

**UNIVERSIDADE FEDERAL DE MINAS GERAIS**

Instituto de Ciências Biológicas

Programa de Pós-graduação em Genética

Maria Luíza Soares Suhadolnik

**FATORES AMBIENTAIS MOLDANDO A ASSOCIAÇÃO ENTRE OS GENES  
ENVOLVIDOS NOS CICLOS DO ARSÊNIO E DO FERRO: uma meta-análise**

Belo Horizonte

2018

Maria Luíza Soares Suhadolnik

**FATORES AMBIENTAIS MOLDANDO A ASSOCIAÇÃO ENTRE OS GENES  
ENVOLVIDOS NOS CICLOS DO ARSÊNIO E DO FERRO: uma meta-análise**

Dissertação apresentada ao curso de pós-graduação do Departamento de Genética, Ecologia e Evolução, Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais, como requisito a obtenção do título de Mestre em Genética.

Orientadora: Andrea Maria Amaral Nascimento

Co-orientador: Francisco Pereira Lobo

BELO HORIZONTE

Fevereiro - 2018

043

Suhadolnik, Maria Luíza Soares.

Fatores ambientais moldando a associação entre os genes envolvidos nos ciclos do arsênio e do ferro: uma meta-análise [manuscrito] / Maria Luíza Soares Suhadolnik. – 2018.

97 f. : il. ; 29,5 cm.

Orientadora: Andrea Maria Amaral Nascimento. Co-orientador: Francisco Pereira Lobo.

Dissertação (mestrado) – Universidade Federal de Minas Gerais, Instituto de Ciências Biológicas. Programa de Pós-Graduação em Genética.

1. Genética. 2. Metanálise. 3. Metagenoma. 4. Arsênio. 5. Ferro. 6. Oxidação. 7. Biotransformação. I. Nascimento, Andrea Maria Amaral. II. Lobo, Francisco Pereira. III. Universidade Federal de Minas Gerais. Instituto de Ciências Biológicas. IV. Título.

CDU: 575



ATA DA DEFESA DE DISSERTAÇÃO

**Maria Luiza Soares Suhadolnik**

266/2018  
entrada  
1º/2016  
CPF:  
097.975.866-18

Às nove horas do dia **27 de fevereiro de 2018**, reuniu-se, no Instituto de Ciências Biológicas da UFMG, a Comissão Examinadora de Dissertação, indicada pelo Colegiado do Programa, para julgar, em exame final, o trabalho intitulado: "**Fatores ambientais moldando a associação entre os genes envolvidos nos ciclos do arsênio e ferro: uma meta-análise**", requisito para obtenção do grau de Mestre em **Genética**. Abrindo a sessão, a Presidente da Comissão, **Dra. Andrea Maria Amaral Nascimento**, após dar a conhecer aos presentes o teor das Normas Regulamentares do Trabalho Final, passou a palavra à candidata, para apresentação de seu trabalho. Seguiu-se a arguição pelos Examinadores, com a respectiva defesa da candidata. Logo após, a Comissão se reuniu, sem a presença da candidata e do público, para julgamento e expedição de resultado final. Foram atribuídas as seguintes indicações:

Prof./Pesq.	Instituição	CPF	Indicação
Dra. Andrea Maria Amaral Nascimento	UFMG	477340606-25	Aprovada
Dr. Francisco Pereira Lobo	UFMG	012273736-34	Aprovada
Dr. Adolf Heinrich Horn	UFMG	000.754547-82	aprovado
Dr. Renan Pedra de Souza	UFMG	064.481.066-01	APROVADO

Pelas indicações, a candidata foi considerada: \_\_\_\_\_  
O resultado final foi comunicado publicamente à candidata pela Presidente da Comissão. Nada mais havendo a tratar, a Presidente encerrou a reunião e lavrou a presente ATA, que será assinada por todos os membros participantes da Comissão Examinadora.  
**Belo Horizonte, 27 de fevereiro de 2018.**

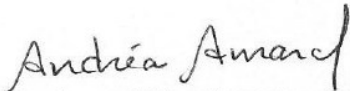
Dra. Andrea Maria Amaral Nascimento - Orientadora Andrea Amaral  
Dr. Francisco Pereira Lobo - Coorientador Francisco P. Lobo  
Dr. Adolf Heinrich Horn Adolf Horn  
Dr. Renan Pedra de Souza Renan Pedra de Souza

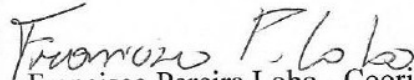



**"Fatores ambientais moldando a associação entre os genes envolvidos  
nos ciclos do arsênio e ferro: uma meta-análise"**

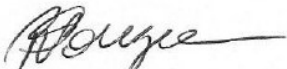
**Maria Luiza Soares Suhadolnik**

Dissertação aprovada pela banca examinadora constituída pelos Professores:

  
Andrea Maria Amaral Nascimento - Orientadora  
UFMG

  
Francisco Pereira Lobo - Coorientador  
Universidade Federal de Minas Gerais

  
Adolf Heinrich Horn  
UFMG

  
Renan Pedra de Souza  
UFMG

Belo Horizonte, 27 de fevereiro de 2018.

## **AGRADECIMENTOS**

À professora Andrea Amaral, pela orientação, amizade e enorme confiança em minha capacidade e neste projeto;

Ao professor Francisco Lobo, pelo suporte e boas ideias;

A Giovanni de Marques Castro, por embarcar neste projeto de última hora e me ensinar toda a bioinformática necessária para executá-lo;

Aos meus amigos do LGM, pelo incentivo, conselhos e por tornar os dias mais felizes;

Aos meus pais e ao meu irmão, pelo incentivo diário e pela paciência;

A todos meus amigos que estiveram presentes durante essa jornada;

Aos membros da Banca avaliadora.

## RESUMO

As relações entre procariotos oxirredutores de As e Fe contribuíram para delinear aspectos geoquímicos da Terra primitiva; e, atualmente, têm grande impacto nos ecossistemas, especialmente favorecendo a mobilização do As e conseqüente contaminação de água e solo. Microrganismos e mecanismos envolvidos nos ciclos do As e Fe são conhecidos. Entretanto, as correlações entre os genes de oxirredução desses elementos, bem como sua abrangência e diversidade ambiental não são bem esclarecidos. Este estudo apresenta a primeira meta-análise da diversidade, frequência relativa e correlações entre os genes envolvidos na biotransformação do As e Fe, e dos fatores físico-químicos que os influenciam. Vinte conjuntos de dados de sequenciamento *shotgun* obtidos de bancos de dados públicos, correspondendo à ambientes ricos em As e/ou Fe foram analisados. Genomas foram montados *de novo*, e as sequências codificantes (CDSs) foram anotadas tendo os bancos de dados Uniprot, RefSeq, Pfam e TIGRFAM como referência. Finalmente, buscou-se na anotação por genes amplamente descritos envolvidos nas reações redox do As e Fe. De 44 genes selecionados, 12 foram encontrados: *aioA* (arsenito oxidase), *arsC* (arseniato redutase detoxificadora), *arrA* (arseniato redutase dissimilatória), *ccoN* e *ccoP* (operon *cco* - complexo citocromo oxidase Cbb3), *cyc1* e *rus* (operon *rus*), *iro* (Fe oxidase), *petA*, *petB* e *petC* (operon *pet* - complexo citocromo bc1) e *omcB* (Fe redutase). Os resultados mostraram que a distribuição ambiental, frequência relativa e correlações entre esses genes estão de acordo com a termodinâmica predita para a especiação do As e Fe, sendo o pH a variável mais relevante. Os genes envolvidos no ciclo do As foram ubíquos, sendo *ArsC* provavelmente a principal enzima responsável pela redução do As(V) no ambiente, atribuindo aos microrganismos resistentes ao As (ARMs) uma posição de destaque na mobilização do As. Outro dado relevante é que, possivelmente, *arrA* esteja sub-representado nos bancos de dados. Dentre os genes envolvidos no ciclo do Fe, o operon *cco* foi ubíquo e o mais frequente, sugerindo um maior *fitness* para seus hospedeiros. *iro* e *cco* parecem estar implicados na oxidação do Fe(II) acompanhada pela imobilização de As(V) em ambientes óxicos. A mobilização de As em consequência da redução de Fe(III) pelo cluster *mtr/omc* depende do pH e das concentrações de As e Fe do ambiente. Finalmente, foram encontrados um grande número de citocromos cujo envolvimento específico na oxirredução do Fe não é esclarecido. Este estudo forneceu informações novas e importantes sobre a mobilização do

As associada aos processos de oxirredução do Fe, sendo que a detoxificação do As predomina sobre as transformações metabólicas no ambiente.

Palavras-chave: Meta-análise, *Shotgun*, Metagenoma, Arsênio, Ferro, oxidação, redução, ARMs, DARBs,

## ABSTRACT

The relationships between prokaryotes performing As and Fe redox reactions contributed to delineate geochemical aspects of the primitive Earth. Currently, they have great impact on ecosystems, specially favouring As mobilization and the consequent water and soil contamination. Microorganisms and mechanisms involved in As and Fe speciation are known, nevertheless, correlations between its oxidation and reduction genes, as well its environmental distribution and diversity are not well established. This work presents the first meta-analysis of the diversity, relative frequencies and correlations between genes involved in As and Fe biotransformations, and the physical-chemical parameters that influence them. Twenty shotgun sequencing datasets, corresponding to As and/or Fe-rich environments, were obtained from public databases and analyzed. Genomes were assembled *de novo* and coding sequences (CDSs) were annotated using Uniprot, RefSeq, Pfam and TIGRFAM as reference databases. Finally, genes with well-described involvement in As and Fe speciation were sought within the annotated datasets. From 44 selected genes, 12 were found: *aioA* (arsenite oxidase), *arsC* (detoxifying arsenate reductase), *arrA* (dissimilatory arsenate reductase), *ccoN* and *ccoP* (*cco* operon – cytochrome oxidase Cbb3 complex), *cyc1* and *rus* (*rus* operon), *iro* (Fe oxidase), *petA*, *petB* and *petC* (*pet* operon - cytochrome bc1 complex) and *omcB* (Fe reductase). The results showed that the environmental distribution, relative frequencies and correlations between these genes correspond to the predicted thermodynamics for the As and Fe speciation, with pH as the most relevant parameter. Genes involved in As cycling were ubiquitous, and ArsC enzyme is likely the main responsible for As(V) reduction in the environment, thus attributing to arsenite resistant microorganisms (ARMs) an important role in As mobilization. Another relevant finding is that *arrA* is possibly under-represented in databases. Among the genes involved in Fe cycling, *cco* operon was ubiquitous and the most frequent, suggesting that its hosts might have an increased fitness. *iro* and *cco* seem to be implicated in Fe(II) oxidation followed by As(V) immobilization in oxic environments. As mobilization due to Fe(III) reduction by *mtr/omc* cluster depend on pH and As and Fe environmental concentrations. Finally, a great number of cytochromes with poorly described roles in Fe redox reactions were found. This work provided new relevant data about As mobilization associated with Fe redox reactions, highlighting that As detoxification processes prevail over the metabolic transformations in the environment.

Keywords: Meta-analysis, Metagenome, Arsenic, Iron, Oxidation, Reduction, ARMs, DARBs.

## LISTA DE ILUSTRAÇÕES

Figura 1 - Esquema gráfico da especiação do As e Fe de acordo com a faixa de pH e a concentração de oxigênio no ambiente	17
Figura 2 - Modelo proposto da oxidação do Fe(II) em <i>Acidithiobacillus ferrooxidans</i> e genes envolvidos	19
Figura 3 - Modelo proposto da redução do Fe(III) em <i>Shewanella spp.</i> e <i>Geobacter spp.</i>	21
Figura 4 - Fluxograma de uma análise metagenômica shotgun	25
Quadro 1 - Conjuntos de dados selecionados para a meta-análise	31
Quadro 2 - Relação de genes envolvidos na oxidação e redução de Fe e As em procariotos	44
Figura 5 - Fluxograma da distribuição dos conjuntos de dados da meta-análise	51
Figura 6 - Classificação dos conjuntos de dados da meta-análise em faixas de pH	53
Figura 7 - Frequência relativa dos genes investigados por tipo de ambiente	58
Figura 8 - Análise fatorial	60
Figura 9 - Distribuição de genes envolvidos na especiação do As e Fe em função do pH e presença/ausência de oxigênio	61
Figura 10 - Coeficientes de correlação (Spearman) entre os genes investigados em ambientes ricos em As (A), em As e Fe (B), em Fe (C) e ambientes referência (D)	64
Figura 11 - Coeficientes de correlação (Spearman) entre os genes investigados em ambientes classificados como muito ácidos (A), ácidos (B), neutros (C) e alcalinos (D).	66

## TABELAS

Tabela 1 – Etapas do processamento das sequencias de cada amostra	55
Tabela 2 - Relação de citocromos encontrados neste estudo que não foram previamente selecionados	57
Tabela 3 - Frequência relativa dos genes envolvidos na transformação de As e Fe por faixa de pH	60

## LISTA DE ABREVIATURAS E SIGLAS

ARM	<i>Arsenate Resistant Microorganism</i> -Microorganismo resistente a arseniato
As	Arsênio
As(III)	Arsenito
As(V)	Arseniato
CAO	Chemoautotrophic As(III) Oxidizer - Oxidador quimioautotrófico de arsenito
CDS	Sequência codificante de proteína
DARP	<i>Dissimilatory Arsenate Respiring Prokaryote</i> - Procarioto redutor dissimilatório de arseniato
DNA	<i>Deoxyribonucleic Acid</i> - ÁcidoDesoxirribonucleioco
Eh	Potencial redox
Fe	Ferro
FRM	<i>Fe Respiring Microorganism</i> - Microorganismo respirador de ferro
H <sub>2</sub> O	Água
H <sub>2</sub> O <sub>2</sub>	Peróxido de hidrogênio
HAOs	<i>Heterotrophic As(III) Oxidizers</i> - Oxidadores heterotróficos de As(III)
HMM	<i>Hidden Markov Model</i> - Modelo Oculto de Markov
O <sub>2</sub>	Oxigênio
OMS	Organização Mundial de saúde
pb	Pares de base
pH	Potencial hidrogeniônico
ppm	Partes por milhão
rRNA	RNA ribossômico
SRA	<i>Sequence Read Archive</i>
WGS	<i>Whole Genome Shotgun</i> – <i>shotgun</i> do genoma total

## SUMÁRIO

1. INTRODUÇÃO .....	15
1.1. Oxidação e redução do Fe e do As por procariotos .....	16
1.1.1. Especiação do Ferro .....	17
1.1.2. Especiação do As.....	22
1.2. Meta-análise de metagenomas <i>shotgun</i> .....	24
2. JUSTIFICATIVA E RELEVÂNCIA .....	27
3. OBJETIVOS.....	28
3.1. Objetivo Geral.....	28
3.2. Objetivos específicos .....	28
4. MATERIAIS E MÉTODOS .....	29
4.1. Busca e seleção de conjunto de dados metagenômicos.....	29
4.2. Análise bioinformática .....	35
4.2.1. Controle de qualidade das sequências.....	36
4.2.3. Predição gênica e anotação funcional .....	41
4.3. Análises estatísticas.....	50
5. RESULTADOS.....	54
5.1. Disponibilidade de metadados como um fator limitante.....	54

5.2. Análise bioinformática .....	54
5.3. Análises estatísticas.....	56
5.3.1. Diversidade e frequência relativa de genes envolvidos nos ciclos do As e Fe.....	56
5.3.2. Interações entre os genes envolvidos nas ciclagens do As e do Fe.....	61
6. DISCUSSÃO.....	68
7. CONCLUSÃO .....	74
8. REFERÊNCIAS BIBLIOGRÁFICAS .....	75
ANEXO 1: Busca por genes envolvidos nas biotransformações do As e Fe nos conjuntos de dados da meta-análise.....	84
ANEXO 2: Análises estatísticas com resultados inconclusivos .....	88

## 1. INTRODUÇÃO

O ferro (Fe), quarto elemento mais abundante na crosta terrestre, é um micronutriente essencial para todas as formas de vida, cuja versatilidade de valência o permite ocupar uma posição central nas reações de oxirredução, tanto no meio ambiente quanto em meio celular (Weber et al., 2006). A biotransformação do Fe é considerada uma das formas mais ancestrais de metabolismo microbiano no planeta, alterando suas características geoquímicas e deixando evidências como as Formações Ferríferas Bandadas, depósito de camadas alternadas de minerais de Fe e sílica que datam do Pré-Cambriano. A oxidação e a redução de espécies de Fe continuam afetando significativamente os ambientes recentes: estão intimamente ligados ao ciclo do Fe processos como a mobilidade de micronutrientes, ciclagem do carbono, e a degradação, transformação e circulação de poluentes orgânicos e inorgânicos (Kappler & Straub, 2005; Weber et al., 2006).

Dentre tais poluentes encontra-se o arsênio (As), metaloide tóxico que figura entre os dez químicos de maior preocupação para a saúde pública, segundo a Organização Mundial de Saúde (WHO, 2016). Assim como o Fe, o As, encontrado em baixas concentrações, é um componente natural da superfície terrestre; e microrganismos também ocupam um papel central em sua ciclagem. A oxidação e redução procariótica de espécies de As torna-o mais ou menos móvel no ambiente, com implicação direta em sua toxicidade para os seres vivos (Oremland & Stolz, 2005). Seu metabolismo também data do Pré-Cambriano (Jackson & Dugas, 2003; Lebrun et al., 2003; Duval et al., 2008; Richey et al., 2009).

A interação entre bactérias redutoras de Fe(III) e arseniato (As(V)) tem um papel central na mobilização do As de sedimentos em águas subterrâneas (Islam et al., 2004, 2005). O consumo a longo prazo de água e alimentos contaminados por As leva ao desenvolvimento de câncer e lesões de pele, além de estar associado com neurotoxicidade e doenças cardiovasculares. Portanto, este processo representa uma ameaça à saúde de milhares de pessoas, principalmente em regiões cuja constituição geológica é naturalmente abundante em As, como a Índia, Bangladesh, Argentina, Chile e Estados Unidos. Ademais, outras fontes do metaloide incluem atividades antrópicas como mineração, *smelting* e alguns processos industriais (WHO, 2016).

Segundo Oremland & Stolz (2005), procariotos redutores de metal podem participar da mobilização do As através da redução simultânea de Fe(III) e As(V), levando à liberação de

arsenito (As(III)) no ambiente; através da redução direta de As(V) por procariotos que realizam redução dissimilatória de arseniato (DARPs) e microrganismos resistentes a arseniato (ARMs); ou através da redução do Fe(III) em minerais, com consequente liberação do As(V) adsorvido em sua superfície (Oremland & Stolz, 2005). Estudos realizados por Islam et al. (2004) mostraram que, nos sedimentos ricos em As de Bengala ocidental, o processo predominante é a liberação de As(V) em consequência da redução de Fe(III) (Islam et al., 2004). As demais possíveis interações entre biotransformadores de Fe e As (interações entre microrganismos transformadores de As e oxidadores de Fe em aerobiose e anaerobiose, e redutores de Fe em aerobiose) ainda não foram estudadas detalhadamente.

A maioria dos estudos para identificar os procariotos e os genes envolvidos nos ciclos do As e Fe, separadamente, é baseado em métodos de cultivo tradicionais (Kappler & Straub, 2005; Weber et al., 2006; Andres & Bertin, 2016). Recentemente, tecnologias de sequenciamento em larga escala têm contribuído para tal caracterização (Kozubal et al., 2012; Fortney et al., 2016; Kurth et al., 2016; Suhadolnik et al., 2017), resultando enorme quantidade de dados pouco explorados. Neste contexto, um estudo mais completo das associações entre os metabolismos microbianos do As e do Fe pode ser realizado a partir da riqueza de dados obtidos de diversos ambientes e localizações, disponíveis em bancos de dados públicos.

### 1.1. Oxidação e redução do Fe e do As por procariotos

As espécies iônicas de As e Fe disponíveis para as reações de oxiredução realizadas por microrganismos são diretamente influenciadas por fatores abióticos. Entre os fatores bióticos, destaca-se o aparato genético dos procariotos presentes na comunidade. A Figura 1 ilustra as espécies de As e Fe predominantes nos ambientes naturais em função dos fatores abióticos presença/ausência de oxigênio, pH, e procariotos biotransformadores de As e Fe.

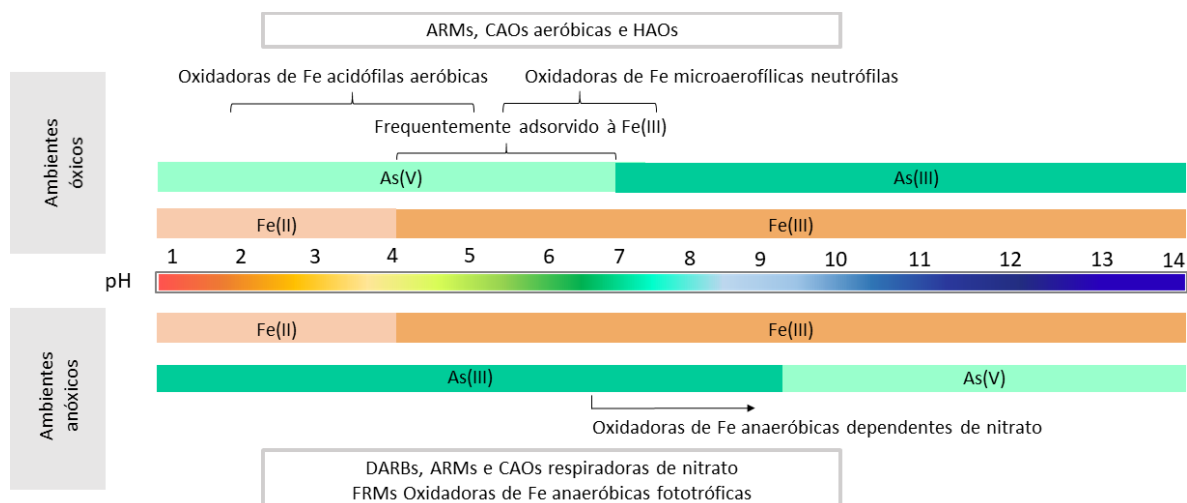


Figura 1: Esquema gráfico da especiação do As e Fe de acordo com a faixa de pH e a concentração de oxigênio no ambiente. As barras verde e laranja indicam as espécies predominantes de As e Fe, respectivamente, em diferentes faixas de pH. Procariotos envolvidos nas reações redox do As e Fe estão indicados no espectro de pH e na concentração de oxigênio encontrados no ambiente. Abreviações: ARMs, microrganismos resistentes à arseniato; CAOs, microrganismos oxidadores quimioautotróficos de arsenito; HAOs, microrganismos oxidadores heterotróficos de arsenito; DARBs, bactérias redutoras de arseniato dissimilatórias; FRMs, microrganismos respiradores de ferro. Fonte: a autora.

### 1.1.1. Especiação do Ferro

Os estados de oxidação Fe(II) e Fe(III) prevalecem na natureza, sendo a especiação química do Fe fortemente influenciada por parâmetros físico-químicos como concentração de oxigênio, pH e potencial redox do meio (Eh). Em ambientes óxicos e  $\text{pH} \geq 4$ , Fe(III) é predominante, reagindo com  $\text{H}_2\text{O}$  em meio aquático e dessa forma apresentando-se dissolvido em baixas concentrações. Em ambientes óxicos e  $\text{pH} < 4$ , a espécie Fe(II) é dominante. De forma similar, em ambiente anóxicos Fe(II) será estável em  $\text{pH} < 4$ , tornando-se um importante capturador de elétrons para processos químicos e biológicos; somente óxidos de manganês e nitrato foram demonstrados capazes de oxidar Fe(II) dissolvido na ausência de oxigênio (Kappler & Straub 2005).

A especiação do Fe mediada por microrganismos se dá através da oxidação de Fe(II) e redução de Fe(III). Oxidadores aeróbicos acidofílicos têm o oxigênio como único acceptor de elétrons termodinamicamente favorável para a oxidação do Fe(II), em  $\text{pH} = 2$ . Oxidadores

aeróbicos neutrofilicos, por outro lado, são capazes de utilizar aceptores de elétrons alternativos ao oxigênio, como nitrito e nitrato em  $\text{pH} = 7$ . Tais oxidadores competem bem com a oxidação química do  $\text{Fe(II)}$ , embora tenham mais sucesso em ambientes microaerófilicos, onde a velocidade desta reação é menor. Outra implicação disto é que, em  $\text{pH}$  neutro, a oxidação do  $\text{Fe(II)}$  para ganho energético pode ocorrer tanto em ambientes óxicos quanto anóxicos (Bonnefoy & Holmes, 2012; Kappler & Straub, 2005).

Oxidadores anaeróbicos de Fe podem ser bactérias fototróficas ou as já mencionadas redutoras de nitrato.  $\text{Fe(II)}$  é doador de elétrons para a fotossíntese de bactérias púrpuras e verdes, que lidam bem com  $\text{Fe(II)}$  dissolvido, mas até o presente não tem demonstrado capacidade de utilizar minerais insolúveis. Atualmente, a oxidação anaeróbica de  $\text{Fe(II)}$  por bactérias fototróficas é considerada uma explicação alternativa ou adicional para a geração das Formações Ferríferas Bandadas, que por muito tempo foram consideradas produto da oxidação química ou microbiana do  $\text{Fe(II)}$  dissolvido nos oceanos com o oxigênio liberado por cianobactérias. Bactérias oxidadoras de  $\text{Fe(II)}$  e redutoras de nitrato em anaerobiose conseguem realizar esta reação porque, em  $\text{pH}$  neutro, todas as espécies de nitrato têm potencial redox mais positivo que o do par  $\text{Fe(II)/Fe(III)}$ , sendo assim aceptoras de elétrons. A maioria destas bactérias cultiváveis necessitam de meio mixotrófico com  $\text{Fe(II)}$  para seu crescimento (Kappler & Straub, 2005).

Procariotos que oxidam Fe são encontrados tanto no domínio Archaea quanto no domínio Bacteria, particularmente no filo Proteobacteria. As vias de oxidação nestes organismos variam muito em relação às proteínas que as compõem, levando à especulação de que esta habilidade tenha surgido independentemente mais de uma vez durante a evolução. Este mecanismo é melhor caracterizado na bactéria acidofílica *Acidithiobacillus ferrooxidans* (grupo I), cujo modelo foi proposto pelo trabalho pioneiro de Ingledew (1982) e posteriormente validado por estudos experimentais e computacionais (Bonnefoy & Holmes, 2012).

Segundo este modelo (Figura 2A), oxidação do  $\text{Fe(II)}$  ocorreria na membrana externa através do citocromo c *Cyc2* nela embebido, que transfere os elétrons para a rusticianina, uma proteína de cobre periplasmática. Os elétrons então podem seguir duas vias distintas: na via *downhill*, passam pelo citocromo c4 *Cyc1* e pelo complexo citocromo oxidase tipo aa3, culminando com a redução do  $\text{O}_2$  em  $\text{H}_2\text{O}$ . Alternativamente, na via *uphill*, da rusticianina os elétrons passam pelo citocromo c4 *CycA1*, pelo complexo *bc1* e quinonas (moléculas

carreadoras de elétrons) associadas à membrana até o complexo NADH1. Esta via não é termodinamicamente favorável, acredita-se que a energia necessária para que ocorra seja obtida do fluxo de elétrons gerado pela diferença de concentração entre os meios intra e extracelulares; essa habilidade parece ser conferida por um complexo *bc1*. Os genes que codificam *Cyc1* (*cyc1*), *Cyc2* (*cyc2*), rusticianina (*rusA*) e a citocromo oxidase tipo aa3 (*coxA*, *coxB*, *coxC*) estão localizados no operon *rus* (Figura 2B) (Bonnefoy & Holmes, 2012; Hedrich et al., 2012).

