibilidade metabólica da barragem de rejeitos e a adaptação ambiental na drenagem ácida de mina

Inferir a função a partir dos metagenomas é uma tarefa computacionalmente mais custosa do que inferir a taxonomia. Essa informação extra não só permite uma melhor utilização dos dados, mas também adiciona uma camada nova de informações que podem ser ecologicamente relevantes.

Ao total, a comunidade microbiana da barragem de rejeitos presente em amostras de água e solo nos três estratos (1M, 15M e sedimento) foram analisadas quanto ao conteúdo total de DNA (*shotgun*), durante duas estações contrastantes (seca e chuvosa), totalizando a construção e análise de 6 metagenomas da barragem. Para representar a drenagem ácida de mina apenas água do período seco pôde ser analisada até o presente momento. Cada metagenoma foi formado por réplicas biológicas coletadas separadamente, mas unidas para essa análise, com o intuito de aumentar a cobertura total de sequências por amostra possibilitando melhorar os resultados da montagem e conseqüentemente da predição gênica. A montagem *de novo* do metagenoma *shotgun* resultou em *contigs* de bom tamanho, com N50 médio de 800 pb. As informações detalhadas com as estatísticas da relação entre sequências filtradas e *contigs* estão resumidas na tabela suplementar 2.

Dos 197.057 genes preditos, apenas cerca de 21.000 puderam ter uma categoria funcional atribuída, isto é, foram confiavelmente anotados em relação a proteínas com funções conhecidas. Muitos genes (~15%) foram relacionados a proteínas preditas com função desconhecida, enquanto o restante aproximadamente 75% não apresentou nenhuma correspondência com as bases de dados Uniprot e KEGG. Apesar do baixo número de proteínas preditas e classificadas nas amostras de sedimento da barragem de rejeitos foi possível obter uma visão geral do papel funcional dessa comunidade e correlacionar esses perfis com os dados de 16S rDNA e dados da literatura. Devido à maior diversidade identificada com as bibliotecas de 16S rDNA, fica clara a importância em garantir a quantificação de unidades

taxonômicas com bibliotecas de *amplicons*. Os rRNAs apresentam maior sensibilidade e são capazes de identificar 80% das bactérias presentes em uma amostra²⁰³. No futuro, um maior esforço de sequenciamento permitirá uma maior resolução das peculiaridades espaço-temporal nesta camada.

De modo geral, os metagenomas foram classificados em 156 subsistemas envolvidos com ampla diversidade de funções metabólicas (Figura 16). Os cinco processos mais abundantes na barragem de rejeitos foram: metabolismo de carbono, biossíntese de aminoácidos, fosforilação oxidativa, ribossomo e biossíntese de aminoacil-tRNA enquanto na drenagem ácida de mina foram metabolismo de carbono, transportadores ABC, metabolismo de purinas, biossíntese de aminoácidos e sistema regulatório de dois componentes. Os metagenomas também apresentaram genes que codificam proteínas para degradação de compostos aromáticos e xenobióticos, como a presença de genes que codificam proteínas envolvidas na produção de catecol e no catabolismo de fenol e benzeno. A partir da análise com DESeq não foi possível encontrar KOs com abundâncias significativamente diferentes entre as amostras de sedimento e água, bem como comparando a estação seca e chuvosa. Isso pode ser explicado pelo fato que diferentes grupos taxonômicos são capazes de desempenhar papéis ecológicos semelhantes.

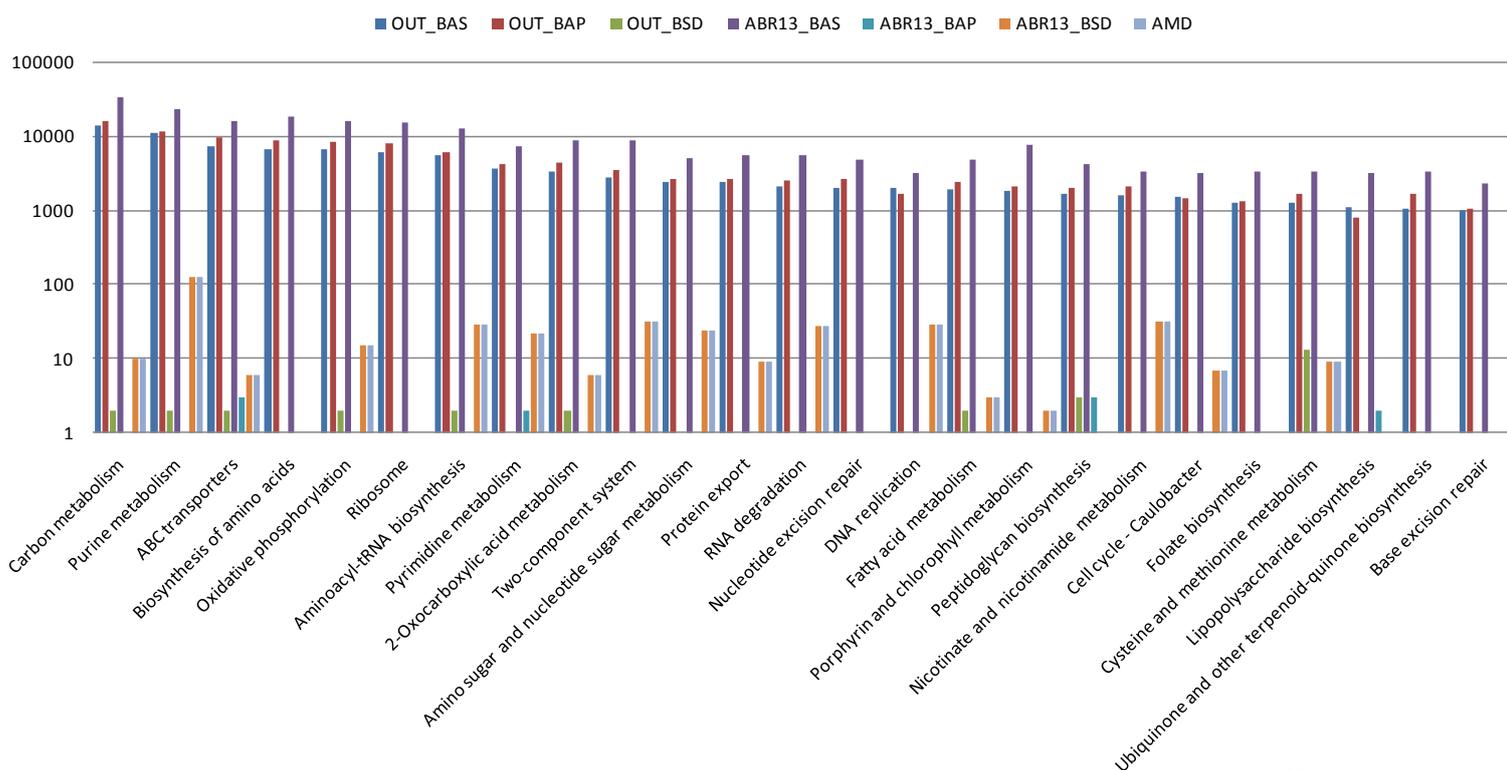


Figura 16 - Resumo das vias mais abundantes identificadas nos metagenomas de barragem de rejeitos e drenagem ácida de mina. Estão representadas as amostras da barragem de rejeitos coletadas em outubro (água de superfície OUT_BAS, água de profundidade OUT_BAP e sedimento OUT_BSD) e abril (água de superfície ABR_BAS, água de profundidade ABR_BAP e sedimento ABR_BSD), e as amostras de água de drenagem ácida de mina (AMD). O eixo x mostra as categorias funcionais do KEGG reportadas para as amostras analisadas. O eixo y representa o número de sequências identificadas por categoria em cada conjunto de amostras.

Na barragem de rejeitos, a importância de um estilo de vida autotrófico foi indicada pela presença de um número significativo de genes envolvidos na via de fixação de carbono. Estes marcadores de fixação de carbono, como genes que codificam as enzimas das vias do ciclo de Calvin e do ciclo redutivo do ácido tricarboxílico, foram diminuindo da superfície para a profundidade. Por outro lado, os genes relacionados aos genes metabólicos foto dependentes (por exemplo, rodopsinas fotoliasas) e à fotossíntese oxigênica do fotossistema II parecem prevalecer nas amostras de sedimento, provavelmente em sua zona fótica superficial. A fotossíntese de cianobactérias, fotossíntese anoxigênica e de redução de sulfato pode ser responsável por uma parte importante da precipitação de carbonato em amostras de água. As subcategorias responsáveis por uma porcentagem substancial de sequências anotadas nas amostras de água estavam relacionadas a processos celulares básicos como: metabolismo e fixação de carbono, replicação de DNA, fosforilação oxidativa, metabolismo de purina, pirimidina, biossíntese de aminoacil-tRNA, metabolismo de alanina, aspartato e glutamato, biossíntese de aminoácidos, transportadores ABC, metabolismo da vitamina B6. No sedimento, outras vias gerais menos abundantes, mas igualmente importantes, estavam relacionadas a manutenção do ciclo celular caulobacter-like, metabolismo de cisteína, metionina e glutatona. A biossíntese de cisteína a partir de sulfato é, aparentemente, o principal mecanismo através do qual o enxofre é reduzido na água. Durante este processo, o sulfato inorgânico é reduzido a sulfito e sulfeto. Todos os genes de redução de sulfato pela via assimilatória foram recuperados (cysDNCH).

Vale lembrar que a barragem de rejeitos é um ambiente oligotrófico, por isso podem ser necessárias outras reações metabólicas que suportem o metabolismo autotrófico. A frequência de marcadores relacionados ao

metabolismo de nitrogênio, tais como genes envolvidos nas vias de desnitrificação (NIRK) diminuem da superfície para o fundo da barragem. No entanto, o potencial para a conversão de nitrogênio tipicamente encontrado em outros ambientes aquáticos (vias da redução de nitrato assimilatória/dissimilatória e de fixação de nitrogênio) não foi identificado no sedimento e parece reduzido na água. Estes resultados não parecem consistentes com a abundância taxonômica baseada em 16S rDNA, em que foram encontradas bactérias fixadoras de nitrogênio, tais como membros da filo nitrospira (gênero *Thermodesulfovibrio*) e Proteobacteria (beta e Gammaproteobacteria provavelmente estavam contribuindo para a nitrificação, bem como desnitrificação).

Por outro lado, foram identificados muitos mecanismos relacionados à baixa disponibilidade de nutrientes no meio, como a utilização de transportadores ABC e sistemas de dois componentes. Para possibilitar a sobrevivência de uma variedade de espécies neste ambiente são necessários mecanismos de adaptação a curto e longo prazo, que visem reduzir a energia metabólica e aumentar a eficiência catabólica. Em ambas épocas do ano (seca e chuvosa), foram identificados genes codificadores de transportadores para absorção de aminoácidos de cadeia ramificada, transportadores de níquel (operon ABCDF), transportadores de zinco (*znuA*, *znuB*, *znuC*) e o sistema completo de transporte de cobalto/níquel (CBIM, *cbiQ*, CBio). Ademais, em resposta a baixa disponibilidade de fósforo na água, foram encontrados genes envolvidos nos mecanismos de reciclagem de fosfato, como por exemplo transportadores com alta afinidade ao fosfato (*psta*, PSTB, PSTC, e *pstS*), transportadores de fosfonato (*Phnc*, *phnD*, *phne*, *phnH*, *phnI*, *phnJ*, PHNL, *phnM*, *phnP*, *phnX*) e reguladores de fosfato (PHOR, *phoA*, PHOP, *phoA*, *phod* e *regX3*).

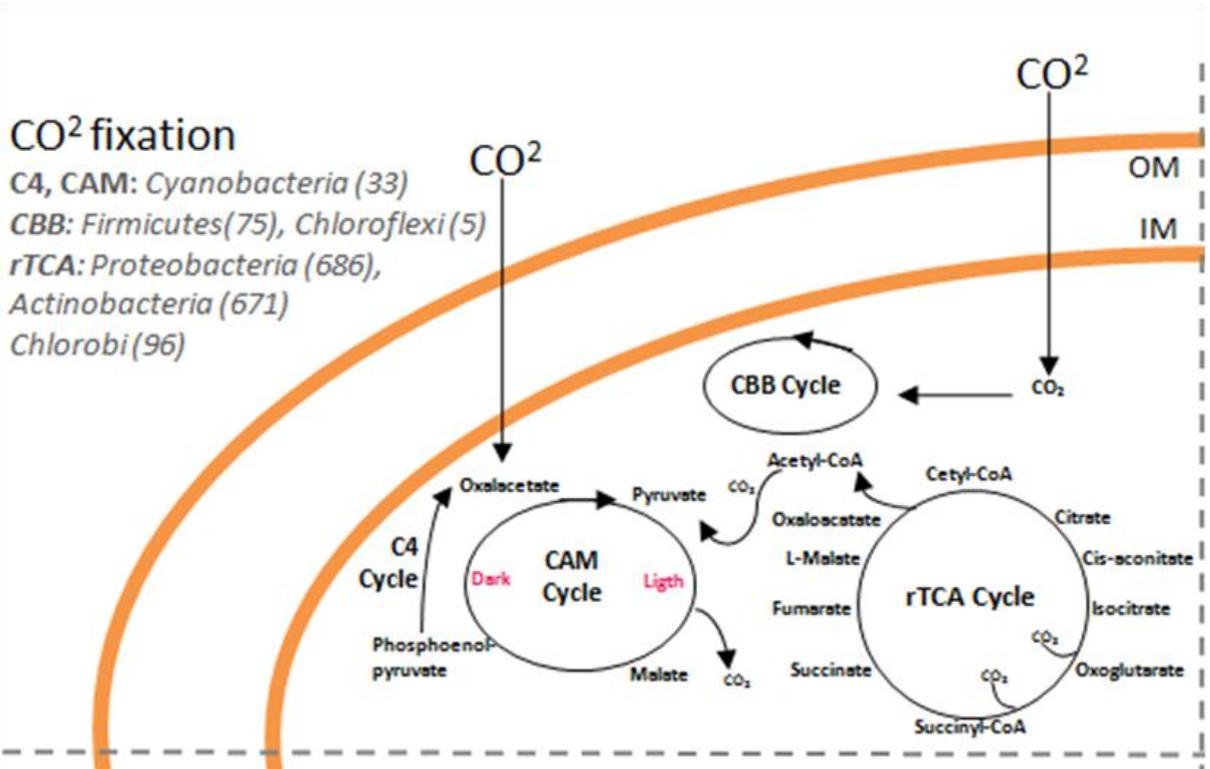
Organismos anaeróbios foram detectados abaixo na água de profundidade e sedimento, onde provavelmente ocorre redução de ferro férrico e sulfato através da via dissimilatória. Essa redução parece estar impulsionada por bactérias redutoras de sulfato das famílias Peptococcaceae, Thermodesulfovibrionaceae, Desulfobacteriaceae, Desulfobulbaceae, Nitrospinaceae. Esse processo é crucial para o desenvolvimento de tecnologias

relacionadas a biorremediação pois conduz a um consumo de prótons, aumentando assim a alcalinidade e a formação de sulfetos. Estes resultados estão de acordo com estudos anteriores em minas de ferro^{12; 204; 205}.

