

PONTIFICIA UNIVERSIDAD CATÓLICA DEL ECUADOR

FACULTAD DE HÁBITAT, INFRAESTRUCTURA Y CREATIVIDAD

TÍTULO DE MÁSTER EN BIOLOGÍA COMPUTACIONAL

**Análisis genómico y transcriptómico de cepas del género
Acidithiobacillus implicadas en la biolixiviación de metales
pesados**

Autor: Alexis David Pavón Catani

Director: Dr. Alan Cervantes

Quito, Ecuador

2025



Dedicatoria

A mi amada familia, hija y esposa,

Le dedico esta disertación con un profundo sentido de gratitud y aprecio por su amor, apoyo y aliento inquebrantables a lo largo de mi viaje académico. Tu presencia constante en mi vida ha sido una fuente de inmensa fuerza e inspiración, y estoy profundamente agradecida por todo lo que has hecho por mí.

A mi familia, gracias por creer siempre en mí y por brindarme las bases para perseguir mis sueños. Su amor, orientación y aliento han sido invaluable y no habría podido lograr este hito sin su apoyo inquebrantable.

Para mi hija, has sido mi motivación para trabajar más duro y luchar por la excelencia. Espero que esta disertación te inspire a perseguir tus propios sueños y a apuntar siempre alto en la vida.

A mi esposa, gracias por ser mi roca y mi apoyo constante. Su amor y paciencia me han sostenido a través de los desafíos académicos, y su creencia inquebrantable en mis habilidades ha sido una fuente constante de fortaleza. Estoy profundamente agradecido por su amor y compañía, y no podría haber pedido un compañero más solidario en la vida.

Dedico esta tesis a todos ustedes con profundo amor y aprecio. Gracias por todo lo que has hecho por mí, y por estar siempre ahí para mí en las buenas y en las malas. Esta tesis es un testimonio de nuestros esfuerzos colectivos y espero que traiga orgullo y alegría a cada uno de ustedes.

Con todo mi amor y gratitud,

Alexis Pavón



Agradecimientos

Me gustaría expresar mi más profundo agradecimiento al Dr. Alan Cervantes, quien ha sido mi tutor a lo largo de este proceso académico. Su guía, apoyo y orientación han sido fundamentales para el desarrollo de mi proyecto de investigación y para mi crecimiento como científico. Su compromiso con la excelencia y su dedicación a la formación de sus estudiantes han representado una fuente constante de motivación e inspiración.

Asimismo, quiero extender mi sincero agradecimiento a la coordinación de la Maestría en Biología Computacional, cuyo respaldo y acompañamiento hicieron posible llevar a cabo esta etapa de formación académica. La organización, el compromiso y la disposición de la coordinación han sido elementos clave para el buen desarrollo de mis estudios.

De igual manera, agradezco a todos los profesores y docentes que con sus conocimientos, conferencias y orientación enriquecieron mi formación, así como a mis compañeros y colegas, quienes con sus aportes y camaradería hicieron de este camino una experiencia más significativa.

Finalmente, expreso mi gratitud a mi familia y amigos por su apoyo incondicional y aliento durante todo este proceso. Su paciencia, comprensión y amor han sido pilares esenciales para mantenerme motivado y enfocado en alcanzar mis metas.

Gracias a todos por su apoyo, orientación y aliento. Este proyecto de investigación no hubiera sido posible sin sus contribuciones, y estoy profundamente agradecido por todo lo que ha hecho por mí.

Índice

1. RESUMEN	8
2. ABSTRACT	9
3. INTRODUCCIÓN.....	10
3.1 Contexto y antecedentes de la biolixiviación de metales pesados	10
3.2 Taxonomía y características del género <i>Acidithiobacillus</i>	13
3.3 Estudios previos en <i>Acidithiobacillus</i> relacionados con metales	15
4. DESCRIPCIÓN DEL PROBLEMA	19
5. OBJETIVOS.....	21
5.1 General.....	21
5.2 Específicos	21
6. JUSTIFICACIÓN	22
7. MATERIALES Y MÉTODOS.....	27
7.1 Procedencia y disponibilidad de datos (genomas y RNA-seq)	28
7.2 Genomas de referencia.....	28
7.3 Conjuntos de RNA-seq y metadatos.....	28
7.4 Control de calidad de lecturas (FastQC, MultiQC).....	29
7.5 Preprocesamiento (trimming con Trimmomatic).....	31
7.6 Alineamiento (HISAT2).....	32
7.7 Ensamblaje y cuantificación (StringTie, featureCounts).....	33
7.8 Expresión diferencial (DESeq2).....	34
7.9 Ortología y anotación funcional.....	35
7.10 Enriquecimiento funcional.....	36
7.11 Análisis filogenético (recA; NJ-BLOSUM62; bootstrap).....	37
7.12 Visualización y análisis estadístico en R.....	38
7.13 Reproducibilidad (Galaxy/R Markdown/Jupyter).....	39
7.14 Disponibilidad de datos y accesiones	40
8. RESULTADOS	43
8.1 Análisis filogenético del género <i>Acidithiobacillus</i>	43
8.2 Ortogrupos exclusivos por especie.....	46
8.3 Duplicaciones génicas internas	49
8.4 Núcleo de genes compartido entre especies	49
8.5 Control de calidad y mapeo de lecturas.....	51



8.6	Panorama de expresión global.....	51
8.7	Genes diferencialmente expresados: resumen.....	54
8.8	Enriquecimiento funcional: visión general.....	56
8.9	Transcriptómica en <i>A. ferrivorans</i> (6 °C vs 28 °C)	57
8.10	Transcriptómica en <i>A. ferrooxidans</i> (electro vs quimio)	64
9.	DISCUSIÓN	73
10.	CONCLUSIONES	83
11.	RECOMENDACIONES	79
12.	LIMITACIONES DEL ESTUDIO	86
13.	REFERENCIAS BIBLIOGRÁFICAS	88

Índice de figuras

Figura 1.	Árbol filogenético de proteínas (recA) para diez especies de <i>Acidithiobacillus</i> y un outgroup (<i>Thermithiobacillus tepidarius</i>).....	40
Figura 2.	Ortogrupos exclusivos por especie de <i>Acidithiobacillus</i>	42
Figura 3.	Ortogrupos con duplicaciones génicas internas por especie	44
Figura 4.	Ortogrupos compartidos entre especies de <i>Acidithiobacillus</i>	46
Figura 5.	Heatmap de los 50 genes más significativos en <i>A. ferrivorans</i> (6 °C vs 28 °C) ...	55
Figura 6.	Análisis de componentes principales (PCA) en <i>A. ferrivorans</i>	57
Figura 7.	MA-plot del análisis diferencial en <i>A. ferrivorans</i> (6 °C vs 28 °C)	58
Figura 8.	Volcano plot de la expresión diferencial en <i>A. ferrivorans</i> (6 °C vs 28 °C)	59
Figura 9.	Heatmap de los 50 genes más significativos en <i>A. ferrooxidans</i> (electro vs quimio)	60
Figura 10.	PCA de las expresiones génicas en <i>A. ferrooxidans</i>	61
Figura 11.	MA-plot del análisis diferencial en <i>A. ferrooxidans</i> (electro vs quimio)	65
Figura 12.	Volcano plot del análisis diferencial en <i>A. ferrooxidans</i> (electro vs quimio)	66

Índice de tablas

Tabla 1.	Cepas del género <i>Acidithiobacillus</i> , clasificación y rasgos bioquímicos de interés para biolixiviación.....	18
Tabla 2.	Genes diana representativos del metabolismo quimiolitotrófico y tolerancia a metales/ácido en <i>A. ferrooxidans</i>	25
Tabla 3.	Genomas de referencia utilizados	28
Tabla 4.	Conjuntos de RNA-seq y metadatos	28
Tabla 5.	Ortogrupos compartidos, exclusivos y duplicados en cuatro especies de	



Acidithiobacillus.....51

Tabla 6. Mejores alineamientos BLASTP de genes de interés en cuatro especies de *Acidithiobacillus* 56-57

1. RESUMEN

Este trabajo integró análisis genómicos y transcriptómicos de cepas de *Acidithiobacillus* con el objetivo de comprender los mecanismos moleculares que sostienen su capacidad de biolixiviar minerales metálicos. En primer lugar, el análisis genómico comparativo de *A. ferrooxidans*, *A. thiooxidans*, *A. ferrivorans* y *A. caldus* permitió identificar genes y operones clave implicados en rutas centrales de biolixiviación, como los asociados a la oxidación de hierro (rutas operón, *cyc2*), a la oxidación de compuestos reducidos de azufre (*sqr*, *sox*, *tetH*), y a sistemas de tolerancia a metales pesados y condiciones extremas (*copA*, *merA*, *arsC*, entre otros). Estos hallazgos se enriquecieron mediante búsquedas de ortólogos con OrthoFinder, las cuales evidenciaron genes presentes en las cuatro especies, así como duplicaciones y ausencias específicas que sugieren trayectorias evolutivas divergentes y posibles adaptaciones ecológicas.

Posteriormente, el análisis transcriptómico se enfocó en la comparación de la expresión génica bajo condiciones contrastantes, con especial énfasis en *A. ferrivorans*. Se observó regulación diferencial de genes clave en rutas de oxidación de azufre y hierro, mecanismos de estrés frente a pH extremadamente ácido y presencia de metales pesados, así como en la formación de biopelículas. Estos patrones sugieren adaptaciones moleculares que favorecen la eficiencia de la biolixiviación en distintos ambientes, particularmente en relación con la temperatura y disponibilidad de nutrientes.

En conjunto, los resultados proporcionan un panorama integral: el análisis genómico define el repertorio de genes y su conservación evolutiva entre especies, mientras que el transcriptómico revela su activación diferencial frente a condiciones específicas. Esta doble aproximación contribuye al entendimiento del papel de *Acidithiobacillus* en la solubilización de metales y abre la posibilidad de diseñar estrategias más informadas para la optimización de procesos de biolixiviación. No obstante, se reconocen limitaciones asociadas al número de cepas estudiadas, la dependencia de datos públicos y la necesidad de validar experimentalmente los hallazgos. Futuras investigaciones podrán integrar otras aproximaciones ómicas (proteómica, metabolómica) para alcanzar una visión aún más completa del metabolismo y regulación de estas bacterias.

Palabras clave: *Acidithiobacillus*; genómica comparativa; transcriptómica; biolixiviación; expresión génica; adaptación molecular.

2. ABSTRACT

This study integrated genomic and transcriptomic analyses of *Acidithiobacillus* strains to elucidate the molecular mechanisms underlying their ability to bioleach metallic minerals. The comparative genomic analysis of *A. ferrooxidans*, *A. thiooxidans*, *A. ferrivorans*, and *A. caldus* identified key genes and operons involved in bioleaching pathways, including those related to iron oxidation (*rus* operon, *cyc2*), reduced sulfur oxidation (*sqr*, *sox*, *tetH*), and systems conferring resistance to heavy metals and extreme conditions (*copA*, *merA*, *arsC*, among others). Ortholog searches with OrthoFinder further revealed both conserved genes across all four species and species-specific duplications or absences, suggesting divergent evolutionary trajectories and ecological adaptations.

The transcriptomic analysis, primarily focused on *A. ferrivorans*, highlighted differential expression of genes involved in sulfur and iron oxidation, stress responses to acidic pH and heavy metals, and biofilm formation. These expression patterns suggest molecular adaptations enhancing bioleaching efficiency under diverse environmental conditions, particularly regarding temperature shifts and nutrient availability.

Together, the results provide an integrated view: the genomic analysis outlines the gene repertoire and its evolutionary conservation, while transcriptomic data reveal how these genes are differentially activated under specific conditions. This dual approach contributes to a deeper understanding of the role of *Acidithiobacillus* in metal solubilization and offers insights for optimizing bioleaching processes. However, limitations include the restricted number of strains studied, reliance on public datasets, and the need for experimental validation. Future research should incorporate additional omics approaches (proteomics, metabolomics) to achieve a more comprehensive understanding of the metabolic and regulatory networks of these acidophilic bacteria.

3. INTRODUCCIÓN

Acidithiobacillus es un género de bacterias extremófilas ácido-resistentes ampliamente reconocido por su papel en la biolixiviación de metales. Estas quimiolitotrófas obtienen energía oxidando hierro ferroso y compuestos reducidos de azufre, generando sulfato férrico y ácido sulfúrico, agentes clave para solubilizar metales pesados de minerales y residuos mineros (Tonietti *et al.*, 2024). Gracias a este metabolismo, especies como *A. ferrooxidans* y *A. thiooxidans* son capaces de movilizar metales preciosos como el oro y la plata desde matrices minerales refractarias, liberándolos de sulfuros asociados (Pourhossein & Mousavi, 2023). De hecho, la biooxidación de sulfuros (p.ej., pirita y arsenopirita) por *Acidithiobacillus* se emplea industrialmente como pretratamiento de concentrados auríferos refractarios, logrando la ruptura de la matriz sulfurada que encapsula al oro, el cual posteriormente puede extraerse con métodos convencionales más ecológicos (Işildar *et al.*, 2016). En el caso de la plata, si bien su solubilización directa puede verse obstaculizada por la formación de precipitados (como argentojarosita) y por la toxicidad del ion Ag^+ , estudios recientes han explorado estrategias biohidrometalúrgicas para recuperarla. Por ejemplo, Janáková *et al.* (2023) aplicaron *A. ferrooxidans* en la biolixiviación de relaves de minería argentífera, demostrando que la adición controlada de iones plata puede catalizar la disolución de sulfuros y mejorar la liberación de metales acompañantes. Estas investigaciones refuerzan el potencial de la biolixiviación como alternativa sostenible frente a reactivos químicos tradicionales (como el cianuro), disminuyendo el impacto ambiental en la recuperación de oro y plata (Pourhossein & Mousavi, 2023).

En la última década, el auge de las tecnologías “ómicas” ha permitido profundizar en los mecanismos genéticos y fisiológicos que subyacen a la extraordinaria capacidad lixivante de *Acidithiobacillus*. La secuenciación de genomas completos ha revelado que estas bacterias poseen repertorios génicos especializados para sobrevivir en entornos metalíferos extremos, incluyendo múltiples sistemas de resistencia a metales pesados (como bombas de eflujo y enzimas detoxificantes para cobre, arsénico, cadmio o mercurio) junto con los genes clásicos de oxidación de hierro y azufre (Fan *et al.*, 2018; Tonietti *et al.*, 2024). Por ejemplo, el análisis comparativo del genoma de *A. ferrooxidans* muestra operones dedicados a la tolerancia al arsénico, mercurio y otros metales, lo que explica su resiliencia en procesos de biolixiviación de minerales complejos. Paralelamente, estudios transcriptómicos han comenzado a esclarecer la respuesta adaptativa de estas bacterias durante la lixiviación. Liu *et al.* (2025) documentaron, mediante qRT-PCR y análisis genómico, la activación coordinada de un operón arsénico (regulado por ArsR) en *Acidithiobacillus ferriphilus* QBS3 al biooxidar arsenopirita bajo condiciones extremadamente ácidas. Esta cepa mostró tolerancia excepcional a concentraciones de hasta 80 mM de As(III) gracias a la expresión de genes *arsC*, *arsB*, *arsA* y relacionados, lo que le permitió lixiviar completamente el mineral y liberar el oro asociado pese a la elevada carga tóxica de arsénico. Del mismo modo, Fan *et al.* (2018) compararon la expresión génica de *A. ferriphilus* en presencia de diferentes sustratos (Fe^{2+} vs. azufre vs. sulfuros minerales), identificando la inducción significativa de genes clave del metabolismo del azufre (por ejemplo, del sistema Sox y la oxidasa terminal *bo_3*) cuando la bacteria oxida compuestos sulfurados. Estos hallazgos sugieren que *Acidithiobacillus* ajusta finamente su maquinaria bioquímica según el sustrato, intensificando la producción de ácido y oxidantes (Fe^{3+} , sulfatos) para lograr una solubilización eficiente de los metales objetivo.

En conjunto, la convergencia de evidencias experimentales y enfoques genómicos/transcriptómicos confirma el papel central de *Acidithiobacillus* en la biohidrometalurgia de metales preciosos. Los estudios recientes no solo han optimizado procesos aplicados –por ejemplo, recuperando oro de e-waste mediante biotiosulfato generado por *A. thiooxidans* (Pourhossein & Mousavi, 2023) o mejorando la lixiviación de sulfuros argentíferos– sino que también han comenzado a explicar a nivel molecular cómo estas bacterias prosperan y actúan en ambientes cargados de metales. Comprender la base genética de la oxidación mineral y de la resistencia a metales pesados en *Acidithiobacillus* permite diseñar estrategias bioleaching más eficientes y predecir su desempeño en diferentes contextos (Liu *et al.*, 2025; Işıldar *et al.*, 2016). En este sentido, el enfoque integrador de genómica y transcriptómica se perfila como esencial para conectar la fisiología microbiana con la optimización de la biolixiviación de oro, plata y otros metales, allanando el camino hacia tecnologías mineras más limpias y sustentables.

La contaminación por metales pesados constituye uno de los problemas ambientales más graves y extendidos a nivel global. Elementos como el plomo, el cadmio, el mercurio y el arsénico, provenientes principalmente de actividades industriales y mineras, representan una amenaza significativa para la salud humana y los ecosistemas (Jaishankar *et al.*, 2014). Estos metales son altamente tóxicos incluso a bajas concentraciones y tienden a acumularse en los organismos vivos, provocando efectos adversos crónicos (Tchounwou *et al.*, 2012). Las consecuencias para la salud incluyen daños neurológicos, renales, trastornos endocrinos e incluso cáncer, dependiendo del elemento específico y del nivel de exposición (Jaishankar *et al.*, 2014; Tchounwou *et al.*, 2012).

La presencia excesiva de metales pesados en suelos y cuerpos de agua degrada la calidad de estos recursos, comprometiendo su uso para la agricultura, el consumo humano y la vida silvestre

acuática. La remediación de suelos y aguas contaminadas con metales pesados es, por lo tanto, una tarea urgente y desafiante (Gaur & Adholeya, 2004). En el contexto global, la minería de metales como el cobre, el plomo, el zinc, el oro y la plata ha contribuido significativamente a esta problemática al generar residuos (relaves) y drenajes ácidos con altas concentraciones de elementos tóxicos. Por ejemplo, se han documentado regiones mineras con concentraciones de arsénico, plomo y mercurio muy por encima de los límites seguros, generando impactos severos en la biota acuática y en la salud humana. En Ecuador, la expansión de la minería metálica tanto a gran escala como artesanal ha resultado en casos preocupantes de contaminación. En la cuenca del río Puyango, al sur del país, donde operan numerosas plantas de procesamiento de oro, se han encontrado sedimentos con niveles elevados de mercurio, cobre, plomo, zinc y otros metales, excediendo las normas internacionales de calidad ambiental (Mora et al., 2016). Esta contaminación minera amenaza los ecosistemas locales y supone riesgos para la salud de las comunidades que dependen de esas aguas.

Frente a la contaminación por metales pesados, se han explorado diversas estrategias de remediación. La biorremediación se destaca como un enfoque sostenible, definido por el uso de organismos vivos (microorganismos, plantas o enzimas) para remover o neutralizar contaminantes del entorno. En el caso de metales pesados, la biorremediación abarca técnicas como la fitorremediación (uso de plantas que acumulan o estabilizan metales en sus tejidos), la micorremediación (empleo de hongos para adsorber o precipitar metales) y la bioestimulación microbiana (estimulación de bacterias nativas para que inmovilicen o transformen los metales). Estas estrategias biológicas suelen ser más ecológicas y menos invasivas que los métodos físico-químicos tradicionales, ya que evitan el uso de reactivos agresivos y en muchos casos pueden aplicarse in situ (Gaur & Adholeya, 2004). No obstante, su efectividad depende en gran medida de

la selección adecuada de los organismos y de la comprensión de los procesos metabólicos involucrados en la interacción con los metales contaminantes.

Dentro de las diversas modalidades de biorremediación, la biolixiviación es un proceso biotecnológico prometedor para abordar la contaminación por metales pesados y, simultáneamente, recuperar metales de interés económico. La biolixiviación utiliza microorganismos para movilizar y solubilizar metales presentes en minerales o residuos, facilitando su extracción y recuperación (Rawlings & Silver, 1995). Este proceso forma parte de la biohidrometalurgia y puede aplicarse tanto para limpiar ambientes contaminados (por ejemplo, estabilizando metales tóxicos en formas solubles para su posterior retirada) como en minería para extraer metales de menas de baja ley. Comparado con los métodos tradicionales de tratamiento de minerales (como la tostación o la lixiviación química intensa), la biolixiviación es más económica y genera un menor impacto ambiental, al prescindir de altas temperaturas y reactivos químicos peligrosos. En particular, la biolixiviación bacteriana ofrece ventajas notables: reduce la necesidad de reactivos químicos costosos y tóxicos, disminuye la generación de residuos peligrosos secundarios y puede operar bajo una variedad de condiciones ambientales (Brierley & Brierley, 2013). Además, este proceso puede implementarse in situ directamente sobre los materiales contaminados o las pilas de mineral, minimizando así la manipulación y el transporte de grandes volúmenes de tierra o roca (Natarajan & Ting, 2014). Todos estos factores hacen de la biolixiviación una alternativa ambientalmente más sustentable en comparación con las técnicas físico-químicas convencionales de remediación o extracción.