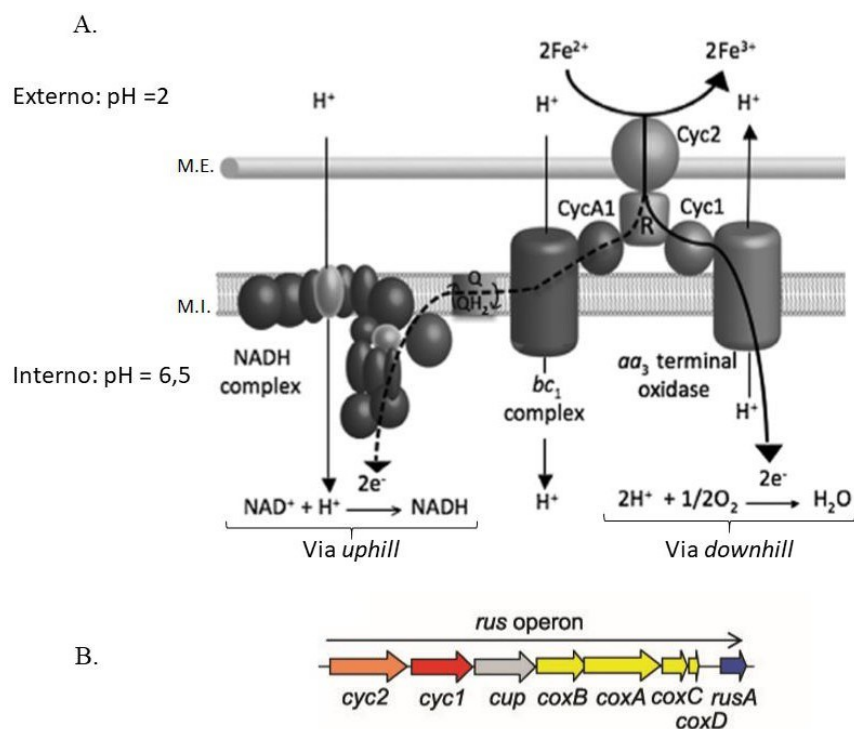


Figura 2. Modelo proposto da oxidação do Fe(II) em *Acidithiobacillus ferrooxidans* e genes envolvidos: a) proteínas envolvidas na oxidação *uphill* e *downhill* do Fe(II). As setas sólidas denotam o influxo de elétrons, a seta pontilhada indica o fluxo de elétrons na via *uphill* de oxidação. Abreviações: R, rusticianina, ME, membrana externa, MI, membrana interna, Q, quinona; b) organização do operon *rus* de *At. ferrooxidans* ATCC 23270. Mais detalhes são descritos no texto. Fonte: adaptada de Bonnefoy & Holmes, 2012 e Hedrich et al., 2012.

Alguns outros mecanismos divergentes já foram relatados. As vias de oxidação de acidithiobacilli dos grupos III (*At. ferrivorans*) e IV (*Acidithiobacillus* sp. JCM 7811) parecem não requerer o operon *rus*, de forma que a oxidação do Fe ocorreria por alguma via ainda não elucidada; uma variante da rusticianina (tipo B) participa do processo, e a proteína Iro (codificada por *iro*) está envolvida de forma mais direta. Em *Rhodopseudomonas palustris* TIE-1 o operon *pio* está implicado na oxidação fototrófica de Fe, este operon contém três

genes que codificam um citocromo *c* (*pioA*, regulado positivamente na presença de Fe(II)), uma proteína de membrana externa (*pioB*) e uma proteína Fe-S de alto potencial redoxsemelhante à Iro (*pioC*). Um último exemplo é o operon *fox* na oxidadora fototrófica *Rhodobacter* sp. SW2, que contém três genes (*foxE*, *foxY*, *foxZ*) que conferem uma atividade de oxidação aumentada (Croal et al., 2007; Jiao et al., 2007; Hedrich et al., 2012).

A redução dissimilatória do Fe(III) (redução do Fe(III) para ganho energético acoplado à oxidação de doadores de elétrons orgânicos ou inorgânicos é considerada uma das formas mais ancestrais de metabolismo microbiano, antecedendo até mesmo as respirações de oxigênio, nitrato e sulfato (Weber et al., 2006). Em procariotos acidófilos esta habilidade é ubíqua, ocorrendo tanto em ambientes anóxicos quanto microóxicos. A respiração dissimilatória de Fe(III) também é amplamente distribuída em microrganismos neutrófilos; entretanto, estes enfrentam um desafio adicional, uma vez que em pH neutro Fe(III) é pouco solúvel. Para superar este obstáculo, procariotos podem valer-se do contato físico com óxidos de Fe(III) para o transporte de elétrons através de *pili* (ou nanofios), podem produzir moléculas transportadoras de elétrons, contornando a necessidade do contato físico, ou então podem produzir moléculas quelantes (sideróforos) para aumentar a solubilidade do Fe(III) (Reyes et al., 2008).

*Shewanella* spp. e *Geobacter* spp. são os organismos modelo para o estudo da redução do Fe(III) por microrganismos e, embora suas vias possuam algumas similaridades, muitos componentes diferem. Nessas bactérias, a redução ocorre fora da célula: os elétrons são transportados de um *pool* de quinonas na membrana citoplasmática para citocromos *c*, e então para uma proteína localizada na membrana externa, presumidamente uma Fe redutase terminal, que ainda não foi identificada (Figura 3) (Weber et al., 2006).

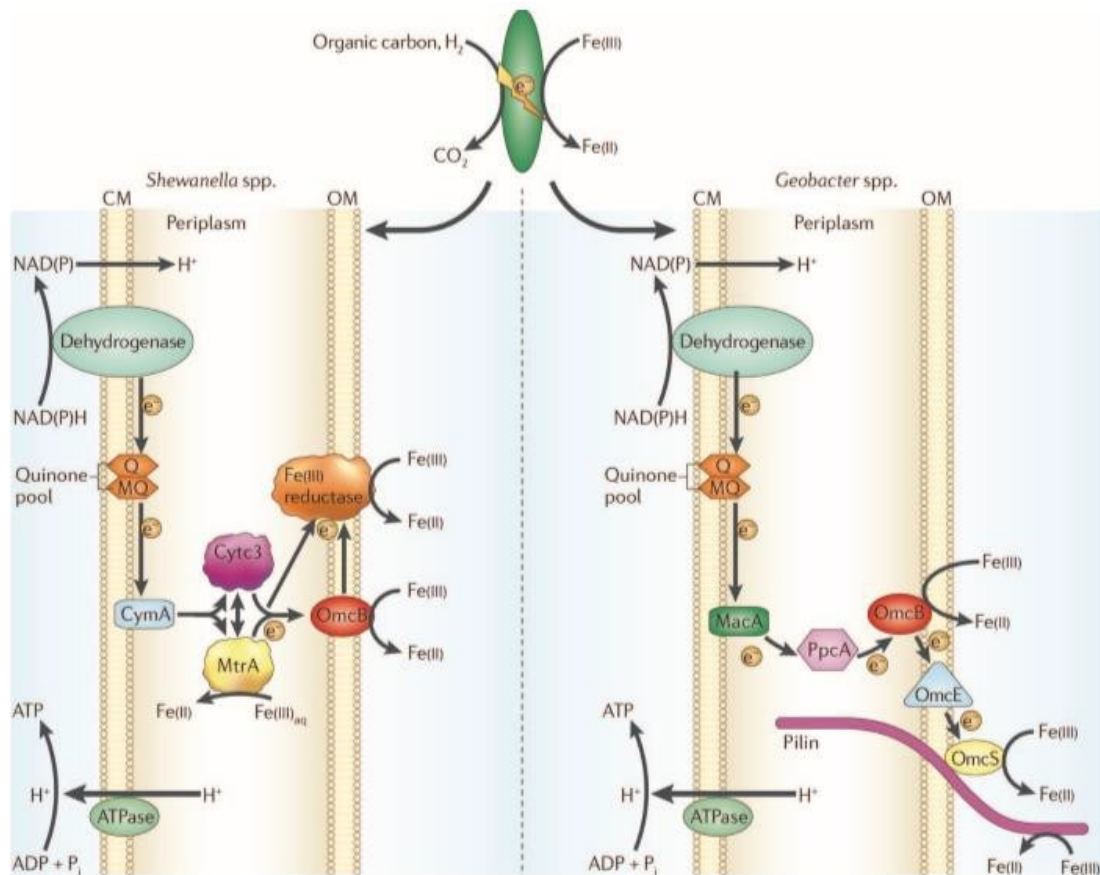


Figura 3. Modelo proposto da redução do Fe(III) em *Shewanella* spp. (esquerda) e *Geobacter* spp. (direita). Abreviações: CM, membrana citoplasmática; OM, membrana externa. Mais detalhes são descritos no texto. Fonte: Weber et al., 2006.

Em *Shewanella* spp., os elétrons são transferidos da menaquinona para CymA, um citocromo tipo c tetraheme localizado na membrana plasmática, e dele para carreadores de elétrons no periplasma. Dois citocromos tipo c, Cyt3 e MtrA, foram identificados como carreadores potenciais até o momento. Acredita-se que Cyt3 possa funcionar como um transportador de elétrons entre os carreadores de elétrons, e que MtrA seja o receptor de elétrons de CymA, transferindo-os para uma proteína de membrana externa. Também é possível que MtrA funcione de forma alternativa como redutase final para Fe(III) solúvel no periplasma. Das proteínas periplasmáticas os elétrons seguem para o citocromo OmcB (anteriormente chamado de MtrC113), localizado na membrana externa, que é capaz de reduzir Fe(III) extracelular diretamente. Curiosamente, estudos mutacionais demonstraram que OmcB não é absolutamente indispensável para a redução do Fe(III) (Weber et al., 2006).

Em *Geobacter* spp., a relação entre os carreadores citoplasmáticos e as proteínas de membrana externa ainda não foram bem estabelecidas. Entre as proteínas já identificadas estão os citocromos periplasmáticos MacA e PpcA; a primeira pode funcionar como carreador intermediário (de forma análoga à MtrA em *Shewanella* spp.), transportando elétrons para outras proteínas citoplasmáticas, e a última está envolvida no transporte de elétrons para proteínas de membrana externa como OmcB. Assim como em *Shewanella* sp., OmcB tem um papel importante na redução do Fe(III) (Weber et al., 2006).

### 1.1.2. Especiação do As

No ambiente, o As pode ocorrer em quatro estados de oxidação: +5, +3, 0 e -3, mas em meio aquático as espécies predominantes são As(V) e As(III) inorgânicos, sendo o último mais móvel e tóxico para os seres vivos que o As(V). Os fatores mais importantes para o controle da especiação química do As são o pH e o Eh. Em condições oxidantes e pH ácido (pH < 6,9), As(V) é a espécie dominante, encontrando-se frequentemente adsorvido à superfície de minerais inorgânicos como ferrihidrita e óxidos de alumínio presentes em argila; enquanto em condições mais básicas predomina o As(III). Em condições redutoras (como a encontrada em sedimentos) e pH < 9,2, espécies não carregadas de As(III) predominam (Smedley & Kinniburgh, 2001).

A especiação microbiana do As de maior relevância para seu ciclo biogeoquímico, fluxo energético e mobilização no ambiente ocorre através das reações de oxidação do As(III) e redução do As(V). Os microrganismos envolvidos são fisiologicamente diversos e ubíquos, e podem ser classificados como microrganismos resistentes à arseniato (ARMs), procariotos que realizam a redução dissimilatória de arseniato (DARPs), e oxidadores heterotróficos ou quimioautotróficos de arsenito (HAOs e CAOs, respectivamente) (Oremland & Stolz, 2003; 2005).

A redução do As(V) por ARMs é uma forma de detoxificação celular. O As(V) é um análogo molecular do fosfato e entra nas células por seus transportadores (Oremland & Stolz, 2003). No citoplasma, As(V) é reduzido pela enzima arseniato redutase, codificada por *arsC*; e então o As(III) é expelido da célula por um transportador específico (ArsB). ArsC pode ser estruturalmente diversa; a família enzimática melhor caracterizada em procariotos até o

momento é aquela do plasmídeo R773 de *Escherichia coli*, que possui um resíduo de cisteína no sítio catalítico e utiliza glutatona reduzida para a catálise (Martin et al., 2001).

DARPs são anaeróbicos e utilizam As(V) como oxidante para a respiração, em uma reação termodinamicamente favorável quando associada à oxidação de matéria orgânica, resultando na produção de As(III). Todas as linhagens bacterianas de DARPs estudadas até o momento podem utilizar doadores de elétrons orgânicos ou inorgânicos (hidrogênio e sulfeto), e não estão restritas à respiração de As(V), podendo utilizar outros aceptores de elétrons para o crescimento (Oremland & Stolz, 2003). A redução dissimilatória do As(V) é catalisada pela enzima arseniato redutase respiratória (Arr), pertencente à família DMSO de enzimas de molybdenum. Arr é um heterodímero em que a subunidade maior, codificada por *arrA*, abriga o sítio catalítico contendo molybdopterinina e um cluster [4Fe-4S], e a subunidade menor contém três a quatro clusters [4Fe-4S] (Andres & Bertin, 2016).

A oxidação heterotrófica do As(III) por HAOs requer carbono orgânico como fonte de energia e é considerada uma forma de detoxificação celular. O As(III) não tem carga em Ph <9,2 e entra na célula através de aqua-gliceroporinas, ficando retido na membrana externa onde é convertido em As(V); acredita-se que assim é menos provável que entre no citoplasma. CAOs, por outro lado, são capazes de usar As(III) como doador de elétrons na redução de nitrato ou oxigênio para a fixação do CO<sub>2</sub>. Tanto em HAOs quanto em CAOs a oxidação do As(III) é catalisada pela enzima arsenito oxidase (Aio) que, tal como Arr, é um heterodímero da família DMSO de enzimas de molybdenum, embora formem clados distintos (Oremland & Stolz, 2005). A subunidade maior é codificada pelo gene *aioA* e abriga o sítio catalítico contendo molybdopterinina e um cluster [3Fe-4S], e a subunidade menor contém um cluster tipo Rieske [2Fe-2S] (Quéméneur et al., 2008). Deve-se ressaltar que uma arsenito oxidase alternativa foi identificada na bactéria quimioautotrófica *Alkalilimnicola ehrlichii* MLHE1. Esta enzima é codificada pelo gene *arxA* e sua atividade foi detectada apenas em anaerobiose; pertence à mesma família enzimática que *aioA* e *arrA*, estando filogeneticamente mais relacionada à última (Zargar et al., 2010).

A história evolutiva destes genes, encontrados tanto em cromossomos quanto plasmídeos, carece de esclarecimentos. A Terra primitiva era abundante em As(III), e a concentração de As(V) aumentou após a alteração do estado de oxidação ambiental causado pelo surgimento da fotossíntese oxigênica. Logo, a linha de pensamento prevalente até o final da década de 1990 sugeria que a habilidade de lidar com o excesso de As(III) evoluiu primeiro (Duval et

al., 2008). Esta hipótese é sustentada por análises filogenéticas realizadas por Lebrun et al. (2003) e Duval et al. (2008) com os genes *aioA* e *arrA*, respectivamente, cujos resultados indicam que *aioA* originou-se antes da separação dos domínios Archaea/Bacteria, e *arrA*, depois, e que este evoluiu a partir de uma polisulfeto redutase, e não de *aioA*.

Por outro lado, Richey et al. (2009) demonstraram que *arrA* pode funcionar bidireccionalmente *in vitro*, e assim sugerem que existe a possibilidade de que este gene também tenha surgido antes da separação dos dois domínios procarióticos; enquanto Oremland et al. (2009) sugerem que *arsA* pode ser ancestral até mesmo à *aioA*. A análise da filogenia de *arsC* realizada por Jackson & Dugas (2003) contribui para a hipótese da redução de As(V) como metabolismo ancestral antecedente à separação dos domínios Archaea/Bacteria. Segundo os autores, a possibilidade de que a vida tenha surgido em fontes hidrotermais submarinas, locais com altas concentrações de As e Fe, torna plausível o desenvolvimento de resistência à As(V) no início da evolução dos microrganismos. Dado que As(V) é um análogo do fosfato, haveria uma forte seleção à favor da habilidade de reduzi-lo (Jackson & Dugas, 2003).

## 1.2. Meta-análise de metagenomas *shotgun*

Entre as abordagens *high-throughput* mais adequadas para investigar a associação entre os metabolismos microbianos do Fe e do As está o sequenciamento *shotgun* do DNA metagenômico. Nesta abordagem (Figura 4), o DNA é extraído de todas as células da comunidade, sem necessidade de cultivo, e então fragmentado em pequenas partes que serão sequenciadas independentemente. As sequências resultantes (*reads*) alinham-se com diferentes *loci* nos inúmeros genomas presentes na amostra; alguns desses *loci* podem corresponder à marcadores genéticos taxonômicos, como o do gene de rRNA 16S, ou a sequências codificadoras de proteínas, que podem fornecer informações sobre funções biológicas. Assim, este método permite explorar simultaneamente a composição taxonômica e as possíveis atividades metabólicas de comunidades microbianas. Embora a presença de um gene não implique que ele esteja sendo expresso durante a amostragem, metagenomas podem fornecer um indicador significativo para as funções microbianas, como sugerido por análises de metatranscriptoma e metagenômica comparativa, as quais indicam que diferenças transcricionais entre comunidades geralmente são refletidas em nível genético (Sharpton, 2014).

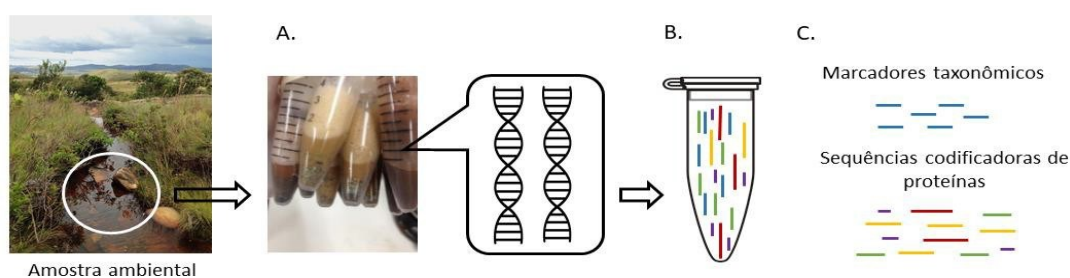


Figura 4. Fluxograma de uma análise metagenômica *shotgun*: a) após a coleta de amostra ambiental, o DNA total é extraído, b) e fragmentado em partes que serão sequenciadas de forma independente, gerando *reads*. c) As *reads* podem corresponder à marcadores taxonômicos, que fornecem informação sobre a composição de espécies da comunidade de microrganismos (microbioma), ou a sequências codificadoras de proteínas, que fornecem informações sobre os processos ecológicos da comunidade. Fonte: a autora.

O sequenciamento *shotgun* de metagenomas não é uma técnica isenta de limitações, mas estas podem ser atenuadas durante o processamento e análise das amostras, a qual é altamente dependente de ferramentas bioinformáticas. Segundo Nayfach & Pollard (2016), seus principais desafios são: i) o desconhecimento do genoma de origem das *reads* geradas e onde neste genoma estas *reads* se mapeiam; ii) a sub-representação de grande parte da diversidade procariótica nos bancos de dados de referência; e iii) os diversos vieses inseridos pelo desenho experimental e parâmetros escolhidos para a análise bioinformática, que dificultam a quantificação acurada da diversidade (Nayfach & Pollard, 2016).

As estratégias de análise para contornar os dois primeiros desafios incluem a montagem das *reads* em genes ou genomas, mapeamento das *reads* (montadas ou não) à bancos de dados de referência usando homologia de sequências, formação de *clusters* de *reads* para descobrir novas famílias proteicas ou novos *taxa*, e diversas combinações destas abordagens; e os principais resultados baseiam-se na contagem de *reads* designadas a funções ou *taxa*. O terceiro desafio pode ser abordado pela escolha de parâmetros para as estatísticas de abundância que tenham significado biológico e que levem em consideração os vieses do trabalho, aspecto especialmente relevante para estudos comparativos entre metagenomas. Felizmente, de acordo com Nayfach & Pollard (2016), a maioria dos vieses experimentais examinados até o momento não foram suficientes para interferir na identificação de processos biológicos. Além disso, Nayfach & Pollard (2016) realizaram uma análise computacional para identificar a intensidade das diferenças introduzidas pelas diversas formas de processamento dos dados, utilizando amostras de diferentes estudos do *Human Microbiome Project*, e mostraram que a maioria dos métodos tem pouco impacto na variação dos perfis funcional e

taxonômico (Nayfach & Pollard, 2016). Por fim, outro aspecto importante para avaliar a relevância estatística da composição funcional e taxonômica de estudos metagenômicos é a cobertura do sequenciamento, que é definida pela fração do metagenoma que foi de fato representada por um conjunto de dados, e que não é determinada apenas pelo tamanho deste, mas principalmente pela complexidade da comunidade amostrada.

Dessa forma, embora ainda careça de um procedimento padrão formal para a análise de conjuntos de dados, a análise comparativa de metagenomas, também chamada de meta-análise, é um campo de estudo da ecologia microbiana muito promissor. Através da escolha de ferramentas e testes estatísticos adequados às características da amostra e à pergunta do trabalho, é possível minimizar os vieses introduzidos pelo desenho experimental e análises computacionais de diferentes conjuntos de dados.

## 2. JUSTIFICATIVA E RELEVÂNCIA

A oxidação e a redução do Fe e do As devido à atividade microbiana são processos biogeoquímicos reconhecidamente relevantes, de grande impacto para a saúde pública e para a compreensão dos processos que ajudaram a moldar as características químicas e físicas da Terra. Nas últimas décadas, diversos trabalhos têm identificado os microrganismos e os mecanismos envolvidos nestas atividades, e grande atenção tem sido dada à mobilização do As por bactérias redutoras de metais em sedimentos de águas subterrâneas, devido ao efeito devastador do consumo de águas contaminadas pelo metaloide sobre a saúde humana em várias áreas do globo. Entretanto, os processos de mobilização do As em associação com o Fe ainda não foram completamente elucidados. Pouco se sabe sobre as relações estabelecidas entre estes dois metabolismos e os genes que os determinam, bem como a abrangência desta importante interação em diferentes tipos de ambientes. Dessa maneira, o presente trabalho propõe elucidar estas questões. Um melhor entendimento da relação entre estes dois metabolismos e os microrganismos implicados poderá auxiliar na elaboração de estratégias mais efetivas para mitigar a contaminação de corpos d'água ricos em Fe e As, além de contribuir para a elucidação das relações microbianas na Terra primitiva.

### 3. OBJETIVOS

#### 3.1. Objetivo Geral

Identificar as possíveis relações entre os genes envolvidos na biotransformação do As e Fe de diferentes ambientes com altas concentrações de As, Fe, ou ambos, através da caracterização funcional de conjuntos de dados metagenômicos disponíveis em bancos de dados públicos.

#### 3.2. Objetivos específicos

1. Realizar a montagem e anotação de genes dos conjuntos de dados obtidos do *Sequence Read Archive* (SRA) e do MG-RAST;
2. Minar genes envolvidos na oxidação e redução do As e do Fe na anotação funcional dos conjuntos de dados;
3. Identificar os genes envolvidos na oxidação e redução do As e do Fe nos diferentes ambientes, bem como suas frequências relativas;
4. Avaliar a relação entre estes genes/metabolismos implicados nos diferentes tipos de ambiente.

## 4. MATERIAIS E MÉTODOS

### 4.1. Busca e seleção de conjunto de dados metagenômicos

A seleção dos conjuntos de dados utilizados neste estudo foi feita buscando por artigos na plataforma Google Acadêmico, e também por sequências públicas depositadas no MG-RAST e no Sequence Read Archive (SRA). No Google Acadêmico, foram buscados artigos combinando as palavras e expressões-chave “arsenic”, “arsenate”, “arsenite”, “iron”, “environmental metagenome/omics”, “comparative metagenome/omics”, “whole genome shotgun” e “WGS”. No MG-RAST foram usadas as palavras-chave “iron”, “arsenic”, “arsenate”, “arsenite”, “As(V)”, “As(III)” e “Fe” em que “sequencing type = shotgun metagenomics”. Por fim, no SRA foi feita uma busca por BioProjects com as mesmas palavras-chave usadas no MG-RAST e filtrando por resultados em que “scope = environmental”.

Após a busca, foram selecionados conjuntos de dados metagenômicos obtidos de ambientes ricos em As, Fe ou ambos, seja por característica natural do ambiente ou por impacto antrópico. Uma área de estudo foi considerada rica em As quando a sua concentração excedia 1,5 ppm, considerada dose letal para humanos (WHO, 2011), e rica em Fe quando sua concentração excedia o limite de 200 ppm recomendados para consumo humano, como estabelecido pela OMS (WHO, 2003). Para compor o grupo controle da análise foram escolhidos conjuntos de dados obtidos de ambientes onde não foram descritos níveis elevados de As e Fe, e que representavam ecossistemas similares ao dos conjuntos de dados enriquecidos com os metais. Foram feitas buscas no Google Acadêmico e SRA, combinando as palavras e expressões-chave “environmental metagenome/omics”, “comparative metagenome/omics”, “whole genome shotgun”, “WGS” e “pristine”.

Outro critério importante para a inclusão de um determinado estudo na meta-análise foi a disponibilidade de metadados no(s) artigo(s) publicado(s) ou na anotação do banco de dados. Estes dados incluíam características físico-químicas do ambiente (ex.: pH, temperatura e concentração de oxigênio) e desenho experimental (plataforma de sequenciamento, sequenciamento *paired-end* ou *single-end*).

As diferentes combinações de palavras-chave, principalmente no Google Acadêmico, por vezes retornaram resultados que não atendiam ao critério do tipo de sequenciamento *shotgun*. Observou-se que a maioria dos estudos em ambientes com as características de interesse tinha como objetivo principal investigar composição taxonômica das comunidades, e, portanto, realizaram somente amplificação do gene de rRNA 16S. Assim, estabeleceu-se como critério de exclusão a realização de amplificação de genes específicos nos metagenomas, ou que realizaram seleção de fragmentos por tamanho. Ao final da seleção foram mantidos 20 estudos, totalizando 40 conjuntos de dados cujas sequências de DNA foram obtidas de forma aleatória, em detrimento de amplicons (Quadro 1).

Quadro 1: Conjuntos de dados selecionados para a meta-análise.