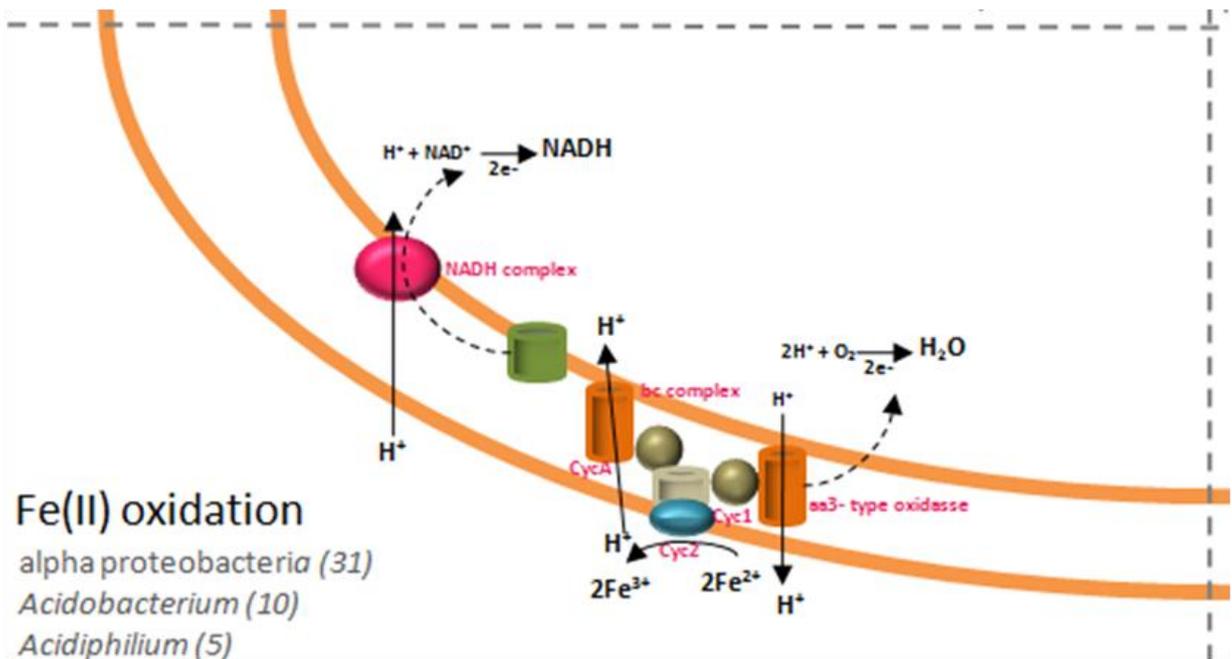
Apenas nas amostras de água foi identificado um repertório gênico potencialmente relacionado a ambientes extremos, como baixo pH (kdpA, clpX e CLPP) e estresse oxidativo (DNAX, clpPX, e groEL). Nas amostras de água do período seco também foi identificado o operon relacionado a síntese de estruturas extras da parede celular (murABCDEFGQ). Vale lembrar que todos os resíduos do processo de mineração de cobre são transportados para a deposição final na barragem de rejeitos. Neste ambiente a oxidação de minerais sulfetados (contendo ferro e enxofre) fica limitada desde que os resíduos permaneçam saturados com água. Em regiões não cobertas, a lixiviação dos rejeitos resulta em um aumento de oxianions em solução (como sulfato, arsenato, molibdato). Ao longo da vida útil da mina ou uma vez que as atividades cessam, os níveis de água represada na barragem caem, criando zonas altamente oxigenadas²⁰⁴ e levando a formação de drenagem ácida de mina dentro do ambiente da barragem de rejeitos. A

pesar de não estarem presentes em grande número na barragem, a abundância dos organismos biolixiviadores poderia aumentar a medida que a barragem de rejeitos vá se tornando saturada. Além disso, pode ocorrer transferência horizontal de genes de resistência, como já relatado em AMD entre *L. ferrodiazotrophum*, *L. rubarum* e *Ferroplasma* em relação a genes de resistência a acidificação (operon gadBC)²⁰⁶. Uma visão geral da relação entre a diversidade funcional e taxonômica encontrada na barragem de rejeitos está representada na Figura 17.

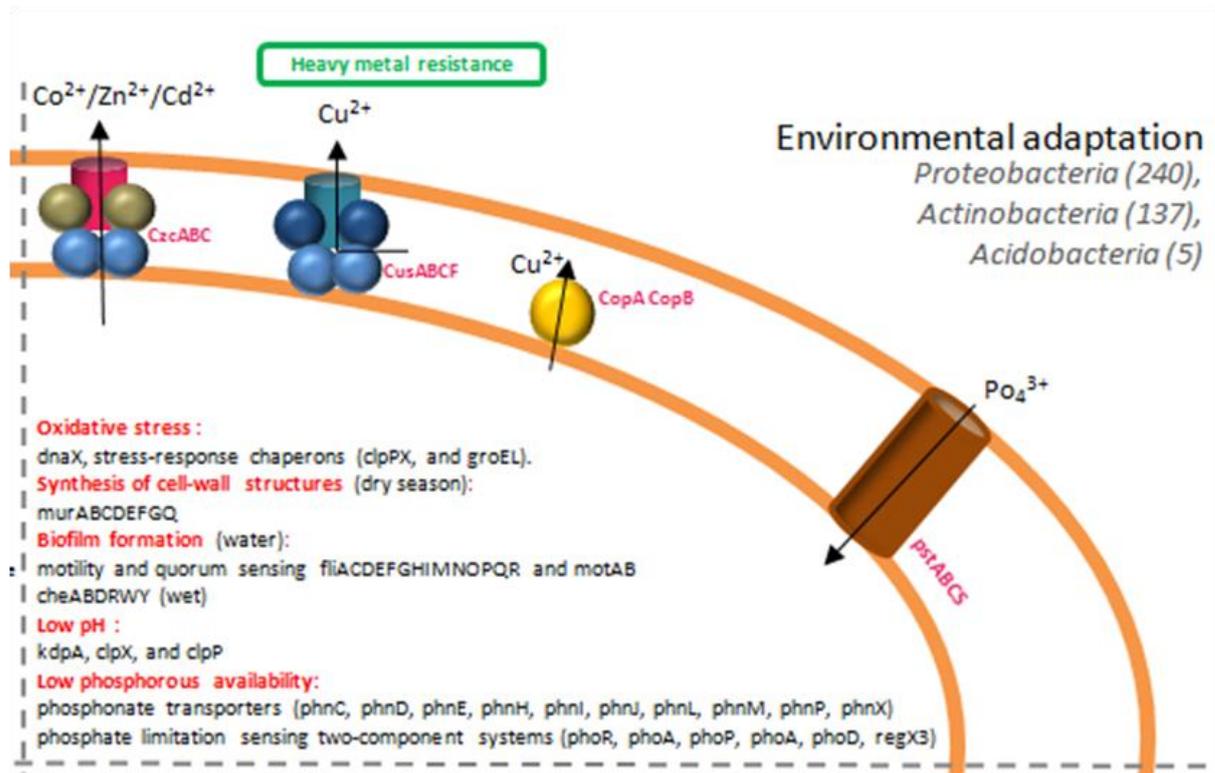
A



B



C



D

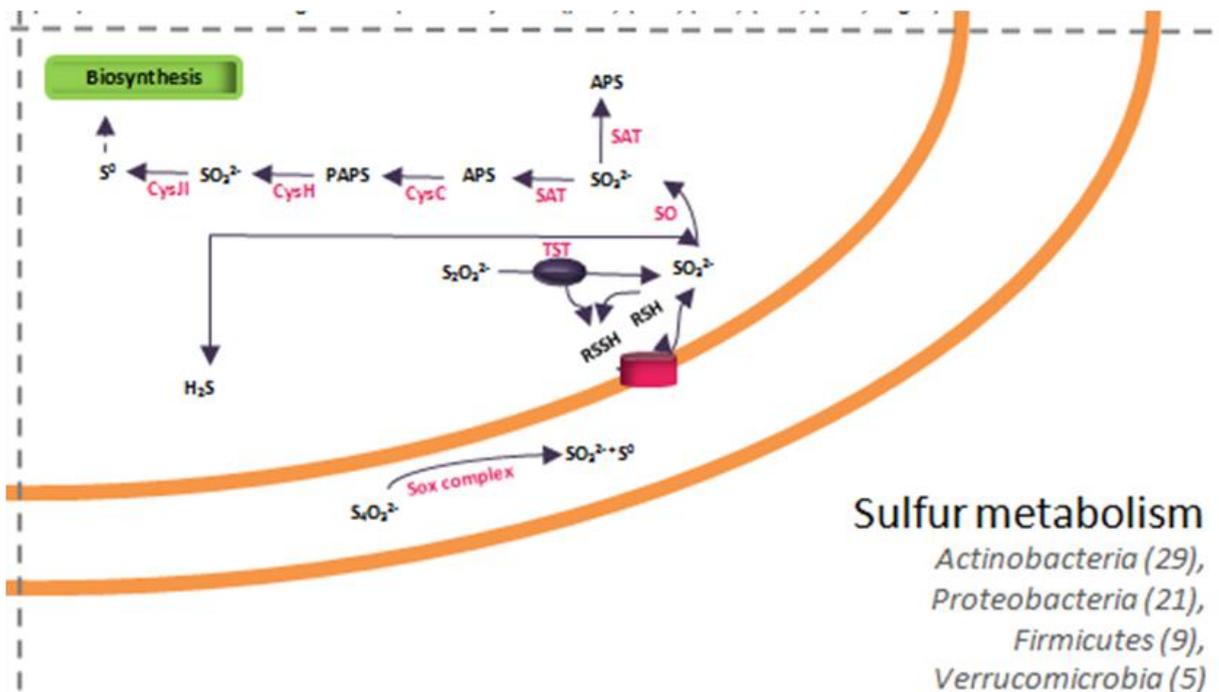


Figura 17 - Visão geral dos principais aspectos funcionais encontrados na barragem de rejeitos. A) fixação de carbono, B) oxidação de ferro, C) adaptação ambiental, D) metabolismo de enxofre. O número entre parênteses apresenta a quantidade de genes completos preditos e atribuídos ao determinado táxon relacionado à via de interesse.

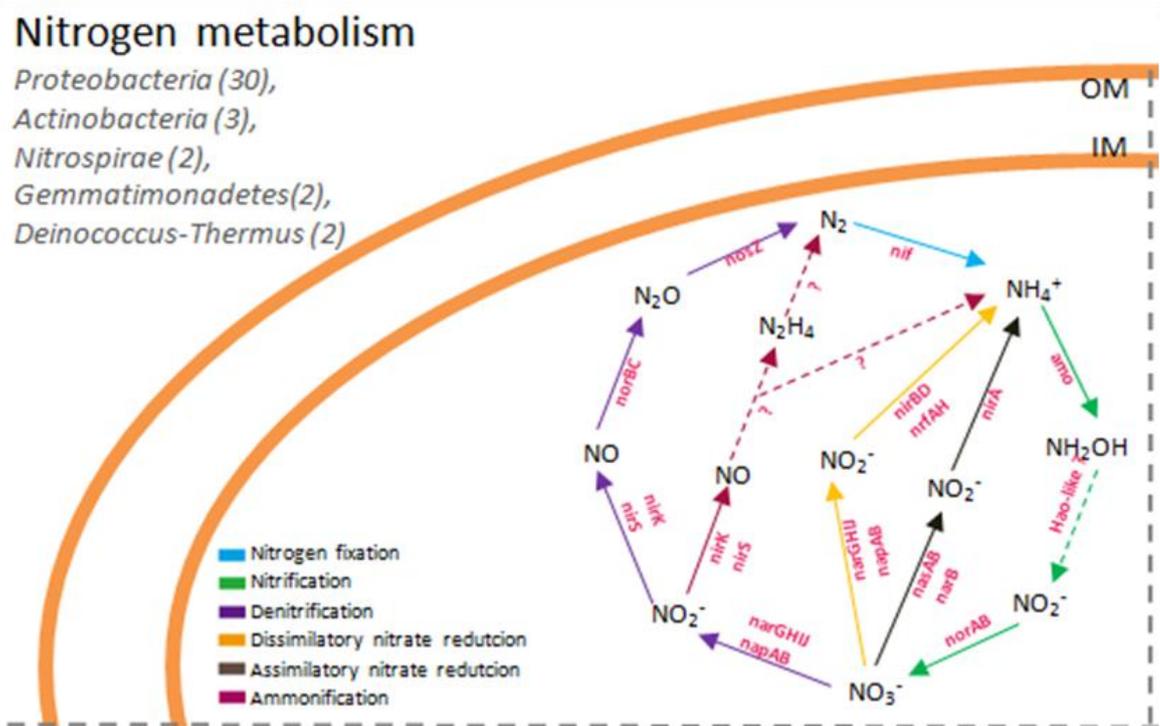
Na drenagem ácida de mina existe o favorecimento de uma comunidade microbiana baseada em quimiolitotrofia, metabolismo geral ligado a oxidação e redução de ferro e enxofre, bem como funções relacionadas aos ciclos do nitrogênio e carbono ¹⁶². A transformação bioquímica primária do ferro nesses sistemas é a oxidação do ferro ferroso, implícita pela predominância de bactérias oxidadoras de ferro quando comparada com a de espécies redutoras de ferro férrico ²⁰⁷. As enzimas relacionadas envolvem uma cadeia de transporte de elétrons em que os elétrons são transferidos das espécies reduzidas para o primeiro componente. Nessas amostras foi encontrada tanto a cyt c que codifica a proteína da membrana exterior do citocromo C, quando a que codifica a membrana periférica (apenas relacionada a oxidação do ferro ferroso). Também foram identificadas as “proteínas azuis” de cobre rusticianina, rusA e sulfocianina, e a Iro (ferro oxidase). As transformações envolvidas no ciclo do enxofre são mais complexas devido à variedade de estados de oxidação adotada por este elemento (politionatos como tiosulfato e tetracionato, enxofre elementar, sulfeto).