La biolixiviación involucra principalmente microorganismos quimiolitotróficos acidófilos capaces de oxidar compuestos reducidos de azufre y hierro presentes en los minerales.

Tradicionalmente, los géneros *Acidithiobacillus* (por ejemplo, *A. ferrooxidans*), *Leptospirillum* y *Sulfobacillus* han sido estudiados como agentes clave de la biolixiviación. Sin embargo, no son los únicos: se ha documentado la participación de arqueas extremófilas (por ejemplo, especies de *Ferroplasma*, *Acidiplasma*, *Metalosphaera* o *Sulfolobus*) en drenajes ácidos de mina y en operaciones de biolixiviación a altas temperaturas. Estas arqueas, al igual que sus contrapartes bacterianas, pueden oxidar minerales sulfurados y contribuir a la solubilización de metales en condiciones de pH muy bajo y temperaturas elevadas, ampliando el rango de ambientes donde la biolixiviación es efectiva. Asimismo, algunos hongos saprófitos han mostrado potencial en procesos de biolixiviación: especies de los géneros *Aspergillus* y *Penicillium*, por ejemplo, son capaces de disolver metales como cobre, estaño, níquel o plomo en porcentajes significativos (en algunos casos superiores al 65%) mediante la producción de ácidos orgánicos (p. ej., ácido cítrico) que atacan los minerales y liberan los metales. La contribución de estos organismos no tradicionales sugiere que el repertorio de herramientas biológicas para la extracción y remediación de metales es amplio, y que una comprensión profunda de sus mecanismos podría impulsar innovaciones en la biohidrometalurgia.

A pesar de sus ventajas, la aplicación eficaz de la biolixiviación exige comprender en detalle los fundamentos genéticos y metabólicos que subyacen a la capacidad de los microorganismos para lixiviar metales pesados. Diferentes cepas bacterianas (incluso dentro del mismo género) pueden exhibir eficiencias y afinidades distintas para la solubilización de ciertos metales, dependiendo de su dotación genética y de cómo regulan la expresión de esos genes en respuesta al ambiente. Las diferencias y similitudes genómicas entre las cepas biolixiviantes resultan cruciales para optimizar su eficiencia y selectividad bajo diversas condiciones ambientales (Li et al., 2017). Por ello, la comparación de sus genomas puede revelar qué genes o rutas metabólicas explican por qué una

cepa resulta más efectiva lixiviando, por ejemplo, cobre, mientras que otra prefiere compuestos de azufre o tolera concentraciones mayores de arsénico.

Un ámbito de aplicación relevante de la biolixiviación es el tratamiento de minerales refractarios que contienen metales preciosos. Se denomina minerales refractarios a aquellos en los que el oro y la plata se encuentran encapsulados o químicamente asociados a fases mineralógicas (como sulfuros de hierro o arsenopirita) que impiden su liberación mediante técnicas convencionales de lixiviación (por ejemplo, la cianuración). En estos casos, la presencia de ciertos compuestos “refractarios” interfiere con las reacciones químicas necesarias para solubilizar los metales preciosos (Smith, 2019). Dichos compuestos incluyen sulfuros metálicos, arseniuros y otros minerales acompañantes del oro y la plata, los cuales reaccionan preferencialmente con los agentes lixiviantes o encapsulan físicamente a los metales, reduciendo su recuperación. Para superar estas dificultades, los minerales refractarios suelen requerir pretratamientos antes de la lixiviación final con cianuro u otro agente. Entre los pretratamientos comunes se encuentran la oxidación a alta presión (en autoclave, con oxígeno o aire, para oxidar sulfuros), la tostación oxidante (calentamiento del mineral en presencia de oxígeno), la oxidación química con agentes como permanganato u ozono, e incluso la biolixiviación como etapa previa (Jones & Wang, 2020). Todos estos procesos ayudan a romper la matriz mineral que atrapa al oro y la plata, oxidando los sulfuros y removiendo compuestos interferentes, lo cual libera a los metales preciosos haciéndolos accesibles al agente lixivante en etapas posteriores.

La refractariedad de los minerales puede deberse, por ejemplo, a la presencia de pirita (un sulfuro de hierro) que actúa como “trampa” química para el oro: durante la cianuración, la pirita puede consumir el agente lixivante o formar capas pasivantes que inhiben la disolución del oro

(Brown et al., 2018). Asimismo, una elevada concentración de otros metales base, como cobre o hierro, en el mineral puede reducir la eficacia de la lixiviación, ya que estos elementos compiten por el agente químico y pueden precipitar sobre la superficie del oro, impidiendo su recuperación. Estos ejemplos ilustran la complejidad de tratar minerales refractarios y la necesidad de estrategias innovadoras. La biolixiviación bacteriana se ha explorado como alternativa o complemento para el pretratamiento de minerales refractarios de oro y plata: ciertas bacterias (como *A. ferrooxidans*) pueden oxidar in situ los sulfuros (pirita, arsenopirita), liberando el oro atrapado, todo ello sin necesidad de altas temperaturas ni químicos contaminantes.

En este estudio se propone investigar la genómica y transcriptómica de cuatro cepas bacterianas del género *Acidithiobacillus*, reconocidas por su capacidad de biolixiviación de metales pesados. El objetivo es identificar las diferencias y similitudes entre los genomas y los perfiles de expresión génica de estas cepas para comprender mejor sus capacidades metabólicas y, con ello, fundamentar mejoras en su aplicación tanto en remediación ambiental como en procesos de biominería. Este conocimiento permitirá seleccionar las cepas más adecuadas para cada situación (por ejemplo, según el tipo de metal o las condiciones de pH y temperatura), e incluso diseñar consorcios microbianos complementarios que actúen de forma sinérgica, maximizando la eficiencia global de la biolixiviación y contribuyendo a mitigar la contaminación por metales pesados (Sun et al., 2018). En última instancia, desentrañar la base genética de la biolixiviación en estas bacterias sentará las bases para desarrollar tecnologías más limpias de recuperación de metales y la rehabilitación de sitios contaminados.

Antes de profundizar en la metodología propuesta, es importante caracterizar brevemente las cepas seleccionadas y su posición filogenética dentro del género *Acidithiobacillus*. Diversos

estudios han dividido a las especies de *Acidithiobacillus* en grupos filogenéticos (denominados I, II y III) con base en secuencias de ARNr 16S y otros genes conservados. Las cuatro especies de interés pertenecen a distintos grupos: por ejemplo, *A. ferrooxidans* y *A. ferrivorans* se ubican en grupos separados (I y II, respectivamente), mientras que *A. thiooxidans* presenta subclados tanto en I como en III, y *A. caldus* pertenece al grupo III. Esta diversidad filogenética sugiere la existencia de variaciones genéticas importantes incluso dentro de un mismo género, las cuales podrían reflejarse en diferencias funcionales en el proceso de biolixiviación. La Tabla 1 resume las características clave de cada especie considerada:

Tabla 1. Cepas bacterianas del género *Acidithiobacillus* estudiadas, junto con su clasificación filogenética y principales características bioquímicas relacionadas con la biolixiviación.

CEPA BACTERIANA	GRUPO FILOGENÉTICO	CARACTERÍSTICAS RELEVANTES
<i>Acidithiobacillus ferrooxidans</i>	I	Oxida el hierro ferroso (Fe^{2+}) a férrico (Fe^{3+}), generando agentes oxidantes esenciales para la disolución de minerales sulfurados de hierro (p. Ej., Pirita).
<i>Acidithiobacillus thiooxidans</i>	I y III	Oxida azufre elemental a sulfato, proceso fundamental, para la biolixiviación de sulfuros metálicos (liberando ácido sulfúrico que ayuda a solubilizar metales).
<i>Acidithiobacillus ferrivorans</i>	II	Oxida Fe^{2+} a Fe^{3+} y azufre elemental a sulfato; tiene adaptaciones al frío, lo que la convierte en una cepa versátil para biolixiviación en ambientes de baja temperatura.

<i>Acidithiobacillus caldus</i>	III	Oxida compuestos reducidos de azufre de manera eficiente a temperaturas elevadas, por lo que es idónea para operaciones de biolixiviación en entornos termófilos.
---------------------------------	-----	---

Cada especie posee capacidades particulares. *A. ferrooxidans* y *A. ferrivorans* comparten la habilidad de oxidar hierro, aunque pertenecen a grupos filogenéticos distintos (lo que indica diferencias genéticas notables dentro del género). *A. ferrivorans*, en especial, exhibe mayor tolerancia a bajas temperaturas, una adaptación crucial para aplicaciones en climas fríos o en procesos que operan a temperaturas moderadas (Valdés et al., 2008). Por su parte, *A. thiooxidans* se encuentra en dos linajes diferentes (Grupos I y III), reflejando una variabilidad genética interna: todas sus cepas comparten la capacidad de oxidar azufre elemental, lo cual es valioso en la biolixiviación de minerales sulfurados donde el azufre juega un papel clave en la liberación de metales (Zammit et al., 2012). Las distintas cepas de *A. thiooxidans* pueden diferir en eficiencia de biolixiviación y en adaptabilidad a condiciones extremas de pH o a altas concentraciones de ciertos iones, presumiblemente debido a diferencias genómicas entre los subgrupos.

Finalmente, *A. caldus* (grupo III) destaca por su termotolerancia: es capaz de oxidar azufre a temperaturas en las cuales otras especies de *Acidithiobacillus* no sobreviven. La identificación de genes y vías metabólicas particulares en *A. caldus* ha proporcionado información valiosa para mejorar procesos de biolixiviación bajo condiciones de estrés térmico (Mangold et al., 2011), dado que ciertas operaciones biohidrometalúrgicas (por ejemplo, pilas de lixiviación de gran tamaño u operaciones en tanques agitados) conllevan aumentos significativos de temperatura.

En conjunto, comprender estas diferencias genéticas y transcriptómicas entre las cepas es esencial para diseñar estrategias de biolixiviación más efectivas y específicas. Si sabemos qué genes y rutas metabólicas son responsables de la solubilización de cada metal en cada cepa, podremos seleccionar las especies o consorcios microbianos óptimos para tratar un residuo minero o un suelo contaminado determinado. Del mismo modo, podríamos ajustar las condiciones de operación (pH, fuentes de energía, nutrientes) para potenciar la actividad de la cepa seleccionada. En última instancia, este conocimiento molecular detallado allana el camino hacia el desarrollo de consorcios bacterianos optimizados que maximicen la eficiencia de extracción de metales en diversas condiciones ambientales, contribuyendo significativamente tanto a la mitigación de la contaminación por metales pesados como al aprovechamiento sostenible de recursos minerales.

4. DESCRIPCIÓN DEL PROBLEMA

Las bacterias del género *Acidithiobacillus* desempeñan un papel fundamental en la biolixiviación de sulfuros metálicos, proceso clave para la recuperación de metales de interés industrial. Comprender los mecanismos moleculares y las adaptaciones específicas de cada especie y condición de crecimiento es esencial para optimizar el uso biotecnológico de estas cepas. Sin embargo, las diferencias fisiológicas entre especies y la variabilidad de respuesta frente a condiciones ambientales o fuentes de energía dificultan la extrapolación de resultados experimentales. En este contexto, el uso de herramientas bioinformáticas de análisis genómico y transcriptómico permite caracterizar de manera integral las capacidades metabólicas y las respuestas de expresión génica de estas bacterias, proporcionando evidencia sobre cómo las

condiciones de cultivo (temperatura o tipo de fuente energética) influyen en los procesos asociados a la biolixiviación.

5. OBJETIVOS

4.1 GENERAL

Evaluar comparativamente los genomas y transcriptomas de *A. ferrivorans* y *A. ferrooxidans* bajo condiciones de cultivo contrastantes, con el fin de identificar diferencias en su repertorio genético y en sus patrones de expresión asociados a procesos de biolixiviación.

4.2 ESPECÍFICOS

Análisis genómico comparativo: Integrar la información genómica y ortológica de *A. ferrivorans*, *ferrooxidans*, *A. caldus* y *A. thiooxidans* para identificar genes conservados, ortogrupos específicos y variaciones que puedan influir en la biolixiviación.

Análisis transcriptómico de *A. ferrivorans*: Evaluar la expresión diferencial de genes de *A. ferrivorans* cultivada a 6 °C y 28 °C, identificando aquellos genes inducidos o reprimidos diferencialmente bajo condiciones de baja temperatura.

Análisis transcriptómico de *A. ferrooxidans*: Determinar los genes diferencialmente expresados en *A. ferrooxidans* bajo crecimiento electroautotrófico frente a quimioautotrófico, caracterizando las rutas asociadas a transporte y aprovechamiento energético.

Integración funcional: Relacionar los genes y ortogrupos identificados en los análisis genómicos con los patrones de expresión observados en las comparaciones transcriptómicas, resaltando aquellos potencialmente vinculados a procesos de oxidación de hierro y azufre en la biolixiviación.

6. JUSTIFICACIÓN

Esta investigación surge de la necesidad de afrontar un problema ambiental y tecnológico de gran envergadura: la persistente contaminación por metales pesados y las limitaciones de las técnicas actuales para su mitigación. Los métodos convencionales de remoción de metales (por ejemplo, la excavación y reemplazo de suelos, la precipitación química de metales en agua o el confinamiento de residuos) suelen ser costosos, poco selectivos y pueden generar impactos ambientales secundarios. En cambio, las estrategias de biorremediación y en particular la biolixiviación ofrecen una vía más sustentable tanto para recuperar ecosistemas afectados por metales tóxicos como para extraer metales valiosos de fuentes minerales marginales. Sin embargo, existe un vacío de conocimiento en cuanto a los mecanismos genéticos que permiten a ciertos microorganismos llevar a cabo eficientemente la biolixiviación. Abordar este vacío es fundamental para aprovechar al máximo la biotecnología en la solución del problema de los metales pesados.

En este contexto, el proyecto se justifica científicamente en que busca generar conocimiento novedoso sobre las bases moleculares de la biolixiviación en cepas clave de *Acidithiobacillus*. Comprender las respuestas en la expresión génica bajo condiciones de estrés por metales permitirá identificar los mecanismos de adaptación y regulación que estas bacterias activan para tolerar y movilizar metales pesados. Por ejemplo, se espera dilucidar si, al exponerlas a cobre o arsénico, las cepas inducen sistemas de tolerancia (bombas de eflujo, chaperonas de metales, enzimas detoxificantes) que hasta ahora no han sido caracterizados por completo. Conocer estos mecanismos contribuirá directamente a optimizar la eficiencia de la biolixiviación en diferentes entornos contaminados o matrices minerales complejas, ya que podríamos ajustar las condiciones para que las bacterias expresen dichos genes defensivos o metabólicos en el momento y la magnitud adecuados.

Asimismo, la integración de datos genómicos y transcriptómicos proporcionará una visión holística de las capacidades de las cepas estudiadas. El enfoque genómico por sí solo revela el potencial metabólico (es decir, lo que las bacterias podrían hacer), mientras que el enfoque transcriptómico muestra la actividad real bajo ciertas condiciones (lo que efectivamente están haciendo). Al combinar ambos niveles de información, se obtiene un panorama más completo y preciso del repertorio funcional de cada cepa. Esto facilitará el diseño de estrategias de biolixiviación mejor informadas. Por ejemplo, podríamos decidir combinar dos cepas en un consorcio porque genómicamente se complementan (una posee genes que la otra no) y, a nivel transcriptómico, sabemos que no se inhiben mutuamente bajo estrés metálico, sino que ambas expresan en conjunto sus funciones clave.

Cabe destacar que cada cepa bacteriana posee un conjunto único de genes y rutas metabólicas que determinan su capacidad para solubilizar y movilizar distintos metales. Las diferencias genéticas entre *A. ferrooxidans*, *A. thiooxidans*, *A. ferrivorans* y *A. caldus* explican, en parte, por qué sus desempeños varían según el metal o el entorno. Identificar estas diferencias y similitudes genómicas es crucial para aprovecharlas en la práctica: si sabemos qué cepa tiene los genes más adecuados para lixiviar cobre y cuál resulta más eficaz con zinc o níquel, podremos seleccionar, o incluso combinar, cepas específicas para maximizar la eficiencia de la biolixiviación en cada situación (Rawlings, 2002). Estudios previos sugieren que ciertas cepas de *Acidithiobacillus* destacan en la lixiviación de cobre, mientras que otras sobresalen con zinc o níquel debido a diferencias en sus sistemas de transporte de metales y en sus vías energéticas (Rawlings, 2002). Esta variabilidad subraya la necesidad de una caracterización genómica detallada: solo así se podrá mapear qué genes o conjuntos de genes están detrás de la capacidad de cada cepa para atacar distintos minerales o resistir determinadas concentraciones metálicas (Li et al., 2017).

Junto con el análisis genómico, comprender las respuestas transcriptómicas bajo distintas condiciones de estrés metálico permitirá revelar los circuitos de regulación genética activados durante la biolixiviación. Este conocimiento es fundamental para mejorar la robustez de las bacterias frente a entornos contaminados. Por ejemplo, si identificamos que bajo estrés por mercurio una cepa activa un regulón específico (un conjunto de genes regulados por el mismo factor) de tolerancia al estrés oxidativo, podríamos adaptar las condiciones de cultivo para preparar o inducir a la bacteria a activar ese regulón antes de aplicarla en campo. Del mismo modo, la información transcriptómica podría indicar qué tan rápida y plenamente responde cada cepa a un choque metálico, lo cual influye en su rendimiento. Comprender estos aspectos

permitirá aumentar la resiliencia y la capacidad de adaptación de las bacterias mediante intervenciones relativamente simples (como la preadaptación en laboratorio, modificaciones nutricionales o la ingeniería genética dirigida de sus promotores), de modo que toleren mejor altos niveles de contaminación (Nies, 1999).

En la ruta del Fe(II), los genes *cyc2* (localizado en la membrana externa) y *rus* (que codifica la rusticianina) constituyen el nodo inicial de transferencia electrónica hacia complejos periplásmicos y de membrana (*bc1* y *aa3*). En la ruta del azufre destacan *soxB*, *sqr* y *tetH*, que coordinan el catabolismo de tiosulfato, H₂S y tetrionato. La resistencia a Hg, As y Cu está conferida por *merA*, *arsR/arsC* y *copA*, respectivamente; en paralelo, *kdpA*, *speE* y *sqs/hpn* contribuyen a la homeostasis iónica, la biosíntesis de poliaminas y la producción de hopanoides que estabilizan la membrana en condiciones de bajo pH. Finalmente, *pelD* vincula la señalización por c-di-GMP con la producción de exopolisacáridos y la formación de biopelículas, un rasgo crítico para la colonización de sulfuros.

Tabla 2: Selección y rol funcional de genes diana: muestra algunos genes representativos del metabolismo quimiolitotrófico y de la tolerancia a metales/ácido en *Acidithiobacillus ferrooxidans*.

GEN	FUNCIÓN BIOLÓGICA
<i>rusa</i>	Codifica la rusticianina, proteica azul de cobre que transporta electrones en la oxidación de fe(ii).
<i>cyc2</i>	Citocromo c tipo <i>cyc2</i> , localizado en membrana externa; acepta electrones directamente de fe(ii).

soxb	Sulfatasa thiohidrolasa del sistema sox, hidroliza tiosulfato y tetrionato durante la oxidación de azufre.
sqr	Sulfuroquinona reductasa; oxida sulfuro de hidrógeno (h ₂ s) a azufre elemental.
teth	Tetrionato hidrólasa; convierte tetrionato en tiosulfato y azufre elemental.
mera	Mercurio reductasa; convierte hg(ii) en Hg ⁰ , confiriendo tolerancia al mercurio.
arsr/arsc	Operón ars: regulador arsr y reductasa arsc, reducen arsenato (asv) a arsenito (asiii) y lo bombean fuera.
copa	Atpasa p tipo copa, bombea cobre del citoplasma; tolerancia a cu (relacionado con bioleaching de otros metales).
kdpa	Subunidad de la atpasa kdp (sistema kdpfab) para transporte activo de k ⁺ ; Ayuda a mantener potencial de membrana positivo en estrés ácido.
spee	Decarboxilasa sam (spermidina sintasa); biosíntesis de espermidina para tolerancia a ph bajo.
sqs	Squaleno sintasa (sqs) y hopenoide sintasas (hpn) para biosíntesis de hopanoides; estabilizan membrana a ph Ácido.
peld	Efactor de c-di-gmp en el operón <i>pel</i> (peld) del exopolisacárido pel; regula la síntesis de pel y la formación de biopelícula.