Base de dados	Projeto/abreviação	Origem das amostras	Categoria ambiental	Plataforma de sequenciamento/configuração das reads	Número de acesso	Referência
Google Acadêmico	Metagenome on soil background with high concentration of arsenic and antimony (ACS)	Solo contaminado por As e Sb	Rico em As	Illumina TruSeq 2000/ Paired-end	Bioproject PRJNA239941 Run SRR1185960	Luo et al., 2014.
Google Acadêmico	Community genomics of hot-springs at Manikaran (HHS)	Tapete microbiano de fontes termais	Rico em As-Fe	Illumina Genome Analyzer II/ Paired-end	Bioproject PRJEB4614 Runs ERR342783 ERR342782	Sangwan et al., 2015.
Google Acadêmico	Socompa Lake Stromatolite Shotgun Metagenomic Sequencing (SLS)	Comunidades microbianas de estromatólitos de altitude	Rico em As	Illumina Genome Analyzer IIx/ Paired-end	Bioproject PRJNA317551 Run SRR3341855	Kurth et al., 2017.
Google Acadêmico	Bangladesh Surface and Ground Water Metagenomes (GWB)	Água subterrânea	Rico em As	Illumina HiSeq 2000/ Paired-end	Bioproject PRJNA260804 Runs SRR1573578 SRR1573565 SRR1564103 SRR1563167	Layton et al. 2014.
Google Acadêmico	CFW	Sedimento de água doce	Rico em Fe	SOLiD v4/Contigs montados	MG-RAST ID 4520540.3	Reis et al., 2016.
Google Acadêmico	MFW	Sedimento de água doce	Rico em As-Fe	SOLiD v4/Contigs montados	MG-RAST ID 4519449.3	Costa et al., 2015.

SRA	Iron sulfur acid spring bacterial and archeal communities from Banff, Canada, to study Microbial Dark Matter (Phase II) - Paint Pots PPA 5.5 (PP)	Fonte ácida rica em Fe e S	Rico em Fe	Illumina HiSeq/2500/Paired-end	Bioproject PRJNA375325 Run SRR5580868	Grasby et al., 2013.
SRA	Soudan Iron Mine Metagenome (SM)	Água de mina de Fe salina terrestre	Rico em Fe	Illumina HiSeq 1000/Paired-end	Bioproject PRJNA340294 Runs SRR4072256 SRR4072255 SRR4072253	Edwards et al., 2006.
SRA	Metagenomic analysis of an Arctic soil profile (APS)	Solo de lago de degelo	Rico em Fe	454 GS FLX Titanium/Single-end	Bioproject PRJNA196185 Runs SRR825190 SRR825189 SRR825188 SRR825158	Lipson et al., 2010.
SRA	Metagenome from Norris Geyser Basin-Beowulf Spring (NGB)	Tapete microbiano de fonte termal	Rico em As	Sanger/contigs montados	Bioproject PRJNA42237 GenBank ADKI00000000	Inskeep et al., 2010.
SRA	Metagenome from Joseph's Coat Hot Spring-Scorodite Spring (JCH)	Tapete microbiano de fonte termal	Rico em As-Fe	Sanger/contigs montados	Bioproject PRJNA42235 GenBank ADKH00000000	Inskeep et al., 2010.

SRA	Yellowstone National Park hot spring microbial community Calcite Springs (CS)	Tapete microbiano de fonte termal	Rico em Fe	Sanger/contigs montados	Bioproject PRJNA42239 Genbank ADKJ00000000	Inskeep et al., 2010.
SRA	Carnoulès Metagenome Project - Environmental Samples (CM)	Sedimento de mina ácida	Rico em As-Fe	454 GS-FLX/Scaffolds montados	Bioproject PRJNA38045 GenBank CABS00000000.1	Bertin et al., 2011
SRA	Time-Series Metagenomics of Lake Lanier, Atlanta, GA (LL)	Água doce	Referência	Illumina Genome Analyzer II/Paired-end	Bioproject PRJNA51219 Runs SRR096389 SRR096388 SRR096387 SRR096386	Oh et al., 2011.
SRA	Forest soil Metagenome (NSP)	Solo de floresta	Referência	Illumina Genome Analyzer/Single-end	Bioproject PRJNA169429 Runs SRR516942 SRR516943 SRR516944 SRR516945 SRR516946 SRR516947	Uroz et al., 2013.

SRA	Oak Ridge Integrated Field Research Challenge (OR-IFRC) Metagenomes (GWOR)	Água subterrânea	Referência	Illumina HiSeq 2000 (SRR*60)/ Single-end Illumina Genome Analyzer Iix (SRR*52 e SRR*55)/ Single-end	Bioproject PRJNA258175 Runs SRR1573960 SRR1573952 SRR1573955	Hemme et al., 2016.
SRA	Arctic permafrost Metagenome (SP)	Permafrost	Referência	Illumina HiSeq 2000/ Paired-end	Bioproject PRJNA266334 Run SRR1653578 SRR1653579	Krivushin et al. 2015.
SRA	Metagenomic study of Alchichica microbialites (AM)	Microbiólitos	Referência	Illumina HiSeq 2000/ Paired-end	Bioproject PRJNA315555 Run SRR3311145	Filinov et al., 2015. Saghai et al., 2016.
SRA	Yellowstone National Park hot spring microbial community Crater Hills (CH)	Tapete microbiano de fonte termal	Referência	Sanger/Contigs montados	Bioproject PRJNA42245 GenBank ADKL00000000.1	Inskeep et al., 2010.
SRA	Yellowstone National Park hot spring microbial Community Mammoth Hot Springs (MHS)	Tapete microbiano de fonte termal	Referência	Sanger/Contigs montados	Bioproject PRJNA42241 GenBank ADKK00000000.1	Inskeep et al., 2010.

## 4.2. Análise bioinformática

A análise bioinformática dos conjuntos de dados compreendeu as seguintes etapas, nesta ordem: controle de qualidade das sequências, montagem de genomas, predição e anotação gênica, e busca pelos genes de interesse no resultado da anotação (Sharpton, 2014, Neyfach & Pollard, 2016). Embora os conjuntos escolhidos já tenham sido processados e analisados em seus estudos de origem, optou-se por trabalhar, quando possível, com sequências brutas neste estudo. Esta estratégia permite o processamento uniforme destas *reads*, assim evitando o viés introduzido por diferentes abordagens bioinformáticas (Neyfach & Pollard, 2016), além de viabilizar a escolha dos melhores parâmetros para responder às perguntas propostas neste estudo.

Os conjuntos de dados disponibilizados como sequências brutas (SRR\* e ERR\*) passaram por todas as etapas da análise; enquanto aqueles disponibilizados como *contigs* (NGB, MHS, JCH, CS, CH, MFW, CFW, CM) não passaram pelo controle de qualidade e montagem de genomas pois foram depositados já montados. Estes entraram na análise a partir da etapa de anotação dos genomas.

Os dados obtidos foram gerados por diferentes plataformas de sequenciamento (Quadro 1), que apresentam diferentes padrões e taxas de erros. Entretanto, estas discrepâncias não representam um viés significativo para a análise, tendo um efeito mínimo na caracterização da composição funcional e taxonômica (Nayfach & Pollard, 2016). O tamanho das *reads* geradas também difere entre as plataformas e pode ser uma fonte de viés, uma vez que sequências muito curtas dificultam a detecção de homologia, principalmente quando sua origem é um microrganismo pouco representado nos bancos de dados. Uma estratégia para minimizar este problema é realizar a montagem de *contigs* e *scaffolds*, abordagem adotada na *pipeline* deste estudo. Buscou-se realizar o processamento de forma homogênea, mas quando necessário os parâmetros foram alterados para melhor se adequar as características dos conjuntos de dados. Nos tópicos seguintes estes comandos serão apresentados em detalhes. As análises foram executadas no Sagarana HPC Cluster, servidor do Instituto de Ciências Biológicas da Universidade Federal de Minas Gerais.

#### 4.2.1. Controle de qualidade das sequências

Os conjuntos de dados que entraram nesta etapa da análise foram gerados pelas plataformas Illumina e Roche 454 (Quadro 1). Após o *download*, a conversão dos arquivos SRA para o formato FASTQ foi feita com a ferramenta SRA Toolkit, usando o seguinte comando em *looping*:

```
for i in /home/malusuhadolnik/projetos/Sleipnir/data/${i}*.sra`; do fastq-dump $i --split-3 -O
~/projetos/Sleipnir/data/fastq/; done
```

Neste *loop*, estabelece-se que para todos os arquivos (designados pela variável genérica “i”) que terminam na extensão “.sra” presentes no diretório “data” do projeto *Sleipnir*, deve-se aplicar o comando `fastq-dump`, que converte a extensão SRA a FASTQ. O comando deve ser expandido a todos os arquivos (argumento `$i`), os arquivos *forward* e *reverse* devem ser separados, quando aplicável (`--split-3`), e “fastq” será o diretório de destino (`-O`) dos arquivos formados (chamados de *output*).

A qualidade das sequências brutas foi avaliada com o programa FastQC (Andrews, 2010):

```
fastqc /home/malusuhadolnik/projetos/Sleipnir/data/SRR* ou ERR*
```

Fastqc é o comando de execução, e o único argumento é o caminho dos conjuntos de dados obtidos do SRA, cujos nomes começam com os códigos SRR ou ERR. O *output* deste comando é um arquivo HTML que sumariza as características das *reads* do conjunto de dados, como o total de sequências, seu tamanho, conteúdo GC, qualidade por base, conteúdo de adaptadores, entre outros.

A remoção dos adaptadores e trimagem de sequências de baixa qualidade foram feitas com o programa Seqclean (Zhbannikov, 2017). Para sequências geradas nas plataformas Illumina, foi usado o comando:

```
seqclean1.10 -1 SRR*/ERR*_R1.fastq -2 SRR*/ERR*_R2.fastq -o SRR*/ERR* -qual -
minlen 30
```

Seqclean1.10 é o comando de execução, -1 e -2 designam os arquivos FASTQ *forward* e *reverse*, respectivamente; -o o prefixo do arquivo *output*; -qual ativa a trimagem de qualidade (a configuração default é 0.01 para o máximo de erro médio e para o máximo de erro permitido nas extremidades) e -minlen determina o tamanho mínimo das *reads*, que foi definido em 30 pb.

E para sequências geradas na plataforma Roche 454 Titanium:

```
seqclean1.10 adaptadores454.fasta SRR*/ERR*.fastq -o SRR*/ERR* -U -adp -qual - minlen 30
```

Como estas haviam sido convertidas ao formato FASTQ, elas foram processadas usando comando para sequências *single end* da Illumina. Nela, o argumento -U designa arquivo fastq single-end, e -adp um arquivo FASTA contendo as sequências dos adaptadores 1 (CCATCTCATCCCTGCGTGTCTCCGAC) e 2 (CCTATCCCCTGTGTGCCTTGGCAGTC) (Mann et al., 2014) da plataforma Roche 454 Titanium, que o programa deve utilizar como referência para a trimagem. Os demais parâmetros foram deixados na configuração *default* do programa. As *reads* duplicadas não foram removidas, porque estas podem ser provenientes de organismos abundantes na população, ao invés de erros de sequenciamento (Nayfach & Pollard, 2016).

Após o fim da execução do Seqclean, os conjuntos de dados limpos foram novamente submetidos ao FASTQC para verificar a eficácia da trimagem.

#### 4.2.2. Montagem *de novo* de genomas

A montagem *de novo* de *contigs* de cada um dos conjuntos de dados foi feita com o programa IDBA-UD (Peng et al., 2012). Este é um programa para montagem de genomas obtidos de metagenomas e de sequenciamentos *single-cell*, que se baseia na estratégia de grafos de Bruijn. Esta é uma forma de representar a estrutura dos dados pautada na sobreposição de porções das *reads*, chamadas de *k-mers*, em que *k* é um número fixo pré-definido, geralmente entre 30-50 nucleotídeos. As *k-mers* são organizadas na forma de um grafo onde os nós/arestas correspondem à (*k*-1) prefixos e sufixos das *k-mers*, conectados por vértices que representam as *k-mers*. Cada *k-mer* é designada à um único nó, de forma

que se alguma *k-mer* se repete não há a criação de um novo nó, e sim a adição de um novo vértice àquele existente. A montagem é feita pela identificação de um caminho Euleriano: aquele que passa por cada vértice uma única vez (Langmead, 2014; Namiki et al., 2012).

Entretanto, esta abordagem pode apresentar alguns problemas. O primeiro deles é a introdução de *k-mers* incorretas devido à erros de sequenciamento; além disso, existe a possibilidade da formação de lacunas, que ocorrem quando *k* tem um valor alto, levando à ausência de algumas *k-mers* principalmente em regiões com baixa cobertura de sequenciamento. Por fim, devido a regiões repetidas ou *reads* errôneas, muitos vértices são adicionados ao grafo, especialmente quando *k* tem um valor baixo (Peng et al., 2012).

O programa escolhido busca contornar estes problemas. Para evitar a formação de *contigs* incorretos, um limite relativo à profundidade do sequenciamento é estabelecido para as *k-mers*, baseado na ideia de que *contigs* muito curtos e com cobertura muito menor que os *contigs* adjacentes tendem a ser erros. As lacunas e os múltiplos vértices são diminuídos ao se estabelecer não apenas um único valor para as *k-mers*, mas uma iteração entre um *k* de valor mínimo à um valor máximo. A cada iteração, os *contigs* construídos atuam como *reads* para a próxima, de forma que as *k-mers* de uma iteração são carregadas para a próxima, na qual a *k-mer* anterior poderia estar ausente. Uma etapa adicional de correção de erros é feita no fim de cada iteração, alinhando as *reads* incorretas aos *contigs* de confiança (com muitas *reads* de suporte).

Para os conjuntos de dados com *reads paired-end*, foram usados os seguintes comandos em looping:

```
CPU=30;
```

```
for FASTQ in /home/malusuhadolnik/projetos/Sleipnir/data/clean/*PE1.fastq do
```

```
fq2fa --merge $FASTQ ${FASTQ%PE1.fastq}PE2.fastq ${FASTQ:50:10}.FASTA
```

```
idba_ud -r ${FASTQ:50:10}.FASTA -o ${FASTQ:50:10} --num_threads $CPU --mink 19
```

```
--maxk 101
```

done

Este *loop* é interpretado da seguinte forma: para todos os arquivos FASTQ *forward* (\*PE1.fastq) presentes no diretório “clean”, aplicar:

- 1) o comando fq2fa da ferramenta fq2fa (Peng et al., 2012), estendendo a todos os arquivos FASTQ (\$FASTQ) o argumento `--merge`, que une todos os arquivos *forward* e *reverse* (\${FASTQ%PE1.fastq}PE2.fastq) de um mesmo SRR/ERR em um único arquivo FASTA (\${FASTQ:50:10}.FASTA), que será o *input* do comando idba\_ud,
- 2) o comando idba\_ud, que executa a montagem dos *contigs*, estendido a todos os arquivos FASTA (-r \${FASTQ:50:10}.FASTA). Foram estabelecidas 30 CPUs (--num\_threads \$CPU) para a execução da montagem, um valor de k mínimo de 19 pb (--mink 19) e um valor de k máximo de 101 (--maxk 101). Os valores mínimo e máximo foram escolhidos tendo como referência o tamanho mínimo (30 pb) e máximo (98 – 120) das *reads* analisadas, de forma que  $k_{min} < 30$  e  $k_{max} < 90$ , 100 ou 120, dependendo das características das *reads* de cada conjunto de dados. Os demais argumentos foram deixados na configuração *default*, incluindo o incremento de k-mer a cada iteração, em que  $k=20$  bp.

Para os conjuntos de dados com *reads single-end*, foram usados os seguintes comandos em *looping*:

```
for FASTQ in /home/malusuhadolnik/projetos/Sleipnir/data/454/*_cleaned_SE.fastq do
fq2fa $FASTQ ${FASTQ:48:12}.FASTA
```

```
idba_ud -l ${FASTQ:48:12}.FASTA -o ${FASTQ:48:12} --num_threads $CPU --mink 19
--maxk 101 done
```

Este *loop* é interpretado de forma análoga ao anterior. Neste foi estabelecido o caminho para os conjuntos de dados *single-end* (\*SE.fastq), e no comando fq2fa não foi usado o argumento `--merge`. O *output* da montagem consiste em um novo diretório criado automaticamente para cada um dos conjuntos de dados, homônimo a estes, contendo um arquivo de alinhamento, grafo e *contig* para cada valor de k-mer, e um arquivo FASTA final com os *scaffolds* e/ou *contigs* formados. Após a montagem, a qualidade desta foi avaliada mapeando os *contigs* ou *scaffolds* montados de volta aos FASTQ originais com o programa

Bowtie2 (Langmead & Salzberg, 2011). Este programa utiliza índices genômicos para acelerar o alinhamento das *reads* às sequências de referência. Sucintamente, índices genômicos são uma forma de reorganizar os caracteres das *reads* para que se tornem mais sucintas, mas sem perder a sensibilidade do alinhamento. Para os conjuntos de dados de *reads paired-end* foram executados os seguintes comandos:

```
bowtie2-build scaffold.fa indexbt2
```

```
bowtie2 -p 7 -x indexbt2 -1 ../../data/clean/$i*cleaned_PE1.fastq -
2
../../data/clean/$i*cleaned_PE2.fastq -S $i.sam samtools view -Sb $i.sam > $i.bam
```

O primeiro comando constrói o index 2 (indexbt2) a partir do arquivo FASTA dos *scaffolds* formados (scaffold.fa), o segundo comando realiza o mapeamento do index 2 de volta aos arquivos FASTQ *forward* e *reverse* das *reads* desmontadas (caminhos indicados pelos argumentos -1 e -2, respectivamente, em que *i\** é uma variável que indica o código do conjunto de dados sendo mapeado), e o *output* (-S) é um arquivo SAM com o nome do conjunto de dados sendo mapeado. O terceiro comando utiliza o aplicativo SamTools (Li et al., 2009) para converter o output do Bowtie2 ao formato BAM, mais facilmente interpretável. Para os conjuntos de dados *single-end* houve a substituição dos argumentos -1 e -2 pelo argumento -U, indicando o caminho do arquivo FASTQ *single-end* do conjunto de dados sendo mapeado.

Os dois conjuntos de dados obtidos do MG-RAST (MFW e CFW) foram gerados pela plataforma SOLiD e estavam disponíveis na forma de *contigs*, montados no programa Metavelvet (Namiki et al., 2012), que assim como o IDBA-UD baseia-se na estratégia da construção de grafos de Brujin para a montagem de genomas a partir de *reads* de metagenomas. O algoritmo do MetaVelvet consiste em quatro etapas. A primeira delas é a montagem de um grafo de Brujin a partir das *reads*; a segunda é a detecção de picos múltiplos na frequência de distribuição das *k-mers*; cada pico do histograma corresponde ao genoma de uma espécie, e cada nó é classificado em um pico. A terceira etapa é a decomposição do grafo principal em sub-grafos, cada um pertencente ao genoma de um organismo, através da identificação de nós que pertencem à um mesmo pico. Nós compartilhados entre dois sub-grafos (quiméricos) são separados; para distinguir nós quiméricos de repetições considera-se a

cobertura da *k-mer* (número de sequências que se sobrepõem naquela posição). A quarta etapa é a montagem de *contigs* (Namiki et al., 2012). Nos estudos originais dos conjuntos de dados MFW (Costa et al., 2015) e CFW (Reis et al., 2016), os *contigs* foram montados usando a configuração padrão recomendada pelos autores ( $kmer = 27$ ).

Os conjuntos de dados NGB, MHS, JCH, CS e CH foram obtidos a partir de biblioteca de clones, gerando fragmentos pareados longos de aproximadamente 800pb, e também foram disponibilizados na forma de *contigs* e *scaffolds*. A montagem dos genomas foi feita com o programa Celera Assembler v.4.0. As amostras analisadas foram obtidas de fontes hidrotermais ( $T = > 65^{\circ}\text{C}$ ) subóxicas/anóxicas ricas em metais e sulfetos. Tais ambientes extremos possuem pouca diversidade taxonômica (Inskeep et al., 2010), de forma que uma boa cobertura de sequenciamento pode ser obtida sem a necessidade de um sequenciamento mais profundo (Rodriguez-R & Konstantinidis, 2014).

#### 4.2.3. Predição gênica e anotação funcional

A predição de genes e anotação funcional de todos os conjuntos de dados após a montagem, incluindo aqueles que já foram depositados na forma de *contigs*, foi feita com o programa Prokka (Seemann, 2014). Este programa utiliza o algoritmo do software Prodigal (Hyatt et al., 2010) para realizar a predição gênica, e a anotação funcional é feita de forma hierárquica.

O programa Prodigal é aplicável à dados metagenômicos, possui um robusto embasamento biológico em propriedades de genomas procarióticos, como o viés de uso de códons de iniciação e motivos de uso de sítios de ligação do ribossomo, e uma estatística confiável para minimizar resultados falso-positivos. Após a identificação das sequências codificantes (CDS), o Prokka utiliza as sequências traduzidas para comparação contra os bancos de dados, começando opcionalmente por um pequeno conjunto de proteínas anotadas, fornecidas pelo usuário; em seguida utiliza proteínas bacterianas contidas no UniProt e RefSeq; e por fim perfis de Modelos Ocultos de Markov (HMMs) de uma série de bancos de dados, como o Pfam e TIGRFAMs.

UniProt é um extenso banco de dados de sequências proteicas anotadas, que atualmente reúne mais de 60 milhões de sequências curadas (The UniProt Consortium, 2016); e o Reference Sequence (RefSeq) é uma coleção de sequências não redundantes e curadas, que representam

o conhecimento atual dos genes conhecidos (Pruitt et al., 2017). Tanto os bancos de dados Pfam e TIGRFAM contém alinhamentos múltiplos de sequências e HMMs, que são modelos de sequência probabilísticos. A função de um modelo de sequência é designar rótulos para cada unidade (letras, palavras, etc.) de uma sequência, assim mapeando uma sequência de observações em uma sequência de rótulos. Dado uma sequência de unidades, HMMs calculam a probabilidade de distribuição sobre possíveis sequências de rótulos, ao fim escolhendo o mais adequado (Jurafsky & Martin, 2017). As versões mais recentes do Pfam e TIGRFAM reúnem 16.712 e 4.488 famílias proteicas, respectivamente. O seguinte comando foi utilizado:

```
for i in SRR*
do
cd $i;

prokka --metagenome contig.fa cd ..

done
```

Para todas as pastas *i* contendo as montagens, aqui exemplificadas de forma genérica por *SRR\**, entrar e executar o programa *prokka*, com o argumento `--metagenome` para indicar o tipo de dado sendo anotado, e utilizando o arquivo FASTA dos *contigs* montados. Após o término, sair da pasta. Quando apropriado, o argumento “*contig.fa*” foi trocado por “*scaffold.fa*”. Entre os *outputs* do programa estão um arquivo TSV com todas características: a identificação do locus, o tipo do transcrito (CDS, rRNA, tRNA, tmRNA, misc\_RNA), o nome do gene que o codifica, EC-number e seu produto; um arquivo TXT com a estatística resumida da análise; e arquivos FASTA de extensão FNA e FFA contendo as sequências de nucleotídeos e aminoácidos do transcrito, respectivamente.

Genes reconhecidamente implicados nas biotransformações do Fe e As foram selecionados *à priori* através de uma extensa revisão bibliográfica (Quadro 2). Estes genes e palavras e expressões-chave relacionadas (Anexo 1) foram minados usando o comando *grep*:

```
grep -i wf grep_lista.txt
home/malusuhadolnik/projetos/Sleipnir/work/assembly/$i*/PROKKA*/PROKKA*.tsv >
$i_mining.txt
```

Este comando busca por padrões (expressões) em um arquivo *input*, e o *output* é a listagem de todas as linhas em que o padrão aparece no *input*. O argumento *-i* faz a busca ignorar distinção entre letras maiúsculas ou minúsculas; *-f* determina que o padrão deve ser obtido do arquivo fornecido, e *-w* estabelece que a correspondência deve ser feita só com palavras inteiras; “*grep\_lista.txt*” é o arquivo que contém a lista de padrões que devem ser procurados nos arquivos TSV da anotação, e o output será registrado em um arquivo novo que termina em “*\_mining.txt*”. *\$i* é uma variável genérica que indica o nome do conjunto de dados.

O gene *arrA*, frequentemente encontrado em procariotos anaeróbicos obrigatórios e facultativos, não foi identificado em nenhum dos conjuntos de dados. Por isso, foi feita uma busca através do alinhamento dos *contigs* e *scaffolds* contra uma lista de sequências FASTA curadas da proteína, adaptada de Suhadolnik et al. (2017). Esta lista inclui sequências provenientes de bactérias cultiváveis e não cultiváveis. Foi usada a ferramenta de alinhamento Blat (Kent, 2002), com a seguinte linha de comando:

```
blat -noHead -maxIntron=0 -minMatch=2 -t=dnax -q=prot $i $datadir$refarrA
$resultdir${name%_}_arrA.psl
```

O argumento *-noHead* cria um arquivo separado por tabulação, *-maxIntron=0* estabelece o tamanho de introns =0 (o gene em questão é procariótico), *-minMatch=2* define como 2 blocos o tamanho mínimo da correspondência para o alinhamento (o tamanho *default* de cada bloco é 5 para proteína), *-t=dnax* estabelece que o tipo de banco de dados utilizado são sequências de DNA traduzidas para proteína, *-q* estabelece que a *query* são sequências de aminoácidos, *\$i* é o caminho para arquivos FASTA utilizados como banco de dados (*contigs* e *scaffolds* de cada um dos conjuntos de dados da meta-análise), *\$datadir\$refarrA* é o caminho para o arquivo contendo as sequências curadas de *arrA*, e por fim *\$resultdir\${name%\_}\_arrA.psl* é o nome dos arquivos *output*.

As sequências FASTA que tiveram identidade com o gene *arrA* foram extraídas com a ferramenta Bedtools (Quinlan, 2010) e então alinhadas contra banco de dados usando a ferramenta BLAST (Altschul et al., 1990), para confirmação de sua identidade. Foram consideradas sequências legítimas aquelas que tiveram cobertura e identidade >60% com as proteínas “dissimilatory/respiratory arsenate reductase” e *e-value* máximo de  $10^{-5}$  (Saghaï et al., 2016) (Anexo 1).