Grande parte do sistema Sox, relacionado a oxidação de compostos reduzidos de enxofre a sulfato sem a formação do intermediário sulfito, foi encontrado. Esse sistema é expresso em mais de 15 genes que codificam vários citocromos e outras proteínas como: o complexo RISC, CydAB (relacionado à *A. thiooxidans*, *A. ferrooxidans* e *A. caldus*), o complexo citocromo d ubiquinol oxidase, bo3 oxidase (cyoABCD) e o complexo bc 1 (relacionado a *A. ferrivorans*). Nas arqueias muitas enzimas que realizam esse processo pertencem à ordem Sulfolobales ²⁰⁸. A principal diferença em relação a outras transformações químicas que envolvem RISCs é que a enzima-chave do processo é uma enxofreoxigenase redutase (Sor), que catalisa a oxidação e redução simultânea de enxofre. Na barragem de rejeitos foram encontradas enzimas típicas de arqueias termofílicas como Sor, teth e DoxDA. A assimilação de enxofre, geralmente na forma de sulfato, parece ser realizada por meio de um sistema de captação de ABC ou sulfato de permease (SulP). Os genes que codificam as enzimas da via para redução de nitrato assimilatório foram mapeadas a diversos táxons bacterianos, principalmente pertencentes a grupos de alfa-proteobactérias.

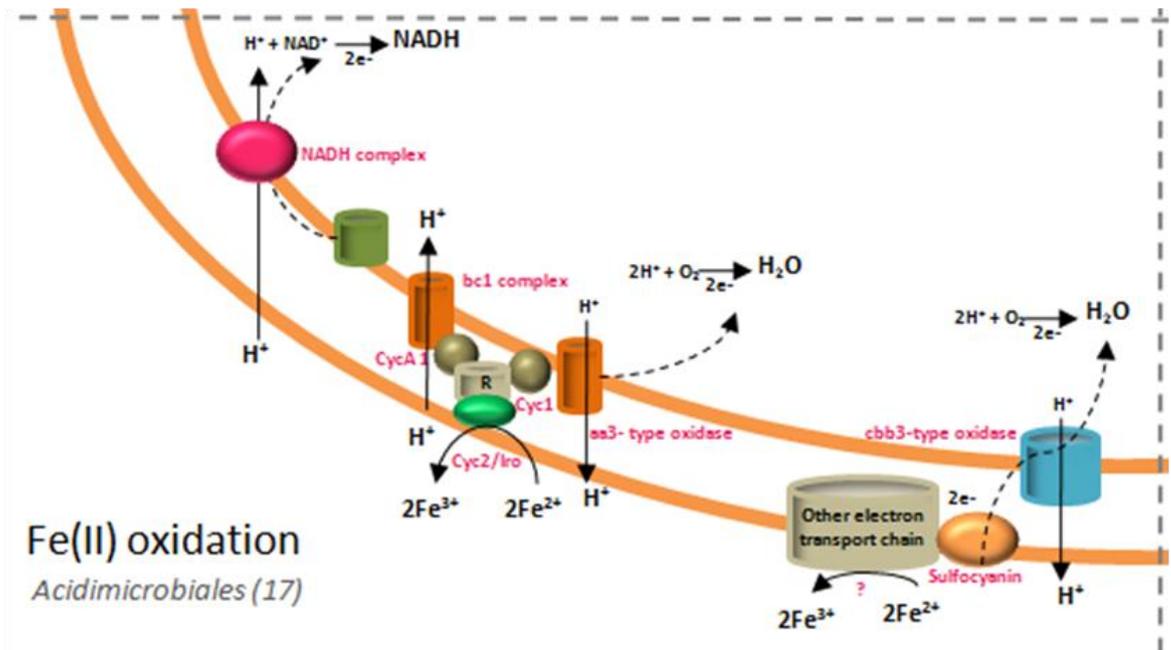
Fontes de carbono são severamente limitadas em AMD, por esse motivo a ciclagem de carbono é uma atividade primordial. Alguns membros da comunidade parecem conter genes de fixação de CO² pelo ciclo de Calvin-Benson-Bassham (CBB) (associado com o gênero *Gallionella*-like), fixação de carbono inorgânico pelo ciclo redutor de ácido tricarboxílico (RTCA) (associado a *Leptospirillum*-like). Fixação de carbono típica de arqueia também foi identificada, como o ciclo do 3-dimetil-hidroxiacetato, 4-hidroxiacetato e via reductiva de acetil-CoA, mas sem associação a nenhum táxon específico.

Além dos mecanismos de resistência a metais pesados e ambientes ácidos já descritos na barragem de rejeitos, também foram encontrados genes relacionados à formação de biofilme, como genes associados com motilidade e *quorum sensing* (*fliACDEFGHIMNOPQR*, *motAB*, *cheABDRWY*). A formação de biofilme pode ser um dos mecanismos de resistência a agentes antimicrobianos, pela adsorção dos antimicrobianos disponíveis no meio pela matriz de polímeros extracelulares, diminuindo sua interação com as células do biofilme²⁰⁹. Uma visão geral da relação entre a diversidade funcional (fixação de carbono, metabolismo de enxofre, oxidação de ferro e adaptação ambiental) e taxonômica encontrada na drenagem ácida de mina está representada na Figura 18.

A



B



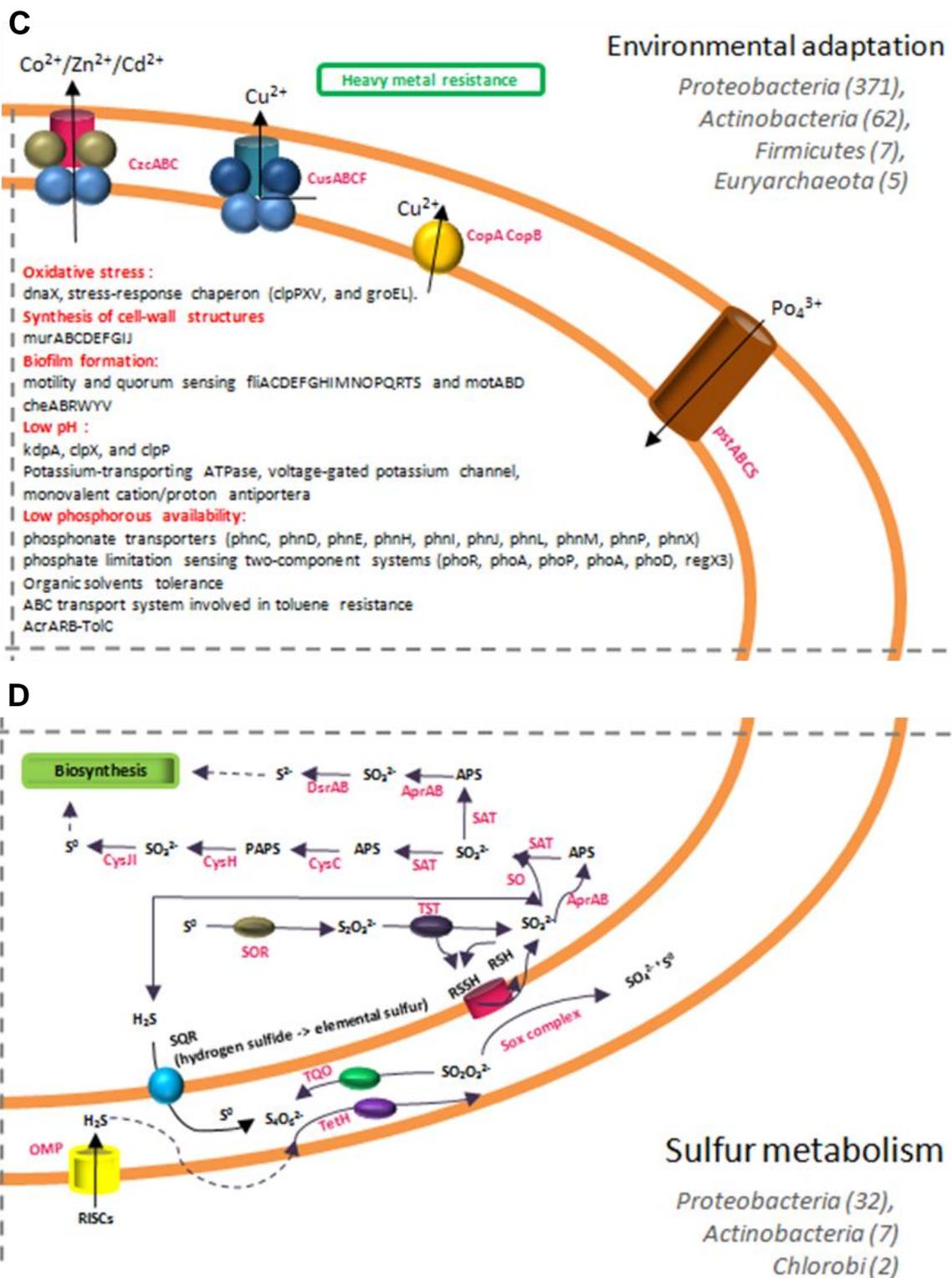


Figura 18 - Visão geral dos principais aspectos funcionais encontrados na drenagem ácida de mina. A) metabolismo de nitrogênio, B) oxidação de ferro, C) adaptação ambiental, D) metabolismo de enxofre. O número entre parênteses representa a quantidade de genes completos preditos e atribuídos ao determinado táxon relacionados à via de interesse.

6- Conclusões e Perspectivas

No presente estudo, demonstrou-se que a barragem e a drenagem, compostas por alta concentração de minerais sulfetados de cobre e ferro, são colonizadas por uma comunidade microbiana diversa, que difere em parte da estrutura das comunidades comumente relatadas em água doce natural, sedimento amazônico, e também em termos de abundância em relação à drenagem ácida de mina de baixo pH. Nos dois ambientes, os rejeitos de mineração da mina de Sossego em conjunto com o aporte de matéria orgânica de seu entorno, parecem suportar o crescimento de microrganismos heterotróficos e acidófilos (tais como *Acidibacillus*, *Acidiphilium*, *Acidobacterium* e bactérias redutoras de sulfato) que provavelmente vivem em lisados e exsudatos (como o carbono orgânico dissolvido) derivados de microrganismos produtores primários, como os autotróficos (ex. *Cyanobacteria* e *Nitrospira*). Esses microrganismos parecem estar adaptados ao ambiente por apresentarem genes de resistência a metais pesados, à baixa disponibilidade de nutrientes e rotas variadas para oxidação / redução de ferro e enxofre.

Até o momento foram gerados como produtos: a padronização de um fluxo de trabalho para análise bioinformática de dados metagenômicos, incluindo ferramentas intermediárias desenvolvidas *in house* para mudanças de formatos, integração dos dados em formato tabular e visualização de alguns resultados (modelos de distribuição, mapas de calor, recrutamento); e uma lista contínua de genes e táxons presentes nas amostras analisadas. Além disso, essa pesquisa revelou um número expressivo de novos genes preditos com função desconhecida. Esses genes confirmam a existência de um acervo metabólico inexplorado que pode estar relacionado a processos de biomineração. A integração de dados genômicos, metagenômicos, ambientais e funcionais (incluindo a transcritômica e metabolômica) é necessária para a elaboração de modelos de funcionamento dessas comunidades, que poderão ser usados no controle de processos de biolixiviação e biorremediação. Neste contexto, o banco de dados gerado a partir deste projeto poderá ser complementado com dados de outros projetos do Laboratório de Informática de Biosistemas e Genômica realizados com amostras provenientes da mina de

Sossego e de outras minas. Assim, modelos *in silico* podem ser criados, com base no conhecimento teórico acumulado, e validados *in vitro*, por exemplo, através da observação do comportamento de sistemas experimentais miniaturizados, ou microcosmos.

7- Considerações finais

Desde o ingresso no Doutorado, além da linha de pesquisa principal mencionada em biomineração, eu atuei como colaboradora em outros estudos de genômica e metagenômica, principalmente na execução e interpretação das análises bioinformáticas. Um pequeno resumo dos trabalhos já publicados encontra-se a seguir:

- Reis MP, Dias MF, Costa PS, Ávila MP, **Leite LR**, de Araújo FM, Salim AC, Bucciarelli-Rodriguez M, Oliveira G, Chartone-Souza E, Nascimento AM. Metagenomic signatures of a tropical mining-impacted stream reveal complex microbial and metabolic networks. *Chemosphere*. 2016 Jul 18;161:266-273. doi: 10.1016/j.chemosphere.2016.06.097.

Neste estudo foram analisados dados de 16S rDNA e *shotgun* a fim de descrever a comunidade bacteriana do sedimento do rio Carrapato (Quadrilátero ferrífero, Minas Gerais, Brasil), um local impactado por altas concentrações de ouro e ferro. Foi revelada uma comunidade diversificada, com predomínio de Proteobacteria (39,4%), Bacteroidetes (12,2%), e Parcubacteria (11,4%), com elevada abundância de bactérias oxidadoras de ferro neutrofilicas, como *Gallionella*, *Sideroxydans* e *Mariprofundus*. A análise funcional revelou genes envolvidos no metabolismo de nitrogênio, ferro e enxofre e em resistência a metais pesados.

- Dias MF, Colturato LF, de Oliveira JP, **Leite LR**, Oliveira G, Chernicharo CA, de Araújo JC. Metagenomic analysis of a desulphurisation system used to treat biogas from vinasse methanisation. *Bioresour Technol.* 2016 Apr;205:58-66. doi: 10.1016/j.biortech.2016.01.007.

O trabalho investigou a resposta da comunidade microbiana às mudanças nas taxas de adição de H₂S em um sistema de dessulfuração microaerada a partir de bibliotecas de 16S rDNA. O objetivo era estabelecer os melhores sistemas para obtenção de biogás a partir da biometanização da vinhaça de cana de açúcar. A eficiência na remoção de H₂S foi maior nos sistemas COD (*chemical oxygen demand*) e DO (*dissolved oxygen*). Nesses sistemas, a maioria das sequências foram classificadas como pertencentes aos filos *Firmicutes*, *Proteobacteria*, *Bacteroidetes*, *Chloroflexi*, e *Synergistetes*, com destaque para as bactérias oxidadoras de sulfeto (SOB), como *Thioalkalimicrobium sp.* O restante da comunidade microbiana permaneceu constante durante todo o experimento.