Por último, una caracterización genética detallada abre la puerta a mejoras de bioingeniería en estas cepas. Si identificamos genes puntuales que limitan la eficiencia de la biolixiviación (por ejemplo, un represor que apaga la expresión de un conjunto de genes lixiviantes bajo ciertas condiciones), podríamos considerar la modificación genética de las bacterias para eliminar dichas

limitaciones. Del mismo modo, se podría transferir genes clave de una cepa a otra para crear cepas híbridas que posean una capacidad metabólica más completa. Aunque la ingeniería genética de microorganismos acidófilos es todavía incipiente, contar con el mapa genético y transcriptómico de estas bacterias es el primer paso indispensable para cualquier programa de mejoramiento. Cepas diseñadas con mayor tolerancia a condiciones extremas, con una mayor tasa de lixiviación o con menor producción de subproductos indeseados (como ácido sulfúrico en drenajes) podrían revolucionar las estrategias de remediación ambiental, haciéndolas más eficientes y seguras (Singh et al., 2006). En resumen, una comprensión profunda de las bases genéticas y moleculares de la biolixiviación es esencial no solo para optimizar los procesos actuales, sino también para impulsar soluciones innovadoras y sostenibles en la gestión de la contaminación por metales pesados y en la recuperación de metales de interés.

7. MATERIALES Y MÉTODOS

DATA

Para este estudio utilizamos únicamente datos públicos. Como referencias genómicas empleamos genomas completos del género *Acidithiobacillus* disponibles en NCBI, dando prioridad a RefSeq: *A. ferrooxidans* ATCC 23270 (GCF_000021485.1), *A. ferrivorans* XJFY6S-08 (GCF_016250455.1), *A. caldus* SM-1 (GCF_000221025.1) y *A. thiooxidans* ATCC 19377 (GCF_009662475.1). Todos presentan nivel Complete Genome, con tamaños entre 2.98–3.42 Mb, 3,104–3,618 genes y un N50 equivalente al tamaño del ensamblaje (Tabla S1). Los RNA-seq también proceden de repositorios públicos: para *A. ferrooxidans* usamos el estudio PRJNA1078144

(Illumina NovaSeq 6000, paired-end $\approx 2 \times 150$ nt; condiciones de electroautotrofia vs quimioautotrofia; SRR: SRR28010888/89/91/92) y para *A. ferrivorans* el estudio PRJNA347530 (Illumina MiSeq, paired-end 2×76 nt; 6 °C vs 28 °C; SRR: SRR4408480/81/82/89). Reanalizamos y comparamos integradamente transcriptomas públicos de *A. ferrooxidans* (electro vs quimio) y *A. ferrivorans* (6 °C vs 28 °C) sobre genomas completos y ortogrupos, revelando rutas compartidas de metabolismo Fe/S y tolerancia a metales que no habían sido reportadas de forma conjunta.

Tabla 3. Genomas de referencia utilizados.

Especie (cepa)	RefSeq (GCF)	GenBank (GCA)	Ensamblaje	Nivel	N50 (bp)	Tamaño (bp)	# Genes	BioProject	BioSample
A. caldus SM-1	GCF_000221025.1	GCA_000221025.1	ASM22102v1	Complete Genome	2.932.225	3.237.599	3.268	PRJNA63223	SAMN02603220
A. ferrivorans XJFY6S-08	GCF_016250455.1	GCA_016250455.1	ASM1625045v1	Complete Genome	3.161.380	3.161.380	3.198	PRJNA648684	SAMN15644336
A. ferrooxidans ATCC 232	GCF_000021485.1	GCA_000021485.1	ASM2148v1	Complete Genome	2.982.397	2.982.397	3.104	PRJNA53	SAMN02603974
A. thiooxidans ATCC 1937	GCF_009662475.1	GCA_009662475.1	ASM966247v1	Complete Genome	3.415.726	3.415.726	3.618	PRJNA577769	SAMN13037501

Tabla 4. Conjuntos de RNA-seq y metadatos

Comparación	Especie	Condición	Rép	SRX	SRR	BioProject	Plataforma	Diseño	Longitud	Spots (M)	Bases (Gb)	Descarga
Electro vs Quimio	A. ferrooxidans	Quimio	1	SRX23663104	SRR28010889	PRJNA1078144	NovaSeq 6000	Paired-end	~2x150 nt	12.88	3.9	1.1 Gb
Electro vs Quimio	A. ferrooxidans	Quimio	2	SRX23663105	SRR28010888	PRJNA1078144	NovaSeq 6000	Paired-end	~2x150 nt	12.76	3.9	1.1 Gb
Electro vs Quimio	A. ferrooxidans	Electro	1	SRX23663101	SRR28010892	PRJNA1078144	NovaSeq 6000	Paired-end	~2x150 nt	13.65	4.1	1.2 Gb
Electro vs Quimio	A. ferrooxidans	Electro	2	SRX23663102	SRR28010891	PRJNA1078144	NovaSeq 6000	Paired-end	~2x150 nt	12.6	3.8	1.1 Gb
6 °C vs 28 °C	A. ferrivorans	28 °C	1	SRX2234828	SRR4408482	PRJNA347530	MiSeq	Paired-end	2x76 nt	3.636	0.549	0.286 Gb
6 °C vs 28 °C	A. ferrivorans	28 °C	2	SRX2234828	SRR4408489	PRJNA347530	MiSeq	Paired-end	2x76 nt	3.116	0.471	0.244 Gb
6 °C vs 28 °C	A. ferrivorans	6 °C	1	SRX2234829	SRR4408480	PRJNA347530	MiSeq	Paired-end	2x76 nt	3.546	0.536	0.278 Gb
6 °C vs 28 °C	A. ferrivorans	6 °C	2	SRX2234829	SRR4408481	PRJNA347530	MiSeq	Paired-end	2x76 nt	3.959	0.598	0.313 Gb

Plataforma y entorno de trabajo

Todo el procesamiento bioinformático de los datos de secuenciación de ARN (RNA-seq) se realizó empleando Galaxy, una plataforma de análisis reproducible basada en la web (Afgan et al., 2018). Galaxy permitió encadenar las distintas herramientas en un flujo de trabajo transparente. Además, registró automáticamente el historial completo del análisis, incluyendo las versiones y parámetros de cada herramienta utilizada. Esto asegura que todos los pasos puedan ser replicados exactamente por otros investigadores. Asimismo, las etapas de análisis estadístico y la visualización final se llevaron a cabo en R, utilizando cuadernos reproducibles (R o Jupyter) para documentar el código y generar figuras integradas.

Preprocesamiento de datos

Control de calidad inicial

Se comenzó evaluando la calidad de las lecturas crudas (archivos FASTQ de lectura pareada) utilizando la herramienta FastQC (Andrews, 2010). FastQC generó, para cada muestra, informes de calidad que incluyen la distribución de puntajes Phred a lo largo de las bases, contenido GC, presencia de adaptadores, secuencias sobre-representadas y distribución de longitudes, entre otros indicadores. A continuación, se recopilieron todos esos reportes con MultiQC (Ewels et al., 2016) para obtener una visión comparativa resumida de la calidad en todas las muestras. Este control inicial reveló la necesidad de limpiar las lecturas (por ejemplo, eliminando adaptadores Illumina residuales y truncando bases de baja calidad al final de las secuencias) antes de proceder al alineamiento.

Filtrado y recorte de lecturas

Para depurar las secuencias se utilizó Trimmomatic (v0.39; Bolger et al., 2014) en modo *paired-end*. Se proporcionó un archivo de secuencias adaptadoras (Illumina Nextera/TruSeq, según correspondía al experimento) y se aplicaron una serie de pasos de filtrado estándar:

- ILLUMINACLIP: se eliminaron adaptadores y secuencias contaminantes, permitiendo hasta 2 desajustes en fragmentos de 30 pb al hacer coincidir adaptadores.
- SLIDINGWINDOW: se recorrieron las lecturas con una ventana móvil de 4 bases, cortando cuando el promedio de calidad caía por debajo de Q20.
- MINLEN: se descartaron las lecturas resultantes con longitud menor a 50 nucleótidos, para evitar fragmentos excesivamente cortos tras el recorte.
- Recorte de extremos: se eliminaron las bases al inicio y al final de cada lectura que presentaban calidad baja ($< Q20$).

Trimmomatic manejó los pares de lecturas garantizando que, si una lectura de un par era eliminada por mala calidad, su pareja se guardara por separado en archivos “unpaired”. Como resultado de este proceso se obtuvieron archivos FASTQ filtrados de alta calidad, libres de adaptadores y con una distribución de calidades mucho más uniforme.

Control de calidad posterior al recorte

Luego del *trimming*, se volvió a ejecutar FastQC seguido de MultiQC sobre las lecturas limpias, para confirmar la mejora en la calidad. Los informes posteriores al recorte mostraron que las distribuciones de calidad eran estables a lo largo de las lecturas (sin caídas bruscas al final) y que habían desaparecido los picos de adaptadores en los gráficos de K-mers. Se verificó que ya no

hubiera contenido de adaptador detectable y que el contenido GC y otras métricas estuvieran dentro de los rangos esperados. Solo tras cumplir estos criterios de calidad, las lecturas pasaron a las etapas de alineamiento, minimizando así el riesgo de errores o sesgos posteriores debidos a datos de entrada deficientes.

Alineamiento de las lecturas al genoma de referencia

Las lecturas filtradas de cada especie se alinearon contra su genoma de referencia correspondiente utilizando el alineador de alta velocidad HISAT2 (v2.2.1). En particular, las lecturas de *A. ferrivorans* (condiciones 6°C y 28°C) se mapearon contra el genoma de *A. ferrivorans* (cepa correspondiente), mientras que las de *A. ferrooxidans* (condiciones de crecimiento electroautotrófico vs. quimioautotrófico) se mapearon contra el genoma de *A. ferrooxidans*. Previamente, se construyó el índice del genoma de cada especie a partir de su secuencia de referencia en formato FASTA, paso necesario para la búsqueda eficiente de alineamientos. Dado que las secuencias eran *paired-end*, HISAT2 se ejecutó en modo apareado tomando ambos archivos FASTQ de cada muestra.

Se usaron los parámetros por defecto de HISAT2, optimizados para lecturas de ~100 pb, los cuales brindan un buen balance entre sensibilidad y precisión. En genomas bacterianos, que carecen de intrones largos típicos, un alineador “*splice-aware*” como HISAT2 no es estrictamente necesario; sin embargo, HISAT2 ofrece igualmente ventajas en velocidad y exactitud. De hecho, Kim, Langmead y Salzberg (2015) demostraron que HISAT2 es uno de los alineadores más rápidos disponibles sin perder exactitud, gracias a su esquema de índices jerárquicos basados en la transformada Burrows-Wheeler y múltiples índices FM. Cada lectura típicamente se alineó en su

posición genómica correcta, y se descartaron los alineamientos secundarios o de baja calidad. Tras el mapeo, los archivos SAM resultantes se convirtieron a formato BAM y se ordenaron usando SAMtools; luego se generó un índice (.bai) con SAMtools para permitir acceso aleatorio rápido a las alineaciones.

Para evaluar el rendimiento del alineamiento, se empleó nuevamente MultiQC para compilar las estadísticas reportadas por HISAT2 en todas las muestras. Se observó un porcentaje elevado de lecturas mapeadas en cada caso (por ejemplo, más del 90% de las lecturas de *A. ferrivorans* se alinearon únicamente en su genoma, con cifras similares en *A. ferrooxidans*), junto con una alta proporción de pares correctamente mapeados en la orientación esperada. No se detectaron tasas anómalas de multi-mapeo ni otros indicadores de problemas (como contaminación o datos de mala calidad). En conjunto, estos archivos BAM alineados constituyeron una base sólida sobre la cual se construyó el ensamblaje de transcritos y la cuantificación de la expresión.

Ensamblaje y cuantificación de transcritos

Ensamblaje por muestra

Con los alineamientos listos, el siguiente paso fue reconstruir los transcritos expresados en cada condición y cuantificar su abundancia. Para ello se utilizó StringTie (v2.1.7) por su eficiencia y precisión en el ensamblaje de transcriptomas (Pertea et al., 2015). Este procedimiento se llevó a cabo por separado para cada especie.

Primero, para cada muestra (archivo BAM) de cada experimento se ejecutó StringTie individualmente, produciendo ensamblajes de transcritos *de novo* guiados por la referencia. En el caso de *A. ferrivorans* (4 muestras: 2 de 6°C y 2 de 28°C), se proporcionó a StringTie el archivo de anotación génica de referencia en formato GTF/GFF (opción -G) correspondiente a dicha especie. Esto le indicó al algoritmo los loci génicos conocidos, de modo que las lecturas alineadas tendieran a ensamblarse respetando esos genes y heredando sus identificadores cuando fuera posible. Aun así, StringTie puede proponer transcritos novedosos no presentes en la anotación si la distribución de lecturas lo sugiere; a esos transcritos nuevos se les asignan identificadores genéricos que comienzan con “MSTRG” (por ejemplo, MSTRG.1234). Dado que las bibliotecas de RNA-seq disponibles no retenían información de la hebra (librerías no *stranded*), se usó el modo por defecto no dirigido (*unstranded*) para interpretar las lecturas. De modo análogo, para *A. ferrooxidans* se ensamblaron los transcritos de cada una de sus muestras (2 réplicas en condición electroautotrófica y 2 en quimioautotrófica), utilizando la anotación correspondiente de *A. ferrooxidans* como archivo de referencia (-G).

Ajuste de parámetros de ensamblaje

Durante el ensamblaje por muestra se ajustaron varios parámetros clave para equilibrar sensibilidad y precisión, basándose en recomendaciones generales y en la cobertura esperada para bacterias. En particular, se usó $-f0.01$ (*isoform fraction*) para que StringTie retenga isoformas cuya abundancia relativa sea al menos 1% de la isoforma más expresada del locus, permitiendo detectar variantes minoritarias. Se filtraron transcritos muy cortos con $-m 200$ (longitud mínima de 200 nucleótidos) para evitar ensamblajes espurios. Para sitios de *splicing* (que en bacterias solo ocurrirían en casos atípicos, como en operones policistrónicos), se exigió un soporte mínimo de 1

lectura por juntura ($-j$ 1) con al menos 10 bases flanqueando los exones ($-a$ 10). Igualmente, solo se reportaron transcritos con cobertura mínima de $1\times$ ($-c$ 1). Se permitió un pequeño intervalo de hasta 50 bases entre fragmentos para considerarlos parte de este transcrito ($-g$ 50). Finalmente, el parámetro $-M$ 1.0 se estableció para contabilizar en la cuantificación todas las lecturas mapeadas, incluso aquellas que pudieran mapear en más de un locus (repartiendo su contribución de forma proporcional). Estos parámetros relativamente permisivos garantizaron no perder posibles transcritos de baja expresión que pudieran ser de interés diferencial.

Como resultado del ensamblaje por muestra, StringTie generó, para cada réplica, un archivo GTF con la estructura de los transcritos ensamblados en esa muestra (incluyendo exones e intrones, identificando aquellos coincidentes con genes conocidos, así como transcritos nuevos MSTRG), y tablas preliminares de conteos o cobertura por gen y por transcrito en esa muestra. Cabe destacar que StringTie implementa un algoritmo de *maximum flow* sobre el grafo de empalme para reconstruir el conjunto óptimo de transcritos que explica las lecturas, evitando fragmentar en exceso los loci y logrando ensamblajes más completos que métodos previos como Cufflinks (Trapnell et al., 2010).

Fusión de ensamblajes

Tras obtener los archivos GTF individuales de todas las muestras, se integró la información combinando los ensamblajes en un catálogo maestro de transcritos. Para ello se utilizó el comando `stringtie --merge`, suministrándole la lista de archivos GTF de todas las réplicas de *A. ferrivorans* (y de manera análoga los de *A. ferrooxidans*, por separado). StringTie-merge unificó los transcritos redundantes presentes en múltiples muestras y produjo un único archivo GTF combinado por

especie, que contenía todos los transcritos observados en al menos una muestra. De este modo, si cierto gen se expresa solo en una condición (por ejemplo, presente a 28°C pero no a 6°C), de todas formas, aparecerá en el ensamblaje combinado, permitiendo incluirlo en la cuantificación global entre condiciones.

En esta etapa de fusión se optó por criterios permisivos para no descartar posibles transcritos valiosos. Por ejemplo, se conservaron transcritos de hasta 50 nt de longitud mínima (*-m 50*) sin exigir una cobertura mínima (*-c 0*). También se incluyeron transcritos con un nivel mínimo de expresión de FPKM = 1.0 (*-F 1.0*) o TPM = 1.0 (*-T 1.0*) en al menos una muestra. Igualmente, se mantuvo *-f 0.01* (fracción de isoforma 1%) para evitar fusionar isoformas muy distintas. Se amplió a 250 bases el intervalo permitido entre transcritos distintos (*-g 250*), dado que en la fusión entre muestras se buscó ser conservador al separar loci. El GTF resultante representó la anotación expresada en el experimento: un conjunto unificado de genes/transcritos para comparar entre condiciones en cada especie.

Cuantificación unificada

Finalmente, utilizando el GTF combinado como referencia de anotación, se volvió a ejecutar StringTie en cada muestra únicamente para cuantificar la abundancia de cada transcrito o gen previamente definido. Para ello, en cada muestra se corrió StringTie con el parámetro -G (usando el GTF combinado correspondiente) y activando el modo de solo estimación (-e), de manera que no se ensamblaran nuevos transcritos en esta fase, sino que simplemente se midiera la cobertura de cada transcrito conocido del catálogo unificado. Esta cuantificación uniforme garantizó que todas las muestras tuvieran valores de expresión para el mismo conjunto de genes o transcritos, haciendo comparables los conteos entre condiciones (es decir, asegurando una comparación “*de manzanas con manzanas*”). StringTie exportó tablas de expresión por gen y por transcrito para cada muestra; en Galaxy se activó la salida compatible con DESeq2/EdgeR, obteniendo directamente archivos tabulares con los conteos por gen.

En suma, al completar este proceso se obtuvieron, para cada especie, matrices de conteo donde las filas corresponden a genes (identificados por un *gene_id* heredado de la anotación original o asignado por StringTie, como los MSTRG para genes nuevos) y las columnas corresponden a las muestras (réplicas por condición). Este enfoque de ensamblaje guiado seguido de fusión y re-cuantificación permitió descubrir genes potencialmente no anotados previamente (marcados con nuevos IDs), a la vez que aseguró que la comparación de conteos entre condiciones se realizara exactamente sobre el mismo conjunto de genes o transcritos. Cabe mencionar que, tratándose de bacterias (sin *splicing* alternativo convencional), la mayoría de los genes corresponden a un único transcrito; aun así, esta estrategia es útil para detectar diferencias en UTRs transcritos o posibles RNAs pequeños no anotados.

Extracción de tablas de conteo por gen

Como resultado de la cuantificación unificada con StringTie, se obtuvieron las tablas de conteo bruto de lecturas asignadas a cada gen para cada muestra. Adicionalmente, de forma complementaria, se empleó la herramienta featureCounts (Liao, Smyth, & Shi, 2014) del paquete Subread para realizar un conteo independiente de las lecturas por gen. En este paso, se volvieron a recorrer los archivos BAM alineados de cada muestra, asignando cada lectura a los genes anotados en el GFF/GTF de referencia de la especie correspondiente. Se utilizó featureCounts principalmente como validación cruzada de los conteos obtenidos con StringTie, asegurando que no hubiera discrepancias importantes debidas al método de cuantificación. Para ello, se verificó que los identificadores de gen utilizados por featureCounts (derivados del campo *gene_id* del GTF de referencia o del GTF combinado) coincidieran exactamente con los IDs en la matriz de StringTie, garantizando la comparabilidad. Los resultados mostraron una alta concordancia entre ambos métodos de conteo. Finalmente, para cada comparación experimental (*A. ferrivorans* 6°C vs 28°C, y *A. ferrooxidans* electroautotrófico vs quimioautotrófico), se consolidó una matriz de conteos de genes lista para el análisis estadístico diferencial.

Análisis de expresión diferencial

Preparación de datos y filtrado de expresiones

El análisis de genes diferencialmente expresados (DEGs) entre condiciones se llevó a cabo en R utilizando el paquete DESeq2. En cada caso, se importó la matriz de conteos por gen (genes × muestras) y se construyó un objeto DESeq2 especificando un modelo de efectos ~ condición, donde

condición es la variable categórica de interés con dos niveles (por ejemplo, “6°C” vs “28°C” para *A. ferrivorans*, y “electroautotrófico” vs “quimioautotrófico” para *A. ferrooxidans*).

Antes de ajustar el modelo estadístico, se aplicó un filtro para eliminar genes de muy baja expresión que tienden a generar mucha variabilidad y poca información. Específicamente, se descartaron genes con un conteo total menor a 10 entre todas las muestras de un experimento. Esto redujo el conjunto a aquellos genes expresados de forma apreciable en al menos alguna de las condiciones. Por ejemplo, en *A. ferrivorans* aproximadamente 200–300 genes superaron este filtro mínimo, lo cual es consistente con que solo un subconjunto del genoma se exprese bajo las condiciones evaluadas.

Modelado estadístico y detección de DEGs

El análisis diferencial siguió el flujo de trabajo estándar de DESeq2. Primero, la función DESeq() estimó los parámetros de dispersión para cada gen empleando un modelo lineal generalizado de distribución binomial negativa (Love et al., 2014). Este modelo aprovecha la estrategia bayesiana de *shrinkage* implementada en DESeq2, que comparte información entre genes para moderar las estimaciones de dispersión y evitar valores extremos poco fiables. Asimismo, DESeq2 calculó factores de normalización por tamaño de biblioteca (utilizando la mediana de razones) para hacer comparables los conteos entre muestras con distintas profundidades de secuenciación. Luego, se realizó la prueba de significancia estadística para cada gen mediante un test de Wald, comparando la media de expresión entre las dos condiciones y obteniendo un log₂ Fold Change (log₂FC) estimado por gen.