Quadro 2: Relação de genes envolvidos na oxidação e redução de Fe e As em procariotos

Gene	Operon/ Cluster	Proteína	Função	Via	Organismo modelo	Localização	Referência
<i>aioA</i>	<i>aio</i>	Subunidade catalítica da arsenito oxidase	Oxidação de As(III) em aerobiose	Oxidação do As(III)	<i>Alcaligenes faecalis</i> <i>Rhizobium NT-26</i> <i>Hydrogenophaga NT-14</i>	Periplasma	Kruger et al., 2013
<i>arxA</i>	<i>arx</i>	Arsenito oxidase	Oxidação de As(III) em anaerobiose	Oxidação do As(III)	<i>Alkalilimnicola ehrlichii</i> MLHE-1,	Periplasma	Zargar et al., 2007; Kruger et al., 2013
<i>arsC</i>	<i>ars</i>	Subunidade catalítica da arseniato redutase	Redução de As(V)	Redução do As(V)	<i>Escherichia coli</i> plasmid R773	Citoplasma	Martin et al., 2001; Kruger et al., 2013
<i>arrA</i>	<i>arr</i>	Subunidade catalítica da arseniato redutase dissimilatória	Redução de As(V) em anaerobiose	Redução do As(V)	<i>Shewanella sp.</i> ANA-3	Periplasma	Saltikov & Newman, 2003; Kruger et al., 2013
<i>cymA</i>	NA	CymA, citocromo tetraheme tipo c	Transfere elétrons do pool de meniquinona para CymA, direta ou indiretamente.	Redução do Fe	<i>Shewanella spp.</i>	Membrana citoplasmática	Weber et al., 2006; Myers & Myers, 1997
<i>mtrA</i>	cluster mtr/omc	MtrA, citocromo decaheme tipo c	Possivelmente transfere elétrons de CymA para uma proteína de membrana externa; ou, alternativamente, atuaria como redutase final para Fe(III) solúvel no citoplasma.	Redução do Fe(III)	<i>Shewanella spp.</i>	Periplasma	Weber et al., 2006

<i>cytC3</i>	NA	CytC3, citocromo tetraheme tipo c	Carreador de elétrons em potencial para redução do Fe(III). Possivelmente funciona como transportador de elétrons entre carreadores.	Redução do Fe(III)	<i>Shewanella spp.</i>	Periplasma	Weber et al., 2006; Tsapin et al., 2001
<i>omcB</i> ( <i>mtrC1</i> 13)	cluster mtr/omc	OmcB, citocromo decaheme	Aceptor de elétrons de proteínas periplasmáticas (pode receber elétrons de MtrA), com potencial de reduzir Fe(III) extracelular diretamente em <i>Shewanella</i> . Papel importante na redução do Fe(III) em <i>Geobacter</i> .	Redução do Fe(III)	<i>Shewanella spp.</i> <i>Geobacter spp.</i>	Parcialmente exposto na superfície celular	Weber et al., 2006
<i>macA</i>	NA	MacA, citocromo	Carreador intermediário de elétrons, similar à MtrA	Redução do Fe(III)	<i>Geobacter spp.</i>	Periplasma	Weber et al., 2006; Butler et al., 2004
<i>ppcA</i>	NA	PpcA, citocromo triheme tipo c	Transporta elétrons para proteínas na membrana externa para redução do Fe(III).	Redução do Fe(III)	<i>Geobacter spp.</i>	Periplasma	Weber et al., 2006; Lloyd et al., 2003
<i>omcE</i>	NA	OmcE, citocromo	Carreador de elétrons de OmcB a OmcS para redução do Fe(III)	Redução do Fe(III)	<i>Geobacter spp.</i>	Parcialmente exposto na superfície celular	Weber et al., 2006
<i>omcS</i>	<i>omcS-omcT'</i>	OmcS, citocromo	Recebe elétrons de OmcE e reduz Fe(III)	Redução do Fe(III)	<i>Geobacter spp.</i>	Parcialmente exposto na superfície celular	Weber et al., 2006; Mehta et al., 2005
<i>mtrD</i>	cluster mtr/omc	MtrD, citocromo decaheme	Interage com citocromos e é necessário para a redução eficaz do Fe(III)	Redução do Fe(III)	<i>Shewanella spp.</i>	Periplasma	Reyes et al., 2010

<i>mtrF</i>	<i>cluster mtr/omc</i>	MtrF, citocromo decaheme	Interage com citocromos e é necessário para a redução eficaz do Fe(III)	Redução do Fe(III)	<i>Shewanella spp.</i>	Membrana externa	Reyes et al., 2010
<i>omcA</i>	<i>cluster mtr/omc</i>	OmcA, citocromo decaheme	Interage com citocromos e é necessário para a redução eficaz do Fe(III)	Redução do Fe(III)	<i>Shewanella spp.</i>	Membrana externa	Reyes et al., 2010
<i>mtrB</i>	<i>cluster mtr/omc</i>	MtrB, proteína de membrana externa	Facilita a troca de elétrons entre membrana externa e OmcB	Redução do Fe(III)	<i>Shewanella spp.</i>	Membrana externa	Reyes et al., 2010
<i>mtrE</i>	<i>cluster mtr/omc</i>	MtrE, proteína de membrana externa	Interage com citocromos e é necessário para a redução eficaz do Fe(III)	Redução do Fe(III)	<i>Shewanella spp.</i>	Membrana externa	Reyes et al., 2010
<i>cyc2</i>	<i>rus</i>	Cyc2 citocromo c	Oxidação do FeII	Oxidação do Fe(II)	<i>Acidithiobacillus ferrooxidans</i>	Membrana externa	Bonnefoy & Holmes, 2012
<i>cyt572</i>	A verificar	Cyc572 citocromo c	Oxida Fe(II)	Oxidação do Fe(II)	<i>Leptospirillum spp.</i>	Membrana externa	Bonnefoy & Holmes, 2012
<i>rusA</i>	<i>rus</i>	Rusticianina A	Recebe elétrons de Cyc2	Oxidação do Fe(II)	<i>Acidithiobacillus ferrooxidans</i>	Periplasma	Bonnefoy & Holmes, 2012
<i>rusB</i>	A verificar	Rusticianina B	Função ainda não caracterizada	Oxidação do Fe(II)	<i>Acidithiobacillus ferrivorans</i> <i>Acidithiobacillus JCM 7811</i>	Periplasma	Bonnefoy & Holmes, 2012
<i>iro</i>	NA	Iro, proteína [Fe-S] de alto potencial	função ferro oxidase proposta	Oxidação do Fe(II)	<i>Acidithiobacillus ferrooxidans</i> <i>Acidithiobacillus ferrivorans</i> <i>Acidithiobacillus JCM 781</i>	Periplasma	Bonnefoy & Holmes, 2012; Kusano et al., 1992
<i>cyt579</i>	A verificar	Cyt579, citocromo c4	Recebe elétrons de Cyt572	Oxidação do Fe(II)	<i>Leptospirillum spp.</i>	Periplasma	Bonnefoy & Holmes, 2012
<i>cyc1</i>	<i>rus</i>	Cyc1 citocromo c4	Transfere elétrons da rusticianina para aa3 oxidase na via <i>downhill</i> de oxidação do Fe.	Oxidação do Fe(II)	<i>Acidithiobacillus ferrooxidans</i>	Periplasma	Bonnefoy & Holmes, 2012

<i>cycA1</i>	<i>petI</i>	CycA1 citocromo c4	Aceptor de elétrons da rusticianina na via <i>uphill</i> de oxidação do Fe(II)	Oxidação do Fe(II)	<i>Acidithiobacillus ferrooxidans</i>	Periplasma	Bonnefoy & Holmes, 2012; Quatrini et al., 2006
<i>ccoN</i>	<i>ccoNOPQ</i>	Subunidade catalítica da Cbb3 citocromo oxidase	Oxidase terminal na via <i>downhill</i> da oxidação do Fe(II)	Oxidação do Fe(II)	<i>Leptospirillum spp.</i>	Membrana plasmática	Bonnefoy & Holmes, 2012; Pitcher & Watmough, 2004
<i>ccoO</i>	<i>ccoNOPQ</i>	Subunidade da Cbb3 citocromo oxidase	Oxidase terminal na via <i>downhill</i> da oxidação do Fe(II)	Oxidação do Fe(II)	<i>Leptospirillum spp.</i>	Membrana plasmática	Bonnefoy & Holmes, 2012; Pitcher & Watmough, 2004
<i>ccoP</i>	<i>ccoNOPQ</i>	Subunidade da Cbb3 citocromo oxidase	Oxidase terminal na via <i>downhill</i> da oxidação do Fe(II)	Oxidação do Fe(II)	<i>Leptospirillum spp.</i>	Membrana plasmática	Bonnefoy & Holmes, 2012; Pitcher & Watmough, 2004
<i>ccoQ</i>	<i>ccoNOPQ</i>	Subunidade da Cbb3 citocromo oxidase	Oxidase terminal na via <i>downhill</i> da oxidação do Fe(II)	Oxidação do Fe(II)	<i>Leptospirillum spp.</i>	Membrana plasmática	Bonnefoy & Holmes, 2012; Pitcher & Watmough, 2004
<i>petA1</i>	<i>petI</i>	Subunidade proteína [Fe-S] tipo Rieske do complexo bc1	Recebe elétrons de CycA1 e transfere para quinonas associadas à membrana na via <i>uphill</i>	Oxidação do Fe(II)	<i>Acidithiobacillus ferrooxidans</i>	Membrana interna	Bonnefoy & Holmes, 2012; Bruscella et al., 2007
<i>petB1</i>	<i>petI</i>	Subunidade citocromo b do complexo bc1	recebe elétrons de CycA1 e transfere para quinonas associadas à membrana na via <i>uphill</i>	Oxidação do Fe(II)	<i>Acidithiobacillus ferrooxidans</i>	Membrana interna	Bonnefoy & Holmes, 2012; Bruscella et al., 2007
<i>petC1</i>	<i>petI</i>	Subunidade citocromo c1 do complexo bc1	recebe elétrons de CycA1 e transfere para quinonas associadas à membrana na via <i>uphill</i>	Oxidação do Fe(II)	<i>Acidithiobacillus ferrooxidans</i>	Membrana interna	Bonnefoy & Holmes, 2012; Bruscella et al., 2007
<i>petC3</i>	<i>petII<sup>2</sup></i>	Subunidade citocromo c1 do complexo bc1	envolvido na transferência de elétrons do enxofre (ou formato) para oxigênio ou Fe(III)	Redução do Fe(III)	<i>Acidithiobacillus ferrooxidans</i>	Membrana interna	Bruscella et al., 2007

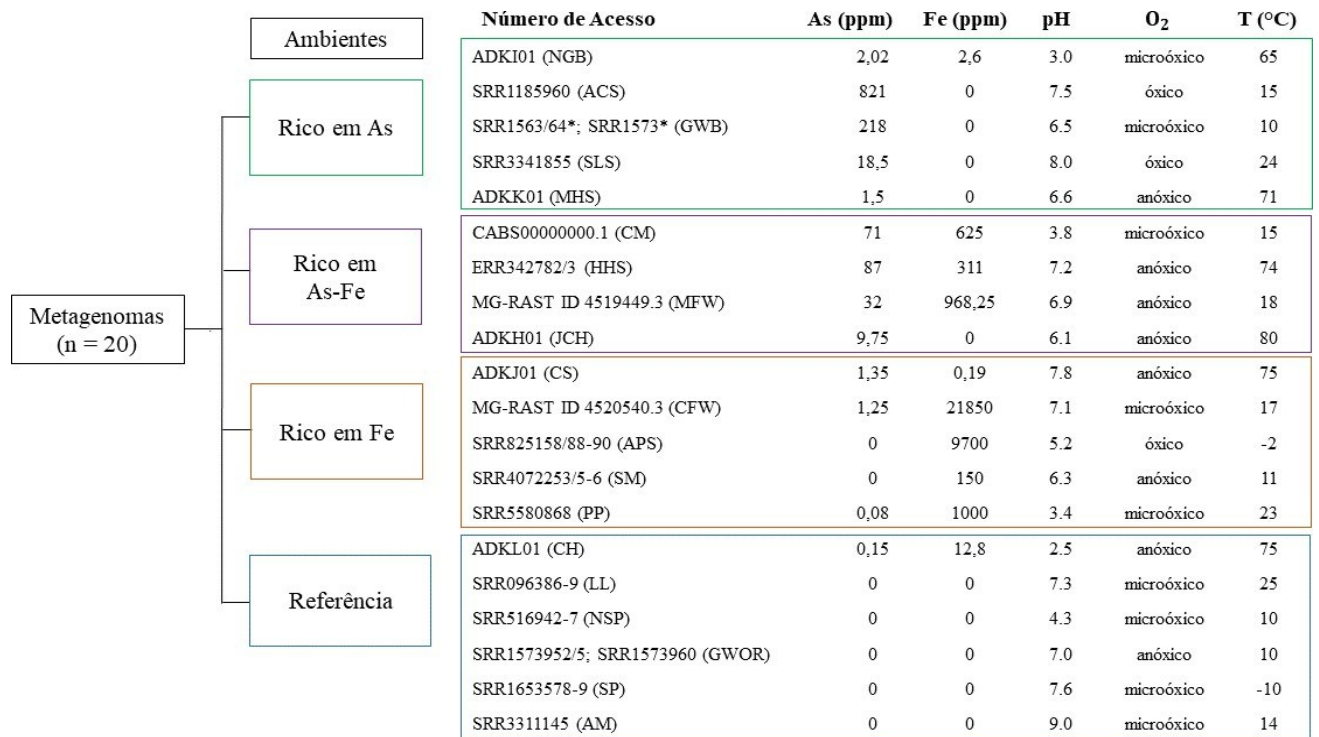
<i>petB2</i>	<i>petII</i>	Subunidade do citocromo b do complexo bc1	envolvido na transferência de elétrons do enxofre (ou formato) para oxigênio ou Fe(III)	Redução do Fe(III)	<i>Acidithiobacillus ferrooxidans</i>	Membrana interna	Bruscella et al., 2007
<i>petA2</i>	<i>petII</i>	Subunidade proteínica [Fe-S]tipo Rieske do complexo bc1	envolvido na transferência de elétrons do enxofre (ou formato) para oxigênio ou Fe(III)	Redução do Fe(III)	<i>Acidithiobacillus ferrooxidans</i>	Membrana interna	Bruscella et al., 2007
<i>coxA</i>	<i>rus</i>	Subunidade da citocromo oxidase tipo aa3	oxidase terminal na via <i>downhill</i> da oxidação do Fe(II)	Oxidação do Fe(II)	<i>Acidithiobacillus ferrooxidans</i>	Membrana interna	Bonnefoy & Holmes, 2012
<i>coxB</i>	<i>rus</i>	Subunidade da citocromo oxidase tipo aa3	oxidase terminal na via <i>downhill</i> da oxidação do Fe(II)	Oxidação do Fe(II)	<i>Acidithiobacillus ferrooxidans</i>	Membrana interna	Bonnefoy & Holmes, 2012
<i>coxC</i>	<i>rus</i>	Subunidade da citocromo oxidase tipo aa3	oxidase terminal na via <i>downhill</i> da oxidação do Fe(II)	Oxidação do Fe(II)	<i>Acidithiobacillus ferrooxidans</i>	Membrana interna	Bonnefoy & Holmes, 2012
<i>pioA</i>	<i>pio</i>	citocromo c	Transporte de elétrons na oxidação fototrófica de Fe(II)	Oxidação do Fe(II)	<i>Rhodopseudomonas palustris</i> TIE-1	Periplasma (possivelmente)	Jiao & Newman, 2007
<i>pioB</i>	<i>pio</i>	proteína de membrana externa	oxidação fototrófica de Fe(II); homólogo à MtrB de <i>S. oneidensis</i> MR-1	Oxidação do Fe(II)	<i>Rhodopseudomonas palustris</i> TIE-1	Membrana externa (possivelmente)	Jiao & Newman, 2007
<i>pioC</i>	<i>pio</i>	proteína [Fe-S] de alto potencial	oxidação fototrófica de Fe(II); similar à Iro de <i>At. ferrooxidans</i>	Oxidação do Fe(II)	<i>Rhodopseudomonas palustris</i> TIE-1	Periplasma (possivelmente)	Jiao & Newman, 2007
<i>foxE</i>	<i>fox</i>	FoxE, citocromo tipo c	oxidação fototrófica de Fe(II)	Oxidação do Fe(II)	<i>Rhodobacter</i> sp. SW2	A determinar	Croal et al., 2007
<i>foxY</i>	<i>fox</i>	FoxY, a caracterizar	Função ainda não caracterizada, possivelmente carreador de elétrons na oxidação fototrófica de Fe(II)	Oxidação do Fe(II)	<i>Rhodobacter</i> sp. SW2	A determinar	Croal et al., 2007

<i>foxZ</i>	<i>fox</i>	FoxZ, a caracterizar	Função ainda não caracterizada, possivelmente uma proteína citoplasmática com função transportadora na oxidação fototrófica de Fe(II)	Oxidação do Fe(II)	<i>Rhodobacter</i> sp. SW2	Citoplasma (Possivelmente)	Croal et al., 2007
-------------	------------	----------------------	---	--------------------	----------------------------	----------------------------	--------------------

### 4.3. Análises estatísticas

A frequência relativa dos genes foi calculada por estudo, ou seja, estudos com réplicas tiveram suas CDS somadas, formando um *pool*, e a frequência do gene foi calculada dividindo-se esta soma pelo número total de CDSs do estudo, classificadas e não classificadas (anotadas como “proteínas hipotéticas”). Segundo Nayfach & Pollard (2016), essa estratégia evita a superestimação de genes conhecidos, cuja proporção varia entre as comunidades. Além disso, foi feito o cálculo da frequência relativa dos genes por característica ambiental. Para isto o parâmetro escolhido foi a concentração de Fe e As nas áreas de estudo, como ilustrado na Figura 2. Os grupos formados foram os seguintes: 1) ambientes ricos somente em As, formado pelos conjuntos de dados NGB, ACS, GWB, MHS e SLS; 2) ambientes ricos tanto em As quanto em Fe (designados As-Fe daqui por diante), composto por JCH, , MFW, CM e HHS; 3) ambientes ricos apenas em Fe, que incluem CFW, APS, SM, CS e PP; e 4) ambientes escolhidos como Referência, com baixas concentrações tanto de As quanto Fe:, CH, LL, NSP, GWOR, SP e AM. Os conjuntos de dados JCH, CS e SM foram classificados com alto teor de Fe neste estudo devido às fases sólidas abundantes nestes locais, como pirita ( $\text{FeS}_2$ ) e goethita ( $\text{FeO}$ ), que infelizmente não foram mensuradas.

Para avaliar a correlação entre os genes envolvidos na biotransformação de As e Fe dentro de cada grupo (Figura 5) foram realizados os testes de Pearson e Spearman, usando uma matriz com os valores de frequência relativa dos genes (Saghäi et al., 2016). Para isto, foi usado o comando *corr* do pacote stats do programa R v. 3.3.1 (R Core Team, 2016). A correlação de Spearman mostrou-se mais adequada que a correlação de Pearson dado a natureza não normal dos dados e à presença de valores extremos (mesmo após transformação logarítmica), que com Pearson podem exacerbar ou subestimar a força das relações entre as variáveis (Mukaka, 2012). A interpretação da força dos coeficientes de correção foi feita de acordo com Mukaka (2012).



Número de Acesso	As (ppm)	Fe (ppm)	pH	O <sub>2</sub>	T (°C)
ADKI01 (NGB)	2,02	2,6	3,0	microóxico	65
SRR1185960 (ACS)	821	0	7,5	óxico	15
SRR1563/64*; SRR1573* (GWB)	218	0	6,5	microóxico	10
SRR3341855 (SLS)	18,5	0	8,0	óxico	24
ADKK01 (MHS)	1,5	0	6,6	anóxico	71
CABS00000000.1 (CM)	71	625	3,8	microóxico	15
ERR342782/3 (HHS)	87	311	7,2	anóxico	74
MG-RAST ID 4519449.3 (MFW)	32	968,25	6,9	anóxico	18
ADKH01 (JCH)	9,75	0	6,1	anóxico	80
ADKJ01 (CS)	1,35	0,19	7,8	anóxico	75
MG-RAST ID 4520540.3 (CFW)	1,25	21850	7,1	microóxico	17
SRR825158/88-90 (APS)	0	9700	5,2	óxico	-2
SRR4072253/5-6 (SM)	0	150	6,3	anóxico	11
SRR5580868 (PP)	0,08	1000	3,4	microóxico	23
ADKL01 (CH)	0,15	12,8	2,5	anóxico	75
SRR096386-9 (LL)	0	0	7,3	microóxico	25
SRR516942-7 (NSP)	0	0	4,3	microóxico	10
SRR1573952/5; SRR1573960 (GWOR)	0	0	7,0	anóxico	10
SRR1653578-9 (SP)	0	0	7,6	microóxico	-10
SRR3311145 (AM)	0	0	9,0	microóxico	14

Figura 5: Fluxograma da distribuição dos conjuntos de dados da meta-análise.

Para avaliar a relação entre a frequência relativa dos genes investigados e os fatores abióticos, os dados foram modelados por regressão linear e modelos lineares generalizados, em que as variáveis resposta foram as frequências relativas dos genes, e as variáveis independentes foram as concentrações de As e Fe (ppm), característica do ambiente em relação à concentração de oxigênio, temperatura e pH. As variáveis resposta não apresentaram distribuição normal, portanto os dados foram transformados. Foram testadas as transformações logarítmica, raiz quadrada, raiz cúbica e Box-Cox.

A transformação mais adequada para cada variável foi avaliada através de análise de histograma. Após a transformação, foram testados os modelos de regressão linear e regressão gama, com as funções *lm* e *glm*, respectivamente. A adequação dos modelos foi avaliada pelos seguintes critérios: existência de pontos de alavancagem, aberrantes e influentes, análise de resíduos, valor de  $R^2$  (para a regressão linear) e teste de *deviance* residual (para a regressão gama). As modelagens dos genes *arsC*, *aioA* e *cco* ficaram bem ajustadas, porém inadequadas, de acordo com a análise de resíduos. Essas regressões são apresentadas no Anexo 2.

Para identificar quais fatores melhor descreveram a variância entre os estudos, foi feita uma análise fatorial utilizando a função *principal* do pacote *psych* v.1.7.8 (Revelle, 2017), testando modelo com e sem rotação varimax. Foram definidos dois fatores de acordo com análise *à priori* da variância explicada pelos autovalores da matriz de correlação, por inspeção visual do seu gráfico de decaimento e pela saída da função *vss* (very simple structure), também do pacote *psych*.

Com o intuito de averiguar a similaridade entre os estudos, foi feita uma análise de agrupamento, testando-se os métodos hierárquico, *k-means* e *fuzzy medias*. Para a análise hierárquica utilizou-se a função *hclust* com matriz de distância euclidiana e *method="ward.D"*; o número de agrupamentos foi definido como 5 após inspeção visual do dendograma; pelo index "ptbserial" (valor máximo do índice) da saída da função *NbClust* (Charrad et al., 2014) e pela análise de componentes principais da solução, dado pela função *cusplot* do pacote *cluster* v.2.0.6 (Maechler et al., 2017). Para a análise por *k-means* utilizou-se a função homônima, e por *fuzzy medias* a função *fanny* do pacote *cluster*; em ambas o número de agrupamentos foi definido como 5 *à priori*, e confirmado por análise de componentes principais das soluções. Entretanto, a análise de agrupamento não mostrou uma divisão clara dos estudos de acordo com suas variáveis. Estes resultados são apresentados no Anexo 2.

Após a interpretação dos resultados da análise fatorial, foi calculada a frequência relativa dos genes estudados por faixa de pH. Novamente as amostras foram agrupadas por estudo, e estes foram divididos por faixa de pH, utilizando como referência a classificação de faixas de pH do solo do Departamento de Agricultura dos Estados Unidos (1993), como ilustrado na Figura 6. A correlação entre os genes investigados por faixa de pH foi calculada como descrito anteriormente. Para aumentar o n amostral por faixa de pH, estas foram reagrupadas arbitrariamente em quatro grupos: i) muito ácido (ultra a extremamente ácido,  $\text{pH} < 4,4$ ,  $n = 5$ ), ii) ácido (fortemente a levemente ácido,  $\text{pH} = 5,1 - 6,5$ ,  $n = 4$ ), neutro (neutro,  $\text{pH} = 6,6 - 7,3$ ,  $n = 6$ ) e alcalino (levemente a fortemente alcalino,  $\text{pH} > 7,4$ ,  $n = 5$ ).

Número de Acesso	pH	As (ppm)	Fe (ppm)	O <sub>2</sub>	T (°C)		Faixa de pH
ADKJ01 (CS)	2,5	149,8	12834	microóxico	75		< 3.5
SRR1563/64*; SRR1573* (GWB)	3	2022,3	2064,3	óxico	65		3.5–4.4
SRR5580868 (PP)	3,4	0,08	1000	microóxico	23		5.1–5.5
ERR342782/3 (HSS)	3,8	71	625	microóxico	15		6.1–6.5
SRR516942-7 (NSP)	4,3	0	0	microóxico	10		6.6–7.3
ADKK01 (MHS)	5,2	0	9700	óxico	-2		7.4–7.8
SRR1185960 (ACS)	6,1	9739,6	>39	microóxico	80		7.9–8.4
SRR4072253/5-6 (SM)	6,3	0	150	anóxico	11		8.5–9.0
SRR1573952/5; SRR1573960 (GWOR)	6,5	218	0	microóxico	10		
SRR3341855 (SLS)	6,6	1498,4	22,3	microóxico	71		
MG-RAST ID 4519449.3 (MFW)	6,9	32	968,25	anóxico	18		
SRR1653578-9 (SP)	7	0	0	anóxico	10		
CABS00000000.1 (CM)	7,1	1,25	21850	microóxico	17		
ADKH01 (JCH)	7,2	87	311	anóxico	74		
SRR096386-9 (LL)	7,3	0	0	microóxico	25		
ADKL01 (CH)	7,5	821	0	óxico	15		
SRR3311145 (AM)	7,6	0	0	microóxico	-10		
ADKI01 (NGB)	7,8	1348,6	189,7	microóxico	75		
SRR825158/88-90 (APS)	8	18,5	0	óxico	24		
MG-RAST ID 4520540.3 (CFW)	9	0	0	microóxico	14		

Figura 6: Classificação dos conjuntos de dados da meta-análise em faixas de pH.

## 5. RESULTADOS

### 5.1. Disponibilidade de metadados como um fator limitante

A obtenção de metadados revelou-se um grande desafio: parâmetros muito bem descritos em alguns trabalhos estavam ausentes em outros, de forma que apenas cinco variáveis ambientais (concentrações de As, Fe, pH, oxigênio e temperatura), presentes em todos os trabalhos, foram incluídos nas análises posteriores. Dessa forma, um total de 20 estudos, realizados com a técnica de sequenciamento *shotgun*, foram selecionados para a meta-análise (Quadro 1). Estes estudos compreenderam amostras ambientais muito diversas: 6 foram obtidas de tapetes microbianos; 5 de água; 4 de solos; 3 de sedimentos e 2 de estromatólitos. A temperatura dos locais de amostragem variou entre -10°C e 80°C; o pH foi de 2,5 a 9, e apresentaram condições óxicas, microóxicas ou anóxicas.

### 5.2. Análise bioinformática

O número de sequências obtidas a cada etapa da análise bioinformática está sumarizado na Tabela 1.

Tabela 1: Etapas do processamento das sequências de cada amostra.