- Laguardia-Nascimento M, Branco KM, Gasparini MR, Giannattasio-Ferraz S, **Leite LR**, Araujo FM, Salim AC, Nicoli JR, de Oliveira G, Barbosa-Stancioli EF. Vaginal Microbiome Characterization of Nelore Cattle Using Metagenomic Analysis. *PLoS One.* 2015 Nov 24;10(11):e0143294. doi: 10.1371/journal.pone.0143294.

Este estudo teve como objetivo investigar o microbioma vaginal em bovinos da raça Nelore, comparando novilhas e vacas, grávidas e não grávidas, usando bibliotecas de 16S rDNA. Os principais filos bacterianos encontrados foram Firmicutes (~40-50%), Bacteroides (~15-25%) e Proteobacteria (~5-25%), enquanto 45-55% das amostras foram representados por apenas dez OTUs: *Aeribacillus*, *Bacteroides*, *Clostridium*, *Ruminococcus*, *Rikenella*, *Alistipes*, *Bacillus*, *Eubacterium* e *Prevotella*. A colonização vaginal parece ser influenciada pela comunidade gastrointestinal, e não foi possível

estabelecer uma correlação entre a influência hormonal e a composição da microbiota. Contudo, observou-se uma tendência para a redução de bactérias em detrimento ao aumento de arqueias em animais prenhes.

- Costa PS, Reis MP, Ávila MP, **Leite LR**, de Araújo FM, Salim AC, Oliveira G, Barbosa F, Chartone-Souza E, Nascimento AM. Metagenome of a Microbial Community Inhabiting a Metal-Rich Tropical Stream Sediment. *PLoS One*. 2015 Mar 5;10(3):e0119465. doi: 10.1371/journal.pone.0119465.

Descrição da composição taxonômica e funcional da comunidade microbiana do sedimento de um córrego de água doce historicamente contaminado por metais pesados (Mina Stream – MSS). A análise *shotgun* revelou uma predominância de proteobactérias e *Parvarcheota*. Os contigs foram recrutados em relação a genomas de referência como *Candidatus Nitrospiradefluvii* e *Nitrosopumilus maritimus*. A presença desses microorganismos foi relacionada ao metabolismo de nitrogênio. A reconstrução funcional também revelou genes relacionados à assimilação de amônia e amonificação. Além disso, uma elevada diversidade metabólica sugeriu que a contaminação por arsênio histórico não parecia afetar a comunidade procariótica.

- Batista DFA, Freitas Neto OC, **Leite LR**, Varani AM, Araujo FMG, Salim A, Almeida AM, Ribeiro SAM, Oliveira GC, Barrow PA, Berchieri Junior A. 2014. Draft genome sequence of *Salmonella enterica* subsp. *enterica* serovar Gallinarum biovar Pullorum strain FCAV198, a Brazilian chicken pathogen. *Genome Announc.* 2(1):e00028-14. doi:10.1128/genomeA.00028-14.

Esse projeto previu a montagem e anotação do genoma de uma estirpe de *Salmonella entérica*, responsável por uma doença sistêmica que leva a altas taxas de mortalidade em aves jovens. A análise comparativa com outros genomas de *Salmonella Pullorum* revelou importantes trechos genômicos correspondentes a elementos móveis e plasmídeos exclusivos da estirpe analisada. Algumas destas regiões podem ser adquiridas por transferência horizontal e potenciais características específicas do biovar, tal como foi observado em outras *Enterobacteriaceae*.

- Pereira Ude P1, Rodrigues Dos Santos A, Hassan SS, Aburjaile FF, Soares Sde C, Ramos RT, Carneiro AR, Guimarães LC, Silva de Almeida S, Diniz CA, Barbosa MS, Gomes de Sá P, Ali A, Bakhtiar SM, Dorella FA, Zerlotini A, Araújo FM, **Leite LR**, Oliveira G, Miyoshi A, Silva A, Azevedo V, Figueiredo HC. Complete genome sequence of *Streptococcus agalactiae* strain SA20-06, A fish pathogen associated to meningoencephalitis outbreaks. Stand Genomic Sci. 2013 May 25;8(2):188-97. doi: 10.4056/sigs.3687314.

O estudo visou descrever as características genômicas e os principais mecanismos moleculares envolvidos na virulência de *Streptococcus agalactiae*, um agente causador de doença neurológica em peixes, mastite em vacas, e sépsia neonatal nos seres humanos. A bactéria foi isolada a partir de amostras de tilápia (*Oreochromis niloticus*) do Nilo após um surto de meningoencefalite no Brasil.

Outras publicações se referem respectivamente à revisão bibliográfica, apresentação em congresso e a construção de uma ferramenta de visualização para monitoramento de computadores bancários:

- Cuadros-Orellana S, **Leite LR**, Smith A, Medeiros JD, Badotti F, et al. (2013) Assessment of Fungal Diversity in the Environment using Metagenomics: a Decade in Review. *Fungal Genom Biol* 3:110. doi:10.4172/2165-8056.1000110
- Guilherme Oliveira ,**Laura Rabelo Leite**, Julliane Dutra, Fernanda Badotti, Sara Cuadros. Compiling Genetic Parts for Synthetic Biology. Conference paper: International Plant and Animal Genome Conference XXII 2014.
- Yussif Barcelos ; Universidade Federal de Minas Gerais ; Flavia Aburjaile ; **Laura R. Leite** ; Solange T. Oliveira, Raquel C. de Melo-Minardi. Combining traditional and high-density visualizations in a dashboard to network health monitoring. *Visual Analytics Science and Technology (VAST)*, 2012 IEEE Conference on. doi: 10.1109/VAST.2012.6400509.

9– Anexos

Tabela suplementar 1 - Descrição dos primers e condições de PCR. A mistura de reação continha a seguinte composição: 1mM de cada primer, 2x KAPA HiFi HotStart ReadyMix e 12.5 ng de DNA metagenômico.

Primers	Sequência (5'-> 3')	DI	# ciclos	D	A	E	EF
341F	CCTAYGGGRBGCASCAG	98° C	35	98° C	56° C	72° C	72° C
806R	GGACTACNNGGTATCTAAT	>30s		>5s	>20s	>20s	5 min

DI – desnaturação inicial; D - desnaturação; A - anelamento; E - extensão; EF – extensão final.

Tabela suplementar 2 - Estatísticas gerais da montagem de bibliotecas *shotgun*

	OUT.BAS	OUT.BAP	OUT.BSD	ABR.BAS	ABR.BAP	ABR.BSD	AMD.JUN_1	AMD.JUN_2
Nº de <i>contigs</i>	13249	8983	121	29118	8698	535	247381	252735
Nº de bases em todos os <i>contigs</i>	10805941	8939228	75369	23408791	6768606	346434	289161051	259561566
N50	779	987	660	779	691	636	1541	1125
Tamanho do maior <i>contigs</i>	123508	373103	2225	234461	107207	2473	796214	795300
Tamanho médio do <i>contigs</i>	557	653	529	546	526	564	601	574
% <i>contigs</i> > 200 b	100	100	88	100	100	93	99.96	99.91
% <i>contigs</i> > 500 b	63	90	56	61	58	81	70.16	67.51
% <i>contigs</i> > 1 kb	13	21	12	13	9	10	21.89	17.14
% <i>contigs</i> > 5 kb	1	1	0	1	1	0	2.01	1.35
Nº de genes preditos	22.522	17.516	117	50.531	14.172	724	47.339	44.131
Nº de genes anotados	1.457	1.840	41	9.767	2.203	117	2.755	2.788
% de genes anotados	6,47	10,50	35,04	19,33	15,54	16,16	5,82	6,32

*N50: Dado um conjunto de sequências de comprimento variável, N50 é definido como o comprimento N em que 50 % da soma de todas as bases das sequências estão contidas no comprimento $L < N$. Por exemplo, no conjunto $L = 2, 2, 2, 3, 3, 4, 8, 8$ o valor de N50 é 4.

OUT_BAS (água superficial da barragem de rejeitos coletada em outubro), OUT_BAP (água de profundidade da barragem de rejeitos coletada em outubro), OUT_BSD (sedimento da barragem de rejeitos coletada em outubro), ABR_BAS (água superficial da barragem de rejeitos coletada em abril), ABR_BAP (água de profundidade da barragem de rejeitos coletada em abril), ABR_BSD (sedimento da barragem de rejeitos coletada em abril), AMD_JUN (água superficial de drenagem ácida de mina)

Tabela suplementar 3 - Estatísticas gerais da montagem de genomas isolados

Genoma	Nº de sequências após o filtro de qualidade	Nº de bases em todos os scaffolds	Tamanho do maior scaffold	Nº de scaffolds	Conteúdo GC %	Classificação taxonômica
A10	9.882.300	602.644	37.750	66	43	Methanosalsum
N19	8.305.741	430.242	31.611	55	43	Methanosalsum
M14	17.527.795	731.904	28.935	79	43	Methanomethylovorans
A04	10.804.334	710.749	53.598	68	36	Chitinophagaceae
B16	15.096.731	1.752.129	108.731	125	37	Chitinophagaceae
J04	9.535.305	1.579.673	54.353	116	37	Chitinophagaceae
K06	16.014.599	1.106.849	59.529	90	37	Chitinophagaceae
L11	13.017.028	1.105.265	66.431	92	37	Chitinophagaceae
P17	14.832.439	2.082.446	71.392	131	37	Chitinophagaceae
HF	10.481.874	3.413.796	288.419	116	36	Hydrotalea flava
ITV01	5.082.656	3.164.591	318.792	72	52	Acidibacillus ferrooxidans

Figura suplementar 2 - Análise de agrupamento de amostras baseada em *Principal Coordinate Analysis (PCoA)*. Representação criada a partir da matriz de abundância taxonômica de OTUs. As figuras geométricas representam as amostras, e estão coloridas de acordo com a descrição do local e época de amostragem. Elipses em preto representam a tendência de agrupamento das amostras.

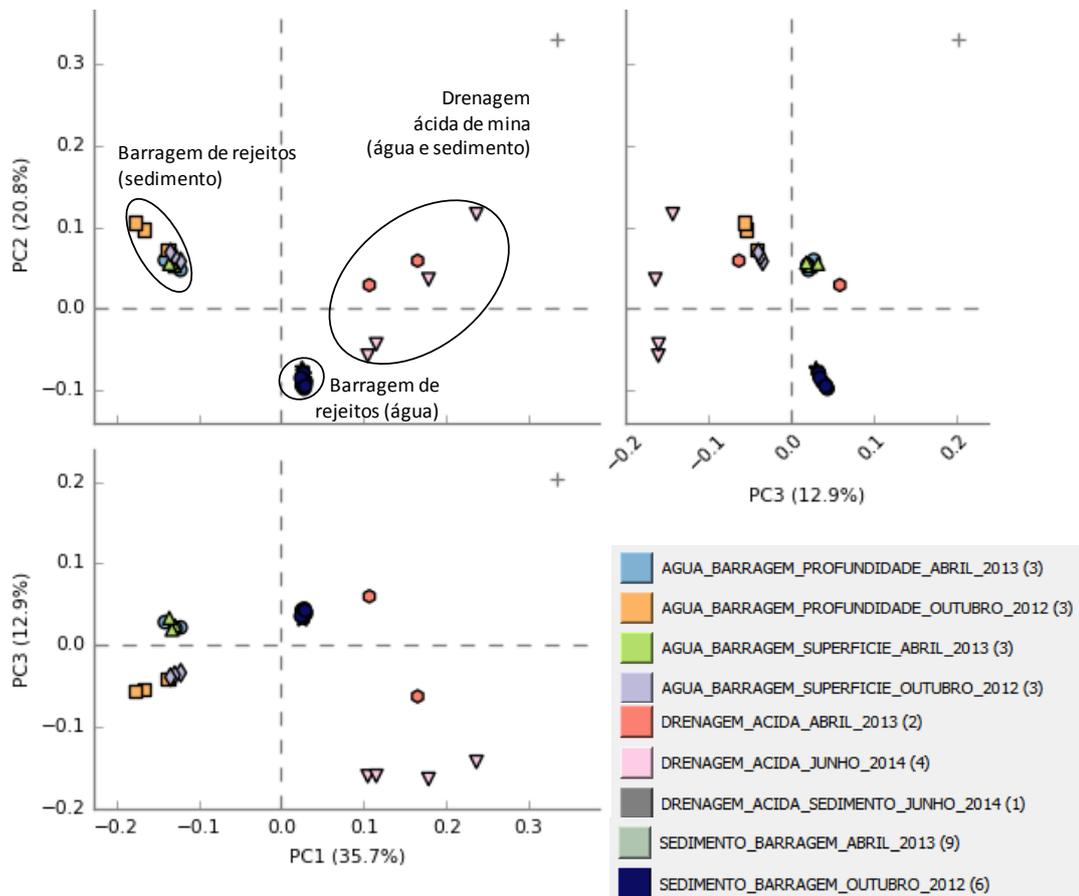
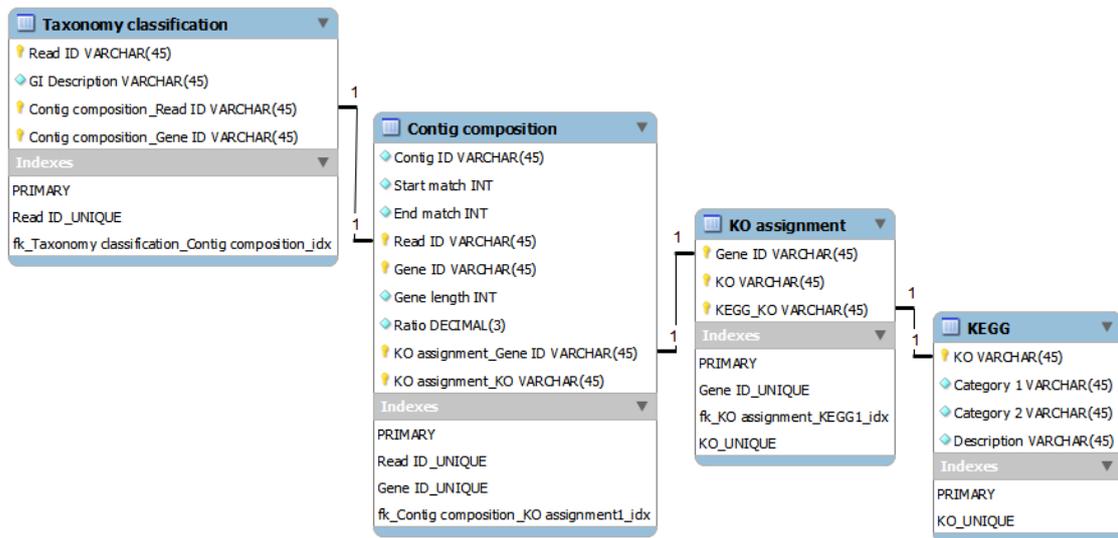


Figura suplementar 3 – Visão geral do banco de dados para integração das classificações taxonômica e funcional nas bibliotecas *shotgun*. A) modelo entidade relacionamento (modelo ER). B) Exemplo dos dados que compõe as tabelas representadas em A.