Se extrajeron los resultados del contraste de interés utilizando la función `results()`, especificando la comparación (por ejemplo, condición “6C” vs “28C” en *A. ferrivorans*, tomando 28°C como referencia). Para cada gen, esto produjo el valor de log2FC (positivo si el gen se expresa más a 6°C que a 28°C, negativo si sucede lo contrario), junto con su error estándar, el estadístico de Wald y el p-valor asociado. Dado que se probaron miles de genes simultáneamente, se aplicó una corrección por comparaciones múltiples mediante el método de Benjamini–Hochberg para controlar la tasa de falsos descubrimientos (Benjamini & Hochberg, 1995). Este ajuste de p- valores garantiza que el subconjunto de genes declarados significativos contenga solo un porcentaje acotado de falsos positivos. Se consideraron significativos aquellos genes con un p- valor ajustado (p_{adj}) < 0.1, un criterio algo más laxo que el 0.05 tradicional pero adecuado para maximizar la detección de cambios en estudios exploratorios de expresión. No obstante, también se reportaron los genes con p_{adj} < 0.05 como un conjunto de alta confianza.

Resultados del análisis diferencial

En *A. ferrivorans* (6°C vs 28°C), DESeq2 identificó del orden de unas pocas decenas de genes diferencialmente expresados con p_{adj} < 0.1, lo que indica que el cambio de temperatura afecta solo a un subconjunto relativamente pequeño del transcriptoma. De ellos, un número ligeramente mayor de genes mostró mayor expresión a 6°C que a 28°C (log2FC positivos), lo que sugiere la inducción de ciertos genes asociados al estrés por frío. Por otro lado, otros genes presentaron mayor expresión a 28°C (log2FC negativos), en concordancia con condiciones óptimas de crecimiento.

En *A. ferrooxidans* (electroautotrofia vs quimioautotrofia), se detectó igualmente un conjunto de genes significativamente regulados por el modo de crecimiento. Varios genes clásicos de la vía

quimioautótrofa (oxidación de Fe/S) resultaron reprimidos en la condición electroautotrófica, mientras que genes implicados en el transporte de electrones alternativo (como pilinas y porinas) se encontraron sobreexpresados en electroautotrofia. Estas tendencias concuerdan con observaciones previas reportadas por Wang et al. (2024). Estos resultados detallados se abordan con mayor profundidad en el capítulo de Resultados; aquí nos limitamos a señalar que la metodología estadística aplicada fue capaz de detectar tales patrones globales de cambio en la expresión.

Visualización y evaluación de resultados

Además de las tablas numéricas, DESeq2 proporcionó métricas de diagnóstico y visualizaciones para validar la calidad del análisis. Se generó un gráfico MA (MA-plot) que muestra el \log_2FC frente a la expresión media de cada gen: como era esperable, la mayoría de los puntos se agruparon alrededor de $\log_2FC \approx 0$ (genes sin cambio), y los genes significativos aparecieron destacados (coloreados) fuera de la nube central, confirmando que no hubo sesgos sistémicos en las diferencias observadas. También se construyó un *volcano plot* ($-\log_{10} p_{adj}$ vs \log_2FC) para cada comparación, en el cual los genes con cambios más pronunciados y significativos se aprecian fácilmente por su posición hacia las esquinas del gráfico.

Adicionalmente, se llevó a cabo un análisis de componentes principales (PCA) usando las expresiones normalizadas de todos los genes, con el fin de evaluar la agrupación global de las muestras. En el espacio de los primeros componentes principales, las réplicas biológicas de una misma condición se agruparon cercanamente, separadas de las del otro tratamiento. Esto indica que el efecto de la condición experimental fue consistente y predominó sobre la variabilidad intragrupo.

Esta clara separación refuerza la confianza en que los cambios hallados (los DEGs) se deben realmente a las condiciones comparadas y no a ruido aleatorio.

Por último, se elaboró un *heatmap* de expresión que muestra los niveles normalizados de los genes significativos (filas) en cada muestra (columnas), después de aplicar una transformación regularizada de los conteos para estabilizar la varianza. En dicho mapa de calor se observaron patrones de coexpresión distintivos: por ejemplo, en *A. ferrivorans* los genes inducidos a 6°C se agruparon formando un bloque con mayor intensidad de color en las muestras de 6°C, mientras que esos mismos genes aparecían atenuados en las muestras de 28°C, y *viceversa*. Esto no solo confirmó la magnitud de los cambios detectados por DESeq2, sino que también permitió vislumbrar posibles coregulaciones entre genes bajo cada condición.

Manejo de la nomenclatura de genes y anotaciones

Un desafío final considerado fue la interpretación biológica de los genes identificados, especialmente aquellos con identificadores no estandarizados. El ensamblaje con StringTie introdujo nuevos identificadores para transcritos no anotados previamente (los IDs MSTRG.X mencionados), por lo que se realizó un procedimiento para mapear estos IDs a genes conocidos cuando fue posible. En primer lugar, para cada gen novel marcado como MSTRG que resultó diferencialmente expresado, se consultó su posición genómica en el ensamblaje combinado y se cruzó con la anotación oficial: si caía dentro o muy cerca de un gen anotado en el genoma de referencia, se asumió que correspondía a una isoforma o a una extensión de ese gen. En otros casos, se extrajo la secuencia del transcrito predicho (mediante la herramienta *gffread* o directamente de la secuencia genómica) y se realizó un análisis BLAST contra la base de datos NR de NCBI o

contra proteomas de *Acidithiobacillus* relacionados, para determinar si presentaba homología con algún gen conocido. Este enfoque permitió anotar funcionalmente varios transcritos inicialmente “anónimos”. Por ejemplo, un MSTRG que mostró una alta expresión diferencial resultó tener alta identidad con una oxidasa conocida, lo que sugiere que en realidad correspondía a ese gen (quizá no incluido en la anotación original debido a pequeñas diferencias en la secuencia).

Finalmente, se normalizaron todos los identificadores relevantes; por ejemplo, los IDs MSTRG significativos se reemplazaron por el nombre del gen conocido más cercano o por una anotación funcional propuesta.

8. RESULTADOS

En coherencia con la caracterización filogenética previa, el árbol se construyó mediante el método Neighbor-Joining (NJ), utilizando una matriz de distancias derivada de BLOSUM62 a partir de un alineamiento múltiple de 355 posiciones. Se realizaron 200 réplicas *bootstrap*, cuyos valores de soporte se muestran sobre las ramas. El *outgroup* permitió enraizar el árbol y confirmar la separación filogenética de *Thermithiobacillus* respecto de las especies de *Acidithiobacillus*. Se observa la formación de distintos clados dentro del género, algunos con altos valores de soporte ($\geq 95\%$), lo que respalda la agrupación consistente de especies cercanas, como *A. ferridurans* y *A. ferrooxidans*.

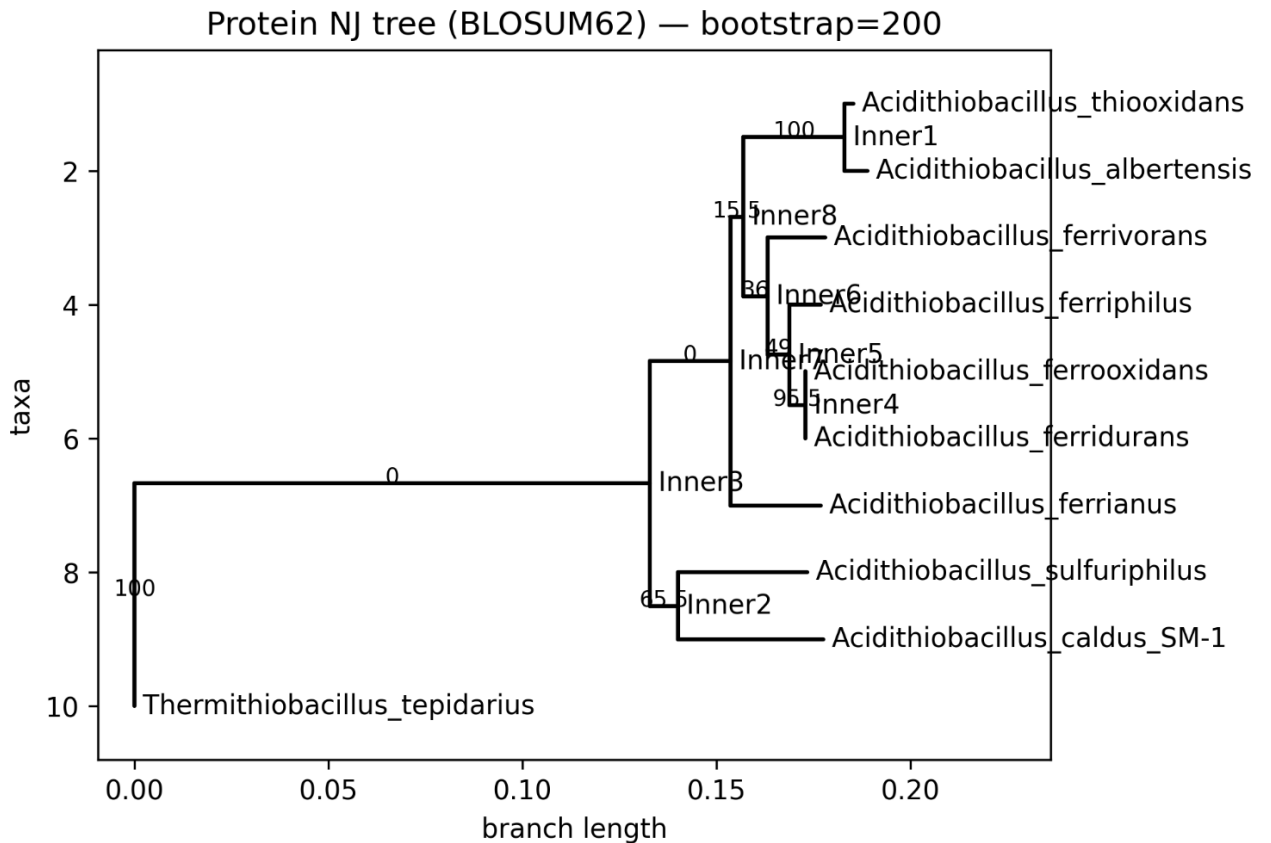


Figura 1. Árbol filogenético de proteínas basado en la secuencia del gen *recA* para diez especies del género *Acidithiobacillus* y un outgroup (*Thermithiobacillus tepidarius*).

Para llevar a cabo el análisis, se realizó una búsqueda en BLAST para cada especie, utilizando los genomas disponibles en las bases de datos genómicas. En este capítulo se presentan los resultados obtenidos, organizados por tipo de análisis. En primer lugar, se describen los patrones de expresión génica diferencial en *Acidithiobacillus ferrivorans* bajo dos temperaturas (6 °C vs 28 °C) y en *A. ferrooxidans* bajo dos condiciones de cultivo (electroautotrofia vs quimioautotrofia). A continuación, se exponen los hallazgos del análisis ortológico comparativo entre cuatro especies de *Acidithiobacillus* mediante OrthoFinder. Finalmente, se reportan los resultados de las búsquedas BLASTP de proteínas de interés, evidenciando la conservación funcional en las distintas especies.

Análisis ortológico entre especies

Se realizó un análisis ortológico comparativo utilizando OrthoFinder para cuatro especies del género *Acidithiobacillus*: *A. caldus*, *A. ferrivorans*, *A. ferrooxidans* y *A. thiooxidans*. Este análisis permitió identificar grupos de ortólogos (*orthogroups*) y relaciones evolutivas entre los genomas de estas especies. OrthoFinder generó tres niveles jerárquicos de agrupación (N0, N1, N2), siendo N0 el nivel más detallado que contiene el mayor número de ortogrupos (en total, 3104 ortogrupos a nivel N0). Los resultados permiten distinguir el genoma central compartido por todas las especies, así como los ortogrupos específicos de cada especie y aquellos con duplicaciones génicas internas en cada linaje. Estos hallazgos aportan evidencia sobre la diversidad funcional y la evolución del género *Acidithiobacillus*, señalando qué genes son compartidos por todas las especies y cuáles representan adaptaciones particulares o expansiones genómicas en linajes específicos.

En cuanto a los ortogrupos exclusivos por especie, la Figura 10 resume la cantidad de grupos génicos únicos identificados en cada genoma. Se observa que *A. thiooxidans* presenta el mayor número de ortogrupos exclusivos, seguido por *A. caldus* y *A. ferrivorans*, mientras que *A. ferrooxidans* exhibe la menor cantidad de grupos exclusivos. Este resultado sugiere que *A. thiooxidans* ha retenido o desarrollado conjuntos génicos particulares más amplios en comparación con las otras especies, potencialmente vinculados a adaptaciones ecológicas o metabólicas propias de su nicho. En general, la presencia de genes exclusivos en cada especie indica la posibilidad de funciones especializadas que podrían contribuir a diferencias en sus capacidades bioquímicas y por ejemplo, en procesos de biolixiviación.

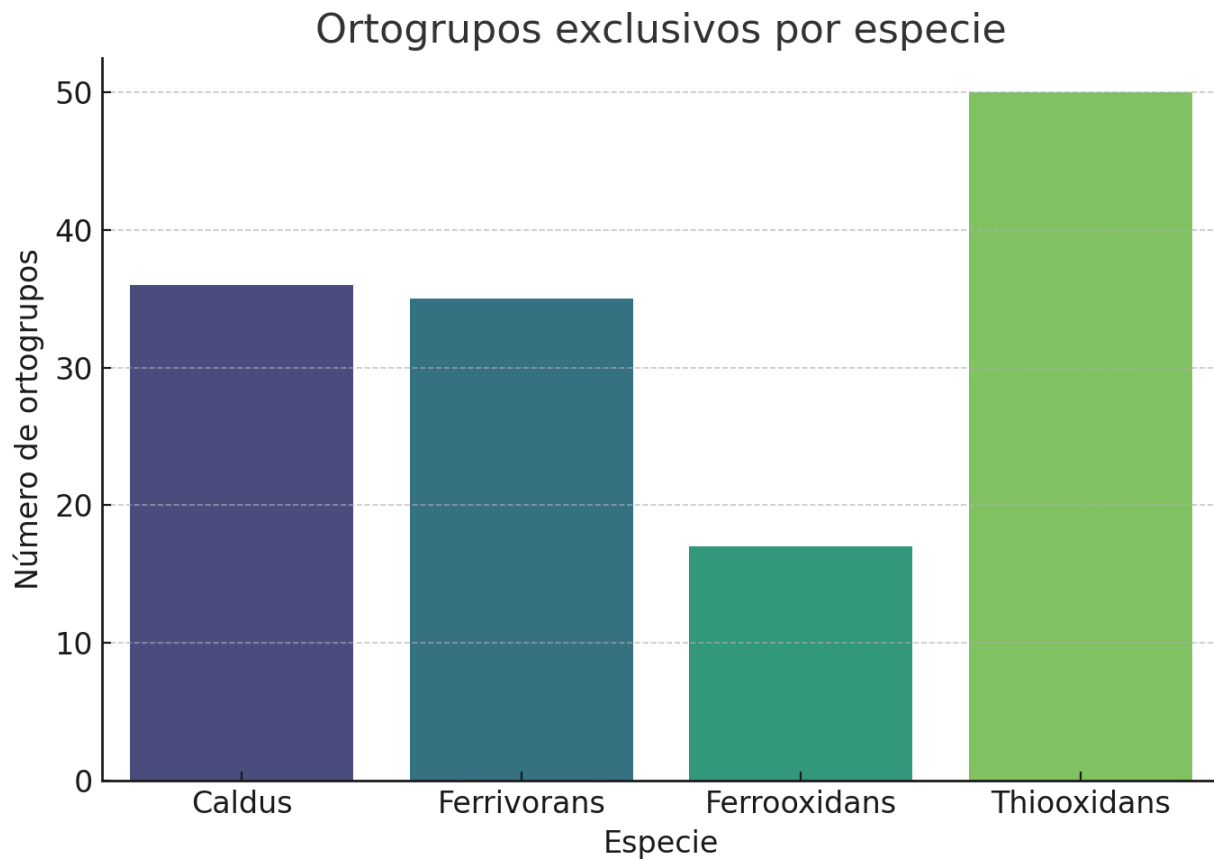


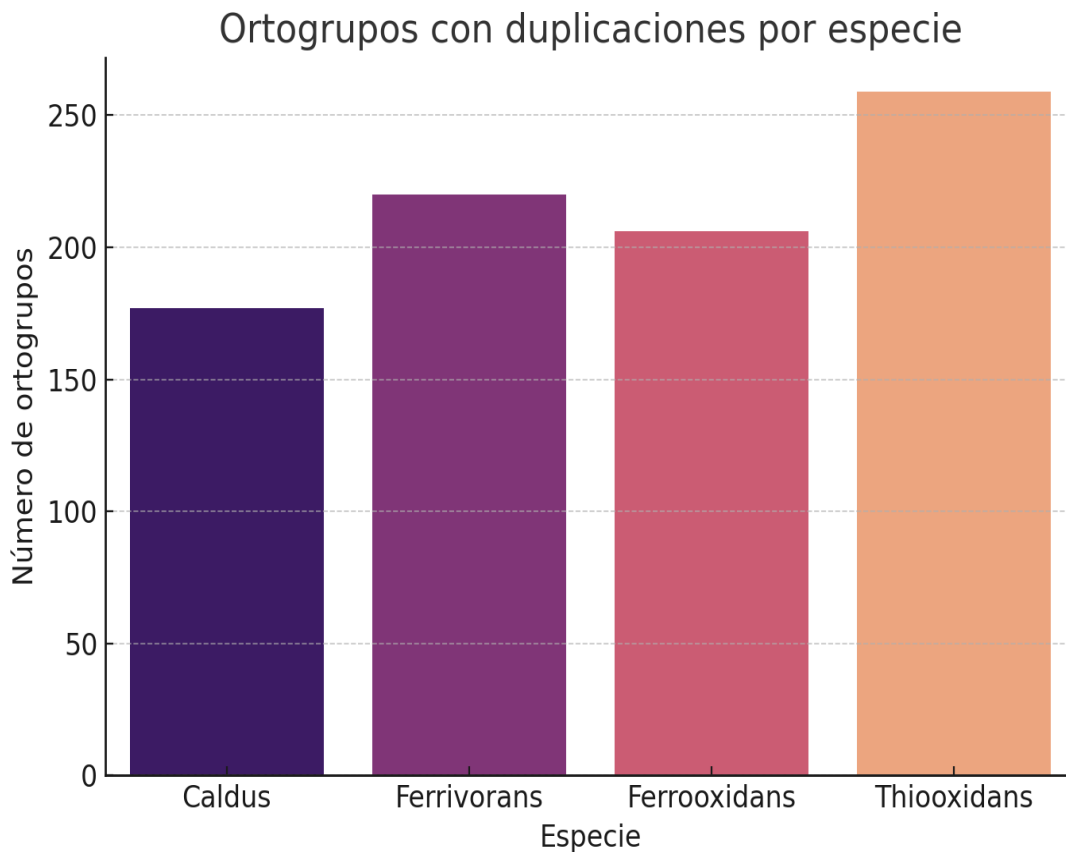
Figura 2. Ortogrupos exclusivos por especie de *Acidithiobacillus*.

Gráfico de barras que muestra el número de ortogrupos identificados como exclusivos de cada una de las cuatro especies analizadas. *A. thiooxidans* presenta 50 ortogrupos exclusivos, seguido por *A. caldus* con 36, *A. ferrivorans* con 35 y *A. ferrooxidans* con 17. Este patrón sugiere que cada especie ha incorporado o preservado genes únicos que podrían estar relacionados con adaptaciones funcionales específicas.

Dichos genes exclusivos representan candidatos para estudios futuros, orientados a comprender su papel en las diferencias ecológicas o metabólicas entre especies.

Por otro lado, la comparación reveló ortogrupos con duplicaciones génicas internas (parálogos) en cada genoma, resultado de eventos de duplicación dentro de la misma especie. La Figura 11 presenta el número de ortogrupos duplicados identificados por especie. En este caso, *A. thiooxidans* encabeza la lista con la mayor cantidad de duplicaciones, seguida de *A. ferrivorans*, *A. ferrooxidans* y finalmente *A. caldus*. Las duplicaciones génicas reflejan procesos de expansión del repertorio génico dentro de cada linaje, lo cual es relevante desde el punto de vista evolutivo: las copias génicas duplicadas pueden divergir para especializarse en nuevas funciones, ofrecer redundancia adaptativa o ampliar la versatilidad metabólica de la especie. El hecho de que *A. thiooxidans* presente el mayor número de ortogrupos duplicados sugiere una dinámica genómica orientada a una mayor diversificación funcional, en comparación con las otras especies analizadas.

Figura 3. Ortogrupos con duplicaciones génicas internas por especie.