<b>Nº de acesso/ do estudo</b>	<b>Abreviação</b>	<b>Nº reads pré- trimagem</b>	<b>Nº reads pós- trimagem</b>	<b>Tamanho das reads</b>	<b>Nº contigs</b>	<b>Eficiência da montagem</b>	<b>Nº de CDS</b>
ADKI01 (NGB)		NA	NA	NA	NA	NA	3773
SRR1185960 (ACS)		61111115	60349042	30-89	423497	16,89%	255417
SRR1563167 (GWB)		4126758	3994952	30-100	23567	18,23%	19762274
SRR1564103 (GWB)		8391525	8078941	30-100	58639	26,53%	43778
SRR1573565 (GWB)		14566759	14064712	30-100	877221	16,61%	68818
SRR1573578 (GWB)		12726811	12350747	30-100	86639	14,94%	52527
SRR3341855 (SLS)		47304781	37190449	30-199	349859	78,22%	307283
ADKK01 (MHS)		NA	NA	NA	NA	NA	1753
CABS0000000.1 (CM)		NA	NA	NA	NA	NA	3454
ERR342782 (HHS)		45102864	45102864	45	116582	47,16%	68071
ERR342783 (HHS)		33788413	33788413	45	81470	44,45%	56541
MG-RAST ID 4519449.3 (MFW)		NA	NA	NA	NA	NA	6460
ADKH01 (JCH)		NA	NA	NA	NA	NA	4600
ADKJ01 (CS)		NA	NA	NA	NA	NA	4064
MG-RAST ID 4520540.3 (CFW)		NA	NA	NA	NA	NA	13641
SRR825158 (APS)		185235	185235	40-630	33874	31,63%	14762
SRR825188 (APS)		216879	216879	40-679	33869	31,62%	14714
SRR825189 (APS)		138794	138794	40-638	20201	40,96%	8400
SRR825190 (APS)		126219	126219	40-755	19877	43,25%	8023
SRR4072253 (SM)		16000000	16000000	113	20440	94,13%	40410
SRR4072255 (SM)		16000000	16000000	113	27616	93,37	39338
SRR4072256 (SM)		16000000	16000000	113	41475	91,67%	40466
SRR5580868 (PP)		89264932	86989303	30-150	2240355	75,66%	2003909
ADKL01 (CH)		NA	NA	NA	NA	NA	2146
SRR096386 (LL)		16047965	14478781	30-100	178323	65,46%	128861
SRR096387 (LL)		14781841	13349080	30-100	161753	67,44%	136583
SRR096388 (LL)		13572457	12132769	30-100	161742	60,13%	125893
SRR096389 (LL)		17102225	14329735	30-100	189087	47,65%	141180
SRR516942 (NSP)		35616572	33277424	30-74	27446	4,10%	9229
SRR516943 (NSP)		33008856	30774403	30-74	8379	1,23%	2548
SRR516944 (NSP)		32029180	29923253	30-74	16219	2,18%	5982
SRR516945 (NSP)		28545299	26595398	30-74	14789	2,09%	4885
SRR516946 (NSP)		33010890	30693130	30-74	5531	1,09%	1617
SRR516947 (NSP)		31541438	29318834	30-74	3589	0,85%	1092

SRR1573952 (GWOR)	66806185	75998753	30-199	605098	3,10%	279698
SRR1573955 (GWOR)	11466495	11454959	30-150	33923	4,75%	19756
SRR1573960 (GWOR)	12388337	11633556	30-150	34143	4,77%	19866
SRR1653578 (SP)	1000000	925158	30-99	550	1,02%	177
SRR1653579 (SP)	1000000	923860	30-99	5668	4,91%	1883
SRR3311145 (AM)	131562915	123010484	30-99	457146	69,67%	349164

NA: Não se aplica. Dados depositados como *contigs/scaffolds* montados.

### 5.3. Análises estatísticas

#### 5.3.1. Diversidade e frequência relativa de genes envolvidos nos ciclos do As e Fe

Dos 44 genes selecionados para meta-análise, dos quais 4 e 40 estão envolvidos nos ciclos do As e Fe, respectivamente (Quadro 2), apenas 12 genes foram encontrados: *aioA* (arsenito oxidase), *arsC* (arseniato redutase detoxificadora), *arrA* (arseniato redutase dissimilatória), *ccoN* e *ccoP* (operon *cco* - complexo citocromo oxidase Cbb3), *cyc1* e *rus* (operon *rus*), *iro* (Fe oxidase), *petA*, *petB* e *petC* (operon *pet* - complexo citocromo bc1) e *omcB* (Fe redutase). A maioria dos citocromos (25/32) investigados não foi encontrada após a anotação. Entretanto, foram encontrados em todos os metagenomas outros genes, não selecionados anteriormente, codificando arseniato redutase (*arsC1* e *arsC2*) e citocromos, cujo envolvimento na redução do As(V) e especificamente na oxidação e redução do Fe ainda são pouco descritos na literatura (Tabela 2). A distribuição e frequência relativa dos genes investigados por ambiente estão ilustradas na Figura 7.

Tabela 2: Relação de citocromos encontrados neste estudo que não foram previamente selecionados.

Operon/Gene	Produto	Anotação Uniprot
<i>app</i>	Citocromo bd-II ubiquinol oxidase	Catalisa a absorção de O <sub>2</sub> dependente de quinol e independente de Na <sup>+</sup>
<i>biol</i>	P450, proteína do grupo heme-tiolato	Catalisa reações diversas, usualmente atua como monooxigenase terminal
<i>caa</i>	Citocromo c oxidase	Catalisa a redução de O <sub>2</sub> em H <sub>2</sub> O
<i>cba</i>	Citocromo c oxidase	Catalisa a redução de O <sub>2</sub> em H <sub>2</sub> O
<i>ccc</i>	Citocromo c-550	Associado ao fotosistema II em cianobacteria;
<i>ccm</i>	Sistema ccm	Maturação do citocromo c em <i>Escherichia coli</i>
<i>ccp</i>	citocromo c 551 peroxidase	Oxida citocromo c e reduz H <sub>2</sub> O <sub>2</sub>
<i>ccs</i>	Sistema ccs	Sistema II da maturação do citocromo c
<i>cox</i>	Citocromo c oxidase alternativa	Catalisa a redução de O <sub>2</sub> em H <sub>2</sub> O
<i>cta</i>	Subunidade citocromo c do complexo citocromo oxidase <i>caa3</i>	Cadeia transportadora de elétrons aeróbica
<i>cyb</i>	Citocromo b diheme	Cadeia transportadora de elétrons
<i>cyc</i>	Citocromo c (homologia com citocromo mitocondrial)	Possivelmente transfere elétrons entre os complexos bc1 e aa3
<i>cycB</i>	Citocromo c	Fixação de nitrogênio ( <i>Bradyrhizobium diazoefficiens</i> ); metabolismo do metanol ( <i>Paracoccus denitrificans</i> )
<i>cycF</i>	Citocromo c monoheme	Cadeia transportadora de elétrons aeróbica
<i>cycM</i>	Citocromo c 552	Possivelmente transfere elétrons entre os complexos bc1 e aa3
<i>cyd</i>	Complexo citocromo d oxidase	Oxidase terminal que catalisa a oxidação de ubiquinol-8 e redução de O <sub>2</sub>
<i>cyo</i>	Citocromo bo ubiquinol oxidase	Componente da cadeia respiratória aeróbica de <i>E. coli</i>
<i>cyt</i>	Citocromo c monoheme	Possivelmente doa elétrons para citocromo oxidase de membrana e nitrato redutase periplasmática ( <i>Nitrosomonas europaea</i> )
<i>fbcH</i>	Complexo citocromo bc1	Característico de respiração anaeróbica
<i>napC</i>	Citocromo c tetraheme	Implicado na transferência de elétrons entre quinol (na membrana) e enzimas periplasmáticas
<i>nirM</i>	Citocromo c 551	Doador de elétrons para a nitrito redutase
<i>nrf</i>	Nitrito redutase	Respiração redutora de nitrito ou óxido nítrico à amônia

<i>psb</i>	Citocromo c -550	Requerido para a estabilização ou função normal do fotosistema II
<i>qcr</i>	Complexo citocromo bc1	Catalisa oxidação de menaquinol e redução do citocromo c na cadeia respiratória
<i>qrcA</i>	Subunidade citocromo c multiheme da menaquinona redutase	Relacionado ao complexo alternativo III, envolvido na redução do sulfato em Deltaproteobacteria
<i>tor</i>	Citocromo c	Participa da cadeia respiratória anaeróbica da trimethylamine-N-oxide reductase
<i>yecJ</i>	Citocromo b 561	Carreador de elétrons na cadeia respiratória
<i>yodB</i>	Citocromo b 561	Carreador de elétrons na cadeia respiratória
<i>yth</i>	Citocromo bd menaquinol oxidase	Compensa pela perda do complexo citocromo aa3 ( <i>Bacillus subtilis</i> 168)

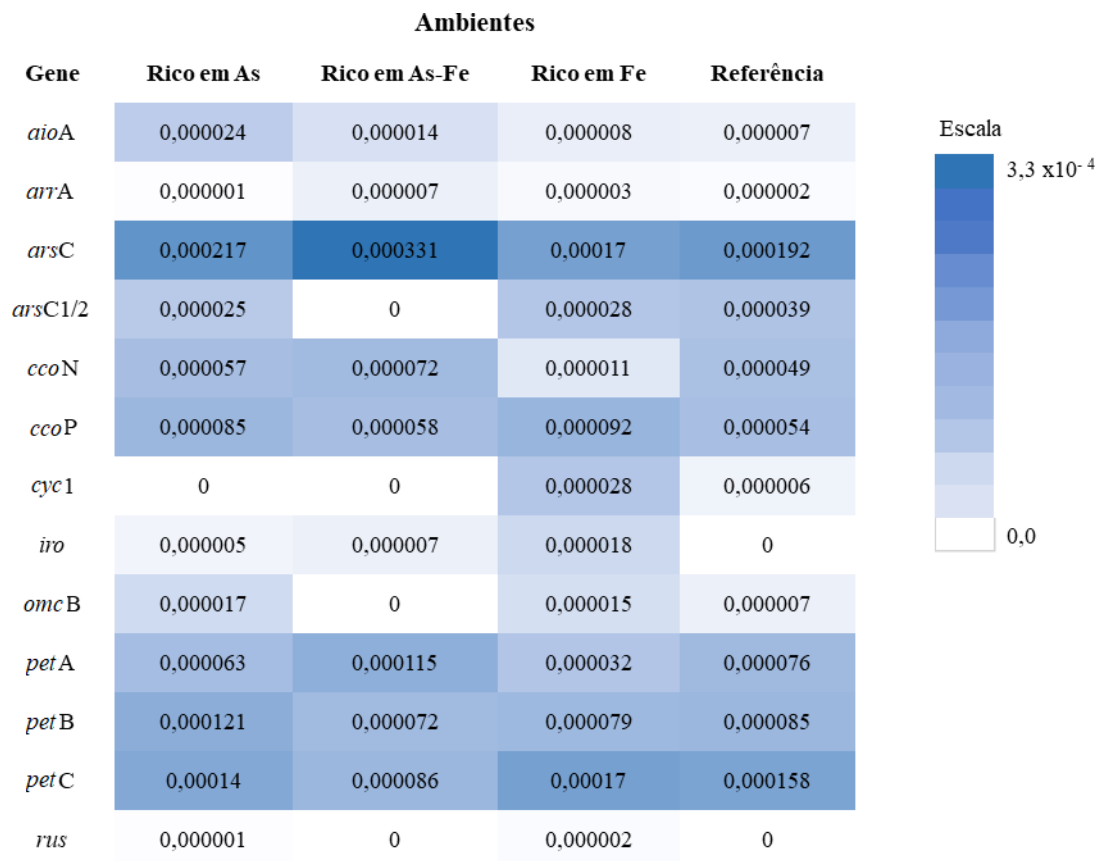


Figura 7: Frequência relativa dos genes investigados por tipo de ambiente

Os genes envolvidos no ciclo do As foram encontrados em todos os tipos de ambientes, exceto para *arsC1/2*, ausentes nos ambientes ricos em As-Fe. Considerando os genes relacionados a redução do As, dois tipos de arseniato redutase detoxificadoras foram identificadas na anotação: *arsC*, correspondendo às bem caracterizadas classes dependentes de thioredoxina/thioredoxina redutase (como a do plasmídeo pI258 de *Staphylococcus aureus*) e glutathiona/glutaredoxina (como a do plasmídeo R773 de *Escherichia coli*), e *arsC1* e *arsC2*, codificando para arseniato redutases acopladas à via redox de mycothiol/mycoredoxina. Deve-se destacar que os genes *arsC1/2* foram descritos, até o momento, apenas na actinobactéria *Corynebacterium glutamicum*, na qual confeririam uma maior vantagem adaptativa a este táxon (Ordóñez et al., 2009, Villadangos et al., 2011). Estes genes são expressos após o contato com As(III) formado após a redução de As(V) pela enzima ArsC, codificada pelo gene *arsC* (da classe thioredoxina), expresso constitutivamente em nível basal.

Em relação aos genes envolvidos na oxiredução do Fe, apenas os operons *cco* e *pet* foram encontrados em todos ambientes investigados, ambos com maior frequência em ambientes ricos em As. O primeiro está implicado na oxidação de Fe(II), enquanto *pet* pode tanto reduzir quanto oxidar espécies de Fe. Os genes *iro* e *rus*, que codificam proteínas que participam da oxidação do Fe(II), não foram encontrados nos ambientes referência, em contraste tiveram maior frequência em ambientes ricos em Fe. Finalmente, genes do operon *rus* (*cyc1* e *rus*) e *omcB* (redução do Fe) não foram encontrados em ambientes com altas concentrações de As-Fe, e apenas *cyc1* não foi encontrado nos ambientes ricos em As.

Nos ambientes ricos em As, o gene envolvido na ciclagem deste metaloide com maior frequência relativa foi *arsC* ( $2,2 \times 10^{-4}$ ), e os genes implicados na ciclagem do Fe foram *petC* ( $1,4 \times 10^{-4}$ ) e *petB* ( $1,2 \times 10^{-4}$ ). Em ambientes ricos em As-Fe, os genes mais frequentes foram *arsC* ( $3,3 \times 10^{-4}$ ), *petA* ( $1,1 \times 10^{-4}$ ) e os genes do operon *cco* ( $6,2 - 7,6 \times 10^{-5}$ ). Em ambientes ricos em Fe, os genes mais abundantes foram *iro* ( $1,8 \times 10^{-4}$ ), *petC* ( $1,7 \times 10^{-4}$ ) e *arsC* ( $1,7 \times 10^{-4}$ ); e nos ambientes selecionados como referência *arsC* ( $1,9 \times 10^{-4}$ ) e *petC* ( $1,6 \times 10^{-4}$ ) predominaram.

Para investigar quais dos fatores ambientais melhor explicam a variância entre os ambientes, foi feita uma análise fatorial. O melhor modelo da análise fatorial foi obtido com dois fatores e rotação varimax. O primeiro fator explica as variáveis concentração de As e pH, e o

segundo fator explica concentração de Fe e temperatura, com cargas fatoriais e coeficientes de 0,78 e 0,6, 0,67 e 0,5; 0,85 e 0,7; e 0,63 e 0,5, respectivamente (Figura 8).

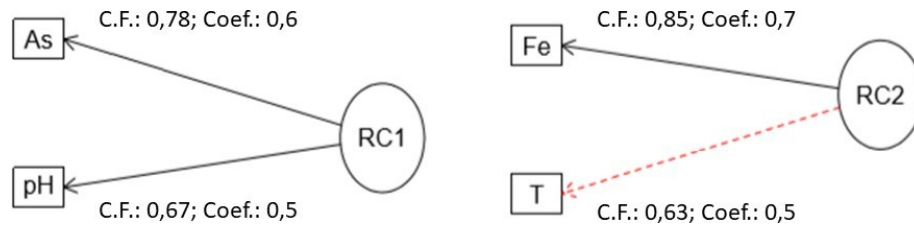


Figura 8: Análise fatorial. O primeiro fator (RC1) explica as variáveis concentração de As (As) e pH (pH), o segundo fator (RC2) explica as variáveis concentração de Fe (Fe) e temperatura (T). Abreviações: C.F, carga fatorial; Coef, coeficiente.

Dada a relevância da variável pH para a variância entre os estudos, também foi calculada a frequência relativa por faixa de pH, os resultados são apresentados na Tabela 3. A abrangência dos genes por faixa de pH é ilustrada na Figura 9. O gene *arsC* e os operons *cco* e *pet* figuraram entre os mais frequentes em todas as faixas de pH, e é notável que a maioria dos genes tiveram sua frequência reduzida à medida em que a alcalinidade do ambiente aumenta. *arrA* e *arsC1/2* foram encontrados em frequência muito menor e em uma faixa de pH mais restrita que *arsC*, estando ausentes em  $\text{pH} > 7,8$ , enquanto *aioA* predominou em pH moderadamente alcalino (7,9 – 8,4). O gene *iro* e o operon *rus* foram mais frequentes em pH ultra ácido ( $< 3,5$ ) e não foram encontrados em  $\text{pH} > 7,4$ . Em contraste, o operon *cco* ocorreu em todas as faixas de pH.

Tabela 3: Frequência relativa dos genes envolvidos na transformação de As e Fe por faixa de pH.

Gene	Ultra ácido	Extremamente ácido	Fortemente ácido	Ligeiramente ácido	Neutro	Ligeiramente alcalino	Moderadamente alcalino	Fortemente e alcalino
<i>aioA</i>	0,000006	0,000013	0	0,000007	0,000022	0,000011	0,000065	0,000007
<i>arrA</i>	0	0,000007	0	0,000009	0	0,000008	0	0
<i>arsC</i>	0,000169	0,000307	0,0005705	0,000187	0,000217	0,000222	0,000044	0,000015
<i>arsC1/2</i>	0,000028	0	0	0,000019	0,000048	0,000034	0	0
<i>cco</i>	0,000097	0,00012	0,001141	0,000102	0,000118	0,000165	0,000109	0,000015
<i>iro</i>	0,000018	0	0	0,000006	0,000001	0	0	0
<i>omcB</i>	0,000017	0	0	0,000001	0,000013	0,000008	0	0

<i>petA</i>	0,00003	0,0001	0,0005705	0,000058	0,000078	0,00011	0	0
<i>petB</i>	0,000094	0,000073	0	0,000075	0,000072	0,000115	0,000044	0,000044
<i>petC</i>	0,000187	0,00008	0	0,000109	0,000151	0,000149	0,000065	0,000015
<i>rus</i>	0,00003	0	0	0,00001	0,000001	0	0	0

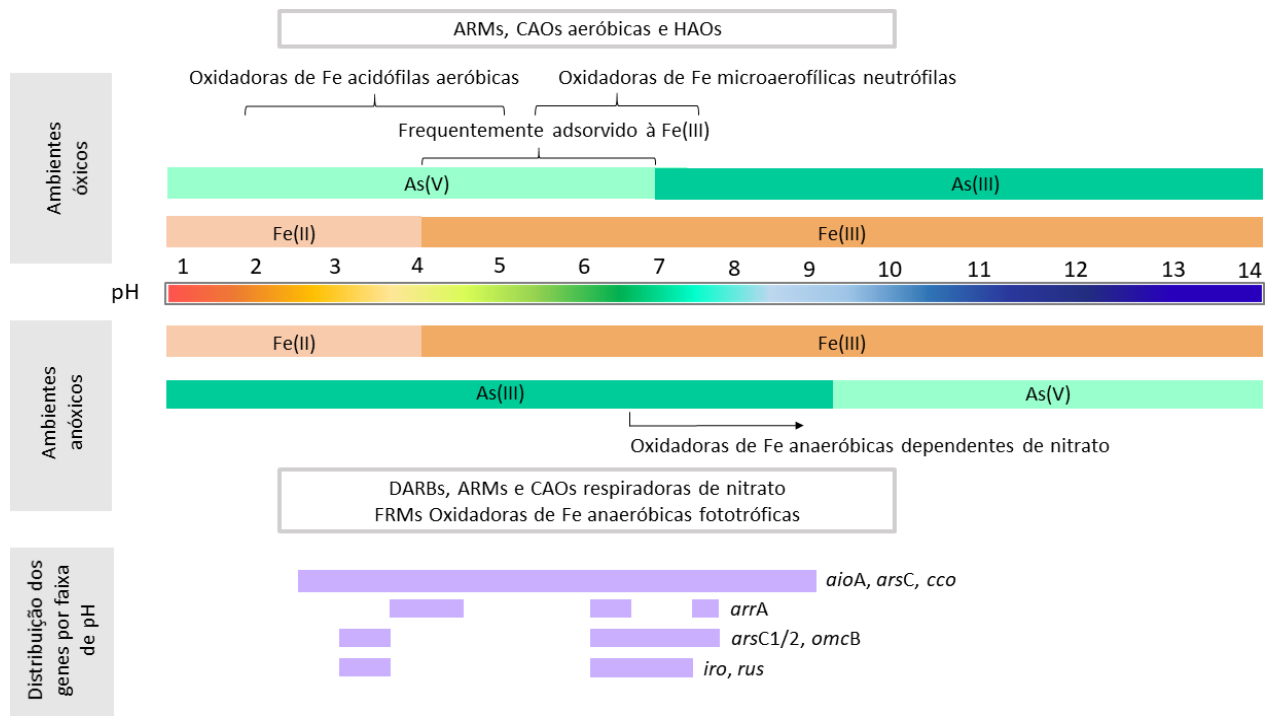


Figura 9: Distribuição de genes envolvidos na especiação do As e Fe em função do pH e presença/ausência de oxigênio.

### 5.3.2. Interações entre os genes envolvidos nas ciclagens do As e do Fe

Os coeficientes de correlação entre os genes investigados nos ambientes ricos em As, em As-Fe, ou Fe e nos ambientes referência são apresentados nas Figuras 10A-D, respectivamente. e suas correlações nos ambientes muito ácido, ácido, neutro e alcalino são apresentadas nas Figuras 11A-D, respectivamente.

Em ambientes ricos em As e em As-Fe, e em  $\text{pH} > 6,6$ , *arsC* apresentou correlação positiva de moderada a alta com *aioA*. Em contraste, o outro gene envolvido na redução do As, codificando para arseniato redutase dissimilatória, *arrA*, não apresentou correlação com o gene de oxidação do As nestes ambientes. A correlação entre *arrA* e *aioA* foi positiva apenas

em  $\text{pH} > 7,4$ . Os genes *arsC* e *arsC1* tiveram correlação positiva em ambiente rico em As, mas *arsC1* não foi encontrado em ambiente rico em As-Fe. Em ambiente rico apenas em Fe, as correlações entre *aioA* e *arsC-arsC1/2* foram positiva e negativa moderada, respectivamente. Finalmente, nas amostras referência, *aioA* teve correlação negligenciável com os genes envolvidos na redução do As, que por sua vez apresentou correlação positiva moderada entre si.

Os genes *ccoP* e *ccoN* codificam subunidades do complexo citocromo oxidase Cbb3, envolvido na oxidação de Fe(II). Como esperado, apresentaram correlação positiva entre si em todos os ambientes e faixas de pH analisadas. O operon *rus* também participa da oxidação do Fe, e compreende os genes *coxABC*, *cyc1*, *cyc2* e *rus*. Embora a anotação dos genes *rus* encontrados não os discriminasse entre *rusA* ou *rusB*, ambos codificam rusticianina, proteína que participa da oxidação do Fe. Não foram identificados na anotação os genes *coxABC* e *cyc2*. É importante ressaltar que *cyc1* estava ausente em ambientes ricos em As e em As-Fe, e *rus* ausente em ambientes ricos em As-Fe e referência. A correlação entre *cyc1* e *rus* foi positiva muito alta em ambientes ricos em Fe e em  $\text{pH} < 4,4$ . Por fim, o gene *iro* codifica uma proteína Fe-S com função ferro oxidase proposta. *cco*, *iro* e *rus* tiveram correlação positiva baixa a alta com os genes que codificam arseniato redutases detoxificadoras em todas as categorias ambientais e faixas de pH em que estavam presentes, e correlação negligenciável à negativa baixa com *aioA* em  $\text{pH} < \sim 6,5$ . Em pH neutro, *iro* teve correlação positiva alta com *aioA*, mas estava ausente em  $\text{pH} > 7,3$ , onde *cco* teve correlação positiva alta com *aioA*. Finalmente, o operon *cco* teve correlação positiva moderada com *aioA* em ambiente rico em As, e o gene *iro* apresentou correlação positiva alta com *aioA* em ambiente rico em As-Fe.

O complexo citocromo bc1 pode atuar tanto na redução quanto na oxidação de espécies de Fe, como observado em *Acidithiobacillus ferrooxidans* (Bruscella et al., 2007). O citocromo bc1 envolvido na via *uphill* de oxidação do Fe(II) é codificado pelo operon *petIABC*, enquanto *petIIABC* está envolvido na transferência de elétrons do enxofre para Fe(III) ou oxigênio. A anotação funcional identificou os genes *petA*, *petB* e *petC* em todos os ambientes, figurando entre os genes de biotransformação do Fe prevalentes nas comunidades. Entretanto, a anotação funcional não continha informação sobre qual operon (*petI* ou *petII*) eles pertenciam. Assim, não foi possível inferir sobre o significado ecológico de suas relações. *petA* e *petC* apresentaram correlação positiva em ambiente rico em Fe, enquanto *petA* e *petB* em ambiente

rico em As-Fe. Porém, estes dados são inconclusivos sobre o tipo de operon *pet* (I ou II) ao qual estes genes possam pertencer.

O gene *omcB* foi o único do cluster *mtr/omc* encontrado nas amostras. Ele codifica um citocromo decaheme que participa da redução do Fe(III), sendo aceitor de elétrons de proteínas periplasmáticas com potencial de reduzir Fe(III) extracelular diretamente em *Shewanella* (Reyes et al., 2010). O gene *omcB* apresentou correlação positiva alta com *arrA* e *arsC-arsC1* em ambientes ricos em As e Fe, respectivamente; ocorrendo na mesma faixa de pH observada para os genes de redução do As(V), embora em um espectro muito reduzido. As correlações entre *omcB* e os genes *arsC*, *arsC1/2* e *arrA* variaram de acordo com a faixa de pH: em pH < 4,4 apresentou correlação positiva alta com *arsC1/2*, e com *arrA* em pH = 5,1 – 6,1. Em pH neutro foi positiva baixa com *arsC1/2*, e em pH > 7,4 foi positiva alta a muito alta com os três genes. Observou-se, ainda, correlação negativa moderada com *aioA* em ambientes ricos em Fe e referência, respectivamente, e positiva moderada em ambientes ricos em As. Destaca-se que *omcB* não foi detectado em ambientes ricos em As-Fe.

No genoma de *Geobacter sulfurreducens* *omcB* está localizado em uma região cromossômica duplicada de 10kb onde se encontra um gene com 79% de identidade, denominado *omcC* (Leang et al., 2003). Ao contrário de *omcB*, *omcC* não possui um papel fundamental na redução do Fe. *omcC* também foi identificado na anotação desta meta-análise, porém em frequência menor, e somente em dois estudos, correspondentes a ambientes rico em As e referência (GWB e AM, respectivamente).