A



B

READ_ID	GIDESC
M01838:15:000000000-AB1Y1:1:1101:17639:1002	
M01838:15:000000000-AB1Y1:1:1101:17545:1003	Myxococcaceae
M01838:15:000000000-AB1Y1:1:1101:16469:1003	
M01838:15:000000000-AB1Y1:1:1101:18399:1003	Streptomyces badius
M01838:15:000000000-AB1Y1:1:1101:13747:1007	
M01838:15:000000000-AB1Y1:1:1101:12658:1011	freshwater metagenome
M01838:15:000000000-AB1Y1:1:1101:15777:1014	Methylophilus
M01838:15:000000000-AB1Y1:1:1101:19007:1014	
M01838:15:000000000-AB1Y1:1:1101:10539:1021	
M01838:15:000000000-AB1Y1:1:1101:15268:1022	

CONTIG	START_MATCH	END_MATCH	READ_ID	GENE_ID	GENE_LENGTH	RATIO
NODE_1_length_796214_cov_39.0756_ID_1	71	218	M01838:15:000000000-AB1Y1:1:1118:4054:6011/1	gene_id_1	627	0.23445
NODE_1_length_796214_cov_39.0756_ID_1	71	218	M01838:15:000000000-AB1Y1:1:1118:4054:6011/2	gene_id_1	627	0.23445
NODE_1_length_796214_cov_39.0756_ID_1	71	205	M01838:15:000000000-AB1Y1:1:1118:10857:14122/2	gene_id_1	627	0.213716
NODE_1_length_796214_cov_39.0756_ID_1	71	251	M01838:15:000000000-AB1Y1:1:2110:15799:5421/2	gene_id_1	627	0.287081
NODE_1_length_796214_cov_39.0756_ID_1	71	213	M01838:15:000000000-AB1Y1:1:2113:7549:10265/2	gene_id_1	627	0.226475
NODE_1_length_796214_cov_39.0756_ID_1	71	218	M01838:15:000000000-AB1Y1:1:2118:9662:16415/2	gene_id_1	627	0.23445
NODE_1_length_796214_cov_39.0756_ID_1	71	163	M01838:15:000000000-AB1Y1:1:1102:25088:16527/1	gene_id_1	627	0.14673
NODE_1_length_796214_cov_39.0756_ID_1	71	163	M01838:15:000000000-AB1Y1:1:1102:25088:16527/2	gene_id_1	627	0.14673
NODE_1_length_796214_cov_39.0756_ID_1	71	199	M01838:15:000000000-AB1Y1:1:1109:25435:19678/2	gene_id_1	627	0.204147
NODE_1_length_796214_cov_39.0756_ID_1	71	218	M01838:15:000000000-AB1Y1:1:1111:4109:9260/2	gene_id_1	627	0.23445

GENE_ID	KO
gene_id_5	K13993
gene_id_11	K15634
gene_id_13	K00951
gene_id_14	K01972
gene_id_15	K03427
gene_id_17	K00571
gene_id_24	K09981
gene_id_25	K01534
gene_id_27	K03892
gene_id_29	K03695

CAT1	CAT2	KO	DESCRIP
Metabolism	Carbon metabolism	K00844	HK; hexokinase [EC:2.7.1.1]
Metabolism	Carbon metabolism	K12407	GCK; glucokinase [EC:2.7.1.2]
Metabolism	Carbon metabolism	K00845	glk; glucokinase [EC:2.7.1.2]
Metabolism	Carbon metabolism	K00886	ppgK; polyphosphate glucokinase [EC:2.7.1.63]
Metabolism	Carbon metabolism	K00874	ADPGK; ADP-dependent glucokinase [EC:2.7.1.147]
Metabolism	Carbon metabolism	K00918	pfkK; ADP-dependent phosphofructokinase/glucokinase [EC:2.7.1.146 2.7.1.147]
Metabolism	Carbon metabolism	K01810	GPI, pgi; glucose-6-phosphate isomerase [EC:5.3.1.9]
Metabolism	Carbon metabolism	K06859	pgi1; glucose-6-phosphate isomerase, archaeal [EC:5.3.1.9]
Metabolism	Carbon metabolism	K15916	pgi-pmt; glucose/mannose-6-phosphate isomerase [EC:5.3.1.9 5.3.1.8]
Metabolism	Carbon metabolism	K00850	pfkA, PFK; 6-phosphofructokinase 1 [EC:2.7.1.11]

8 - Referências

- 1 MINERAÇÃO., B. I. B. D. **Relatório Anual.**
<http://www.ibram.org.br/sites/1400/1457/00000383.pdf> 2012.
- 2 ADIMB. Valor Econômico.
<http://www.adimb.com.br/site/admin/inc/clipping/244.pdf>, 2012.
- 3 ENERGIA, M. D. M. E. Relatório Técnico 23 Perfil da Mineração de Cobre.
http://www.mme.gov.br/portalmme/opencms/sgm/galerias/arquivos/plano_duo_decenal/a_mineracao_brasileira/P14_RT23_Perfil_da_Mineraçao_de_Cobre.pdf, 2009.
- 4 BIZZI, L. A., SCHOBENHAUS, C., VIDOTTI, R.M., GONÇALVES, J.H. **Depósitos Minerais no Tempo Geológico e Épocas Metalogênicas.** CPRM. http://www.cprm.gov.br/publique/media/cap_VII_b.pdf 2003
- 5 ROCIO, M. A. R., DA SILVA, M.M., CARVALHO, P.S.L., CARDOSO, J.G.R. **Perspectivas atuais da indústria de cobre no Brasil.** SETORIAL, B. http://www.bndes.gov.br/SiteBNDES/export/sites/default/bndes_pt/Galerias/Arquivos/conhecimento/bnset/set3611.pdf 2011.
- 6 JOHNSON, B. D. **Biomining goes underground:** Nature Geoscience. 8: 165-166 p. 2015.
- 7 GADD, G. M. Metals, minerals and microbes: geomicrobiology and bioremediation. **Microbiology**, v. 156, n. Pt 3, p. 609-43, Mar 2010. ISSN 1465-2080. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/20019082> >.
- 8 BAKER, B. J.; BANFIELD, J. F. Microbial communities in acid mine drainage. **FEMS Microbiol Ecol**, v. 44, n. 2, p. 139-52, May 2003. ISSN 1574-6941. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/19719632> >.
- 9 HAWKES, R. B. et al. *Ferroplasma cupricumulans* sp. nov., a novel moderately thermophilic, acidophilic archaeon isolated from an industrial-scale chalcocite bioleach heap. **Extremophiles**, v. 10, n. 6, p. 525-30, Dec 2006. ISSN 1431-0651. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/16721487> >.
- 10 GOLYSHINA, O. V. et al. *Acidiplasma aeolicum* gen. nov., sp. nov., a euryarchaeon of the family Ferroplasmaceae isolated from a hydrothermal pool, and transfer of *Ferroplasma cupricumulans* to *Acidiplasma cupricumulans* comb. nov. **Int J Syst Evol Microbiol**, v. 59, n. Pt 11, p. 2815-23, Nov 2009. ISSN 1466-5026. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/19628615> >.

- 11 CORAM-ULIANA, N. Y., VAN HILLE, R.P., KOHR, W.J., HARRISON, S.T.L. **Development of a method to assay the microbial population in heap bioleaching operation:** Hydrometallurgy. 83: 237-244 p. 2006.
- 12 DIABY, N. et al. Microbial communities in a porphyry copper tailings impoundment and their impact on the geochemical dynamics of the mine waste. **Environ Microbiol**, v. 9, n. 2, p. 298-307, Feb 2007. ISSN 1462-2912. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/17222129> >.
- 13 MENDONÇA, A. M., RABELO, L.F., ROSA, M.A.N., FONSECA, R.A., MACHADO, L.C.R. **Influência da qualidade de água de processo na flotação da mina de Sossego.** http://searchentmme.yang.art.br/download/2011/flota%C3%A7%C3%A3o/2162%20%20Anderson.%20M.%20Mendon%C3%A7a_Lidiane.%20F.%20Rabelo_Marcos.%20A.%20N.%20Rosa_Ronaldo.%20A.%20Fonseca_Luis.C.R.Machado%20%20INFLU%C3%8ANCIA%20DA%20QUALIDADE%20DE%20%20C3%81GUA%20DE%20PROCESSO%20NA%20FLOTA%C3%87%C3%83O%20DA%20MINA%20DE%20SOSSEGO.pdf 2011.
- 14 VASCONCELOS, Y. **Bactérias Mineradoras.** http://revistapesquisa.fapesp.br/wp-content/uploads/2012/10/092-095_extracaominerios_200.pdf 2012.
- 15 DIABY, N. et al. Temporal evolution of bacterial communities associated with the in situ wetland-based remediation of a marine shore porphyry copper tailings deposit. **Sci Total Environ**, v. 533, p. 110-21, Nov 2015. ISSN 1879-1026. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/26151655> >.
- 16 LIU, J. et al. Correlating microbial diversity patterns with geochemistry in an extreme and heterogeneous environment of mine tailings. **Appl Environ Microbiol**, v. 80, n. 12, p. 3677-86, Jun 2014. ISSN 1098-5336. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/24727268> >.
- 17 GILLER, K. E., WITTER, E., MCGRATH, S.P. **Toxicity of heavy metals to microorganisms and microbial processes in agricultural soils: a review:** Soil Biol. Biochem. 30: 1389–1414 p. 1998.
- 18 G.M., G. Microbial influence on metal mobility and application for bioremediation. <http://linkinghub.elsevier.com/retrieve/pii/S0016706104000060>, 2004.
- 19 LEE, M. R.; CORREA, J. A. Effects of copper mine tailings disposal on littoral meiofaunal assemblages in the Atacama region of northern Chile. **Mar Environ Res**, v. 59, n. 1, p. 1-18, Feb 2005. ISSN 0141-1136. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/15325133> >.
- 20 SHARMA, R. S., AL-BUSAIDI, T.S. **Groundwater pollution due to a tailings dam:** Engineering Geology. 60: 235-244 p. 2001.

- 21 KALINA, M., WHEELER, W.N. **Ecological perspectives in restoring mine waste management areas**: *Procedia Environmental Sciences*. 9: 90-95 p. 2011.
- 22 BILLITON. **Creating Value through innovation biotechnology in mining**. <http://www.imm.org.uk/gilbersonpaper.htm>, 2001.
- 23 LI, Y. et al. A review of the structure, and fundamental mechanisms and kinetics of the leaching of chalcopyrite. **Adv Colloid Interface Sci**, v. 197-198, p. 1-32, Sep 2013. ISSN 1873-3727. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/23791420> >.
- 24 RAWLINGS, D. E. Heavy metal mining using microbes. **Annu Rev Microbiol**, v. 56, p. 65-91, 2002. ISSN 0066-4227. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/12142493> >.
- 25 AGARWAL, S. K. **Advanced Environmental Biotechnology**. COMPUTERS, P. http://books.google.com.br/books?id=iSckhBR4FGcC&printsec=frontcover&hl=pt-BR&source=gbs_ge_summary_r&cad=0#v=onepage&q&f=false 2005.
- 26 A., S. **Microorganisms involved in bioleaching and nucleic acid-based molecular methods for their identification and quantification**. Microbial processing of metal sulfides DONATI, E. R., SAND, W. Springer 2007.
- 27 BRUNE, K. D.; BAYER, T. S. Engineering microbial consortia to enhance biomining and bioremediation. **Front Microbiol**, v. 3, p. 203, 2012. ISSN 1664-302X. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/22679443> >.
- 28 ROHWERDER, T. et al. Bioleaching review part A: progress in bioleaching: fundamentals and mechanisms of bacterial metal sulfide oxidation. **Appl Microbiol Biotechnol**, v. 63, n. 3, p. 239-48, Dec 2003. ISSN 0175-7598. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/14566432> >.
- 29 VALENZUELA, L. et al. Genomics, metagenomics and proteomics in biomining microorganisms. **Biotechnol Adv**, v. 24, n. 2, p. 197-211, 2006 Mar-Apr 2006. ISSN 0734-9750. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/16288845> >.
- 30 CRAIG, J., SCOTT, S. **Sulphide phase equilibria**. **Sulfide mineralogy**. http://www.minsocam.org/msa/OpenAccess_Publications/MSA_Special_Papers/MSA_SP_3/MSA_SP3_199-210.pdf: Mineral Soc Am Rev Mineral 1974.
- 31 OLIVEIRA, C., DUARTE, H.A. **Disulphide and metal sulphide formation on the reconstructed surface of chalcopyrite: a DFT study**: *Appl Surf Sci*. 257: 1319–1324 p. 2010.

- 32 NORRIS, P. R., KELLY, D.P. **Dissolution of pyrite (FeS₂) by pure and mixed cultures of some acidophilic bacteria.**: FEMS Microbiol. 4: 143–146 p. 1980.
- 33 PECCIA, J. et al. Development and application of small-subunit rRNA probes for assessment of selected Thiobacillus species and members of the genus Acidiphilium. **Appl Environ Microbiol**, v. 66, n. 7, p. 3065-72, Jul 2000. ISSN 0099-2240. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/10877807> >.
- 34 MARTINEZ, P.; VERA, M.; BOBADILLA-FAZZINI, R. A. Omics on bioleaching: current and future impacts. **Appl Microbiol Biotechnol**, v. 99, n. 20, p. 8337-50, Oct 2015. ISSN 1432-0614. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/26278538> >.
- 35 DOPSON, M.; LINDSTROM, E. B. Potential role of thiobacillus caldus in arsenopyrite bioleaching. **Appl Environ Microbiol**, v. 65, n. 1, p. 36-40, Jan 1999. ISSN 1098-5336. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/9872756> >.
- 36 MCGUIRE M.M., E. K. J., BANFIELD J.F., HAMERS R.J. **Kinetics, surface chemistry, and structural evolution of microbially mediated sulfide mineral dissolution.** : Geochim. 65: 1243–1258 p. 2001.
- 37 BOBADILLA FAZZINI, R. A. et al. Stoichiometric modeling of oxidation of reduced inorganic sulfur compounds (Riscs) in Acidithiobacillus thiooxidans. **Biotechnol Bioeng**, v. 110, n. 8, p. 2242-51, Aug 2013. ISSN 1097-0290. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/23436458> >.
- 38 HUGUES, P. et al. **Continuous bioleaching of a pyrite concentrate in stirred reactors:** Population Hydrometallurgy 94: 34-41 p. 2008.
- 39 JOHNSON, D. B. **Biomining: an established and dynamic biotechnology:** Microbiology Indonesia. 6: 189-193 p. 2013.
- 40 BONNEFOY, V.; HOLMES, D. S. Genomic insights into microbial iron oxidation and iron uptake strategies in extremely acidic environments. **Environ Microbiol**, v. 14, n. 7, p. 1597-611, Jul 2012. ISSN 1462-2920. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/22050575> >.
- 41 VALDÉS, J. et al. Acidithiobacillus ferrooxidans metabolism: from genome sequence to industrial applications. **BMC Genomics**, v. 9, p. 597, 2008. ISSN 1471-2164. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/19077236> >.
- 42 ILBERT, M.; BONNEFOY, V. Insight into the evolution of the iron oxidation pathways. **Biochim Biophys Acta**, v. 1827, n. 2, p. 161-75, Feb 2013. ISSN 0006-3002. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/23044392> >.