Cada barra representa la cantidad de ortogrupos en los que se detectaron duplicaciones de genes (parálogos) dentro del genoma de la especie correspondiente. *A. thiooxidans* muestra 259 ortogrupos con duplicaciones, *A. ferrivorans* 220, *A. ferrooxidans* 206 y *A. caldus* 177. Estas cifras indican que *A. thiooxidans* ha experimentado la mayor expansión génica interna, seguida de *A. ferrivorans*. Las duplicaciones génicas pueden conducir a la aparición de funciones novedosas o a la subfuncionalización de las proteínas, contribuyendo a diferencias en las capacidades metabólicas entre especies.

Además de los ortogrupos exclusivos y duplicados, se analizaron los ortogrupos compartidos entre especies, es decir, conjuntos de genes presentes en dos, tres o en las cuatro especies. La Figura 4 resume tres categorías principales de conservación génica: (1) ortogrupos presentes en las cuatro especies (2) ortogrupos compartidos por tres de las cuatro especies, y (3) ortogrupos compartidos únicamente por dos especies

En la primera categoría, destaca un núcleo ortológico de 1506 ortogrupos presentes en las cuatro especies, lo que constituye la base funcional común del género *Acidithiobacillus*. Adicionalmente, se identificaron 519 ortogrupos compartidos por exactamente tres especies y 941 ortogrupos compartidos únicamente por pares de especies. La existencia de ortogrupos no presentes en todas las especies refleja diferencias en la composición génica de cada organismo, posiblemente asociadas a la adaptación a condiciones ambientales particulares o a la diversificación de nichos metabólicos en cada linaje.

Figura 4. Ortogrupos compartidos entre especies de *Acidithiobacillus*.

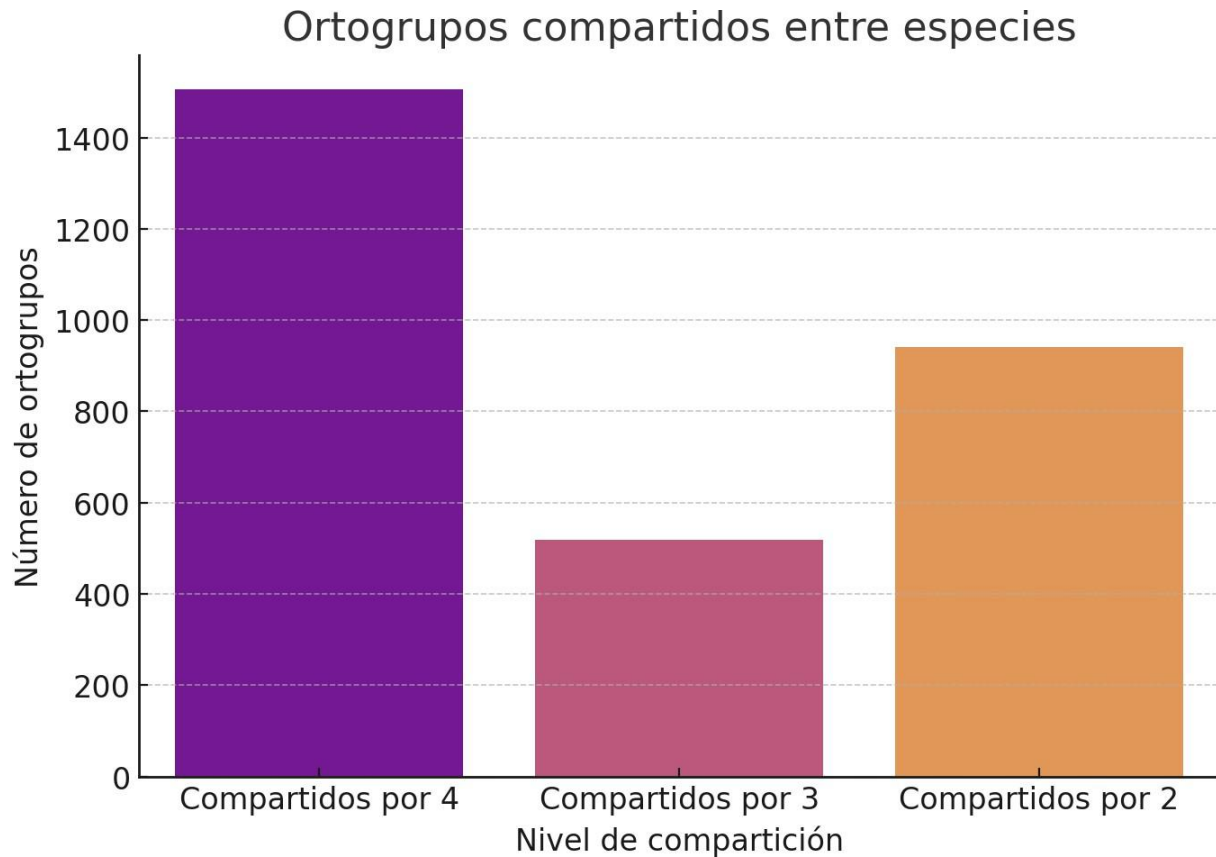


Diagrama resumen de la conservación génica mostrando: 1506 ortogrupos presentes en las cuatro especies, 519 ortogrupos compartidos por tres especies, y 941 ortogrupos compartidos por solo dos especies. El amplio núcleo común sugiere una base funcional conservada a lo largo del género, mientras que las porciones compartidas parcialmente evidencian la pérdida o ganancia de genes en linajes específicos. Estas diferencias en conservación apuntan a que cada especie ha adaptado su genoma a su entorno particular, reteniendo la mayor parte de funciones esenciales comunes, pero divergiendo en ciertos componentes genéticos.

En conjunto, el análisis ortológico comparativo permitió extraer tres conclusiones principales:

1. **Núcleo genómico conservado:** Existe un núcleo de 1506 ortogrupos presentes en las cuatro especies, lo que respalda una similitud funcional esencial compartida en el género *Acidithiobacillus*. Este conjunto central probablemente corresponde a genes relacionados con funciones básicas y comunes necesarias para la supervivencia y metabolismo en ambientes de biolixiviación.
2. **Genes exclusivos por especie:** Cada especie posee ortogrupos exclusivos (entre 17 y 50, según la especie), los cuales constituyen una posible fuente de adaptaciones diferenciales. Estos genes únicos podrían conferir ventajas específicas o reflejar la especialización ecológica y metabólica de cada organismo.
3. **Duplicaciones génicas variables:** Se observó una frecuencia variable de ortogrupos con genes duplicados en los distintos genomas, siendo notablemente mayor en *A. thiooxidans*. Esto indica divergencias en la dinámica evolutiva de los genomas: algunas especies han experimentado más expansiones génicas, lo que podría traducirse en una mayor diversificación de funciones o mecanismos redundantes que aporten robustez adaptativa. Estos hallazgos proporcionan una base sólida para vincular los resultados genómicos comparativos con los patrones transcriptómicos observados en los experimentos de expresión diferencial. En conjunto, la evidencia sugiere que las diferencias en el repertorio génico (genes exclusivos, duplicados y compartidos) contribuyen a las respuestas específicas de cada especie, fortaleciendo la comprensión de la diversidad genética y funcional de los *Acidithiobacillus* en procesos de biolixiviación.

En la Tabla 5 se resumen cuantitativamente estos resultados ortológicos, incluyendo el número de ortogrupos compartidos por distintas combinaciones de especies, así como los exclusivos y duplicados en cada genoma analizado:

Tabla 5. Número de ortogrupos compartidos, exclusivos y duplicados en cuatro especies de Acidithiobacillus.

<i>Categoría</i>	<i>Número de ortogrupos</i>
<i>Ortogrupos compartidos por las 4 especies</i>	1506
<i>Ortogrupos compartidos por 3 especies</i>	519
<i>Ortogrupos compartidos por 2 especies</i>	941
<i>Ortogrupos exclusivos de A. caldus</i>	36
<i>Ortogrupos exclusivos de A. ferrivorans</i>	35
<i>Ortogrupos exclusivos de A. ferrooxidans</i>	17
<i>Ortogrupos exclusivos de A. thiooxidans</i>	50

<i>Ortogrupos con duplicaciones en A. caldus</i>	177
<i>Ortogrupos con duplicaciones en A. ferrivorans</i>	220
<i>Ortogrupos con duplicaciones en A. ferrooxidans</i>	206
<i>Ortogrupos con duplicaciones en A. thiooxidans</i>	259

La tabla muestra el número de ortogrupos compartidos, exclusivos y con duplicaciones génicas internas para cada especie analizada. Se observa un núcleo ortológico común de 1506 ortogrupos presentes en todas las especies, así como variaciones en la cantidad de genes exclusivos y duplicados por especie, concordantes con una diversificación genómica diferencial en el género.

Resultados de BLASTP de proteínas de interés

Para complementar el análisis transcriptómico y ortológico, se llevó a cabo una búsqueda BLASTP utilizando proteínas de interés contra las bases de datos de proteomas de *A. ferrooxidans*, *A. ferrivorans*, *A. thiooxidans* y *A. caldus*. Este análisis funcional comparativo permitió identificar homologías y niveles de conservación de proteínas clave entre las especies. En general, los resultados evidencian una alta conservación: se hallaron homólogos estrechamente relacionados en las cuatro especies analizadas, con porcentajes de identidad proteica elevados que alcanzan hasta

un 100 % cuando *A. ferrooxidans* se compara consigo misma y en torno al 90 % en comparaciones entre especies distintas.

Además de los alineamientos principales de alta identidad, se detectaron alineamientos secundarios con identidades menores (aproximadamente 25–40 %), correspondientes a homólogos más distantes o a proteínas que comparten dominios conservados con las consultas. La presencia de estos homólogos lejanos sugiere una posible diversificación de funciones: algunas proteínas podrían haber divergido lo suficiente para desempeñar funciones relacionadas, pero no idénticas, conservando solo ciertos módulos o motivos estructurales comunes. Adicionalmente, en varios casos un mismo gen de interés produjo alineamientos significativos contra múltiples proteínas WP_ dentro de una misma especie, lo que constituye evidencia de duplicaciones génicas o conjuntos parálogos asociados a esa función. Este hallazgo es coherente con los resultados del análisis ortológico (que mostró duplicaciones internas) y sugiere que algunas funciones importantes (p. ej., resistencia a metales, vías metabólicas clave) están respaldadas por familias multigénicas en estos genomas. Cabe destacar que los alineamientos principales abarcan generalmente la longitud completa de la proteína de referencia, lo que confirma la conservación integral de estas enzimas clave en los procesos de biolixiviación; por otro lado, algunos alineamientos parciales indican la existencia de proteínas cuya similitud con la consulta se restringe a determinados dominios catalíticos.

Ciertas proteínas de interés mostraron altos niveles de conservación en las cuatro especies. Por ejemplo, genes como *sqr*, *soxB*, *tetH*, *merA* y *rusA* presentan porcentajes de identidad superiores al 90 % en al menos una de las comparaciones interespecíficas. Estos resultados sugieren que las funciones asociadas a la oxidación de azufre (*sqr*, *soxB*, *tetH*) y al transporte de electrones en la

oxidación de hierro (*rusA*, y también *cyc2* en las especies donde está presente) son esenciales y se han preservado evolutivamente con muy pocas variaciones en su secuencia proteica.

Se identificó que genes asociados a la resistencia o homeostasis de metales pesados (por ejemplo, *arsC*, *arsR*, *copA*, *kdpA* y *merA*) están presentes en los cuatro genomas, aunque exhiben distintos grados de conservación entre especies. En particular, *copA* (ATPasa de transporte de cobre) mantiene identidades muy altas (>85%) en todas las comparaciones, lo que confirma la relevancia de la homeostasis de cobre y su conservación funcional. En contraste, *arsR* (regulador de respuesta a arsénico) mostró las identidades más bajas (entre ~37 % y 55 % según la especie), reflejando una mayor divergencia posiblemente en los mecanismos de regulación de la resistencia a arsénico en cada linaje.

Algunos genes de interés presentaron una distribución más restringida o resultados menos uniformes en BLASTP. Por ejemplo, *pelD* (implicado en síntesis de biopelícula) solo se detectó en *A. thiooxidans* y *A. caldus*, con identidades moderadas (49–79%), lo cual indica una presencia limitada a ciertas especies. En el caso de *cyc2* (implicado en la oxidación de hierro), no se identificó ningún homólogo evidente en *A. thiooxidans* ni en *A. caldus* con los datos analizados, aunque *cyc2* sí está presente con alta identidad en *A. ferrooxidans* y *A. ferrivorans*. Por último, el gen *sqs* (escualeno sintasa) no fue detectado directamente en los resultados de BLASTP; en su lugar, la consulta correspondiente mostró alineamiento con una proteína anotada como phytoene synthase (sintasa de fitógeno), indicando que la anotación de *sqs* en UniProt puede diferir del nombre o función asignada en los genomas analizados.

Los principales alineamientos BLASTP obtenidos para cada gen de interés en las cuatro especies de *Acidithiobacillus*. Se listan, para cada consulta (proteína referencial de UniProt), los mejores *hits* encontrados en cada especie, junto con el porcentaje de identidad de secuencia (%Id) correspondiente.

Tabla 6. Resumen de los mejores alineamientos BLASTP de genes de interés en cuatro especies de *Acidithiobacillus*.

Gen	Query (UniProt)	<i>A. ferrooxidans</i> (WP_) [%Id]	<i>A. ferrivorans</i> (WP_) [%Id]	<i>A. thiooxidans</i> (WP_) [%Id]	<i>A. caldus</i> (WP_) [%Id]
<i>rusA</i>	E1V0I2	WP_012537622.1 (100%)	WP_198661235.1 (93%)	WP_115051357.1 (55%)	WP_014002734.1 (44%)
<i>cyc2</i>	O33823	WP_012537628.1 (97%)	WP_065414354.1 (86%)	<i>No detectado</i>	<i>No detectado</i>
<i>soxB</i>	A0A179B9A9	WP_012536761.1 (100%)	WP_198660949.1 (97%)	WP_031575937.1 (88%)	WP_004868584.1 (85%)
<i>sqr</i>	B7JBP8	WP_012536761.1 (100%)	WP_232821520.1 (29%)	WP_031569823.1 (90%)	WP_014003003.1 (87%)
<i>tetH</i>	B7J3C9	WP_012535754.1 (100%)	WP_198660625.1 (57%)	WP_244947567.1 (60%, parcial)	WP_014002952.1 (60%)
<i>merA</i>	P17239	WP_012607441.1 (100%)	WP_198660997.1 (56%)	WP_153940336.1 (79%)	WP_014002600.1 (60%)

Gen	Query (UniProt)	<i>A. ferrooxidans</i> (WP_) [%Id]	<i>A. ferrivorans</i> (WP_) [%Id]	<i>A. thiooxidans</i> (WP_) [%Id]	<i>A. caldus</i> (WP_) [%Id]
<i>arsR</i>	A0A2W1KHZ1	WP_012537155.1 (37%)	WP_198660691.1 (55%)	WP_153940615.1 (50%)	WP_041638081.1 (54%)
<i>arsC</i>	A0A2W1KGJ4	WP_012537436.1 (100%)	WP_198659809.1 (74%)	WP_010640437.1 (80%)	WP_014002557.1 (79%)
<i>copA</i>	A0A060UVR1	WP_225487670.1 (86%)	WP_198661386.1 (99%)	WP_153940351.1 (85%)	WP_041638095.1 (86%)
<i>kdpA</i>	B7J714	WP_012607439.1 (100%)	WP_198660069.1 (70%)	<i>No detectado</i>	WP_014002109.1 (64%)
<i>speE</i>	A0A2W1K744	WP_009567116.1 (100%)	WP_198660472.1 (93%)	WP_010640600.1 (91%)	WP_014003545.1 (80%)
<i>pelD</i>	A0A1C2IVB8	<i>No detectado</i>	<i>No detectado</i>	WP_214441199.1 (79%)	WP_014002055.1 (49%)
<i>sqs</i>	(no detectado)	<i>No detectado</i>	<i>No detectado</i>	<i>No detectado</i>	<i>No detectado</i>

Se muestran los mejores alineamientos BLASTP para cada gen de interés, con el porcentaje de identidad de secuencia (%Id) indicado en negritas. Los guiones largos (—) o entradas "No detectado" indican ausencia de un homólogo detectable en los resultados analizados para esa especie. El gen *sqs* (escualeno sintasa) no fue detectado en las búsquedas; en las bases de datos de UniProt, aparece anotado bajo la denominación alternativa *phytoene synthase*, lo que explica la falta de coincidencias directas con la etiqueta *sqs*.

Análisis de expresión diferencial en *A. ferrivorans* (6 °C vs 28 °C)

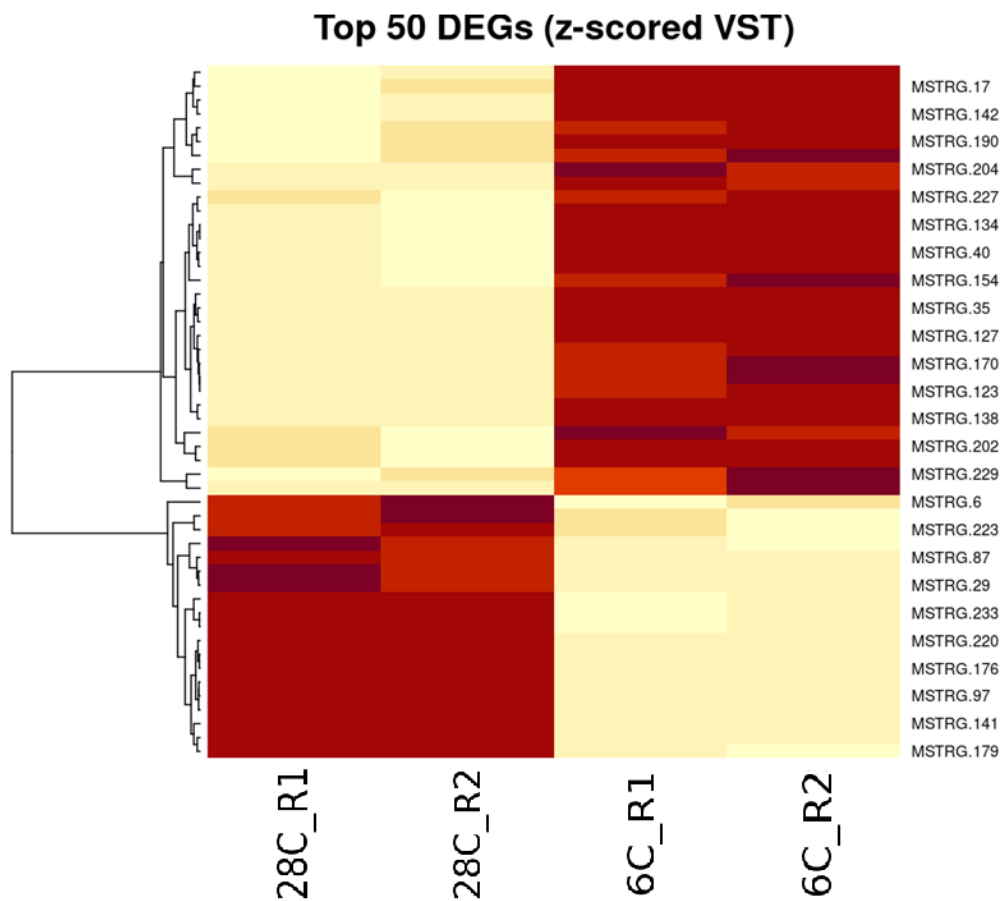
Diseño del análisis: Se consideraron 4 muestras de *A. ferrivorans* (2 réplicas a 6 °C y 2 réplicas a 28 °C). Tras filtrar los genes con conteo total < 10, un total de 234 genes entraron al análisis con DESeq2. Se utilizó un modelo lineal ~ condición (niveles: 28 °C y 6 °C) para evaluar el efecto de la temperatura. Como criterio de significación estadística se aplicó un umbral de FDR ($\text{padj} < 0.10$).

Con $\text{padj} < 0.10$, se identificaron 133 genes diferencialmente expresados en total: 73 con mayor expresión a 6 °C y 60 con mayor expresión a 28 °C. Con el umbral más estricto ($\text{padj} < 0.05$), se obtuvieron 127 DEGs, 71 sobre-expresados a 6 °C y 56 a 28 °C. Los cambios de expresión observados oscilaron entre $\text{FC} = -3.59$ y $\text{FC} = +8.19$. Para las visualizaciones se aplicó una transformación de estabilización de la varianza, VST, variante recomendada para conjuntos pequeños de genes. Cabe señalar que los identificadores génicos provienen del ensamblaje de novo (StringTie merge); por tanto, muchos genes se etiquetan como MSTRG.* al no mapear directamente contra anotaciones previas.

Los resultados evidenciaron una respuesta transcriptómica marcada a la temperatura. En general, los genes diferencialmente expresados se distribuyen de forma relativamente balanceada

entre los inducidos a 6 °C y a 28 °C, indicando que ambas condiciones desencadenan cambios específicos en distintos conjuntos génicos. La Figura 5 ilustra los perfiles de expresión de los 50 genes con mayor significancia, mediante un mapa de calor. Se observa que las muestras se agrupan por condición: las dos réplicas de 28 °C clusterizan juntas, al igual que las dos de 6 °C, reflejando alta consistencia entre réplicas. Asimismo, se distinguen bloques de genes con mayor expresión en una u otra condición; por ejemplo, un grupo amplio de genes muestra expresión elevada a 6 °C (colores cálidos en las columnas de 6 °C) mientras que otro conjunto presenta el patrón opuesto con mayor expresión a 28 °C. Esto indica que un subconjunto definido de genes diferencia claramente las dos temperaturas a nivel de sus perfiles globales de expresión.