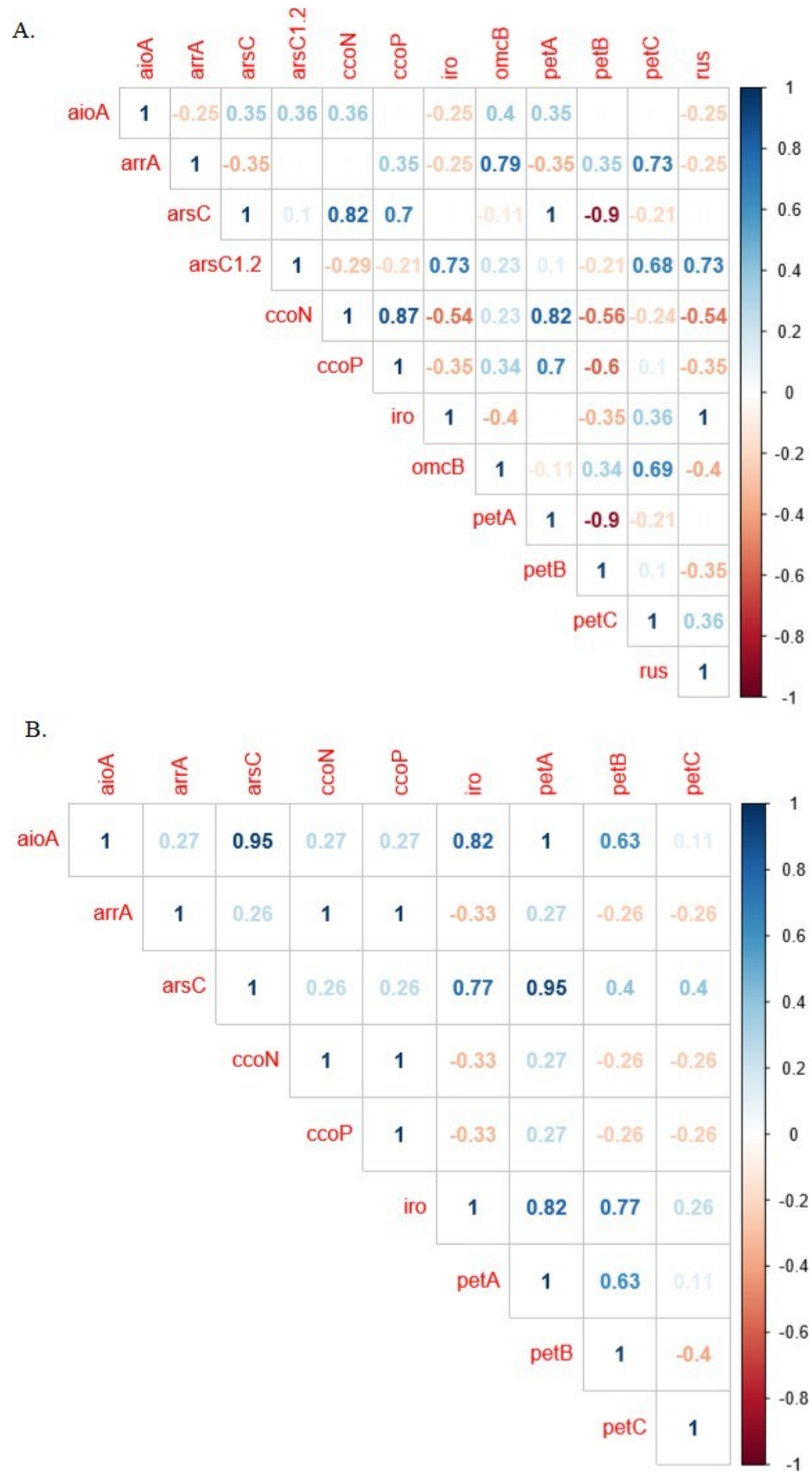


Figura 10: Coeficientes de correlação (Spearman) entre os genes investigados em ambientes ricos em As (A), em As e Fe (B), em Fe (C) e ambientes referência (D).

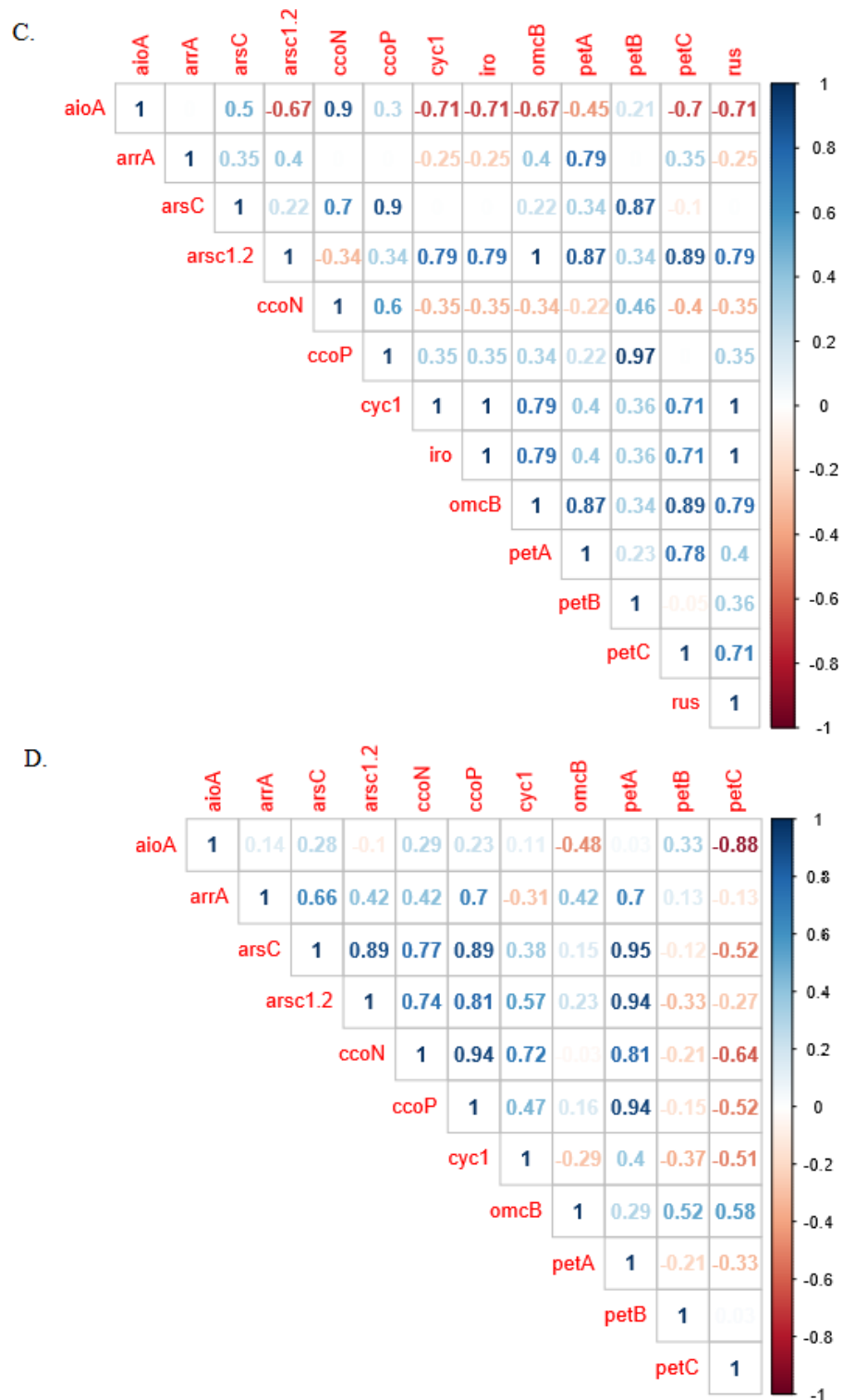


Figura 10: Coeficientes de correlação (Spearman) entre os genes investigados em ambientes ricos em As (A), em As e Fe (B), em Fe (C) e ambientes referência (D).

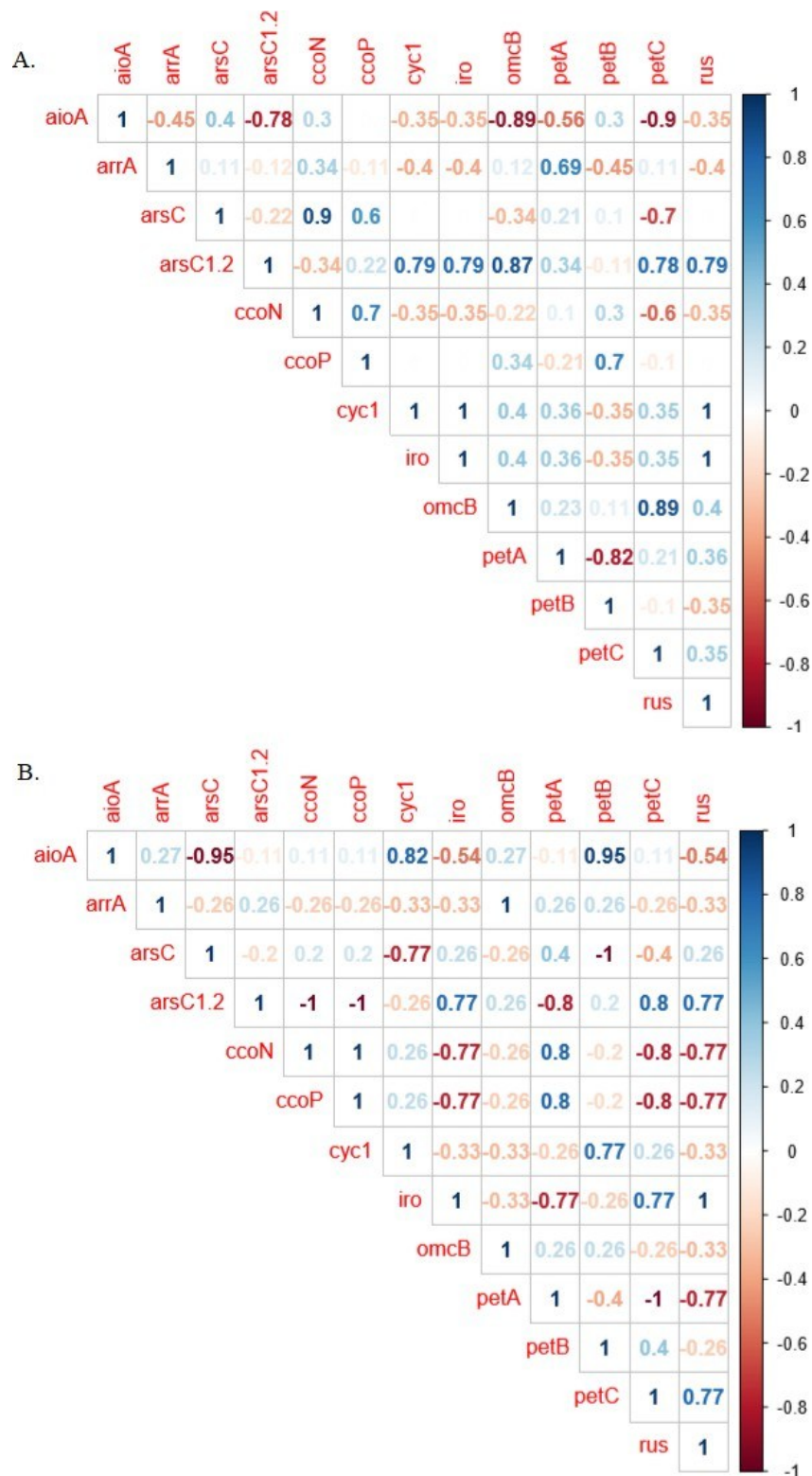


Figura 11: Coeficientes de correlação (Spearman) entre os genes investigados em ambientes classificados como muito ácidos (A), ácidos (B), neutros (C) e alcalinos (D).

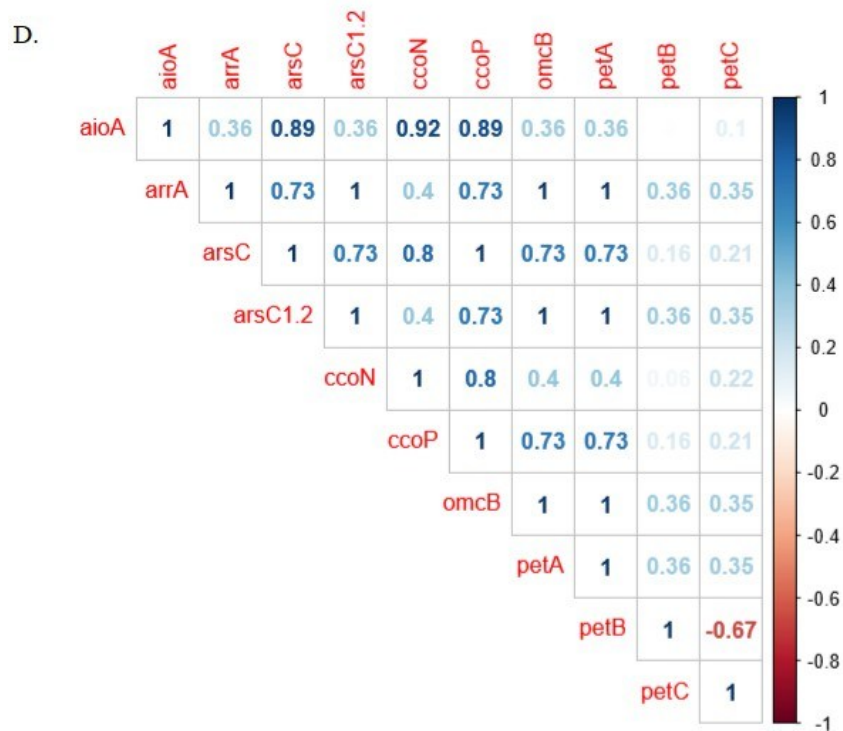
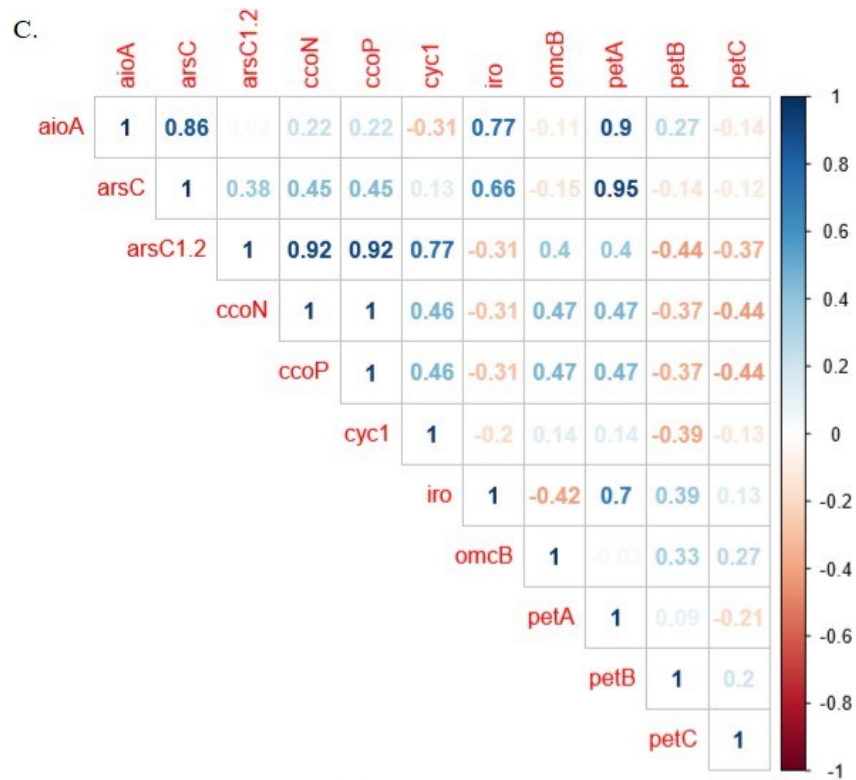


Figura 11: Coeficientes de correlação (Spearman) entre os genes investigados em ambientes classificados como muito ácidos (A), ácidos (B), neutros (C) e alcalinos (D).

## 6. DISCUSSÃO

As interações entre procariotos que participam da ciclagem do As e Fe contribuíram para moldar as características geoquímicas do planeta desde seu surgimento na Terra primitiva. Estes microrganismos exercem grande impacto nos ambientes naturais, principalmente participando da mobilização de As e conseqüente contaminação da água para consumo humano. Embora sejam conhecidos genes e mecanismos implicados na oxirredução do As e Fe, as suas correlações, diversidade gênica e abrangência ambiental não são completamente elucidadas. Este estudo é a primeira meta-análise da diversidade, frequência relativa e correlações entre os genes envolvidos na ciclagem do As e Fe, e os fatores físico-químicos que os influenciam. Esta abordagem revelou a ubiquidade e alta frequência do gene *arsC* e do operon *cco*, incluindo aqueles ambientes com pHs e temperaturas extremas. Estes dados corroboram a ancestralidade do gene *arsC*, destacando a relevância dos ARMs na mobilização do metaloide.

Em condições óxicas, como a da atmosfera atual (Jackson & Dugas, 2003), e em pH ácido a neutro, As(V) predomina em meio aquoso (Smedley & Kinniburgh, 2001) (Figura 1). Assim, espera-se que genes envolvidos na redução do As, *arsC*, *arsC1/2* e *arrA*, ocorram com maior frequência nestas condições. Isso de fato foi observado para *arsC*. Dentre os genes envolvidos na biotransformação do As, *arsC* foi o que apresentou maior frequência em todos os ambientes. Além disso, cobriu amplo espectro de pH (de ultra ácido a moderadamente alcalino) e temperatura (-10°C a 80°C), predominando onde a disponibilidade de As(V) dissolvido é maior e diminuindo gradativamente em meio alcalino.

As correlações positivas moderada entre *arsC* e *aioA* e *arsC1/2*, e a correlação negativa baixa de *arrA* com *aioA*, indicam que em ambiente rico em As ocorre uma intensa atividade microbiana de redução. Estes dados, aliados à frequência relativa dos genes estudados, sugerem que a arseniato redutase detoxificadora, codificada por *arsC*, é a principal enzima responsável por catalisar esta reação. Além disso, a meta-análise revelou que nos demais ambientes ricos em Fe, As-Fe e referência *arsC* foi também o gene mais frequente, e seus coeficientes de correlação corroboram a hipótese que *ArsC* é a principal enzima responsável pela redução do As no ambiente. Esses dados fornecem um *insight* sobre os possíveis mecanismos de oxirredução do As predominantes na Terra primitiva, uma vez que a maioria dos conjuntos de dados de ambiente rico em As-Fe possuem características similares às

condições onde a vida provavelmente se originou (altas concentrações de Fe, As e anaerobiose). As enzimas de oxiredução do As são evolutivamente muito antigas, e possivelmente surgiram antes mesmo da separação dos domínios Archaea e Bacteria. Uma hipótese para a origem da vida na Terra sustenta que ela se deu em fontes hidrotermais submarinas, ambientes naturalmente abundantes em As e Fe. Nesse contexto, haveria uma pressão seletiva para o desenvolvimento de sistemas de detoxificação do As(V), um análogo do fosfato. O acúmulo de As(V) só seria possível após o surgimento da fotossíntese oxigênica, que aumentou o estado de oxidação da atmosfera terrestre primitiva (Lebrun et al., 2002; Jackson & Dugas, 2003; Duval et al., 2008). Dessa forma, é plausível que *aioA* tenha surgido mais cedo na evolução, devido a abundância de substrato (As(III)), seguido pelos sistemas de redução do As(V). Análise filogenética do gene *arrA* sugere fortemente que seu surgimento ocorreu após a separação dos domínios Archaea e Bacteria, sendo assim posterior ao surgimento de *arsC* (Duval et al., 2008). Os resultados obtidos neste estudo corroboram esta hipótese, uma vez que a correlação entre *aioA* e *arrA* foi negligenciável.

Devido à ampla dispersão do As na crosta terrestre, sua alta toxicidade para todos os organismos e a antiguidade das suas vias de oxirredução, não é surpreendente que os genes envolvidos na ciclagem do As sejam ubíquos. De fato, além de *arsC*, os genes *aioA* e *arrA* também foram encontrados em todas as categorias ambientais. *aioA* exibiu frequência relativa similar nos quatro tipos de ambiente, sendo mais frequente em pH moderadamente alcalino (7,9 – 8,4). Estes dados estão de acordo com a dinâmica de especiação do As, em que As(III) predomina em pH > 7 em ambientes óxicos. A ocorrência dos genes *arsC*<sub>1/2</sub> apenas em ambientes ricos em As-Fe e em uma faixa de pH mais restrita que *arsC*, sugere que a arseniato redutase codificada por estes genes seja evolutivamente mais recente que *ArsC*.

O gene *arrA* foi descrito pela primeira vez por Saltikov & Newman (2003) na bactéria respiradora de As(V) *Shewanella* sp. ANA-3, e a maioria dos estudos sobre o papel das DARBs nas reações redox do As usaram técnicas de cultivo. Nas últimas décadas, a introdução de técnicas metagenômicas para o estudo de DARBs não cultiváveis revelou uma enorme diversidade do gene *arrA*, ainda pouco explorada (Song et al., 2009; Costa et al., 2014; Suhadolnik et al., 2017). Assim, o fato de que sequências de *arrA* foram detectadas somente após uma busca refinada, alinhando-se sequências curadas do gene (incluindo sequências de bactérias não cultiváveis) contra os *contigs* e *scaffolds* dos conjuntos de dados, pode ser explicado pela sub-representação da diversidade de *arrA* nos bancos de dados

disponíveis, ou porque o gene foi anotado incorretamente. Erros na anotação do *arrA* tornam-se evidentes quando sequências do gene, com atividade comprovada experimentalmente, são alinhadas contra bancos de dados. Estas sequências frequentemente se alinham com alta identidade a sequências anotadas genericamente como “molybdopterin oxidoreductase” (*arrA* faz parte da enorme superfamília DMSO de enzimas molybdenum), ou com sequências parciais da subunidade A da enzima dimethyl sulfoxide respiratória. Dessa forma, é possível que existam mais cópias deste gene nos conjuntos de dados desta meta-análise. Entretanto, deve-se destacar que o gene *arrA* foi menos frequente (aproximadamente 200 vezes) e ocorreu em uma faixa de pH mais restrita que *arsC*, ausente em  $\text{pH} > 7, 8$ . Assim, é possível que este gene seja raro em ambientes naturais, e então a redução de As(V) por *ArsC* seria o mecanismo predominante. Interessantemente, estes achados estão de acordo com recente estudo realizado por Crognale et al. (2017), em que o gene *arrA* não foi encontrado em comunidades de água doce geotermiais ricas em As, mas *arsC* sim.

Tanto em ambientes óxicos quanto anóxicos, Fe(II) dissolvido predomina em  $\text{pH} \leq 4$  (Kappler & Straub, 2005). Dessa forma, espera-se encontrar genes envolvidos na oxidação do Fe em maior frequência nesta faixa de pH. Isso foi observado para o gene *iro* e para o operon *rus*, cujas frequências relativas foram maiores na faixa de  $\text{pH} < 3,5$  (ultra ácido) e diminuiu à medida em que o pH aumentou, e ausente em  $\text{pH} > 7,4$ . As vias de oxidação do Fe(II) por estes dois genes foram propostas a partir de estudos com linhagens de *Acidithiobacillus sp.*, proteobactérias acidófilas extremas. Interessantemente, nesta meta-análise *iro* e *rus* foram restritos a ambientes microaerofílicos. Dessa forma, nos conjuntos de dados selecionados, as bactérias que os abrigam poderiam ser microaerofílicas neutrófilas ou anaeróbicas facultativas. Os índices de correlação e a frequência relativa destes genes nas quatro categorias de ambientes analisadas indicam que o processo de oxidação do Fe(II) mediado pelo operon *rus* é mais relevante em ambientes ricos em Fe do que nos demais, e que a oxidação por *iro* predomina apenas em ambientes ricos em As-Fe.

O operon *cco*, por outro lado, foi encontrado em todas as faixas de pH, embora sua frequência relativa decresceu à medida que o pH aumentou. É interessante que *cco* tenha sido encontrado em três dos quatro ambientes classificados como óxicos, uma vez que ele é expresso em baixa pressão de oxigênio (Bertini et al., 2006). Além disso, sua presença em conjunto de dados classificado como óxico e alcalino (APS), onde Fe(II) dissolvido é muito pouco disponível, indica que o complexo Cbb3 citocromo oxidase pode ser relevante para bactérias que ocupam nichos onde o Fe é um micronutriente limitante. Finalmente, os genes deste operon foram

encontrados em todos os tipos de ambientes, e dentre os genes de oxidação do Fe(II) investigados em maior frequência relativa. Assim, estas observações sugerem que *cco* pode conferir um maior *fitness* às bactérias que o abrigam, tornando-as capazes de oxidar Fe(II) em uma variedade de condições ambientais.

De acordo com a termodinâmica do As e do Fe, em ambientes óxicos a interação mais provável entre os processos de oxirredução destes elementos seria a oxidação do Fe(II) em fase aquosa em Fe(III) sólido, que seria capaz de adsorver As(V) proveniente da oxidação do As(III) em pH > 4, imobilizando-o, como ilustrado na Figura 1. Em pH < 4 há intensa oxidação de Fe(II). Entretanto, neste pH existe uma grande tendência à solubilização tanto do Fe quanto do As (Smedley & Kinniburgh, 2001, Kappler & Straub, 2005), sugerindo que os processos de oxidação As(III) e consequente imobilização do As(V) não ocorreriam em pH < 4. Os dados biológicos obtidos nesta meta-análise concordam com a dinâmica de especiação de As e Fe, uma vez que se observou uma alta correlação entre os genes de oxidação desses dois elementos em ambientes ricos em As e em As-Fe. Notavelmente, a meta-análise revelou que os genes *cco* e *iro* estão envolvidos na oxidação do Fe em ambientes óxicos ricos em As e As-Fe, respectivamente. Deve-se, ainda, ressaltar que em pH < ~ 6,5 estas reações não estariam associadas, uma vez que *cco*, *iro* e *rus* tem correlação negligenciável à negativa baixa com *aiOA*. Além disso, uma observação importante a partir das correlações foi que a reação catalisada por *iro* seria predominante em pH neutro (6,6 - 7,3) e por *cco* em pH > 7,3. O operon *rus* possivelmente não está envolvido nesse processo, uma vez que não foi encontrado em ambientes ricos em As-Fe, e se mostrou pouco relevante em ambientes ricos em As. Estes dados, juntamente com a alta frequência relativa de *arsC*, indicam comunidades dinâmicas em relação ao ciclo do As, sendo capazes de realizar tanto a redução quanto oxidação do metaloide.

Em ambientes anóxicos, a redução do Fe(III) em estruturas cristalinas está associada com a mobilização de As no meio, através da redução do As(V) adsorvido à estrutura (Oremland & Stolz, 2005). O mecanismo de redução do Fe(III) que envolve o gene *omcB* é o melhor caracterizado até o momento. Reyes et al. (2010) observaram que, na bactéria *Schewanella* sp. ANA-3, o cluster *mtr/omc* é inativo em aerobiose. De fato, *omcB* foi mais frequente em ambiente anóxico (SP), mas também foi encontrado em ambientes microóxicos e óxico (GWB). Assim, estudos posteriores devem ser realizados para ampliar o conhecimento sobre a regulação do operon *omc*.