- 43 QUATRINI, R. et al. Extending the models for iron and sulfur oxidation in the extreme acidophile *Acidithiobacillus ferrooxidans*. **BMC Genomics**, v. 10, p. 394, 2009. ISSN 1471-2164. Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/19703284> >.
- 44 GADD, G. M. **Microbial influence on metal mobility and application for bioremediation**: *Geoderma*. 122: 109-119 p. 2004.
- 45 ALMÁRCEGUI, R. J. et al. New copper resistance determinants in the extremophile *acidithiobacillus ferrooxidans*: a quantitative proteomic analysis. **J Proteome Res**, v. 13, n. 2, p. 946-60, Feb 2014. ISSN 1535-3907. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/24380576> >.
- 46 _____. Response to copper of *Acidithiobacillus ferrooxidans* ATCC 23270 grown in elemental sulfur. **Res Microbiol**, v. 165, n. 9, p. 761-72, Nov 2014. ISSN 1769-7123. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/25041950> >.
- 47 VERA, M.; SCHIPPERS, A.; SAND, W. Progress in bioleaching: fundamentals and mechanisms of bacterial metal sulfide oxidation--part A. **Appl Microbiol Biotechnol**, v. 97, n. 17, p. 7529-41, Sep 2013. ISSN 1432-0614. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/23720034> >.
- 48 LOVLEY, D. R. Cleaning up with genomics: applying molecular biology to bioremediation. **Nat Rev Microbiol**, v. 1, n. 1, p. 35-44, Oct 2003. ISSN 1740-1526. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/15040178> >.
- 49 MUYZER, G.; STAMS, A. J. The ecology and biotechnology of sulphate-reducing bacteria. **Nat Rev Microbiol**, v. 6, n. 6, p. 441-54, Jun 2008. ISSN 1740-1534. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/18461075> >.
- 50 QUALITY, S. O. M. D. O. E. **Fundamental Principles of Bioremediation**. [http://www.deq.state.ms.us/MDEQ.nsf/0/EDAFD869AB77B3B386256D47005C3109/\\$file/Bioremediation.pdf](http://www.deq.state.ms.us/MDEQ.nsf/0/EDAFD869AB77B3B386256D47005C3109/$file/Bioremediation.pdf) 1998.
- 51 COSTA, M. C. et al. Wine wastes as carbon source for biological treatment of acid mine drainage. **Chemosphere**, v. 75, n. 6, p. 831-6, May 2009. ISSN 1879-1298. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/19201010> >.
- 52 SU, Y. et al. [Rice straw and sewage sludge as carbon sources for sulfate-reducing bacteria treating acid mine drainage]. **Huan Jing Ke Xue**, v. 31, n. 8, p. 1858-63, Aug 2010. ISSN 0250-3301. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/21090305> >.
- 53 BARBOSA, L. P. et al. Nickel, manganese and copper removal by a mixed consortium of sulfate reducing bacteria at a high COD/sulfate ratio. **World J**

- Microbiol Biotechnol**, Apr 2014. ISSN 1573-0972. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/24710619> >.
- 54 BARTON, L. L.; FAUQUE, G. D. Biochemistry, physiology and biotechnology of sulfate-reducing bacteria. **Adv Appl Microbiol**, v. 68, p. 41-98, 2009. ISSN 0065-2164. Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/19426853> >.
- 55 KLEIN, M. et al. Multiple lateral transfers of dissimilatory sulfite reductase genes between major lineages of sulfate-reducing prokaryotes. **J Bacteriol**, v. 183, n. 20, p. 6028-35, Oct 2001. ISSN 0021-9193. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/11567003> >.
- 56 GREIN, F. et al. Unifying concepts in anaerobic respiration: insights from dissimilatory sulfur metabolism. **Biochim Biophys Acta**, v. 1827, n. 2, p. 145-60, Feb 2013. ISSN 0006-3002. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/22982583> >.
- 57 DANIEL, R. The soil metagenome--a rich resource for the discovery of novel natural products. **Curr Opin Biotechnol**, v. 15, n. 3, p. 199-204, Jun 2004. ISSN 0958-1669. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/15193327> >.
- 58 KUNIN, V. et al. A bioinformatician's guide to metagenomics. **Microbiol Mol Biol Rev**, v. 72, n. 4, p. 557-78, Table of Contents, Dec 2008. ISSN 1098-5557. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/19052320> >.
- 59 THOMAS, T.; GILBERT, J.; MEYER, F. Metagenomics - a guide from sampling to data analysis. **Microb Inform Exp**, v. 2, n. 1, p. 3, 2012. ISSN 2042-5783. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/22587947> >.
- 60 DELMONT, T. O. et al. Accessing the soil metagenome for studies of microbial diversity. **Appl Environ Microbiol**, v. 77, n. 4, p. 1315-24, Feb 2011. ISSN 1098-5336. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/21183646> >.
- 61 BARTRAM, A. K. et al. Generation of multimillion-sequence 16S rRNA gene libraries from complex microbial communities by assembling paired-end illumina reads. **Appl Environ Microbiol**, v. 77, n. 11, p. 3846-52, Jun 2011. ISSN 1098-5336. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/21460107> >.
- 62 LAZAREVIC, V. et al. Metagenomic study of the oral microbiota by Illumina high-throughput sequencing. **J Microbiol Methods**, v. 79, n. 3, p. 266-71, Dec 2009. ISSN 1872-8359. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/19796657> >.
- 63 HERNANDEZ, D. et al. De novo bacterial genome sequencing: millions of very short reads assembled on a desktop computer. **Genome Res**, v. 18, n. 5, p.

- 802-9, May 2008. ISSN 1088-9051. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/18332092> >.
- 64 OULAS, A. et al. Metagenomics: tools and insights for analyzing next-generation sequencing data derived from biodiversity studies. **Bioinform Biol Insights**, v. 9, p. 75-88, 2015. ISSN 1177-9322. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/25983555> >.
- 65 SINGH, A. H. et al. Discovering functional novelty in metagenomes: examples from light-mediated processes. **J Bacteriol**, v. 191, n. 1, p. 32-41, Jan 2009. ISSN 1098-5530. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/18849420> >.
- 66 RAES, J.; FOERSTNER, K. U.; BORK, P. Get the most out of your metagenome: computational analysis of environmental sequence data. **Curr Opin Microbiol**, v. 10, n. 5, p. 490-8, Oct 2007. ISSN 1369-5274. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/17936679> >.
- 67 KUNIN, V. et al. A bioinformatician's guide to metagenomics. **Microbiol Mol Biol Rev**, v. 72, n. 4, p. 557-78, Table of Contents, Dec 2008. ISSN 1098-5557. Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/19052320> >.
- 68 SCHMIEDER, R.; EDWARDS, R. Quality control and preprocessing of metagenomic datasets. **Bioinformatics**, v. 27, n. 6, p. 863-4, Mar 2011. ISSN 1367-4811. Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/21278185> >.
- 69 PATEL, R. K.; JAIN, M. NGS QC Toolkit: a toolkit for quality control of next generation sequencing data. **PLoS One**, v. 7, n. 2, p. e30619, 2012. ISSN 1932-6203. Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/22312429> >.
- 70 BOLGER, A. M.; LOHSE, M.; USADEL, B. Trimmomatic: a flexible trimmer for Illumina sequence data. **Bioinformatics**, v. 30, n. 15, p. 2114-20, Aug 2014. ISSN 1367-4811. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/24695404> >.
- 71 HUSON, D. H. et al. MEGAN analysis of metagenomic data. **Genome Res**, v. 17, n. 3, p. 377-86, Mar 2007. ISSN 1088-9051. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/17255551> >.
- 72 STARK, M. et al. MLTreeMap--accurate Maximum Likelihood placement of environmental DNA sequences into taxonomic and functional reference phylogenies. **BMC Genomics**, v. 11, p. 461, 2010. ISSN 1471-2164. Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/20687950> >.

- 73 SCHMIDT, H. A.; VON HAESELER, A. Maximum-likelihood analysis using TREE-PUZZLE. **Curr Protoc Bioinformatics**, v. Chapter 6, p. Unit 6.6, Mar 2007. ISSN 1934-340X. Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/18428792> >.
- 74 WU, M.; EISEN, J. A. A simple, fast, and accurate method of phylogenomic inference. **Genome Biol**, v. 9, n. 10, p. R151, 2008. ISSN 1474-760X. Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/18851752> >.
- 75 GUINDON, S. et al. Estimating maximum likelihood phylogenies with PhyML. **Methods Mol Biol**, v. 537, p. 113-37, 2009. ISSN 1064-3745. Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/19378142> >.
- 76 TEELING, H. et al. TETRA: a web-service and a stand-alone program for the analysis and comparison of tetranucleotide usage patterns in DNA sequences. **BMC Bioinformatics**, v. 5, p. 163, Oct 2004. ISSN 1471-2105. Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/15507136> >.
- 77 WOYKE, T. et al. Symbiosis insights through metagenomic analysis of a microbial consortium. **Nature**, v. 443, n. 7114, p. 950-5, Oct 2006. ISSN 1476-4687. Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/16980956> >.
- 78 WU, Y. W. et al. MaxBin: an automated binning method to recover individual genomes from metagenomes using an expectation-maximization algorithm. **Microbiome**, v. 2, p. 26, 2014. ISSN 2049-2618. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/25136443> >.
- 79 DIAZ, N. N. et al. TACOA: taxonomic classification of environmental genomic fragments using a kernelized nearest neighbor approach. **BMC Bioinformatics**, v. 10, p. 56, 2009. ISSN 1471-2105. Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/19210774> >.
- 80 RUP, L. The human microbiome project. **Indian J Microbiol**, v. 52, n. 3, p. 315, Sep 2012. ISSN 0046-8991. Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/23997318> >.
- 81 LANGILLE, M. G. et al. Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. **Nat Biotechnol**, v. 31, n. 9, p. 814-21, Sep 2013. ISSN 1546-1696. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/23975157> >.
- 82 TRUONG, D. T. et al. MetaPhlan2 for enhanced metagenomic taxonomic profiling. **Nat Methods**, v. 12, n. 10, p. 902-3, Oct 2015. ISSN 1548-7105. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/26418763> >.
- 83 TU, Q.; HE, Z.; ZHOU, J. Strain/species identification in metagenomes using genome-specific markers. **Nucleic Acids Res**, v. 42, n. 8, p. e67, Apr 2014.

ISSN 1362-4962. Disponível em: <
<https://www.ncbi.nlm.nih.gov/pubmed/24523352>>.

- 84 NAMIKI, T. et al. MetaVelvet: an extension of Velvet assembler to de novo metagenome assembly from short sequence reads. **Nucleic Acids Res**, v. 40, n. 20, p. e155, Nov 2012. ISSN 1362-4962. Disponível em: <
<https://www.ncbi.nlm.nih.gov/pubmed/22821567>>.
- 85 LASERSON, J.; JOJIC, V.; KOLLER, D. Genovo: de novo assembly for metagenomes. **J Comput Biol**, v. 18, n. 3, p. 429-43, Mar 2011. ISSN 1557-8666. Disponível em: <
<https://www.ncbi.nlm.nih.gov/pubmed/21385045>>.
- 86 PENG, Y. et al. Meta-IDBA: a de Novo assembler for metagenomic data. **Bioinformatics**, v. 27, n. 13, p. i94-101, Jul 2011. ISSN 1367-4811. Disponível em: <
<https://www.ncbi.nlm.nih.gov/pubmed/21685107>>.
- 87 BROWN, C. T. CAREER: ABI: Assembling Extremely Large Metagenomes. <http://ged.msu.edu/downloads/2012-career-nsf-final.pdf>, 2012.
- 88 SOGIN, M. L. et al. Microbial diversity in the deep sea and the underexplored "rare biosphere". **Proc Natl Acad Sci U S A**, v. 103, n. 32, p. 12115-20, Aug 2006. ISSN 0027-8424. Disponível em: <
<http://www.ncbi.nlm.nih.gov/pubmed/16880384>>.
- 89 MOUNT, D. W. Using the Basic Local Alignment Search Tool (BLAST). **CSH Protoc**, v. 2007, p. pdb.top17, 2007. Disponível em: <
<http://www.ncbi.nlm.nih.gov/pubmed/21357135>>.
- 90 BADGER, J. H.; OLSEN, G. J. CRITICA: coding region identification tool invoking comparative analysis. **Mol Biol Evol**, v. 16, n. 4, p. 512-24, Apr 1999. ISSN 0737-4038. Disponível em: <
<https://www.ncbi.nlm.nih.gov/pubmed/10331277>>.
- 91 FRISHMAN, D. et al. Combining diverse evidence for gene recognition in completely sequenced bacterial genomes. **Nucleic Acids Res**, v. 26, n. 12, p. 2941-7, Jun 1998. ISSN 0305-1048. Disponível em: <
<https://www.ncbi.nlm.nih.gov/pubmed/9611239>>.
- 92 SOLOVYEV, V. **Automatic Annotation of Microbial Genomes and Metagenomic Sequences**. SALAMOV, A. In *Metagenomics and its Applications in Agriculture, Biomedicine and Environmental Studies* (Ed. R.W. Li): Nova Science Publishers: 61-78 p. 2011.
- 93 BESEMER, J.; BORODOVSKY, M. GeneMark: web software for gene finding in prokaryotes, eukaryotes and viruses. **Nucleic Acids Res**, v. 33, n. Web Server issue, p. W451-4, Jul 2005. ISSN 1362-4962. Disponível em: <
<https://www.ncbi.nlm.nih.gov/pubmed/15980510>>.