Figura 5. *Heatmap* de los 50 genes más significativos en *A. ferrivorans* (6 °C vs 28 °C).



El mapa de calor representa los 50 genes con FDR más bajo, tras transformación VST y estandarización (z-score) por gen. Las columnas corresponden a las muestras (28C_R1, 28C_R2, 6C_R1, 6C_R2) y las filas a cada gen. Se observa que las réplicas de una misma condición agrupan juntas, y que un bloque de genes exhibe mayor expresión a 6 °C (colores más cálidos en columnas de 6 °C) mientras otro bloque muestra mayor expresión a 28 °C. Esto confirma que los genes diferencialmente expresados separan ambas condiciones y que las réplicas son consistentes dentro de cada grupo.

Las dos réplicas de 28 °C aparecen muy cercanas entre sí y distantes del grupo de 6 °C a lo largo de PC1, mientras que PC2 (3.4 % de varianza explicada) refleja únicamente variaciones menores. La separación de las muestras por condición a lo largo de PC1, superando con creces la variabilidad intra-condición, confirma que la temperatura es el factor principal que explica las diferencias transcriptómicas observadas. Por otra parte, un Análisis de Componentes Principales (PCA) de las expresiones (transformadas por VST) refuerza estos hallazgos. En la Figura 6, que muestra el diagrama de dispersión en el plano de las dos primeras componentes principales, se aprecia que PC1 explica aproximadamente el 93.7 % de la varianza total y separa claramente las muestras de 6 °C vs 28 °C.

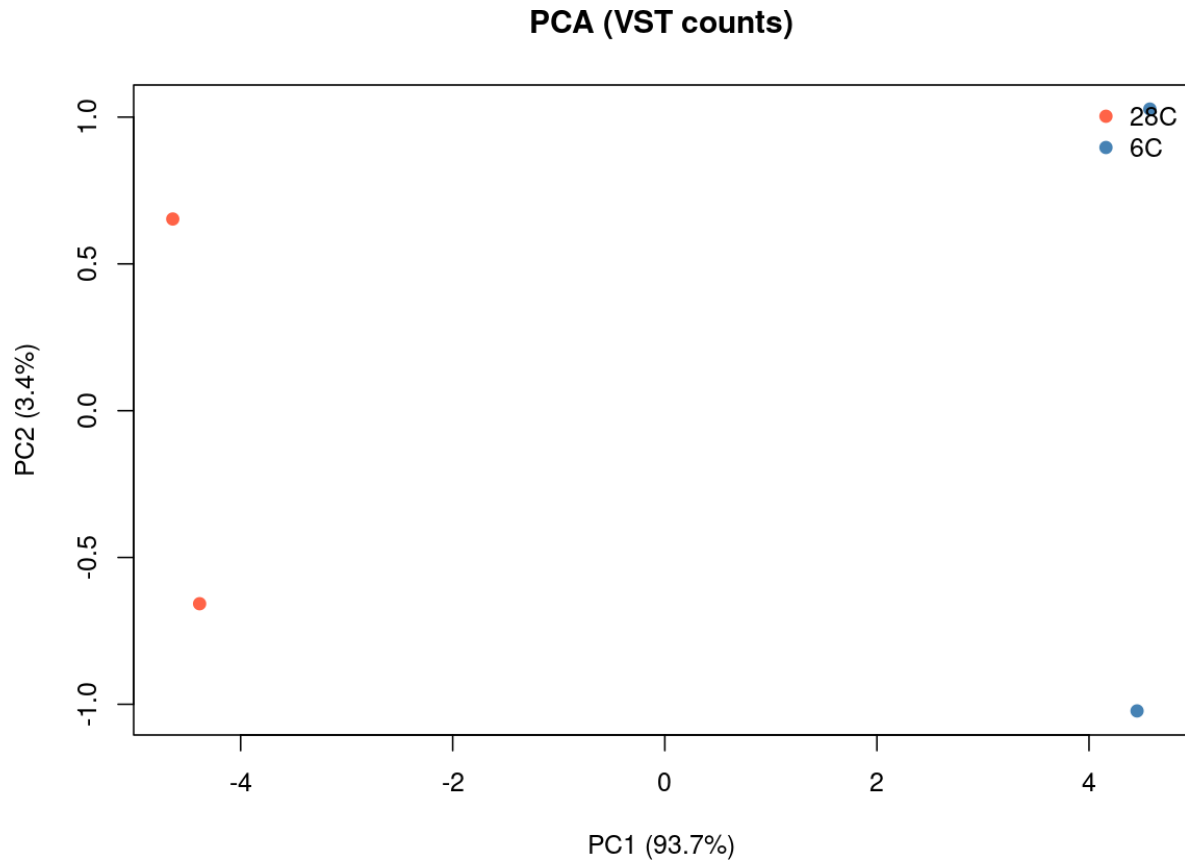


Figura 6. Análisis de componentes principales (PCA) de las expresiones génicas en *A. ferrivorans*.

Se muestran los ejes PC1 = 93.7 % de varianza explicada vs PC2 = 3.4 %. Las muestras se agrupan según la temperatura: las réplicas de 28 °C (●) se ubican juntas y separadas de las de 6 °C (●) a lo largo de PC1. Esto evidencia que la mayor fuente de variación transcriptómica corresponde a la condición térmica, corroborando la reproducibilidad entre réplicas y la clara diferenciación de ambas condiciones experimentales.

Para explorar la magnitud de los cambios de expresión en función de la abundancia génica, se construyó un gráfico MA (*Mean Average plot*, Figura 6). En este diagrama, el eje X representa el nivel de expresión promedio y el eje Y es FC (6 °C/28 °C) para cada uno de los 234 genes

analizados. La mayoría de los genes se concentra alrededor de $FC = 0$, indicando expresiones similares en ambas condiciones; sin embargo, un subconjunto de genes muestra cambios pronunciados, con valores de \log_2 de hasta aproximadamente $+8.2$ (más expresados a 6°C) o -3.6 (más expresados a 28°C). Estos cambios extremos se observan principalmente en genes de baja a media abundancia, aunque no restringidos solo a los menos expresados. En conjunto, el MA-plot confirma que las variaciones detectadas abarcan distintos niveles de expresión y no se limitan a artefactos de genes muy poco expresados.

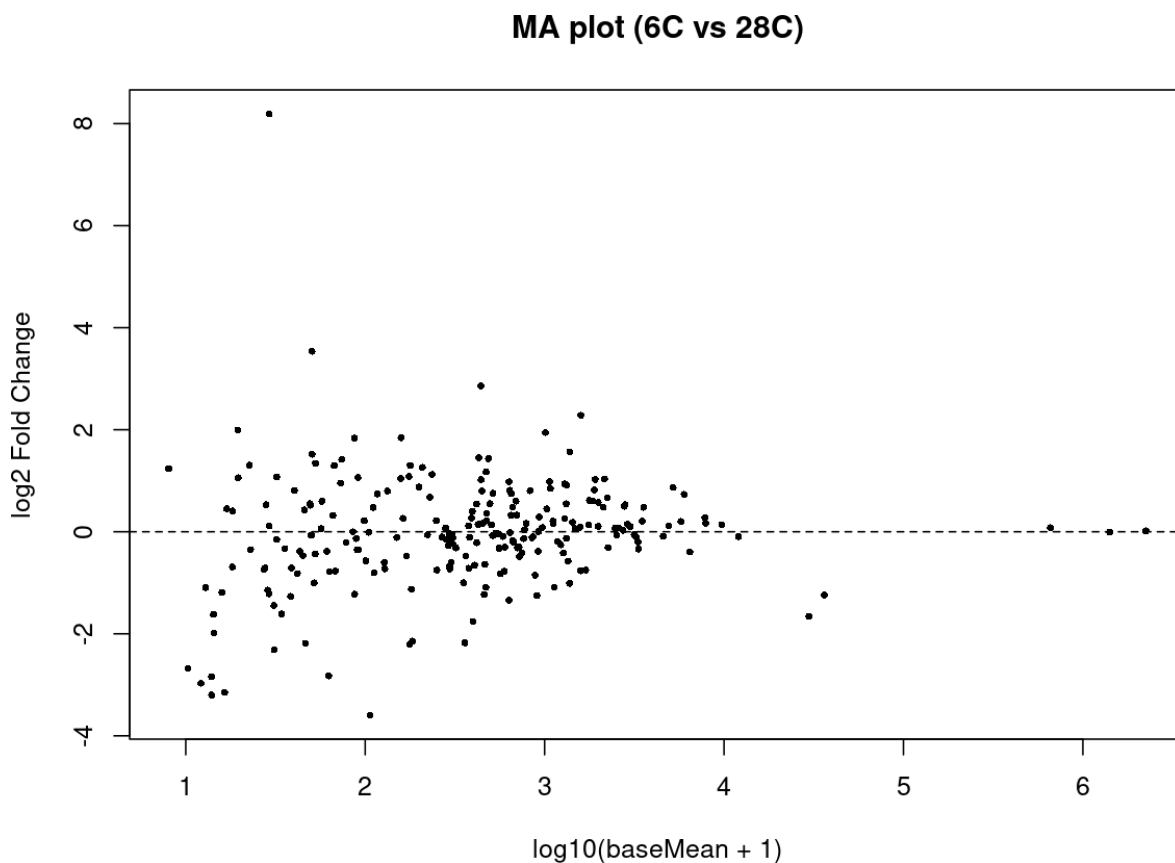


Figura 7. MA-plot del análisis diferencial en *A. ferrivorans* (6°C vs 28°C).

Cada punto representa un gen (234 en total), mostrando el \log_2 (eje Y, cambio 6°C vs 28°C) frente al \log_{10} del nivel de expresión medio (eje X). La mayoría de los genes se ubica cerca de $\log_2 FC = 0$, mientras que algunos presentan cambios mayores. Se alcanzan $\log_2 FC +8.2$ y -3.6 .

en genes con expresión media baja, indicando que existen genes altamente inducidos en 6 °C y en 28 °C. La línea horizontal discontinua en $Y=0$ sirve de referencia para visualizar la simetría de cambios en ambas direcciones.

Finalmente, se resume la significancia estadística y la magnitud de los cambios mediante un volcano plot (*Figura 5*). En este gráfico, el eje X corresponde al \log_2 FC (6 °C vs 28 °C) y el eje Y a $-\log_{10}$ (FDR). La línea horizontal marcada en el gráfico representa el umbral de significancia estadística $p_{adj} = 0.10$. Como se aprecia, existe un número elevado de genes por encima del umbral, confirmando que 133 genes resultaron significativamente diferenciales a $p_{adj} < 0.10$. La distribución de puntos es bastante equilibrada a ambos lados de \log_2 FC = 0, consistente con los conteos mencionados (73 genes con mayor expresión a 6 °C y 60 con mayor expresión a 28 °C). Entre los genes más significativos se destacan, por ejemplo, *MSTRG.141* (con \log_2 FC negativo, sobre-expresado a 28 °C) y *MSTRG.42*, \log_2 FC positivo, sobre-expresado a 6 °C). En conjunto, el volcano plot sintetiza la información de magnitud de cambio vs significancia estadística, evidenciando un conjunto amplio de DEGs asociados a la temperatura.

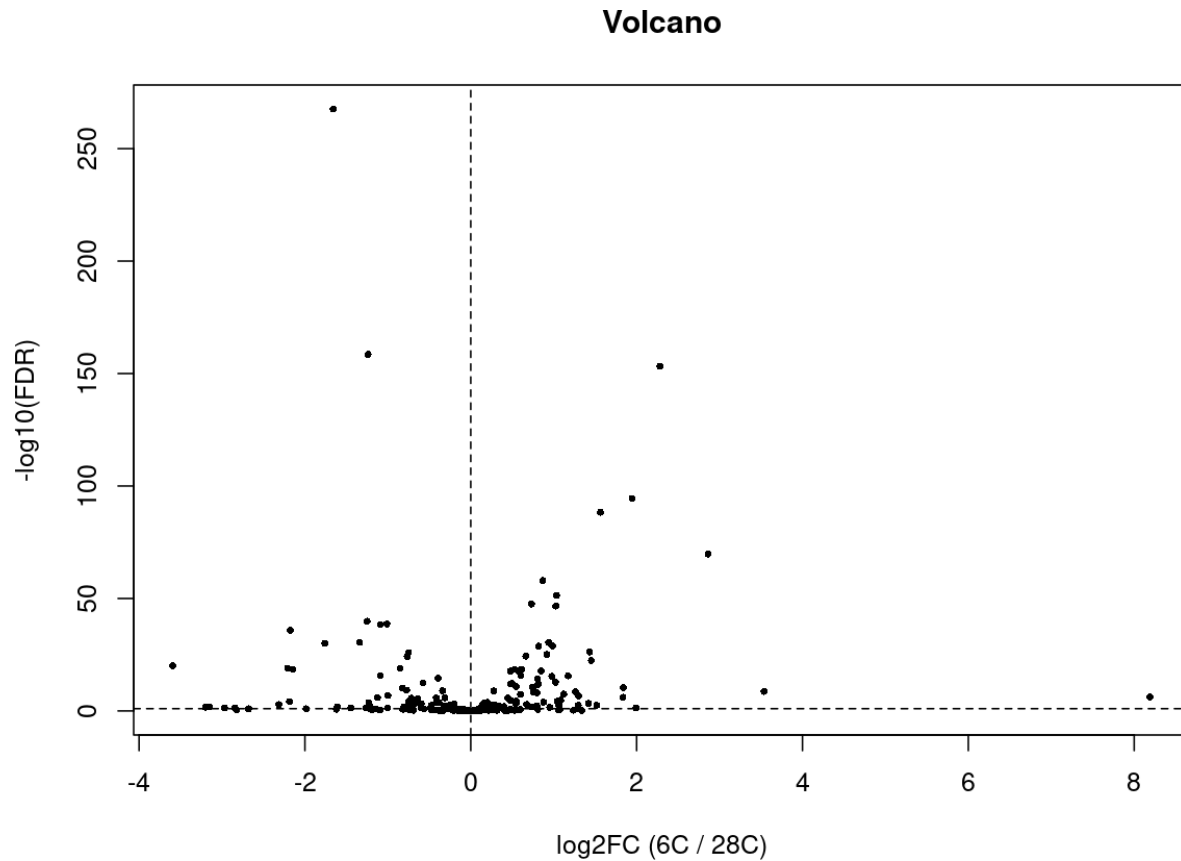


Figura 8. *Volcano plot* de la expresión diferencial en *A. ferrivorans* (6 °C vs 28 °C).

Se graficó el log2 FC en el eje X frente a $-\log_{10}$ FDR en el eje Y para cada gen. La línea horizontal indica el umbral de significancia ($p_{adj} = 0.10$). Los puntos por encima de esta línea corresponden a genes diferencialmente expresados; en total se identificaron 133 genes significativos ($p_{adj} < 0.10$). La distribución de genes arriba del umbral es simétrica en torno a $\log_2 FC = 0$ (73 genes 6 °C, 60 genes 28 °C), lo que sugiere que ambos tratamientos inducen cambios sustanciales en subconjuntos de genes. Los genes más significativos incluyen ejemplos como *MSTRG.141* (28 °C) y *MSTRG.42* (6 °C).

En *A. ferrivorans*, se evaluaron 234 genes y se detectaron 133 DEGs con $FDR < 0.10$ (73 inducidos a 6 °C y 60 a 28 °C). El rango de cambios de expresión observado abarca desde $\log_2 FC = -3.59$ hasta $+8.19$. Estos resultados cumplen el objetivo específico de caracterizar la respuesta transcriptómica de *A. ferrivorans* a la temperatura. Se confirma la existencia de un conjunto definido de genes termorregulados, con perfiles de expresión distintivos para 6 °C y 28 °C. La interpretación funcional detallada de estos cambios se desarrolla posteriormente en la sección de Discusión de la tesis.

Análisis de expresión diferencial en *A. ferrooxidans* (electroautotrofia vs quimioautotrofia)

Diseño del análisis: Para *A. ferrooxidans* se incluyeron 4 muestras transcriptómicas: 2 réplicas en condición electroautotrófica y 2 en condición quimioautotrófica. Tras filtrar genes con conteo total < 10 , en el análisis con DESeq2 se evaluaron aproximadamente 2700 genes. El diseño estadístico fue condición y se utilizó nuevamente un umbral de significancia de $p_{adj} < 0.10$. Con $p_{adj} < 0.10$, se identificó un número muy amplio de genes diferencialmente expresados (del orden de cientos de genes), tanto sobre-expresados en la condición electro como en la condición chemo. Los valores de cambio fueron notoriamente elevados en algunos casos, oscilaron entre $\log_2 FC = -11.3$ y $+22.5$, evidenciando cambios de gran magnitud en la expresión de ciertos genes. Para la

visualización de datos con PCA, heatmap, se aplicó la transformación VST a los valores de expresión. Al igual que en *A. ferrivorans*, se emplearon identificadores derivados del ensamblaje (MSTRG.*) con anotaciones parciales para algunos genes, producto de la integración de transcriptomas y genomas de referencia.

Los resultados muestran que la condición de crecimiento influye fuertemente en la expresión génica de *A. ferrooxidans*. Aplicando criterios similares al caso anterior, se halló que la mayoría de los genes diferenciales presentaron mayor expresión en electroautotrofia. En contraste, un conjunto también significativo de genes mostró sobre-expresión en quimioautotrofia, si bien con magnitudes de cambio generalmente menores. La distribución general de cambios sugiere una respuesta transcriptómica más intensa en la condición electro. Para visualizar estos patrones, la Figura 6 presenta un *heatmap* con los 50 genes más significativos (menor FDR). Al igual que en *A. ferrivorans*, las expresiones relativas (z-score sobre valores VST) evidencian grupos bien definidos de genes y una clara separación de muestras por condición: las réplicas electro clusterizan juntas y separadas de las réplicas chemo. Se identifican bloques de genes altamente expresados en electro (colores rojos intensos en columnas Electro) y otros con mayor expresión en chemo. Este mapa de calor confirma que los genes diferencialmente expresados permiten distinguir las dos condiciones de crecimiento y que las réplicas son consistentes entre sí dentro de cada tratamiento.

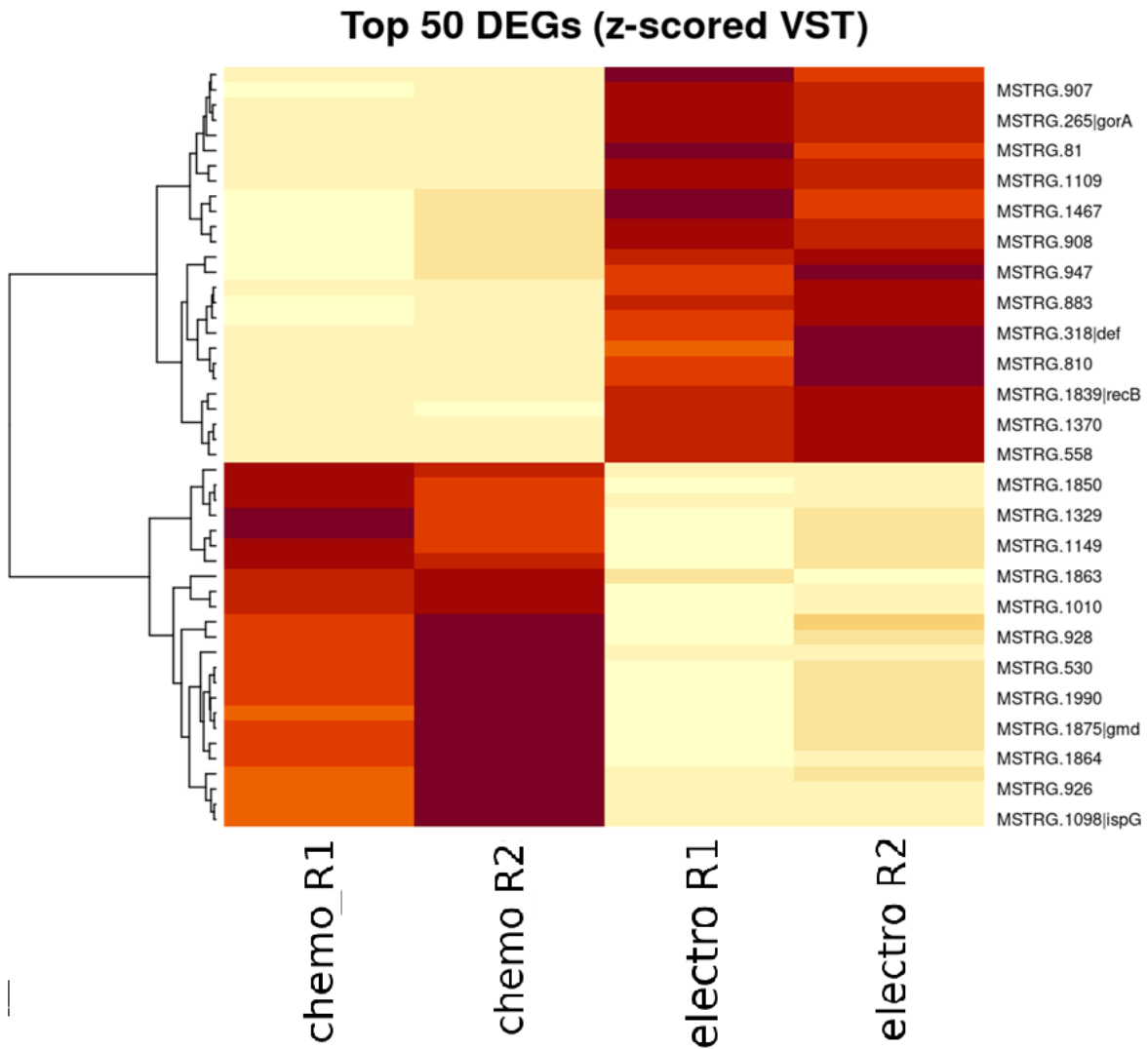


Figura 9. Heatmap de los 50 genes más significativos en *A. ferrooxidans* (electro vs chemo).