Atualmente discute-se qual seria o mecanismo predominante na mobilização do As(V) em águas subterrâneas por procariotos redutores de As(V) e/ou Fe(III) (Zobrist et al., 2000; Islam et al., 2005, Campbell et al., 2006, Reyes et al., 2010). Uma das hipóteses sugere que a redução de Fe(III) levaria à dissolução de óxidos de Fe contendo o metaloide, tornando-o disponível no meio para a redução por procariotos. DARBs seriam os principais microrganismos implicados nessa reação, porque observa-se que muitas delas também possuem a capacidade de reduzir estruturas cristalinas de Fe(III), como *Geobacter* sp. Outra hipótese que vem ganhando força nos últimos anos é de que a redução do As(V) independente da redução de Fe(III) poderia levar à liberação de As dos minerais sólidos (Oremland & Stolz, 2005, Reyes et al., 2010). Esta última foi corroborada pelo trabalho de Reyes et al (2010), no qual constatou-se que a deleção do *cluster mtr/omc* de *Shewanella* sp. ANA-3 não afeta sua capacidade de redução dissimilatória do As(V), catalisada por ArrA. Assim, a mobilização de As(V) por DARBs independe da redução de Fe(III) (Reyes et al., 2010). No presente estudo, a meta-análise revelou a ausência de *omcB* em ambientes ricos em As-Fe e correlação negligenciável com *arrA* em ambientes ricos em Fe. Isto poderia ser explicado pela independência entre a redução dissimilatória do As(V) e redução do Fe(III) pelo *cluster mtr/omc*, como sugerido por Reyes et al. (2010). Entretanto, em ambientes ricos em As, a correlação positiva alta entre os dois indica que nestas comunidades os processos poderiam estar associados. Uma possível explicação reside na dinâmica de especiação do As e Fe nesses ambientes e a pressão seletiva que ela exerce sobre ARMs e DARBs, e pela capacidade redutora de Fe(III) observada em muitas DARBs. Em ambientes ricos em As, presumivelmente há uma maior mobilização de espécies do metaloide do que em ambientes ricos em As-Fe, onde a presença em abundância de Fe(III) em estrutura cristalina contribuiria para a adsorção e imobilização de As(V). Dessa forma, em ambientes ricos em As, haveria uma maior pressão seletiva para a manutenção tanto de ARMs quanto DARBs. Nesse cenário, esses microrganismos competiriam por substratos energéticos, e a capacidade de obter energia por diferentes vias se tornaria vantajosa. ARMs não utilizam As(V) em seu metabolismo, mas o retiram do meio como forma de detoxificação, e uma vez que são abundantes nas comunidades, esse processo deve ser realizado intensamente. DARBs conhecidas podem utilizar uma grande diversidade de aceptores de elétrons além do As(V), de forma que a redução desta espécie poderia proporcionar um maior *fitness* a estas bactérias, que poderiam assim ocupar diferentes nichos ecológicos. Portanto, dado que a redução dissimilatória do As(V) é uma característica muito vantajosa, mas a disponibilidade de As(V) em meio aquoso pode se tornar limitante, a concomitante capacidade de obtê-lo à partir da dissolução de

estruturas cristalinas de Fe (III) às quais estava adsorvido proporciona um recurso à mais para sobrevivência das DARBs. Assim, conjectura-se que ambientes ricos em As favoreceriam a co-seleção dos genes *arrA* e do operon *omc*.

Um dado chave deste estudo foi que *omcB* manteve uma correlação diferente entre os genes envolvidos na redução do As(V) em cada uma das restritas faixas de pH em que co-ocorrem (ultra e ligeiramente ácido, neutro e ligeiramente alcalino, Figura 5). Além disso, a alta correlação positiva de *omcB* com as enzimas arseniato redutases detoxificadoras em ambientes ricos em Fe indica que ARMs ao invés de DARBs podem ter um papel relevante na mobilização do As neste ambiente. Em síntese, os resultados desta meta-análise sugerem que a mobilização de As em consequência da redução de Fe(III) pelo cluster *mtr/omc* depende do pH e altas concentrações de As e Fe, e que tanto DARBs quanto ARMs podem estar implicadas no processo.

A ausência de determinados citocromos previamente selecionados também fornece informações sobre os principais genes e mecanismos envolvidos na ciclagem do As e Fe em ambientes naturais. Não foram encontrados nenhum dos genes dos operons envolvidos na oxidação fototrófica do Fe(II), *fox* e *pio*. Isso pode ser explicado pela pouca importância do processo para o ciclo global do Fe, como sugerido por Weber et al. (2006). A oxidação fototrófica é limitada pela penetração máxima da luz em solos e sedimentos, que alcança até 200 µm, e pela solubilidade do Fe(II), uma vez que os microrganismos que executam essa reação não são capazes de dissolver estruturas cristalinas (Weber et al., 2006). Além disso, a ausência do citocromo c tetraheme codificado por *cymA* corrobora a hipótese de que *arrA* pode ser, de fato, pouco abundante: em *Shewanella* sp., ele é necessário para a redução dissimilatória do As(V) catalisada por ArrA (Murphy & Saltikov, 2007).

Não é surpreendente o grande número de citocromos c encontrados (Tabela 2) que não haviam sido previamente selecionados. A participação de tais proteínas nas reações redox do Fe é bem documentada na literatura (Tsapin et al., 2001; Shi et al., 2007; Amouric et al., 2011.; Hedrich et al., 2011). Entretanto, a identificação de quais tipos de citocromos c estão diretamente implicados nessas biotransformações é elusiva. Um mesmo gênero bacteriano pode apresentar mecanismos e genes distintos para oxidação do Fe(II), como ocorre em linhagens *Acidithiobacillus* sp. (Amouric et al., 2011.; Hedrich et al., 2011); e a presença de numerosos citocromos c no genoma de *Geobacter sulfurreducens* sugere a existência de múltiplas vias de transporte de elétrons (Weber et al., 2006) para a redução do Fe(III).

## 7. CONCLUSÃO

A meta-análise do presente trabalho revelou que a distribuição ambiental, frequência relativa e correlações entre genes envolvidos nos ciclos do As e Fe acompanham elegantemente a termodinâmica predita para a especiação dos dois elementos, e o pH revelou-se o fator ambiental mais importante para esta dinâmica. Destaca-se a relevância das ARMs para a mobilização do As, inclusive em águas subterrâneas ricas em As e Fe, papel antes atribuído majoritariamente à DARBs, que catalisariam a reação mediante a redução do Fe(III). Além disso, os dados corroboram a ancestralidade do gene *arsC*. Entretanto, a real história evolutiva dos genes de oxirredução do As, assim como o principal mecanismo de redução do As(V) no ambiente, pode estar sendo obscurecida por uma possível sub-representação e anotação incorreta do gene *arrA* nos bancos de dados. Assim como *arsC*, o operon *cco* ocorreu em uma variedade de condições ambientais, conferindo um maior *fitness* às bactérias que o abrigam. A imobilização de As(V) em ambientes óxicos em decorrência da oxidação do Fe(II) parece ser guiada pelos genes *iro* e do operon *cco*. Por fim, estudos futuros devem focar na bioprospecção da diversidade do gene *arrA* no ambiente e em estratégias para aprimorar sua identificação e anotação; e também na identificação de outros tipos de citocromos e proteínas envolvidas na oxirredução do Fe. A diversidade de citocromos c encontrada neste estudo requer investigações posteriores para ampliar o conhecimento destas proteínas envolvidas no metabolismo do Fe.

## 8. REFERÊNCIAS BIBLIOGRÁFICAS

ALTSCHUL, S. F. *et al.* Basic local alignment search tool. **Journal of Molecular Biology**, [s. l.], v. 215, n. 3, p. 403-410, 1990.

ANDRES, J.; BERTIN, P. N. The microbial genomics of arsenic. **FEMS Microbiology Reviews**, [s. l.], v. 39, n. 1, p. 1-24, 2016.

ANDREWS, S. **FastQC**: a quality control tool for high throughput sequence data. [S. l.: s. n.], 2010. Disponível em: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>. Acesso em: 16 mar. 2026.

BERTIN, P. N. *et al.* Metabolic diversity among main microorganisms inside an arsenic-rich ecosystem revealed by meta- and proteo-genomics. **The ISME Journal**, [s. l.], v. 5, n. 12, p. 1735-1747, 2011.

BONNEFOY, V.; HOLMES, D. S. Genomic insights into microbial iron oxidation and iron uptake strategies in extremely acidic environments. **Environmental Microbiology**, [s. l.], v. 14, n. 6, p. 1597-1611, 2012. DOI 10.1111/j.1462-2920.2011.02626.x.

BRUSCELLA, P. *et al.* Differential expression of two bc1 complexes in the strict acidophilic chemolithoautotrophic bacterium *Acidithiobacillus ferrooxidans* suggests a model for their respective roles in iron or sulfur oxidation. **Microbiology**, [s. l.], v. 153, n. 1, p. 102-110, 2007. DOI 10.1099/mic.0.2006/000067-0.

BUTLER, J. E. *et al.* MacA, a Diheme c-Type Cytochrome Involved in Fe(III) Reduction by *Geobacter sulfurreducens*. **Journal of Bacteriology**, [s. l.], v. 186, n. 12, p. 4042-4045, 2004.

CAMPBELL, K. M. *et al.* Simultaneous Microbial Reduction of Iron(III) and Arsenic(V) in Suspensions of Hydrous Ferric Oxide. **Environmental Science and Technology**, [s. l.], v. 40, n. 19, p. 5950-5955, 2006.

CHARRAD, M. *et al.* NbClust: An R Package for Determining the Relevant Number of Clusters in a Data Set. **Journal of Statistical Software**, [s. l.], v. 61, n. 6, p. 1-36, 2014. Disponível em: <http://www.jstatsoft.org/v61/i06/>. Acesso em: 16 mar. 2026.

COSTA, P. S. *et al.* Bacteria and Genes Involved in Arsenic Speciation in Sediment Impacted by Long-Term Gold Mining. **PLoS ONE**, [s. l.], v. 9, n. 4, p. e95655, 2014.

COSTA, P. S. *et al.* Metagenome of a Microbial Community Inhabiting a Metal-Rich Tropical Stream Sediment. **PLoS ONE**, [s. l.], v. 10, n. 3, p. e0119465, 2015.

CROAL, L. S. *et al.* The Genetics of Geochemistry. **Annual Review of Genetics**, [s. l.], v. 38, p. 175-202, 2004.

CROGNALE, S. *et al.* Phylogenetic Structure and Metabolic Properties of Microbial Communities in Arsenic-Rich Waters of Geothermal Origin. **Frontiers in Microbiology**, [s. l.], v. 8, art. 2468, 2017.

DUVAL, S. *et al.* Enzyme phylogenies as markers for the oxidation state of the environment: the case of respiratory arsenate reductase and related enzymes. **BMC Evolutionary Biology**, [s. l.], v. 8, art. 206, 2008.

EDWARDS, R. A. *et al.* Using pyrosequencing to shed light on deep mine microbial ecology. **BMC Genomics**, [s. l.], v. 7, art. 57, 2006. DOI 10.1186/1471-2164-7-57.

FILINOV, A. *et al.* Dynamics of internal waves generated by mountain breeze in Alchichica Crater Lake, Mexico. **Geofísica Internacional**, [s. l.], v. 54, n. 1, p. 21-30, 2015.

GRASBY, S. E. *et al.* The Paint Pots, Kootenay National Park, Canada — a natural acid spring analogue for Mars. **Canadian Journal of Earth Sciences**, [s. l.], v. 50, n. 1, p. 94–108, 2013.

HEDRICH, S.; SCHLÖMANN, M.; JOHNSON, D. B. The iron-oxidizing proteobacteria. **Microbiology**, [s. l.], v. 157, n. 6, p. 1551–1564, 2011.

HEMME, C. L. *et al.* Lateral gene transfer in a heavy metal-contaminated-groundwater microbial community. **mBio**, [s. l.], v. 7, n. 2, e02234-15, 2016. DOI 10.1128/mBio.02234-15.

HYATT, D. *et al.* Prodigal: prokaryotic gene recognition and translation initiation site identification. **BMC Bioinformatics**, [s. l.], v. 11, art. 119, 2010. Disponível em: <http://www.biomedcentral.com/1471-2105/11/119>. Acesso em: 16 mar. 2026.

INGLEDEW, J. W. Thiobacillus ferrooxidans. The bioenergetics of an acidophilic chemolithotroph. **Biochimica et Biophysica Acta (BBA) - Reviews on Bioenergetics**, [s. l.], v. 683, n. 2, p. 89–117, 1982.

INSKEEP, W. P. *et al.* Metagenomes from High-Temperature Chemotrophic Systems Reveal Geochemical Controls on Microbial Community Structure and Function. **PLoS ONE**, [s. l.], v. 5, n. 3, e9773, 2010. DOI 10.1371/journal.pone.0009773.

ISLAM, F. S. *et al.* Role of metal-reducing bacteria in arsenic release from Bengal delta sediments. **Nature**, [s. l.], v. 430, n. 6995, p. 68-71, 2004.

ISLAM, F. S. *et al.* Interactions between the Fe(III)-reducing bacterium *Geobacter sulfurreducens* and arsenate, and capture of the metalloid by biogenic Fe(II). **Applied and Environmental Microbiology**, [s. l.], v. 71, n. 12, p. 8642–8648, 2005.

JACKSON, C. R.; DUGAS, S. L. Phylogenetic analysis of bacterial and archaeal *arsC* gene sequences suggests an ancient, common origin for arsenate reductase. **BMC Evolutionary Biology**, [s. l.], v. 3, art. 18, 2003.

JIAO, Y.; NEWMAN, D. The *pio* Operon Is Essential for Phototrophic Fe(II) Oxidation in *Rhodospseudomonas palustris* TIE-1. **Journal of Bacteriology**, [s. l.], v. 189, n. 5, p. 1765-1773, 2007.

JURAFSKY, D.; MARTIN, J. H. Hidden Markov Models. **In: JURAFSKY, D.; MARTIN, J. H. Speech and Language Processing**. 3. ed. [S. l.]: Pearson Education, 2017. cap. 9. (Draft Edition).

KAPPLER, A.; STRAUB, K. L. Geomicrobiological Cycling of Iron. **Reviews in Mineralogy and Geochemistry**, [s. l.], v. 59, n. 1, p. 85-108, 2005. DOI 10.2138/rmg.2005.59.5.

KOZUBAL, M. A. *et al.* Microbial iron cycling in acidic geothermal springs of Yellowstone National Park: integrating molecular surveys, geochemical processes, and isolation of novel Fe-active microorganisms. **Frontiers in Microbiology**, [s. l.], v. 3, art. 109, 2012.

KRIVUSHIN, K. *et al.* Two metagenomes from Late Pleistocene northeast Siberian permafrost. **Genome Announcements**, [s. l.], v. 3, n. 1, e01380-14, 2015. DOI 10.1128/genomeA.01380-14.

KRUGER, M. C. *et al.* Bacterial metabolism of environmental arsenic — mechanisms and biotechnological applications. **Applied Microbiology and Biotechnology**, [s. l.], v. 97, n. 9, p. 3827–3841, 2013. Disponível em: <https://doi.org/10.1007/s00253-013-4838-5>. Acesso em: 16 mar. 2026.

KURTH, D. *et al.* Arsenic metabolism in high altitude modern stromatolites revealed by metagenomic analysis. **Scientific Reports**, [s. l.], v. 7, art. 1024, 2017.

KUSANO, T. *et al.* Molecular cloning of the gene encoding Thiobacillus ferrooxidans Fe(II) oxidase. High homology of the gene product with HiPIP. **The Journal of Biological Chemistry**, [s. l.], v. 267, n. 16, p. 11242-11247, 1992.

LANGMEAD, B. **De Bruijn Graph assembly**. Baltimore: Johns Hopkins University, 2014. Recurso didático. Disponível em: [https://www.cs.jhu.edu/~langmea/resources/lecture\\_notes/assembly\\_dbg.pdf](https://www.cs.jhu.edu/~langmea/resources/lecture_notes/assembly_dbg.pdf). Acesso em: 16 mar. 2026.

LANGMEAD, B.; SALZBERG, S. L. Fast gapped-read alignment with Bowtie 2. **Nature Methods**, [s. l.], v. 9, n. 4, p. 357–359, 2012. DOI 10.1038/nmeth.1923.

LAYTON, A. C. *et al.* Metagenomes of microbial communities in arsenic- and pathogen-contaminated well and surface water from Bangladesh. **Genome Announcements**, [s. l.], v. 2, n. 6, e01170-14, 2014. DOI 10.1128/genomeA.01170-14.

LEANG, C.; COPPI, M. V.; LOVLEY, D. R. OmcB, a c-type polyhemeryochrome, involved in Fe(III) reduction in *Geobacter sulfurreducens*. **Journal of Bacteriology**, [s. l.], v. 185, n. 7, p. 2096-2103, 2003.

LEBRUN, E. *et al.* Arsenite Oxidase, an Ancient Bioenergetic Enzyme. **Molecular Biology and Evolution**, [s. l.], v. 20, n. 5, p. 686–693, 2003.

LIPSON, D. A. *et al.* Metagenomic Insights into Anaerobic Metabolism along an Arctic Peat Soil Profile. **PLoS ONE**, [s. l.], v. 8, n. 5, e64659, 2013. DOI 10.1371/journal.pone.0064659.

LLOYD, J. R. *et al.* Biochemical and genetic characterization of PpcA, a periplasmic c-type cytochrome in *Geobacter sulfurreducens*. **Biochemical Journal**, [s. l.], v. 369, n. 1, p. 153-161, 2003. DOI 10.1042/bj20020597.

LUO, J. *et al.* Metagenomic Approach Reveals Variation of Microbes with Arsenic and Antimony Metabolism Genes from Highly Contaminated Soil. **PLoS ONE**, [s. l.], v. 9, n. 10, e108185, 2014. DOI 10.1371/journal.pone.0108185.

MAECHLER, M. *et al.* **Cluster**: Cluster Analysis Basics and Extensions. Versão 2.0.6. [S. l.: s. n.], 2017. R package.

MARTIN, P. *et al.* Insights into the structure, solvation, and mechanism of ArsC arsenate reductase, a novel arsenic detoxification enzyme. **Structure**, [s. l.], v. 9, n. 11, p. 1071–1081, 2001.

MEHTA, T. *et al.* Outer Membrane c-Type Cytochromes Required for Fe(III) and Mn(IV) Oxide Reduction in *Geobacter sulfurreducens*. **Applied and Environmental Microbiology**, [s. l.], v. 71, n. 12, p. 8634-8641, 2005.

MUKAKA, M. M. Statistics Corner: A guide to appropriate use of Correlation coefficient in medical research. **Malawi Medical Journal**, [s. l.], v. 24, n. 3, p. 69-71, 2012.

MURPHY, J. N.; SALTIKOV, C. W. The *cymA* Gene, Encoding a Tetraheme c-Type Cytochrome, Is Required for Arsenate Respiration in *Shewanella* Species. **Journal of Bacteriology**, [s. l.], v. 189, n. 6, p. 2283–2290, 2007.

MYERS, C. R.; MYERS, J. M. Cloning and sequence of *cymA*, a gene encoding a tetraheme cytochrome c required for reduction of iron(III), fumarate, and nitrate by *Shewanella putrefaciens* MR-1. **Journal of Bacteriology**, [s. l.], v. 179, n. 4, p. 1143-1152, 1997.

NAMIKI, T. *et al.* MetaVelvet: an extension of Velvet assembler to de novo metagenome assembly from short sequence reads. **Nucleic Acids Research**, [s. l.], v. 40, n. 20, e155, 2012. DOI 10.1093/nar/gks678.

NAYFACH, S.; POLLARD, K. S. Toward Accurate and Quantitative Comparative Metagenomics. **Cell**, [s. l.], v. 166, n. 5, p. 1103-1116, 2016.

OH, S. *et al.* Metagenomic Insights into the Evolution, Function, and Complexity of the Planktonic Microbial Community of Lake Lanier, a Temperate Freshwater Ecosystem. **Applied and Environmental Microbiology**, [s. l.], v. 77, n. 17, p. 6000–6011, 2011.

ORDÓÑEZ, E. *et al.* Arsenate Reductase, Mycothiol, and Mycoredoxin Concert Thiol/Disulfide Exchange. **The Journal of Biological Chemistry**, [s. l.], v. 284, n. 22, p. 15107-15116, 2009.

OREMLAND, R. S.; STOLZ, J. F. The Ecology of Arsenic. **Science**, [s. l.], v. 300, n. 5621, p. 939-949, 2003.

OREMLAND, R. S.; STOLZ, J. F. Arsenic, microbes and contaminated aquifers. **Trends in Microbiology**, [s. l.], v. 13, n. 2, p. 45–49, 2005.

OREMLAND, R. S. *et al.* Arsenic in the evolution of Earth and extraterrestrial ecosystems. **Geomicrobiology Journal**, [s. l.], v. 26, n. 7, p. 522–536, 2009.

PENG, Y. *et al.* IDBA-UD: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. **Bioinformatics**, [s. l.], v. 28, n. 11, p. 1420-1428, 2012. DOI 10.1093/bioinformatics/bts174.

PITCHER, R. S.; WATMOUGH, N. J. The bacterial cytochrome cbb3 oxidases. **Biochimica et Biophysica Acta (BBA) – Bioenergetics**, [s. l.], v. 1655, n. 1, p. 388-399, 2004.

PRUITT, K. D.; TATUSOVA, T.; MAGLOTT, D. R. NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. **Nucleic Acids Research**, [s. l.], v. 35, n. suppl. 1, p. D61-D65, 2007.

QUARTINI, R. *et al.* Insights into the iron and sulfur energetic metabolism of *Acidithiobacillus ferrooxidans* by microarray transcriptome profiling. **Hydrometallurgy**, [s. l.], v. 83, n. 1–4, p. 263-272, 2006.

QUÉMÉNEUR, M. *et al.* Diversity surveys and evolutionary relationships of *aoxB* genes in aerobic arsenite-oxidizing bacteria. **Applied and Environmental Microbiology**, [s. l.], v. 74, n. 14, p. 4567–4573, 2008.

QUINLAN, A. R.; HALL, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. **Bioinformatics**, [s. l.], v. 26, n. 6, p. 841–842, 2010.

R CORE TEAM. **R**: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing, 2013. Disponível em: <http://www.R-project.org/>. Acesso em: 16 mar. 2026.

REIS, M. P. *et al.* Metagenomic signatures of a tropical mining-impacted stream reveal complex microbial and metabolic networks. **Chemosphere**, [s. l.], v. 161, p. 266–273, 2016.

REVELLE, W. **psych**: Procedures for Personality and Psychological Research. Versão 1.7.8. Evanston: Northwestern University, 2017. R package. Disponível em: <https://CRAN.R-project.org/package=psych>. Acesso em: 16 mar. 2026.

REYES, C.; LLOYD, J. R.; SALTIKOV, C. W. Geomicrobiology of iron and arsenic in anoxic sediments. In: \_\_\_\_\_ (org.). **Arsenic Contamination of Groundwater: Mechanisms, Analysis, and Remediation**. Hoboken: John Wiley and Sons, 2008. p. 123-146.

REYES, C.; MURPHY, J. N.; SALTIKOV, C. W. Mutational and gene expression analysis of *mtrDEF*, *omcA* and *mtrCAB* during arsenate and iron reduction in *Shewanella* sp. ANA-3. **Environmental Microbiology**, [s. l.], v. 12, n. 7, p. 1878–1888, 2010. DOI 10.1111/j.1462-2920.2010.02192.x

RICHEY, C. *et al.* Respiratory arsenate reductase as a bidirectional enzyme. **Biochemical and Biophysical Research Communications**, [s. l.], v. 382, n. 2, p. 298-302, 2009.

RODRIGUEZ-R, L. M.; KONSTANTINIDIS, K. T. Estimating coverage in metagenomic data sets and why it matters. **The ISME Journal**, [s. l.], v. 8, n. 11, p. 2349–2351, 2014. DOI 10.1038/ismej.2014.76.

SAGHAÏ, A. *et al.* Comparative metagenomics unveils functions and genome features of microbialite-associated communities along a depth gradient. **Environmental Microbiology**, [s. l.], v. 18, n. 12, p. 4990–5004, 2016. DOI 10.1111/1462-2920.13456.

SALTIKOV, C. W.; NEWMAN, D. K. Genetic identification of a respiratory arsenate reductase. **Proceedings of the National Academy of Sciences of the United States of America (PNAS)**, [s. l.], v. 100, n. 19, p. 10983–10988, 2003.

SANGWAN, N. *et al.* Arsenic rich Himalayan hot spring metagenomics reveal genetically novel predator–prey genotypes. **Environmental Microbiology Reports**, [s. l.], v. 7, n. 6, p. 812–823, 2015. DOI 10.1111/1758-2229.12297.

SEEMANN, T. Prokka: rapid prokaryotic genome annotation. **Bioinformatics**, [s. l.], v. 30, n. 14, p. 2068–2069, 2014. DOI 10.1093/bioinformatics/btu153.

SHARPTON, T. J. An introduction to the analysis of shotgun metagenomic data. **Frontiers in Plant Science**, [s. l.], v. 5, art. 209, 2014. DOI 10.3389/fpls.2014.00209.

SHI, L. *et al.* Respiration of metal (hydr)oxides by *Shewanella* and *Geobacter*: a key role for multihaem c-type cytochromes. **Molecular Microbiology**, [s. l.], v. 65, n. 1, p. 12–20, 2007. DOI 10.1111/j.1365-2958.2007.05783.x.

SMEDLEY, P. L.; KINNIBURGH, D. G. A review of the source, behaviour and distribution of arsenic in natural waters. **Applied Geochemistry**, [s. l.], v. 17, n. 5, p. 517–568, 2002.

SOIL SURVEY DIVISION STAFF. **Soil survey manual**. Washington, DC: U.S. Department of Agriculture, 1993. (Handbook, 18). cap. 3.

SONG, B. *et al.* Molecular methods to detect and monitor dissimilatory arsenate-respiring bacteria (DARB) in sediments. **FEMS Microbiology Ecology**, [s. l.], v. 68, n. 1, p. 108–117, 2009. DOI 10.1111/j.1574-6941.2009.00657.x.

SUHADOLNIK, M. L. S. *et al.* Novel arsenic-transforming bacteria and the diversity of their arsenic-related genes and enzymes arising from arsenic-polluted freshwater sediment. **Scientific Reports**, [s. l.], v. 7, art. 11231, 2017.

THE UNIPROT CONSORTIUM. UniProt: the universal protein knowledgebase. **Nucleic Acids Research**, [s. l.], v. 45, n. D1, p. D158–D169, 2017.

TSAPIN, A. I. *et al.* Identification of a Small Tetraheme Cytochrome c and a Flavocytochrome c as Two of the Principal Soluble Cytochromes c in *Shewanella oneidensis* Strain MR1. **Applied and Environmental Microbiology**, [s. l.], v. 67, n. 7, p. 3236–3244, 2001.

UROZ, S. Functional Assays and Metagenomic Analyses Reveals Differences between the Microbial Communities Inhabiting the Soil Horizons of a Norway Spruce Plantation. **PLoS ONE**, [s. l.], v. 8, n. 2, e55929, 2013. DOI 10.1371/journal.pone.0055929.

VILLADANGOS, F. *et al.* *Corynebacterium glutamicum* survives arsenic stress with arsenate reductases coupled to two distinct redox mechanisms. **Molecular Microbiology**, [s. l.], v. 82, n. 4, p. 998–1014, 2011. DOI 10.1111/j.1365-2958.2011.07882.x.

WEBER, K. A.; ACHENBACH, L. A.; COATES, J. D. Microorganisms pumping iron: Anaerobic microbial iron oxidation and reduction. **Nature Reviews Microbiology**, [s. l.], v. 4, n. 10, p. 752-764, 2006. DOI 10.1038/nrmicro1490.

WORLD HEALTH ORGANIZATION (WHO). **Arsenic in Drinking-water**: Background document for development of WHO Guidelines for Drinking-water Quality. Geneva: WHO, 2011.

WORLD HEALTH ORGANIZATION (WHO). **Fact Sheet**: Arsenic. Geneva: WHO, 2016. Disponível em: <http://www.who.int/mediacentre/factsheets/fs372/en/>. Acesso em: 16 mar. 2026.

WORLD HEALTH ORGANIZATION (WHO). **Iron in Drinking-water**: Background document for development of WHO Guidelines for Drinking-water Quality. Geneva: WHO, 2003.