- 94 GARDNER, P. P. The use of covariance models to annotate RNAs in whole genomes. **Brief Funct Genomic Proteomic**, v. 8, n. 6, p. 444-50, Nov 2009. ISSN 1477-4062. Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/19833700> >.
- 95 TSAI, M. C. et al. Long noncoding RNA as modular scaffold of histone modification complexes. **Science**, v. 329, n. 5992, p. 689-93, Aug 2010. ISSN 1095-9203. Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/20616235> >.
- 96 HAFT, D. H.; SELENGUT, J. D.; WHITE, O. The TIGRFAMs database of protein families. **Nucleic Acids Res**, v. 31, n. 1, p. 371-3, Jan 2003. ISSN 1362-4962. Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/12520025> >.
- 97 FINN, R. D. et al. The Pfam protein families database: towards a more sustainable future. **Nucleic Acids Res**, v. 44, n. D1, p. D279-85, Jan 2016. ISSN 1362-4962. Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/26673716> >.
- 98 BORATYN, G. M. et al. Domain enhanced lookup time accelerated BLAST. **Biol Direct**, v. 7, p. 12, 2012. ISSN 1745-6150. Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/22510480> >.
- 99 BUCHFINK, B.; XIE, C.; HUSON, D. H. Fast and sensitive protein alignment using DIAMOND. **Nat Methods**, v. 12, n. 1, p. 59-60, Jan 2015. ISSN 1548-7105. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/25402007> >.
- 100 ALTSCHUL, S. F. et al. Basic local alignment search tool. **J Mol Biol**, v. 215, n. 3, p. 403-10, Oct 1990. ISSN 0022-2836. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/2231712> >.
- 101 FONG, C. et al. PSAT: a web tool to compare genomic neighborhoods of multiple prokaryotic genomes. **BMC Bioinformatics**, v. 9, p. 170, 2008. ISSN 1471-2105. Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/18366802> >.
- 102 SERRES, M. H.; RILEY, M. Gene fusions and gene duplications: relevance to genomic annotation and functional analysis. **BMC Genomics**, v. 6, p. 33, 2005. ISSN 1471-2164. Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/15757509> >.
- 103 WU, J.; HU, Z.; DELISI, C. Gene annotation and network inference by phylogenetic profiling. **BMC Bioinformatics**, v. 7, p. 80, 2006. ISSN 1471-2105. Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/16503966> >.

- 104 WONG, D. C.; SWEETMAN, C.; FORD, C. M. Annotation of gene function in citrus using gene expression information and co-expression networks. **BMC Plant Biol**, v. 14, p. 186, 2014. ISSN 1471-2229. Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/25023870> >.
- 105 KANEHISA, M. et al. KEGG as a reference resource for gene and protein annotation. **Nucleic Acids Res**, v. 44, n. D1, p. D457-62, Jan 2016. ISSN 1362-4962. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/26476454> >.
- 106 KONWAR, K. M. et al. MetaPathways: a modular pipeline for constructing pathway/genome databases from environmental sequence information. **BMC Bioinformatics**, v. 14, p. 202, 2013. ISSN 1471-2105. Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/23800136> >.
- 107 MEYER, F. et al. The metagenomics RAST server - a public resource for the automatic phylogenetic and functional analysis of metagenomes. **BMC Bioinformatics**, v. 9, p. 386, 2008. ISSN 1471-2105. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/18803844> >.
- 108 ANGIUOLI, S. V. et al. CloVR: a virtual machine for automated and portable sequence analysis from the desktop using cloud computing. **BMC Bioinformatics**, v. 12, p. 356, 2011. ISSN 1471-2105. Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/21878105> >.
- 109 KULTIMA, J. R. et al. MOCAT2: a metagenomic assembly, annotation and profiling framework. **Bioinformatics**, Apr 2016. ISSN 1367-4811. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/27153620> >.
- 110 ARUMUGAM, M. et al. SmashCommunity: a metagenomic annotation and analysis tool. **Bioinformatics**, v. 26, n. 23, p. 2977-8, Dec 2010. ISSN 1367-4811. Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/20959381> >.
- 111 LOZUPONE, C. et al. UniFrac: an effective distance metric for microbial community comparison. **ISME J**, v. 5, n. 2, p. 169-72, Feb 2011. ISSN 1751-7370. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/20827291> >.
- 112 SANGWAN, N.; XIA, F.; GILBERT, J. A. Recovering complete and draft population genomes from metagenome datasets. **Microbiome**, v. 4, p. 8, 2016. ISSN 2049-2618. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/26951112> >.
- 113 SAEED, I.; TANG, S. L.; HALGAMUGE, S. K. Unsupervised discovery of microbial population structure within metagenomes using nucleotide base composition. **Nucleic Acids Res**, v. 40, n. 5, p. e34, Mar 2012. ISSN 1362-4962. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/22180538> >.

- 114 ULTSCH A, M. F. **ESOM-Maps: tools for clustering, visualization, and classification with emergent SOM.** . <https://www.uni-marburg.de/fb12/datenbionik/pdf/pubs/2005/ultsch05esom>: Germany: Data Bionics Research Group, University of Marburg 2005.
- 115 DICK, G. J. et al. Community-wide analysis of microbial genome sequence signatures. **Genome Biol**, v. 10, n. 8, p. R85, 2009. ISSN 1474-760X. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/19698104> >.
- 116 ALBERTSEN, M. et al. Genome sequences of rare, uncultured bacteria obtained by differential coverage binning of multiple metagenomes. **Nat Biotechnol**, v. 31, n. 6, p. 533-8, Jun 2013. ISSN 1546-1696. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/23707974> >.
- 117 NIELSEN, H. B. et al. Identification and assembly of genomes and genetic elements in complex metagenomic samples without using reference genomes. **Nat Biotechnol**, v. 32, n. 8, p. 822-8, Aug 2014. ISSN 1546-1696. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/24997787> >.
- 118 IMELFORT, M. et al. GroopM: an automated tool for the recovery of population genomes from related metagenomes. **PeerJ**, v. 2, p. e603, 2014. ISSN 2167-8359. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/25289188> >.
- 119 KANG, D. D. et al. MetaBAT, an efficient tool for accurately reconstructing single genomes from complex microbial communities. **PeerJ**, v. 3, p. e1165, 2015. ISSN 2167-8359. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/26336640> >.
- 120 STEPANAUSKAS, R. Single cell genomics: an individual look at microbes. **Curr Opin Microbiol**, v. 15, n. 5, p. 613-20, Oct 2012. ISSN 1879-0364. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/23026140> >.
- 121 RINKE, C. et al. Obtaining genomes from uncultivated environmental microorganisms using FACS-based single-cell genomics. **Nat Protoc**, v. 9, n. 5, p. 1038-48, May 2014. ISSN 1750-2799. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/24722403> >.
- 122 WOYKE, T. et al. One bacterial cell, one complete genome. **PLoS One**, v. 5, n. 4, p. e10314, 2010. ISSN 1932-6203. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/20428247> >.
- 123 MCCULLOCH, J. A. **Single cell genomics.** CHAPARRO, J. P. <http://sbmicrobiologia.org.br/wp-content/uploads/2015/09/Revista17.pdf> *Microbiologia* in foco 2011.
- 124 CHITSAZ, H. et al. Efficient de novo assembly of single-cell bacterial genomes from short-read data sets. **Nat Biotechnol**, v. 29, n. 10, p. 915-21, Oct 2011.

- ISSN 1546-1696. Disponível em: <
<https://www.ncbi.nlm.nih.gov/pubmed/21926975> >.
- 125 BANKEVICH, A. et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. **J Comput Biol**, v. 19, n. 5, p. 455-77, May 2012. ISSN 1557-8666. Disponível em: <
<http://www.ncbi.nlm.nih.gov/pubmed/22506599> >.
- 126 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. 1420-8, Jun 2012. ISSN 1367-4811. Disponível em: <
<https://www.ncbi.nlm.nih.gov/pubmed/22495754> >.
- 127 BINNEWIES, T. T. et al. Ten years of bacterial genome sequencing: comparative-genomics-based discoveries. **Funct Integr Genomics**, v. 6, n. 3, p. 165-85, Jul 2006. ISSN 1438-793X. Disponível em: <
<http://www.ncbi.nlm.nih.gov/pubmed/16773396> >.
- 128 MEWIS, K.; TAUPP, M.; HALLAM, S. J. A high throughput screen for biomining cellulase activity from metagenomic libraries. **J Vis Exp**, n. 48, 2011. ISSN 1940-087X. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/21307835> >.
- 129 DALL'AGNOL, H. et al. Draft Genome Sequence of "Acidibacillus ferrooxidans" ITV01, a Novel Acidophilic Firmicute Isolated from a Chalcopyrite Mine Drainage Site in Brazil. **Genome Announc**, v. 4, n. 2, 2016. ISSN 2169-8287. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/26988062> >.
- 130 VASZITA, E., SZABÓ, J., & GRUIZ, K. **Complex leaching of metal-sulfide-containing mine waste and soil in microcosms.** <https://hungary.pure.elsevier.com/hu/publications/complex-leaching-of-metal-sulfide-containing-mine-waste-and-soil-> Land Contamination and Reclamation. 17: 463-471 p. 2009.
- 131 REDDY, U. A. et al. Biomarkers of oxidative stress in rat for assessing toxicological effects of heavy metal pollution in river water. **Environ Sci Pollut Res Int**, v. 22, n. 17, p. 13453-63, Sep 2015. ISSN 1614-7499. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/25940462> >.
- 132 ROMO, E. et al. Bacterial consortium for copper extraction from sulphide ore consisting mainly of chalcopyrite. **Braz J Microbiol**, v. 44, n. 2, p. 523-8, 2013. ISSN 1678-4405. Disponível em: <
<http://www.ncbi.nlm.nih.gov/pubmed/24294251> >.
- 133 CAPORASO, J. G. et al. Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. **ISME J**, v. 6, n. 8, p. 1621-4, Aug 2012. ISSN 1751-7370. Disponível em: <
<http://www.ncbi.nlm.nih.gov/pubmed/22402401> >.

- 134 PYLRO, V. S. et al. Brazilian Microbiome Project: revealing the unexplored microbial diversity--challenges and prospects. **Microb Ecol**, v. 67, n. 2, p. 237-41, Feb 2014. ISSN 1432-184X. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/24173537> >.
- 135 MAGOČ, T.; SALZBERG, S. L. FLASH: fast length adjustment of short reads to improve genome assemblies. **Bioinformatics**, v. 27, n. 21, p. 2957-63, Nov 2011. ISSN 1367-4811. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/21903629> >.
- 136 EDGAR, R. C. Search and clustering orders of magnitude faster than BLAST. **Bioinformatics**, v. 26, n. 19, p. 2460-1, Oct 2010. ISSN 1367-4811. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/20709691> >.
- 137 CAPORASO, J. G. et al. QIIME allows analysis of high-throughput community sequencing data. **Nat Methods**, v. 7, n. 5, p. 335-6, May 2010. ISSN 1548-7105. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/20383131> >.
- 138 HUSON, D. H. et al. MEGAN analysis of metagenomic data. **Genome Res**, v. 17, n. 3, p. 377-86, Mar 2007. ISSN 1088-9051. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/17255551> >.
- 139 ZHU, W.; LOMSADZE, A.; BORODOVSKY, M. Ab initio gene identification in metagenomic sequences. **Nucleic Acids Res**, v. 38, n. 12, p. e132, Jul 2010. ISSN 1362-4962. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/20403810> >.
- 140 BOUTET, E. et al. UniProtKB/Swiss-Prot, the Manually Annotated Section of the UniProt KnowledgeBase: How to Use the Entry View. **Methods Mol Biol**, v. 1374, p. 23-54, 2016. ISSN 1940-6029. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/26519399> >.
- 141 TATUSOV, R. L. et al. The COG database: an updated version includes eukaryotes. **BMC Bioinformatics**, v. 4, p. 41, Sep 2003. ISSN 1471-2105. Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/12969510> >.
- 142 ASHBURNER, M. et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. **Nat Genet**, v. 25, n. 1, p. 25-9, May 2000. ISSN 1061-4036. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/10802651> >.
- 143 FERNANDES, G. R. et al. A procedure to recruit members to enlarge protein family databases--the building of UECOG (UniRef-Enriched COG Database) as a model. **Genet Mol Res**, v. 7, n. 3, p. 910-24, 2008. ISSN 1676-5680. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/18949709> >.

- 144 LANGMEAD, B. et al. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. **Genome Biol**, v. 10, n. 3, p. R25, 2009. ISSN 1474-760X. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/19261174> >.
- 145 LI, H. et al. The Sequence Alignment/Map format and SAMtools. **Bioinformatics**, v. 25, n. 16, p. 2078-9, Aug 2009. ISSN 1367-4811. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/19505943> >.
- 146 QUINLAN, A. R. BEDTools: The Swiss-Army Tool for Genome Feature Analysis. **Curr Protoc Bioinformatics**, v. 47, p. 11.12.1-34, 2014. ISSN 1934-340X. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/25199790> >.
- 147 WU, Y. W.; SIMMONS, B. A.; SINGER, S. W. MaxBin 2.0: an automated binning algorithm to recover genomes from multiple metagenomic datasets. **Bioinformatics**, v. 32, n. 4, p. 605-7, Feb 2016. ISSN 1367-4811. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/26515820> >.
- 148 PARKS, D. H. et al. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. **Genome Res**, v. 25, n. 7, p. 1043-55, Jul 2015. ISSN 1549-5469. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/25977477> >.
- 149 GRAPHPAD SOFTWARE, I. A. **What you can conclude when two error bars overlap (or don't)?**: <http://www.graphpad.com/faq/viewfaq.cfm?faq=1362> p. 2008.
- 150 LOVE, M. I.; HUBER, W.; ANDERS, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. **Genome Biol**, v. 15, n. 12, p. 550, 2014. ISSN 1474-760X. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/25516281> >.
- 151 PACE, N. R. A molecular view of microbial diversity and the biosphere. **Science**, v. 276, n. 5313, p. 734-40, May 1997. ISSN 0036-8075. Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/9115194> >.
- 152 WOESE, C. R.; FOX, G. E. Phylogenetic structure of the prokaryotic domain: the primary kingdoms. **Proc Natl Acad Sci U S A**, v. 74, n. 11, p. 5088-90, Nov 1977. ISSN 0027-8424. Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/270744> >.
- 153 TSAI, Y. C. et al. Resolving the Complexity of Human Skin Metagenomes Using Single-Molecule Sequencing. **MBio**, v. 7, n. 1, p. e01948-15, 2016. ISSN 2150-7511. Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/26861018> >.