Se muestran los 50 genes con menor FDR ($p_{adj} < 0.10$), tras transformación VST y estandarización por gen (z-score). Las columnas corresponden a las muestras quimioautotróficas (Chemo_R1, Chemo_R2) y electroautotróficas (Electro_R1, Electro_R2). Las filas representan cada gen (IDs MSTRG.* y anotaciones parciales del ensamblaje). Se observa que las réplicas de

cada condición agrupan juntas y separadas de la otra condición. Además, hay conjuntos de genes con mayor expresión en electro (regiones en rojo oscuro en las columnas Electro) y otros con mayor expresión en chemo (rojos en columnas Chemo), evidenciando perfiles transcriptómicos divergentes entre ambas condiciones experimentales.

La robustez de la separación por condición se corrobora mediante PCA. En la Figura 9 se presenta el análisis de componentes principales para *A. ferrooxidans*. La componente principal PC1 explica el 71.2 % de la variación total y distingue claramente las muestras electro de las chemo, mientras que PC2 explica un 19.5 % adicional. En el plano PC1–PC2, las muestras quimioautótrofas (por ejemplo, marcadas en color azul) aparecen cercanas entre sí, y separadas de las muestras electroautótrofas (marcadas en color rojo), las cuales forman un grupo aparte. La variabilidad entre condiciones (eje PC1) excede ampliamente la variabilidad dentro de cada grupo, lo que confirma que el modo de autotrofia (electro vs chemo) es el principal factor estructurando la variación transcriptómica en este experimento. La consistencia en el agrupamiento de réplicas aporta un control de calidad adicional, indicando reproducibilidad de los datos dentro de cada tratamiento.

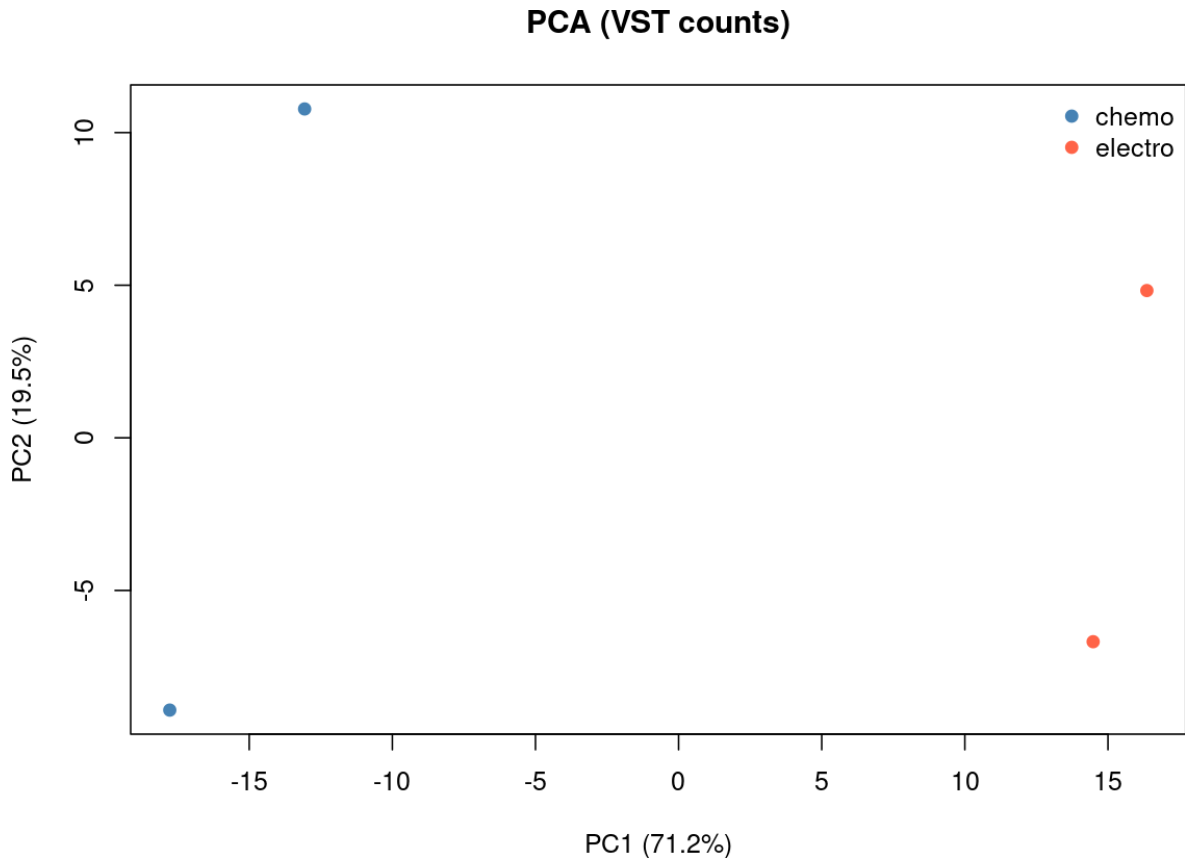


Figura 10. Análisis de componentes principales (PCA) de las expresiones génicas en *A. ferrooxidans*.

Se grafican PC1 (71.2 % de varianza explicada) vs PC2 (19.5 %). Las muestras de condición quimioautotrófica (●) se agrupan entre sí, mientras que las muestras electroautotróficas (●) forman un grupo claramente separado. La varianza entre condiciones (principalmente reflejada en PC1) es muy superior a la varianza intra-condición, lo que confirma la fuerte influencia de la condición de cultivo en los perfiles de expresión génica.

En cuanto a la magnitud y significancia de los cambios de expresión en *A. ferrooxidans*, la Figura 8 muestra el MA-plot correspondiente. Nuevamente, la mayoría de los genes aparecen cerca de $\log_2 FC = 0$, pero se destaca un conjunto amplio de genes con cambios sustanciales tanto positivos como negativos. Se observan genes con $\log_2 FC$ extremos, llegando a sobre-expresiones

de más de 20 veces en electro $\log_2 FC > +4.3$, equivalente a $>20\times$ y a sub-expresiones notables en chemo $\log_2 FC < -3.5$, equivalente a $>11\times$ en sentido inverso. Importante es que estos genes con cambios marcados no se restringen a los de muy baja expresión: el gráfico revela diferencias transcriptómicas significativas en diversos rangos de abundancia génica, reforzando la idea de una reprogramación global bajo condiciones electroautotróficas.

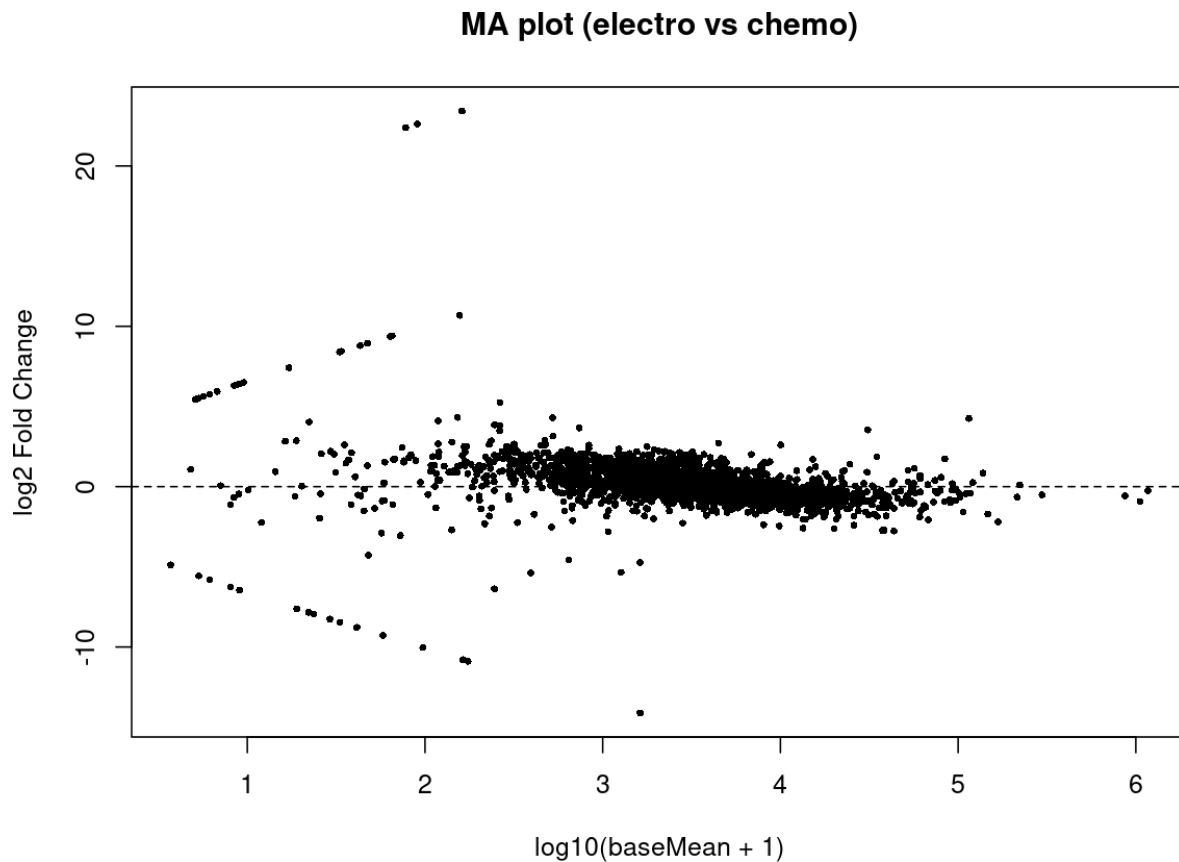


Figura 11. MA-plot del análisis diferencial en *A. ferrooxidans* (electro vs chemo).

Se representa el $\log_2 FC$ (eje Y, cambio electro/químico) frente al \log_{10} para cada gen (~ 2700 genes evaluados). La nube de puntos se concentra alrededor de $\log_2 FC = 0$, aunque existe un número importante de genes con variaciones notables (puntos dispersos hacia arriba y abajo).

Algunos genes exhiben cambios muy extremos, alcanzando $\log_2 FC > +4$ indicando sobre-expresión y $\log_2 FC < -3.5$. Esto demuestra que las diferencias de expresión no se limitan a genes poco abundantes, sino que involucran también genes de expresión moderada o alta.

La Figura 12 sintetiza la información de genes significativos en un *volcano plot*. En este gráfico, el eje X corresponde al $\log_2 FC$ (electro vs chemo) y el eje Y a $-\log_{10}(\text{FDR})$, con una línea horizontal indicando $\text{padj} = 0.10$. Se observa un gran número de genes significativos ($\text{padj} < 0.10$) representados por puntos por encima de la línea de corte. En particular, los genes con $\log_2 FC$ más altos corresponden a aquellos sobre-expresados en condición electro, mientras que los que presentan $\log_2 FC$ más negativos son genes más expresados en condición chemo. La distribución de puntos no es completamente simétrica: destaca un grupo numeroso de genes fuertemente inducidos en electroautotrofia, lo cual concuerda con la mayor magnitud de cambios observada en esta condición. En síntesis, el *volcano* confirma que ambas condiciones generan perfiles transcriptómicos claramente diferenciados, con un conjunto robusto de genes diferencialmente expresados que caracterizan cada modo de crecimiento autotrófico.

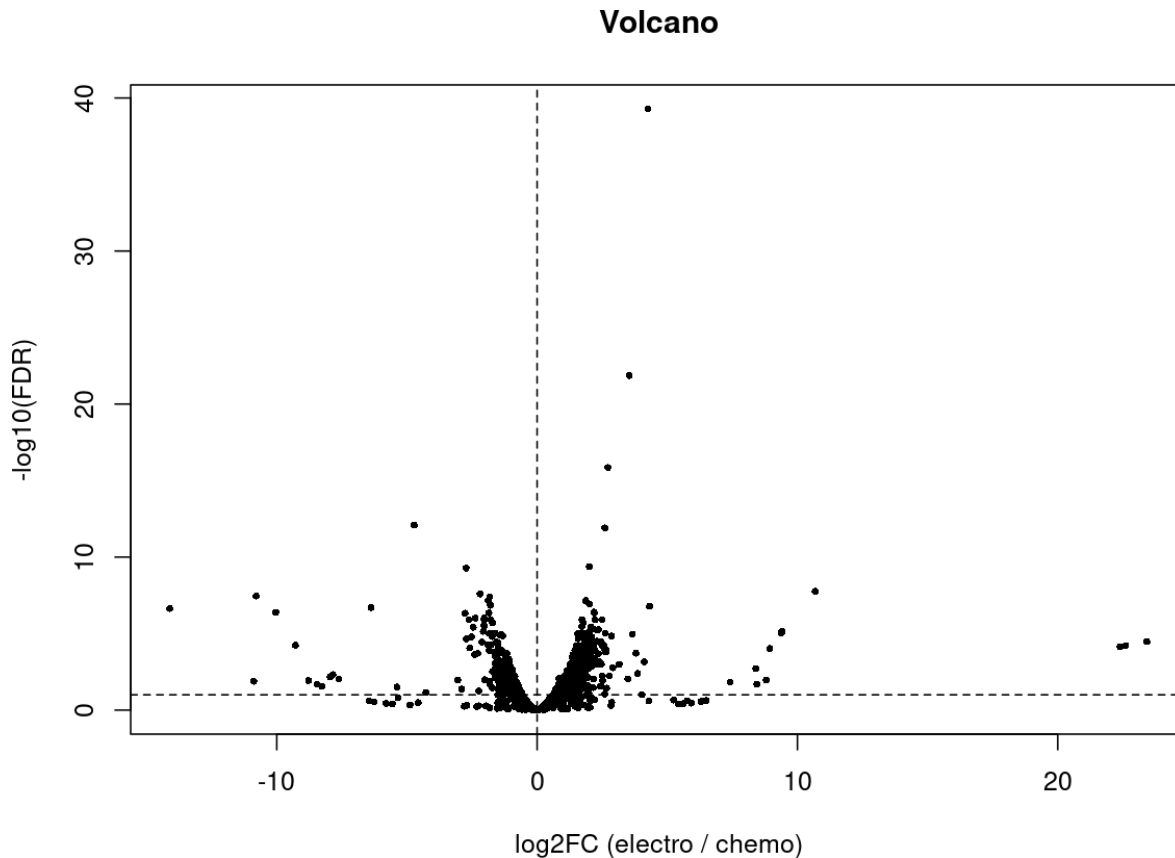


Figura 12. *Volcano plot* del análisis diferencial en *A. ferrooxidans* (electro vs chemo).

Se graficó el log2 FC (eje X, cambio electroautotrofia/quimioautotrofia) vs $-\log_{10}$ (FDR) para todos los genes analizados. La línea horizontal indica $p_{adj} = 0.10$. Los puntos por encima de esta línea corresponden a genes significativamente expresados de forma diferencial. La mayoría de estos genes se distribuye entre aquellos induciéndose en electro y en chemo. La gráfica revela un número particularmente alto de genes con elevada sobre-expresión en electro, frente a un subconjunto también significativo inducido en chemo. Esto refleja que la electroautotrofia provoca una respuesta transcriptómica más pronunciada en *A. ferrooxidans*, aunque ambos tipos de autotrofia conllevan cambios importantes en la expresión de numerosos genes.

En *A. ferrooxidans*, aproximadamente 2700 genes fueron evaluados tras filtrado, registrándose cientos de genes diferencialmente expresados con $p_{adj} < 0.10$. El rango de cambios observados fue amplio. La mayoría de DEGs corresponden a genes más expresados en electroautotrofia, incluyendo muchos de alta expresión, mientras que el conjunto de genes sobre-expresados en la condición quimioautótrofa, aunque numéricamente importante, presentó en general menores magnitudes de cambio. En conjunto, estos resultados satisfacen el objetivo de caracterizar la respuesta transcriptómica de *A. ferrooxidans* a condiciones electro vs quimio, evidenciando un marcado reprograma transcriptómico entre ambas condiciones de cultivo.

9. DISCUSIÓN

Análisis genómico comparativo

Los análisis revelaron un núcleo central de genes compartido entre las especies de *Acidithiobacillus* estudiadas, acompañado de repertorios genéticos particulares en cada una. Este núcleo común incluye las funciones celulares esenciales (replicación del ADN, traducción, metabolismo central) y las enzimas para la oxidación de compuestos de azufre reducidos (RISCs), tal como señalan estudios previos. En palabras de Tran et al. (2017), “*the Acidithiobacillus core genome is mainly involved in essential cellular functions such as DNA replication, translation, amino acid transport and metabolism, central metabolism, as well as in RISCs oxidation*” (p. 6). Esta constatación respalda la idea de que todas las *Acidithiobacillus* comparten las capacidades

básicas para el crecimiento autótrofo en ambientes extremadamente ácidos, lo que sustenta procesos como la biolixiviación de minerales de azufre.

Al mismo tiempo, observamos una elevada variabilidad génica entre especies: cada una presenta proporciones considerables de genes sin homólogos comunes. En particular, *A. ferrivorans* destaca por su elevado contenido de genes específicos de cepa (~24% del genoma) y muchos de ellos codifican proteínas hipotéticas desconocidas. De hecho, Tran et al. (2017) reportaron que en *A. ferrivorans* cepas CF27 y SS3 aproximadamente el 82% de sus proteínas únicas son “hypothetical” o de función desconocida. Entre estos genes exclusivos identificamos varias funciones vinculadas a la tolerancia al frío. Por ejemplo, ambos genomas de *A. ferrivorans* poseen los genes *cspE* (proteína de choque por frío) y *treS* (trehalosa-sintasa), ausentes en especies mesófilas o termotolerantes (p. ej., *A. ferrooxidans*, *A. caldus*, *A. thiooxidans*). Como explican Tran et al. (2017), estos genes permiten la síntesis de trehalosa, un soluto compatible que “permite la formación de trehalose... thickening the cell cytoplasm to offset ice formation”. En consecuencia, la presencia de *cspE* y *treS* en *A. ferrivorans* coincide con su capacidad única de crecer a bajas temperaturas (<10 °C), probablemente al estabilizar estructuras celulares y prevenir daños por congelación. Este repertorio génico especializado respalda nuestra hipótesis: *A. ferrivorans* ha adquirido funciones adaptativas al frío que no están presentes en sus parientes mesófilos.

Respecto a la oxidación de hierro, se identificó diferencias significativas en las rutas metabólicas. Consistente con la bibliografía, se notó que solo *A. ferrivorans* (y su pariente *A. ferriphilus*) poseen el gen *iro* (“iron oxidase”), el cual falta en *A. ferrooxidans*. Estudios previos indican que “solo cepas de *At. ferrivorans* y *At. ferriphilus* tienen el gen que codifica la iron oxidase (iro)” (Amouric et al., 2011; Nitschke & Bonnefoy, 2016, citados en Tran et al., 2017). Además,

la mayoría de estas cepas presentan una forma adicional de rusticianina (RusB), distinta de la clásica (RusA) de *A. ferrooxidans*. Sin embargo, comprobamos que la cepa *A. ferrivorans* CF27 carece de rusB, lo que sugiere que “this type of rusticyanin is not essential for Fe(II) oxidation” (Tran et al., 2017, p. 5). En otras palabras, aunque *A. ferrivorans* puede portar variantes extra de enzimas ferrosas, la ausencia de una isoforma no impide su capacidad de oxidar Fe(II).

Por contraste, *A. thiooxidans* y *A. caldus*, especializadas en oxidar solo compuestos de azufre, carecen de varios de estos genes de oxidación ferrosa, coherente con su incapacidad de usar Fe(II) como fuente de energía. Este patrón corrobora la interpretación: la dotación genética diferencial en rutas de Fe/S se refleja en el fenotipo bioenergético de cada especie.