ZARGAR, K. *et al.* Identification of a novel arsenite oxidase gene, *arxA*, in the haloalkaliphilic, arsenite-oxidizing bacterium *Alkalilimnicola ehrlichii* strain MLHE-1. **Journal of Bacteriology**, [s. l.], v. 192, n. 14, p. 3755–3762, 2010.

ZHBANNIKOV, I. Y. *et al.* SeqyClean: A Pipeline for High-throughput Sequence Data Preprocessing. In: ACM INTERNATIONAL CONFERENCE ON BIOINFORMATICS, COMPUTATIONAL BIOLOGY, AND HEALTH INFORMATICS, 8., 2017, Boston. **Proceedings [...]**. New York: ACM, 2017. p. 407-416.

ZOBRIST, J. *et al.* Mobilization of Arsenite by Dissimilatory Reduction of Adsorbed Arsenate. **Environmental Science and Technology**, [s. l.], v. 34, n. 22, p. 4747-4753, 2000.

**ANEXO 1: Busca por genes envolvidos nas biotransformações do As e Fe nos conjuntos de dados da meta-análise.**

A. Lista de palavras e expressões-chave utilizadas para a busca.

arsA	ccmF	cycA
"3.6.3.16"	ccmH	cydA
arsenical	ccoN	"1.10.3.14"
arsB	ccoP	cydB
arsC	ccoP1	cydX
"1.20.4.1"	ccoP2	cyoA
arsenate	ccp	cyoC
arsC	"1.11.1.5"	cyoD
"3.1.3.48"	ccpA	cyt
"2.8.4.2"	ccsA	iro
arsD	ccsB	"1.16.3"
arsR	coxM	"iron oxidase"
cbaA	coxN	macA
"1.9.3.1"	ctaC	mtrA
cytochrome	ctaD	mtrB
cbaB	ctaE	"2.7.13.3"
cc4	ctaF	omcB
cccA	ctaG	petA
ccmA	cybB	"1.10.2.2"
"3.6.3.41"	cybC	petB
ccmE	cyc1	petC

"1.10.9.1"	mtrF
rus	omcA
"rusticyanin"	omcE
dmsB	omcS
"dimethyl sulfoxide reductase"	pioA
dmsC	pioB
dmsA	pioC
"1.8.5.3"	ppcA
dorA	rusA
"1.7.2.3"	rusB
ynfE	
"1.8.99"	
ynfF	
qcrA	
"Rieske iron-sulfur"	
"Rieske 2Fe-2S iron- sulfur "	
aioA	
aioB	
aroA	
aroB	
aoxB	
arxA	
arx	
arrA	
arrB	
arsenite	
arsenic	
coxA	
coxB	
coxC	
cymA	
foxE	
foxY	
foxZ	
mtrD	
mtrE	

B. Quadro A1: Resultado do alinhamento BLAST de sequências identificadas como *arrA*, após alinhamento contra sequências curadas.

Dataset	Contig ID	Blast hit	E-value	Query Cover	%ID	Accession
ERR342782	contig-43_113:3680-3888	Anaerobic dimethyl sulfoxide reductase chain A, partial [Salmonella enterica subsp. enterica serovar Alachua str. R6-377]	3,00E-40	98	96	EHC43720.1
ERR342782	contig-43_13186:578-807	formate dehydrogenase subunit alpha, partial [Bacteroidetes bacterium HGW-Bacteroidetes-12]	4,00E-35	98	83	PKP49537.1
ERR342782	contig-43_63:1525-1721	anaerobic dimethyl sulfoxide reductase chain a [Salmonella enterica subsp. enterica serovar Cerro FSL R8-0235].	5,00E-37	97	97	KCU93213.1
ERR342782	contig-43_80917:14-184	arsenate respiratory reductase, partial [uncultured bacterium].	3,00E-17	70	97	AIZ48368.1
ERR342783	contig-43_12:25689-25885	anaerobic dimethyl sulfoxide reductase chain a [Salmonella enterica subsp. enterica serovar Cerro FSL R8-0235].	5,00E-37	97	97	KCU93213.1
SRR1573952	contig-101_124417:17-216	molybdopterin oxidoreductase, molybdopterin binding subunit [Burkholderiales bacterium 70-64]	5,00E-18	97	60	OJX08343.1
SRR1573952	contig-101_248807:19-344	dissimilatory arsenate reductase, partial [uncultured bacterium].	2,00E-34	66	78	BAM71901.1
SRR3311145	contig-97_12143:1197-1417	molybdopterin oxidoreductase [Ectothiorhodospira sp. PHS-1]	8,00E-32	98	79	WP_008932021.1

SRR3311145	contig-97_12143:1626-1942	dissimilatory arsenate reductase, partial [uncultured bacterium].	4,00E-51	92	81	AEK98605.1
SRR3311145	contig-97_372862:34-276	dissimilatory arsenate reductase, partial [uncultured bacterium]	6,00E-32	81	86	AEK98605.1
SRR3311145	contig-97_3958:1532-2481	arsenate respiratory reductase, partial [Shewanella sp. HAR-4]	0	87	94	AAU11842.1
SRR3311145	contig-97_3958:3296-3870	dehydrogenase [Shewanella sp. W3-18-1]	7,00E-132	99	99	WP_011790217.1
SRR4072253	contig-101_14679:5-389	dissimilatory arsenate reductase, partial [uncultured bacterium]	9,00E-80	1,00E+02	88	AFM73914.1
SRR4072253	contig-101_4292:1-525	respiratory arsenate reductase catalytic subunit, partial [[Bacillus] selenitireducens MLS10]	2,00E-71	99	67	AAQ19491.1
SRR4072253	contig-101_4292:758-1300	dehydrogenase [Desulfitibacter sp. BRH_c19]	7,00E-85	99	72	KUO49382.1
SRR4072253	contig-101_6529:0-677	arsenate respiratory reductase, partial [Desulfosporosinus sp. Y5]	1,00E-106	89	76	ABB02056.1
SRR4072256	contig-101_17332:0-599	arsenate respiratory reductase, partial [Desulfosporosinus sp. Y5]	3,00E-90	87	74	ABB02056.1
SRR4072256	contig-101_27166:2-425	arsenate respiratory reductase, partial [Desulfosporosinus sp. Y5]	2,00E-78	100	85	ABB02056.1
SRR4072256	contig-101_40654:99-302	arsenate respiratory reductase, partial [Desulfosporosinus sp. Y5]	3,00E-18	75	78	ABB02056.1
SRR4072256	contig-101_5:77719-79145	dehydrogenase [Desulfovibrio sp. S3730MH75].	0.0	99	83	OEU64902.1
SRR4072256	contig-101_5:79474-80162	dehydrogenase [Desulfovibrio sp. S3730MH75]	3,00E-133	99	82	OEU64902.1
SRR4072256	contig-101_5874:203-921	respiratory arsenate reductase [Bacillus boroniphilus JCM 21738]	4,00E-90	73	70	GAE44118.1

## ANEXO 2: Análises estatísticas com resultados inconclusivos

A. Regressão gene arsC

Call: glm(formula = dados5[, 8] ~ dados5[, 5] + dados5[, 7], family = Gamma(link = log))

Deviance Residuals:

Min	1Q	Median	3Q	Max
-1.0587	-0.2835	-0.1053	0.1468	1.0803

Coefficients:

Estimate	Std. Error	t value	Pr(> t )	
(Intercept)	-8.371159	0.586302	-14.278	2.53e-09 ***
dados5[, 5]	-0.067739	0.091252	-0.742	0.4711 #pH
dados5[, 7]	0.017219	0.005801	2.968	0.0109 * #temperatura

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1 (Dispersion parameter for Gamma family taken to be 0.2826963)

Null deviance: 5.9686 on 15 degrees of freedom Residual deviance: 3.3147 on 13 degrees of freedom AIC: -242.38

Number of Fisher Scoring iterations: 8

> r.squaredLR(gama6)

[1] 0.4596613

```
attr("adj.r.squared")
```

```
[1] -1.360352e-07
```

```
> pchisq(3.31, 13, lower.tail=FALSE) #função de adequação do modelo
```

```
[1] 0.9965648 #não rejeitamos H0, então o modelo proposto é mais adequado em relação ao saturado
```

```
> round(exp(cbind(coef = coef(gama6), confint(gama6))),3) Waiting for profiling to be done...
```

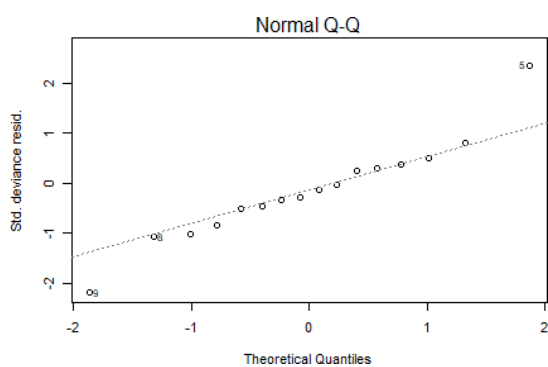
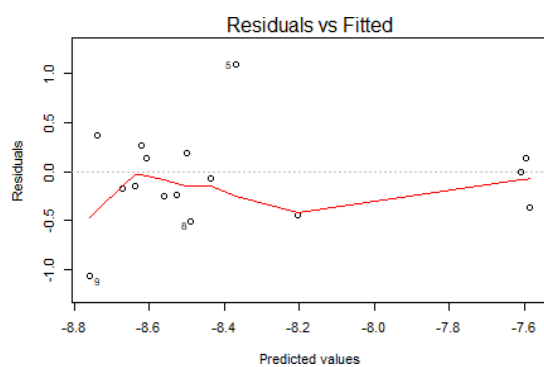
```
coef 2.5 % 97.5 %
```

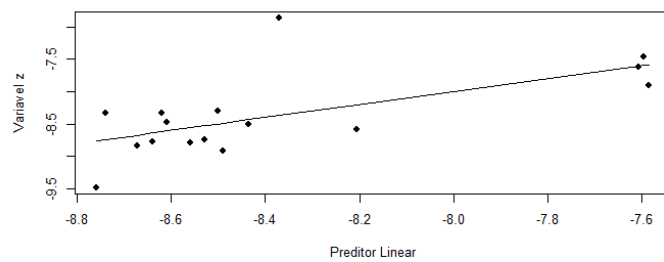
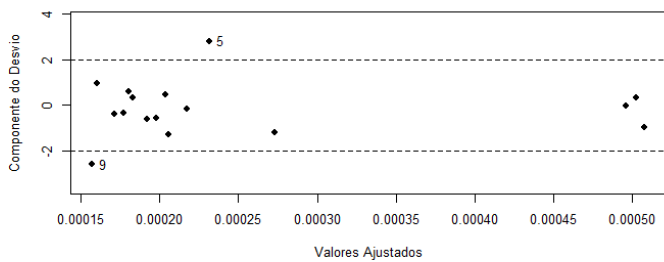
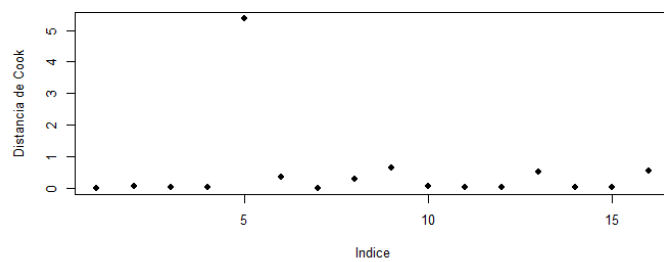
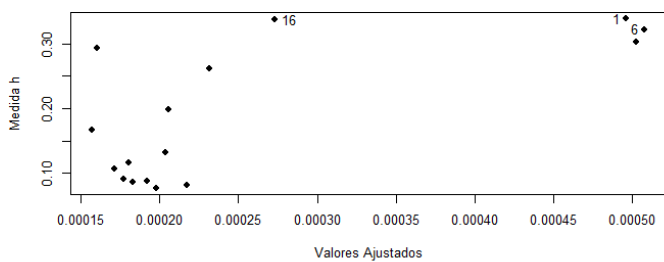
```
(Intercept) 0.000 0.000 0.001
```

```
dados5[, 5] 0.935 0.792 1.096
```

```
dados5[, 7] 1.017 1.006 1.030
```

```
#Análise de resíduos
```





B. Regressão operon *cco*

##Transformação Box-Cox, sem as observações 15 e 18 (temperaturas negativas)

Call: lm(formula = dados2[, 10] ~ dados2[, 3] + dados2[, 5])

Residuals:

Min

1Q Median

3Q

Max -2.24642 -0.48206 -0.08531 0.63604 2.00006

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) -9.782650 0.967737 -10.11 4.33e-08 \*\*\*

dados2[, 3] -0.002919 0.001489 -1.96 0.06882 . [As]

dados2[, 5] 0.490060 0.153622 3.19 0.00609 \*\* [pH] ---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.176 on 15 degrees of freedom

Multiple R-squared: 0.446,

Adjusted R-squared: 0.3722

F-statistic: 6.039 on 2 and 15 DF, p-value: 0.01191

```
shapiro.test(rstudent(lm_boxcox))
```

Shapiro-Wilk normality test

```
data: rstudent(lm_boxcox)
```

W = 0.98385, p-value = 0.9809 #não rejeitamos H0, ou seja, a distribuição é normal.

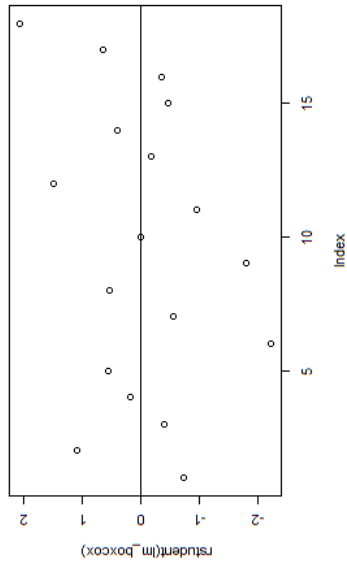
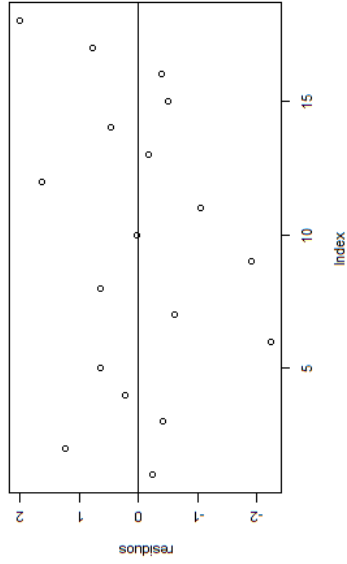
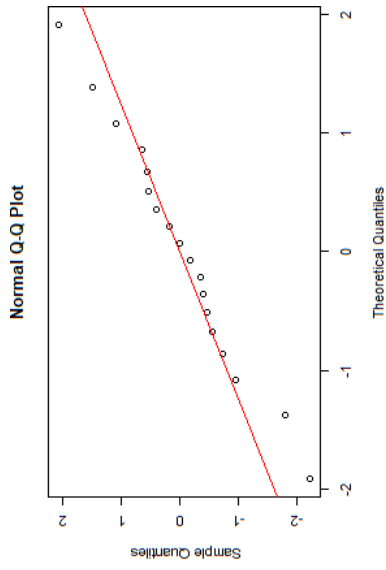
#Ponto de influência: observação n.1 e n.13

```
vif(lm_boxcox)
```

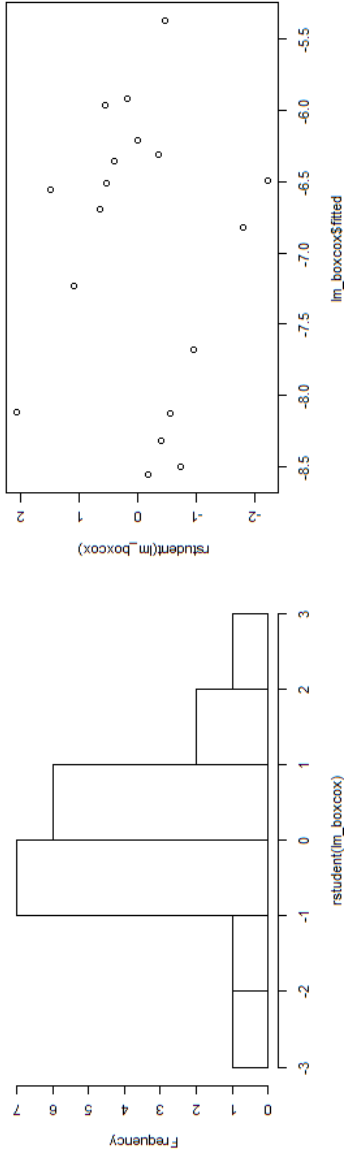
```
dados2[, 3] dados2[, 5]
```

```
1.037475 1.037475
```

#Análise de resíduos



**Histogram of rsstudent(lm\_boxcox)**



C. Regressão gene *aioA*

#### raiz quadrada gama

Call: glm(formula = dados2[, 8] ~ dados2[, 3], family = Gamma(link = log))

Deviance Residuals:

Min

1Q Median

3Q

Max -1.8659 -1.8534 -0.3572 0.4090 2.0633

Coefficients:

Estimate Std. Error t value Pr(&gt;|t|)

(Intercept) -5.387310 0.299767 -17.972 6.05e-13 \*\*\*

dados2[, 3] -0.003203 0.001563 -2.049 0.0553 . #[As] ---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for Gamma family taken to be 1.602164)

Null deviance: 38.430 on 19 degrees of freedom

Residual deviance: 35.094 on 18 degrees of freedom

AIC: -179.22

Number of Fisher Scoring iterations: 7

&gt; r.squaredLR(gama2)

[1] 0.1076173

attr("adj.r.squared")

[1] -1.14675e-05

&gt;pchisq(35, 18, lower.tail=FALSE)

[1] 0.009452403

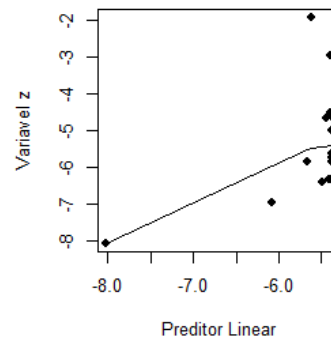
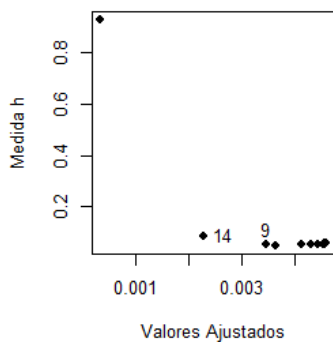
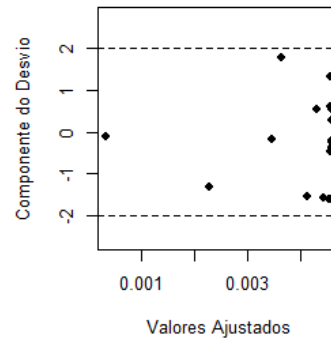
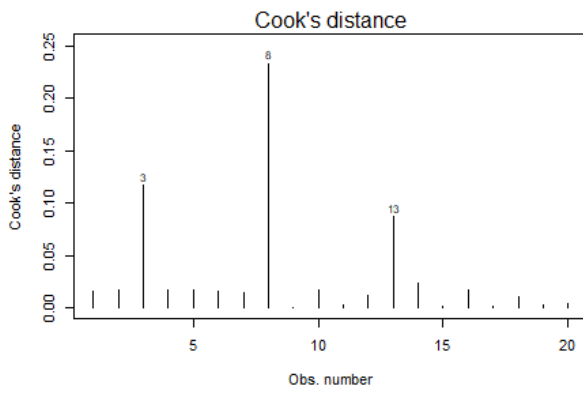
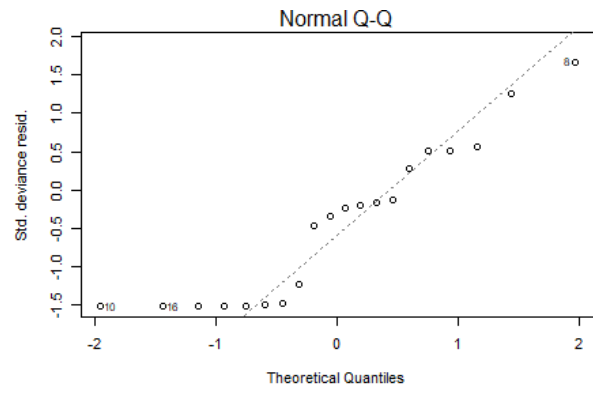
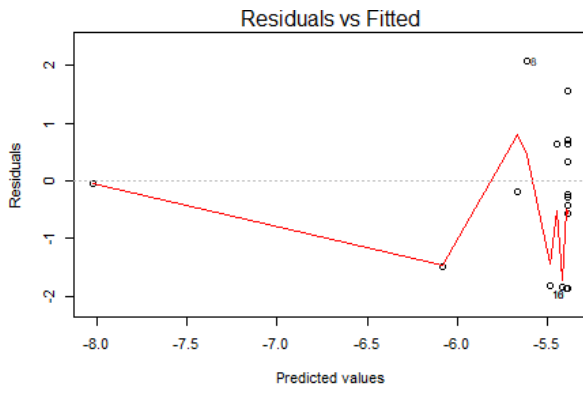
Waiting for profiling to be done...

coef 2.5 % 97.5 %

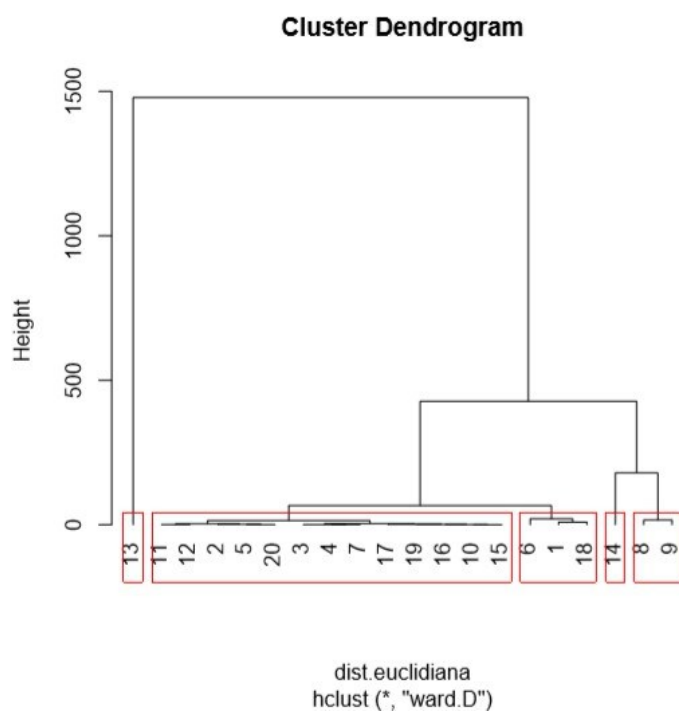
(Intercept) 0.005 0.003 0.009

dados1[, 3] 0.997 0.994 1.002

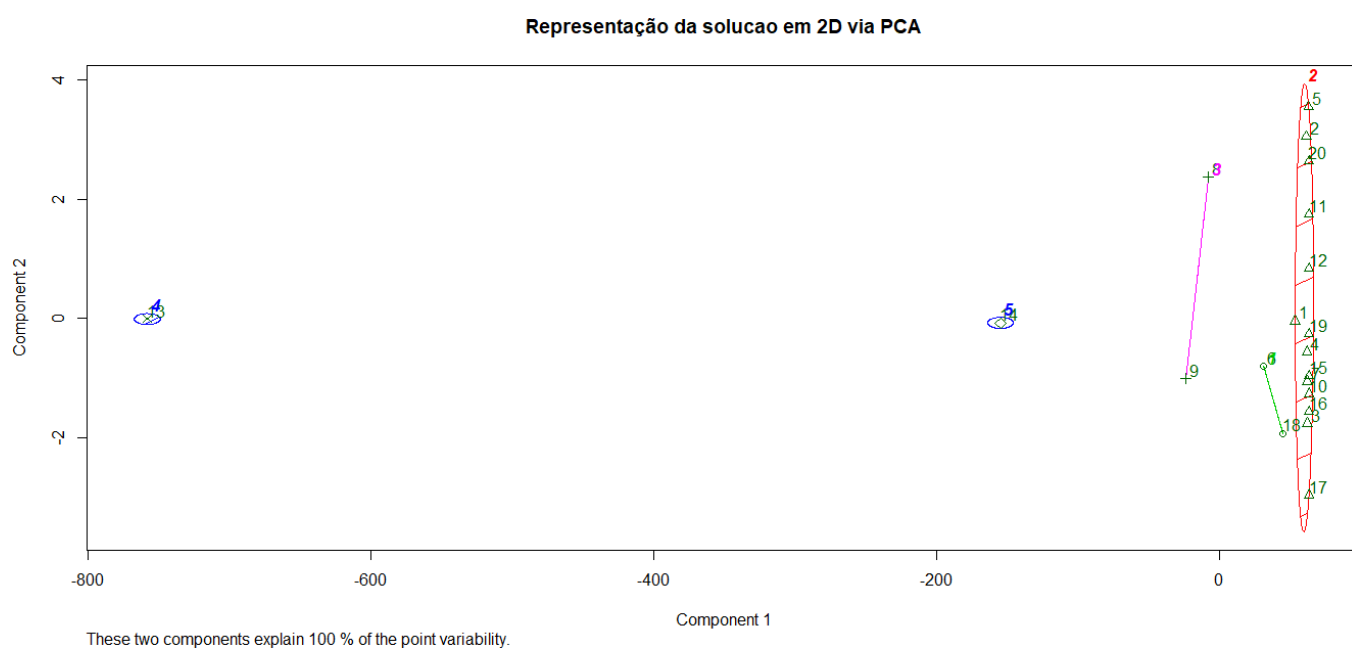
### #Análise de resíduos



## D. Análise de cluster



## 1. Dendrograma da solução hierárquica.



## 2. PCA da solução por k means.

```

> fuzzymedias
Fuzzy Clustering object of class 'fanny' :
m.ship.expon.      2
objective          31.05636
tolerance          1e-15
iterations         31
converged          1
maxit              500
n                  20
Membership coefficients (in %, rounded):
  [,1] [,2] [,3] [,4] [,5]
[1,]  49  43   5   0   2
[2,]  14  82   3   0   1
[3,]   7  91   1   0   0
[4,]   5  93   1   0   0
[5,]  13  83   3   0   1
[6,]  70  16  11   1   3
[7,]   5  94   1   0   0
[8,]   8   5  84   0   2
[9,]   6   4  87   0   3
[10,]  4  94   1   0   0
[11,]  7  91   2   0   1
[12,]  5  93   1   0   0
[13,]  0   0   0 100   0
[14,]  0   0   0   0 100
[15,]  4  95   1   0   0
[16,]  5  93   1   0   0
[17,] 10  86   2   0   1
[18,] 87   9   3   0   1
[19,]  4  95   1   0   0
[20,] 10  87   2   0   1
Fuzzyness coefficients:
dunn_coeff normalized
0.8009693 0.7512116
Closest hard clustering:
[1] 1 2 2 2 2 1 2 3 3 2 2 2 4 5 2 2 2 1 2 2

```

3. Solução por fuzzy medias. Probabilidade de alocação das observações em cada um dos grupos.