- 154 GERLACH, W.; STOYE, J. Taxonomic classification of metagenomic shotgun sequences with CARMA3. **Nucleic Acids Res**, v. 39, n. 14, p. e91, Aug 2011. ISSN 1362-4962. Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/21586583> >.
- 155 SHABALINA, S. A. et al. Selective constraint in intergenic regions of human and mouse genomes. **Trends Genet**, v. 17, n. 7, p. 373-6, Jul 2001. ISSN 0168-9525. Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/11418197> >.
- 156 SRIDHAR, J. et al. Junker: an intergenic explorer for bacterial genomes. **Genomics Proteomics Bioinformatics**, v. 9, n. 4-5, p. 179-82, Oct 2011. ISSN 2210-3244. Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/22196361> >.
- 157 EUGENE, V. K. M., Y.G. **Computational Approaches in Comparative Genomics**. <http://www.ncbi.nlm.nih.gov/books/NBK20260/>: Kluwer Academic 2003.
- 158 VALDIVIA-GRANDA, W. The next meta-challenge for Bioinformatics. **Bioinformatics**, v. 2, n. 8, p. 358-62, 2008. ISSN 0973-2063. Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/18685725> >.
- 159 BORI, J. et al. Geochemistry and environmental threats of soils surrounding an abandoned mercury mine. **Environ Sci Pollut Res Int**, v. 23, n. 13, p. 12941-53, Jul 2016. ISSN 1614-7499. Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/26996905> >.
- 160 ROBERTSON, G. P. et al. The biogeochemistry of bioenergy landscapes: carbon, nitrogen, and water considerations. **Ecol Appl**, v. 21, n. 4, p. 1055-67, Jun 2011. ISSN 1051-0761. Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/21774413> >.
- 161 BRASIL, C. N. D. M. A. C. **Resolução nº 357, de 17 de março de 2005**. . Dispõe sobre a classificação dos corpos de água e diretrizes ambientais para o seu enquadramento, bem como estabelece as condições e padrões de lançamento de efluentes, e dá outras providências. Diário Oficial da República Federativa do Brasil, Brasília, DF, 18 mar 2005.
- 162 MÉNDEZ-GARCÍA, C. et al. Microbial diversity and metabolic networks in acid mine drainage habitats. **Front Microbiol**, v. 6, p. 475, 2015. ISSN 1664-302X. Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/26074887> >.
- 163 AL-RUB, A. **Biosorption of copper on Chlorella vulgaris from single binary and ternary metal aqueous solutions**. F.A., E.-N., M.H., ASHOUR, I., AL-MARZOUQI, M. Process Biochem. 41: 457-464 p. 2006.

- 164 NUHOGLU, Y. **Thermodynamic and kinetic studies for environmentally friendly Ni(II) biosorption using waste pomace of olive oil factory** MALKOC, E. *Bio-resource technology*. 100: 2375-2380 p. 2009.
- 165 MELLO, J. W. V. **Origem e Controle do Fenômeno Drenagem Ácida de Mina.** DUARTE, J. W. V., LADEIRA, ANA CLÁUDIA Q. <http://qnesc.sbg.org.br/online/cadernos/08/06-CTN4.pdf>: *Cadernos Temáticos de Química Nova na Escola*. 8: 24-29 p. 2014.
- 166 STEGER, H. F. Oxidation of sulphide minerals--I: determination of ferrous and ferric iron in samples of pyrrhotite, pyrite and chalcopyrite. **Talanta**, v. 24, n. 4, p. 251-4, Apr 1977. ISSN 0039-9140. Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/18962075> >.
- 167 BAKER, B. J.; BANFIELD, J. F. *Microbial communities in acid mine drainage*. ISSN 2.
- 168 SRI LAKSHMI SUNITA, M. et al. Molecular identification of arsenic-resistant estuarine bacteria and characterization of their ars genotype. **Ecotoxicology**, v. 21, n. 1, p. 202-12, Jan 2012. ISSN 1573-3017. Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/21879358> >.
- 169 MÉNDEZ-GARCÍA, C. et al. Microbial stratification in low pH oxic and suboxic macroscopic growths along an acid mine drainage. **ISME J**, v. 8, n. 6, p. 1259-74, Jun 2014. ISSN 1751-7370. Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/24430486> >.
- 170 TUNDISI, J. G. **Limnologia.** TUNDISI, T. M. https://books.google.com/books?hl=pt-BR&lr=&id=6LcWDAAAQBAJ&oi=fnd&pg=PT5&ots=8INWuWJMi-&sig=Da002OAE4iZdn77_tMdvxCI-Bw4: Oficina de Textos, São Paulo 2008.
- 171 ALMEIDA, D. F. **Weight-of-evidence on environmental impact assessment of metal contaminated sediments in the São Francisco River (Três Marias - Minas Gerais - Brazil) - a case study.** MARTINS, A. H. e TUNDISI, J. G.: *Journal brasileiro de Biology*. 71: 961- 973 p. 2011.
- 172 SMITH, J. L. **The Significance of soil biomass estimates.** PAUL, E. A. I. B., J.M. e STOTTZKY, G. *Soil Biochemistry* 357-396 p. 1990.
- 173 KUANG, J. et al. Predicting taxonomic and functional structure of microbial communities in acid mine drainage. **ISME J**, v. 10, n. 6, p. 1527-39, Jun 2016. ISSN 1751-7370. Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/26943622> >.

- 174 MARGALEF, R. **Homage to Evelyn Hutchinson, or why is there an upper limit to diversity**. Connecticut Academy of Arts and Sciences. 14: 211-235 p. 1972.
- 175 MAGURRAN, A. E. **Measuring biological diversity**. Oxford, Blackwell Science: 256 p. 2004.
- 176 MORRIS, E. K. et al. Choosing and using diversity indices: insights for ecological applications from the German Biodiversity Exploratories. **Ecol Evol**, v. 4, n. 18, p. 3514-24, Sep 2014. ISSN 2045-7758. Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/25478144> >.
- 177 STYLE="FONT-SIZE:12.0PT, S. et al. **Measuring biological diversity**. Blackwell Publishing: 256 p. 2006.
- 178 _____ . **Methods for fitting dominance/diversity curves**. <http://onlinelibrary.wiley.com/doi/10.2307/3235896/pdf>> **Journal of Vegetation Science**. 2: 35-46 p. 1991.
- 179 ZIEGLER, S. et al. Oxygen-dependent niche formation of a pyrite-dependent acidophilic consortium built by archaea and bacteria. **ISME J**, v. 7, n. 9, p. 1725-37, Sep 2013. ISSN 1751-7370. Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/23619304> >.
- 180 HUA, Z. S. et al. Ecological roles of dominant and rare prokaryotes in acid mine drainage revealed by metagenomics and metatranscriptomics. **ISME J**, v. 9, n. 6, p. 1280-94, Jun 2015. ISSN 1751-7370. Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/25361395> >.
- 181 NEWTON, R. J. et al. A guide to the natural history of freshwater lake bacteria. **Microbiol Mol Biol Rev**, v. 75, n. 1, p. 14-49, Mar 2011. ISSN 1098-5557. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/21372319> >.
- 182 PERNTHALER, J. et al. Predator-specific enrichment of actinobacteria from a cosmopolitan freshwater clade in mixed continuous culture. **Appl Environ Microbiol**, v. 67, n. 5, p. 2145-55, May 2001. ISSN 0099-2240. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/11319094> >.
- 183 LI, M.; GU, J. D. The diversity and distribution of anammox bacteria in the marine aquaculture zones. **Appl Microbiol Biotechnol**, Jul 2016. ISSN 1432-0614. Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/27368740> >.
- 184 MADRID, V. M. et al. Phylogenetic diversity of bacterial and archaeal communities in the anoxic zone of the Cariaco Basin. **Appl Environ Microbiol**, v. 67, n. 4, p. 1663-74, Apr 2001. ISSN 0099-2240. Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/11282619> >.

- 185 NUNES, I. et al. Coping with copper: legacy effect of copper on potential activity of soil bacteria following a century of exposure. **FEMS Microbiol Ecol**, Aug 2016. ISSN 1574-6941. Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/27543319> >.
- 186 RANJAN, K. et al. Forest-to-pasture conversion increases the diversity of the phylum Verrucomicrobia in Amazon rainforest soils. **Front Microbiol**, v. 6, p. 779, 2015. ISSN 1664-302X. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/26284056> >.
- 187 DUNFIELD, P. F. et al. Methane oxidation by an extremely acidophilic bacterium of the phylum Verrucomicrobia. **Nature**, v. 450, n. 7171, p. 879-82, Dec 2007. ISSN 1476-4687. Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/18004300> >.
- 188 ERIKSTAD, H. A.; BIRKELAND, N. K. Draft Genome Sequence of "Candidatus Methylacidiphilum kamchatkense" Strain Kam1, a Thermoacidophilic Methanotrophic Verrucomicrobium. **Genome Announc**, v. 3, n. 2, 2015. ISSN 2169-8287. Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/25745002> >.
- 189 YIN, H. et al. An integrated insight into the response of sedimentary microbial communities to heavy metal contamination. **Sci Rep**, v. 5, p. 14266, 2015. ISSN 2045-2322. Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/26391875> >.
- 190 BUECHE, M.; JUNIER, P. Effect of organic carbon and metal accumulation on the bacterial communities in sulphidogenic sediments. **Environ Sci Pollut Res Int**, v. 23, n. 11, p. 10443-56, Jun 2016. ISSN 1614-7499. Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/26780045> >.
- 191 LOVLEY, D. R. Bioremediation of organic and metal contaminants with dissimilatory metal reduction. **J Ind Microbiol**, v. 14, n. 2, p. 85-93, Feb 1995. ISSN 0169-4146. Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/7766214> >.
- 192 IINO, T. et al. Candidatus Methanogranum caenicola: a novel methanogen from the anaerobic digested sludge, and proposal of Methanomassiliicoccaceae fam. nov. and Methanomassiliicoccales ord. nov., for a methanogenic lineage of the class Thermoplasmata. **Microbes Environ**, v. 28, n. 2, p. 244-50, 2013. ISSN 1347-4405. Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/23524372> >.
- 193 KATO, S.; HASHIMOTO, K.; WATANABE, K. Methanogenesis facilitated by electric syntrophy via (semi)conductive iron-oxide minerals. **Environ Microbiol**, v. 14, n. 7, p. 1646-54, Jul 2012. ISSN 1462-2920. Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/22004041> >.

- 194 NEWTON, R. J. et al. Phylogenetic ecology of the freshwater Actinobacteria acI lineage. **Appl Environ Microbiol**, v. 73, n. 22, p. 7169-76, Nov 2007. ISSN 0099-2240. Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/17827330> >.
- 195 MICHAUD, L. et al. Snow surface microbiome on the High Antarctic Plateau (DOME C). **PLoS One**, v. 9, n. 8, p. e104505, 2014. ISSN 1932-6203. Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/25101779> >.
- 196 COLE, J. R. et al. Ribosomal Database Project: data and tools for high throughput rRNA analysis. **Nucleic Acids Res**, v. 42, n. Database issue, p. D633-42, Jan 2014. ISSN 1362-4962. Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/24288368> >.
- 197 SOERGEL, D. A. et al. Selection of primers for optimal taxonomic classification of environmental 16S rRNA gene sequences. **ISME J**, v. 6, n. 7, p. 1440-4, Jul 2012. ISSN 1751-7370. Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/22237546> >.
- 198 YANG, B.; WANG, Y.; QIAN, P. Y. Sensitivity and correlation of hypervariable regions in 16S rRNA genes in phylogenetic analysis. **BMC Bioinformatics**, v. 17, p. 135, 2016. ISSN 1471-2105. Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/27000765> >.
- 199 FRANZOSA, E. A. et al. Sequencing and beyond: integrating molecular 'omics' for microbial community profiling. **Nat Rev Microbiol**, v. 13, n. 6, p. 360-72, Jun 2015. ISSN 1740-1534. Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/25915636> >.
- 200 SCHIRALDI, C.; GIULIANO, M.; DE ROSA, M. Perspectives on biotechnological applications of archaea. **Archaea**, v. 1, n. 2, p. 75-86, Sep 2002. ISSN 1472-3646. Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/15803645> >.
- 201 LOMANS, B. P. et al. Isolation and characterization of Methanomethylivorans hollandica gen. nov., sp. nov., isolated from freshwater sediment, a methylotrophic methanogen able to grow on dimethyl sulfide and methanethiol. **Appl Environ Microbiol**, v. 65, n. 8, p. 3641-50, Aug 1999. ISSN 0099-2240. Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/10427061> >.
- 202 SEKHON RANDHAWA, K. K.; RAHMAN, P. K. Rhamnolipid biosurfactants-past, present, and future scenario of global market. **Front Microbiol**, v. 5, p. 454, 2014. ISSN 1664-302X. Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/25228898> >.
- 203 KLINDWORTH, A. et al. Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies.

- Nucleic Acids Res**, v. 41, n. 1, p. e1, Jan 2013. ISSN 1362-4962. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/22933715> >.
- 204 DOLD, B. et al. Low molecular weight carboxylic acids in oxidizing porphyry copper tailings. **Environ Sci Technol**, v. 39, n. 8, p. 2515-21, Apr 2005. ISSN 0013-936X. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/15884343> >.
- 205 KOCK, D.; SCHIPPERS, A. Quantitative microbial community analysis of three different sulfidic mine tailing dumps generating acid mine drainage. **Appl Environ Microbiol**, v. 74, n. 16, p. 5211-9, Aug 2008. ISSN 1098-5336. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/18586975> >.
- 206 GUO, J. et al. Horizontal gene transfer in an acid mine drainage microbial community. **BMC Genomics**, v. 16, p. 496, 2015. ISSN 1471-2164. Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/26141154> >.
- 207 TAN, G. L. et al. Seasonal and spatial variations in microbial community structure and diversity in the acid stream draining across an ongoing surface mining site. **FEMS Microbiol Ecol**, v. 70, n. 2, p. 121-9, Nov 2009. ISSN 1574-6941. Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/19678846> >.
- 208 FRIEDRICH, C. G. Physiology and genetics of sulfur-oxidizing bacteria. **Adv Microb Physiol**, v. 39, p. 235-89, 1998. ISSN 0065-2911. Disponível em: < <https://www.ncbi.nlm.nih.gov/pubmed/9328649> >.
- 209 BURMØLLE, M. et al. Enhanced biofilm formation and increased resistance to antimicrobial agents and bacterial invasion are caused by synergistic interactions in multispecies biofilms. **Appl Environ Microbiol**, v. 72, n. 6, p. 3916-23, Jun 2006. ISSN 0099-2240. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/16751497> >.