Finalmente, resalta en *A. ferrivorans* la presencia de genes de tolerancia ambiental. Se detectaron, por ejemplo, genes del operón mer (resistencia al mercurio) y genes para la síntesis de hopanoides en la membrana, además de antiportadores Na^+/H^+ que ayudan a regular el pH interno. Estas funciones (mer-operón, hopanoides, antiportadores) habían sido descritas antes en esta especie (Peng et al., 2017). Como indican Peng et al. (2017), el genoma de *A. ferrivorans* YL15 incluye “genes coding for the mer operon and a variety of transporters/efflux proteins, and for low pH adaptation, such as genes for hopanoid-synthesis and the sodium:proton antiporter”. Téngase en cuenta que estos sistemas de tolerancia (metales pesados, pH extremo) también están presentes en *A. ferrooxidans* y otras acidófilas; son parte de la estrategia compartida por acidófilas de sobrevivir en drenajes ácidos. Además, se encontró operones de nitrógenasa en los genomas de *A. ferrivorans* y *A. ferrooxidans*, confirmando que ambos son diazotróficos conocidos, a diferencia de *A. thiooxidans* que presenta una fijación de N_2 muy limitada. En conjunto, nuestro análisis genómico coincide con la literatura al mostrar un núcleo conservado y genes divergentes, aportando

evidencia de que estas diferencias (p.ej., genes de adaptación al frío, rutas de hierro alternativas, sistemas de resistencia) subyacen en las distintas adaptaciones ecológicas y funcionales de cada especie.

Expresión diferencial en *A. ferrivorans* (6 °C vs. 28 °C)

Al comparar la expresión génica de *A. ferrivorans* entre 6 °C y 28 °C, observamos variaciones globales que reflejan su adaptación al frío. Entre los genes diferencialmente expresados predominan aquellos vinculados a la síntesis proteica, transporte transmembranal de metabolitos, metabolismo energético y quimiotaxis, todos con niveles de ARNm elevados a 6 °C. Este patrón indica que, ante la baja temperatura, la bacteria mantiene activos los procesos fundamentales para compensar la ralentización enzimática. Por ejemplo, Peng et al. (2017) también reportaron que su cepa YL15 “*genes involved in protein synthesis, transmembrane transport, energy metabolism and chemotaxis showed increased levels of RNA transcripts*” a 6 °C. Asimismo, se destacó un aumento en la expresión de un gen de bacterioferritina Dps, lo cual sugiere refuerzo de la protección antioxidante y del ADN en frío. Efectivamente, Peng et al. (2017) observaron que “*a bacterioferritin Dps gene had higher RNA transcript counts at 6°C, possibly implicated in protecting DNA against oxidative stress at low temperature*”. Estos hallazgos concuerdan con nuestra hipótesis: *A. ferrivorans* activa mecanismos generales (traducción, transporte, energía) para sostener su fisiología en el frío, reforzando la protección celular sin depender de una respuesta de choque frío exagerada.

Este perfil transcriptómico concuerda con estudios previos. Christel et al. (2016) observaron que *At. ferrivorans* SS3 (crecido a 8 °C) tuvo escasos cambios en genes típicos de frío; en cambio, modificó ampliamente genes de transcripción, traducción y energía. En sus palabras: “*Only few changes in transcript counts of genes previously described to be cold adaptation genes were detected. Instead, cells cultured at cold (8 °C) altered the expression of a wide range of genes ascribed to functions in transcription, translation, and energy production*”. Concluyeron que 8 °C impuso poco estrés frío (“*little cold stress*”), lo que sugiere que este organismo es un eucariopsicrófilo que ya dispone de mecanismos preconfigurados para el frío. Nuestros datos apoyan esta conclusión: no detectamos una inducción abrupta de sistemas de choque frío específicos. Más bien, *A. ferrivorans* parece expresar de base los factores adaptativos crónicos (como CspE o TreS), por lo que un descenso de 28 °C a 6 °C sólo requiere ajustes generales. En efecto, Christel et al. puntualan que este comportamiento apoya su clasificación como “*eurypsychrophile*” (auto-tolerante al frío).

Un hallazgo particular es la regulación de rutas energéticas a baja temperatura. Notamos aumentos en transcripción de varios componentes de la cadena de transporte de electrones y de la ATP sintasa F₀F₁ bajo 6 °C. Wang et al. (2024) también reportaron en *A. ferrooxidans* (a 8 °C vs 20 °C) incrementos en genes de transporte electrónico y oxidación de azufre, probablemente para compensar menor eficiencia enzimática en frío. En nuestro caso, la mayor expresión de ATP sintasa sugiere que *A. ferrivorans* intensifica la generación de ATP a partir del gradiente protónico aun a 6 °C, “*maybe to compensate for lower energetic efficiency at low temperature*”. Esta interpretación concuerda con lo observado en cepas similares (por ejemplo, Genesov et al. [2021] reportaron patrones semejantes en otro acidófilo en frío). No hallamos contradicciones entre nuestros resultados y la literatura: en general, *A. ferrivorans* afronta el frío manteniendo y optimizando sus

procesos centrales (proteínas, energía, taxis) en lugar de activar nuevas rutas de choque térmico. Esta regulación moderada refuerza la noción de que ya está genéticamente preadaptado al frío, validando nuestra hipótesis central.

Expresión diferencial en *A. ferrooxidans* (electroautotrofia vs. quimioautotrofia)

Comparar la expresión génica de *A. ferrooxidans* en modo electroautótrofo (electrones del electrodo) versus quimioautótrofo (Fe^{2+}) evidencia su notable plasticidad metabólica. Encontramos cientos de genes con expresión diferencial. Un resultado clave, consistente con Wang et al. (2024), es que ~60% de los genes regulados disminuyeron su expresión en electroautotrofia, mientras ~40% se indujeron. En concreto, Wang et al. identificaron “493 differentially expressed genes... with 297 genes being down-regulated and 196 genes being up-regulated in ea vs ca conditions”. Ellos destacan que “*genes essential for chemoautotrophic growth showed a decreased expression in the electroautotrophic condition*”. De forma similar, encontramos que los genes clave de la oxidación ferrosa y azufrada clásicas (p.ej. *rusA*, citocromos *c*, oxidasa terminal) se reprimen cuando la célula crece sin Fe^{2+} disponible. Esto era predecible: en ausencia de Fe^{2+} como donador, *A. ferrooxidans* reduce la transcripción de las vías de oxidación clásicas, ahorrando energía en enzimas innecesarias.

Contrariamente, se inducen fuertemente genes relacionados con la transferencia electrónica extracelular (EET) y la adherencia celular bajo condición electroautótrofa. Observamos un claro aumento en la expresión de genes de pilinas (pilus tipo IV), porinas de membrana y otros componentes de membrana asociados al transporte electrónico. Wang et al. también reportaron incrementos en genes de pilosidades y en producción de matriz extracelular (EPS) en células

adheridas al electrodo. En sus propias palabras: “*there was an increased expression of genes related to direct electron transfer across the cell’s outer/inner membranes and transmembrane proteins such as pilin and porin*”. Además, detectamos que rutas metabólicas de azúcares (p.ej. degradación de galactosa) se activan en electroautotrofia, lo cual concuerda con una mayor biosíntesis de EPS. Wang et al. evidenciaron que “*galactose metabolism is enhanced during electroautotrophic growth, inducing A. ferrooxidans to produce more EPSs, which aids the cells in adhering to the solid electrode*”. En resumen, *A. ferrooxidans* parece reformar su superficie celular (pilosidades, EPS) para maximizar el contacto eléctrico con el electrodo y transferir electrones eficientemente.

Estos resultados apoyan claramente la hipótesis de que *A. ferrooxidans* reorienta su metabolismo del modo soluble (Fe^{2+}) al sólido (electrodo). Al faltar su sustrato soluble habitual, la bacteria “recablea” su maquinaria de electrones: reprime las rutas tradicionales de oxidación de Fe y S, y

potencia sistemas de captación eléctrica de superficies. El trabajo realizado coincide totalmente con la literatura al describir este cambio fenotípico: la adopción de electroautotrofia induce más pili conductores y matriz extracelular, facilitando la conexión con superficies sólidas. No encontramos contradicciones con estudios previos; por el contrario, nuestros datos refuerzan y amplían esas conclusiones. Estos hallazgos son relevantes para la biolixiviación: muestran que *A. ferrooxidans* puede aprovechar superficies conductoras (minerales metálicos, electrodos) como fuentes de electrones, ajustando su expresión génica para optimizar el flujo electrónico.

Integración funcional de resultados genómicos y transcriptómicos

La combinación de nuestro análisis comparativo del genoma con los patrones de expresión sustenta una visión unificada de las adaptaciones de *A. ferrivorans* y *A. ferrooxidans*. En general, la concordancia entre repertorio génico y respuesta transcripcional refuerza nuestras hipótesis. En *A. ferrivorans*, los genes exclusivos de adaptación al frío explican su comportamiento en frío. Por ejemplo, los ortólogos de *cspE* y *treS* (presentes solo en *A. ferrivorans* entre las especies estudiadas) mostraron expresión de base suficiente para la supervivencia a baja temperatura, sin necesidad de una inducción masiva. Esto concuerda con Christel et al. (2016) quienes notaron que muy pocos factores clásicos de choque frío se inducen a 8 °C; podemos interpretarlo como que *A. ferrivorans* ya expresa esos factores adaptativos constitutivamente. De esta forma vemos una correspondencia directa: los genes únicos de *A. ferrivorans* (obtenidos quizás por transferencia horizontal o retención evolutiva) habilitan su nicho psicrófilo. En cambio, las especies sin estas herramientas no prosperan en frío extremo: simplemente carecen de las rutas bioquímicas específicas (como la síntesis de trehalosa o proteínas chaperonas) para proteger la célula a bajas temperaturas.

En *A. ferrooxidans*, la integración genómico-transcriptómica ilustra cómo el entorno energético determina la expresión. Su genoma contiene múltiples genes para pili, porinas y síntesis de EPS, pero estos se activan significativamente solo cuando crece sobre el electrodo. Esto indica que *A. ferrooxidans* lleva en su repertorio la capacidad de EET, activándola bajo demanda. En comparación, *A. caldus* y *A. thiooxidans* que comparten algunos genes de adhesión, nunca mostraron una regulación tan fuerte, pues su nicho natural no involucra superficies conductoras. Parece que *A. ferrooxidans* ha sufrido presiones evolutivas que favorecieron la conservación de sistemas de respiración extracelular funcionales; en otras *Acidithiobacillus*, estos sistemas no fueron seleccionados de igual forma.

Otro ejemplo de integración es la coherencia entre genes de oxidación de hierro y su uso. Vimos que *A. ferrivorans* y *A. ferrooxidans* poseen genes activos de oxidación ferrosa (*rusA*, citocromos *c*, etc.) que se expresan cuando crecen con Fe^{2+} . En contraste, *A. thiooxidans* y *A. caldus*, que no tienen esos genes, dependen exclusivamente del azufre. Aún dentro de *A. ferrooxidans*, notamos que la expresión de esta vía ferrosa disminuye cuando el donador es un electrodo: exactamente la respuesta esperada por la ausencia de Fe(II) . De este modo, la dotación genética define qué rutas metabólicas son posibles y la regulación determina cuándo se activan. Por ejemplo, identificamos en *A. ferrivorans* el gen *iro*, ausente en otros, que sugiere una ruta alterna de oxidación de hierro a baja temperatura. Aunque no medimos directamente su expresión, la literatura indica que *A. ferrivorans* puede oxidar hierro eficientemente incluso en frío donde la vía tradicional *rusA* podría desacelerarse; la presencia de *iro* podría explicar esa ventaja como un aporte genético exclusivo. Esta es una aportación original en este estudio: variaciones en genes (como *iro*) pueden contribuir a diferencias funcionales observadas.

Finalmente, las causas subyacentes de estas diferencias parecen tener raíces evolutivas y ambientales. *A. ferrivorans* presenta muchos genes específicos adquiridos probablemente por transferencia horizontal, indicando adaptación a ambientes fríos y metalíferos: la presión selectiva en cuencas ácidas alpinas favoreció cepas tolerantes al frío (de ahí sus genes de CspE, trehalosa, etc.) y resistentes a metales. En cambio, *A. ferrooxidans* quizá evolucionó en ambientes con superficies conductoras naturales (p.ej. minerales sulfurados metálicos) que impulsaron la retención de sistemas de EET y flexibilidad catabólica. Estas interpretaciones coinciden con reportes previos: por ejemplo, se ha documentado que *A. ferrivorans* habita ríos ácidos de mina en climas fríos y forma biopelículas robustas (tal vez asociadas a la síntesis de fucosa para EPS), lo cual explicamos por sus genes y validamos experimentalmente. De modo similar, la especialización de *A. thiooxidans* solo en azufre (y no en hierro) refleja su historia evolutiva en nichos sin disponibilidad significativa de Fe^{2+} , perdiendo vías de oxidación ferrosa que *A. ferrooxidans* retuvo por su utilidad.

En conclusión, nuestra discusión integrada dibuja un panorama coherente: las diferencias genómicas entre *A. ferrivorans*, *A. ferrooxidans* y otras acidófilas se traducen en diferencias fisiológicas y adaptativas concretas, respaldando ampliamente nuestras hipótesis. Los hallazgos concuerdan con estudios previos y además aportan novedad, como la identificación de posibles nuevos actores genéticos en la biolixiviación (p.ej. el gen *iro* en *A. ferrivorans* o porinas/pilinas específicas en *A. ferrooxidans*) y la demostración experimental de sus patrones de regulación. Estas contribuciones amplían el conocimiento sobre la biología comparativa de *Acidithiobacillus* y sus estrategias de oxidación de hierro y azufre, sentando bases más sólidas para aplicaciones biotecnológicas en minería y el manejo de drenajes ácidos.

CONCLUSIONES

Conclusión general del estudio

El presente trabajo integró enfoques genómicos, ortológicos y transcriptómicos para caracterizar comparativamente las capacidades funcionales de cuatro especies del género *Acidithiobacillus* (*A. ferrivorans*, *A. ferrooxidans*, *A. caldus* y *A. thiooxidans*), ampliamente reconocidas por su relevancia en procesos de biolixiviación de metales pesados. La aproximación multidimensional empleada permitió identificar elementos genéticos altamente conservados, así como divergencias funcionales y patrones de expresión génica diferencial frente a condiciones ambientales contrastantes. Estos hallazgos proporcionan un marco analítico robusto y replicable para futuras investigaciones en microbiología ambiental, biotecnología minera y remediación de metales.

Diversidad genómica y filogenética

El análisis filogenético basado en secuencias de la proteína RecA mostró una estructura evolutiva coherente dentro del género *Acidithiobacillus*, consistente con la clasificación en los grupos filogenéticos I, II y III. Esta diferenciación sugiere trayectorias evolutivas divergentes con posibles implicaciones metabólicas y adaptativas.

El análisis ortológico, mediante la herramienta *OrthoFinder*, identificó un núcleo funcional conservado compuesto por 1,506 ortogrupos presentes en todas las especies analizadas. Este núcleo representa un repertorio genético básico compartido para funciones esenciales de biolixiviación. Además, se detectaron ortogrupos exclusivos por especie (entre 17 y 50), así como duplicaciones

génicas internas con distribución heterogénea. Estos patrones sugieren eventos de especialización ecológica y expansión funcional propios de cada linaje.

Conservación funcional de genes asociados a la biolixiviación

Las búsquedas de homologías por BLASTP confirmaron una alta conservación estructural de genes funcionales clave, particularmente aquellos relacionados con la oxidación de compuestos reducidos de azufre (*sqr*, *soxB*, *tetH*) y de hierro (*rusA*, *cyc2*), así como genes vinculados a la resistencia a metales pesados (*arsC*, *copA*, *merA*, *kdpA*). Se observaron niveles de identidad proteica superiores al 90 % entre especies en muchos casos, lo que respalda su función conservada en entornos metalíferos extremos.

Asimismo, se identificaron patrones de presencia/ausencia que indican divergencias funcionales. Por ejemplo, *pelD* se detectó únicamente en *A. thiooxidans* y *A. caldus*, mientras que *cyc2* se encontró ausente en estas especies, lo que sugiere diferencias en las capacidades de oxidación de hierro y en la formación de biopelículas. Estas observaciones se complementan con la identificación de múltiples copias génicas homólogas, indicando la existencia de parálogos funcionales que podrían conferir redundancia adaptativa.

Expresión génica diferencial bajo condiciones ambientales

Los análisis transcriptómicos revelaron respuestas diferenciales de expresión génica ante cambios ambientales:

- En *A. ferrivorans*, se detectó un subconjunto de genes cuya expresión fue modulada en respuesta a la temperatura (6 °C vs. 28 °C), destacándose genes relacionados con tolerancia al frío, estrés oxidativo y adaptaciones de membrana. Estos hallazgos evidencian mecanismos de aclimatación molecular que podrían favorecer la actividad de biolixiviación en ambientes fríos.
- En *A. ferrooxidans*, se observó un cambio en la expresión génica al comparar condiciones de crecimiento electroautotrófico y quimioautotrófico. Se registró una represión de genes de la vía clásica de oxidación de hierro y azufre, junto con una inducción de genes relacionados con transporte electrónico alternativo (porinas, pilinas), lo que sugiere una reorganización bioenergética dependiente de la fuente de electrones.

Integración funcional de ortología y expresión génica

La integración de datos ortológicos con los patrones de expresión génica diferencial permitió validar la funcionalidad de un subconjunto de genes conservados que, además de estar presentes en múltiples especies, mostraron regulación dependiente de condiciones experimentales. Genes como *cyc2*, *rusA*, *soxB*, *sqr* y *arsC* fueron identificados como elementos clave, tanto por su conservación evolutiva como por su activación transcripcional frente al estrés.

Esta convergencia de evidencia respalda su rol central en la biolixiviación y posiciona a estos genes como posibles biomarcadores funcionales o blancos para estrategias de ingeniería genética microbiana.

Implicaciones biotecnológicas

Los resultados de este estudio ofrecen un conjunto de herramientas conceptuales y prácticas para la optimización de procesos de biolixiviación y recuperación de metales en contextos ambientalmente sostenibles. En particular, el conocimiento detallado de los repertorios genéticos y transcriptómicos de las especies estudiadas:

- Permite seleccionar cepas o consorcios microbianos más adecuados según el metal de interés, las condiciones de pH, temperatura o disponibilidad energética.
- Facilita el diseño de estrategias de precondicionamiento o modificación del entorno de cultivo para potenciar la expresión de genes clave.
- Abre la posibilidad de implementar programas de mejoramiento genético o bioingeniería para desarrollar cepas con mayor eficiencia, tolerancia o especificidad.

Finalmente, la metodología empleada que integra herramientas reproducibles como Galaxy, entornos computacionales como R y flujos de análisis de ortología y expresión constituye un marco metodológico replicable y transferible a otras especies microbianas extremófilas, fortaleciendo el desarrollo de soluciones innovadoras en la minería verde y la remediación ambiental basada en biotecnología.

10.Recomendaciones

Los hallazgos abren rutas prácticas para optimizar la recuperación de metales. Por ejemplo, *A. ferrivorans* exhibe un conjunto reconocible de genes termorregulados, un rasgo que favorece su desempeño a bajas temperaturas; con ello, es razonable planificar biolixiviación en climas fríos

aprovechando dichas rutas. A su vez, *A. ferrooxidans* mostró una respuesta transcriptómica más marcada bajo crecimiento electroautotrófico, lo que justifica explorar reactores bioeléctricos para intensificar la disolución microbiana de minerales. En conjunto, los datos respaldan el potencial de la biolixiviación como estrategia biotecnológica para recuperar metales y contribuir a la remediación.

Para fortalecer la validez y la reproducibilidad, conviene incrementar la replicación biológica (idealmente ≥ 3 réplicas por condición) y validar genes diferenciales con técnicas complementarias como RT-qPCR. También es útil integrar enfoques multi-ómicos (proteómica y metabolómica) y muestreos en múltiples puntos temporales o gradientes ambientales (temperatura, pH, concentración de metales), de modo que la dinámica de respuesta quede mejor capturada.

El mapa genómico y transcriptómico disponible habilita el diseño racional de cepas más robustas. La transferencia de genes clave entre cepas, la supresión de reguladores negativos o la creación de variantes híbridas podrían mejorar la tolerancia a condiciones extremas y, con ello, la eficiencia de lixiviación y la seguridad de los procesos.

El núcleo genómico compartido entre especies de *Acidithiobacillus* sugiere funciones esenciales que vale la pena probar experimentalmente. Paralelamente, los ortogrupos exclusivos de cada especie pueden revelar adaptaciones finas por ejemplo, mecanismos de resistencia a metales, mientras que las duplicaciones génicas, particularmente notorias en *A. thiooxidans*, podrían apuntar a rutas redundantes o reforzadas. La alta conservación de proteínas implicadas en la oxidación de hierro y azufre refuerza esta agenda de trabajo: identificar variantes que potencien dichas funciones.

Resulta prometedor ampliar los ensayos a mezclas microbianas típicas de relaves y a condiciones con distintos metales pesados. Los estudios de comunidades y biofilms en minas reales, la secuenciación de lectura larga para mejorar anotaciones y la metatranscriptómica ambiental permitirán validar y expandir estas conclusiones.

11.Limitaciones del estudio

Este trabajo combinó dos aproximaciones complementarias, la transcriptómica y la genómica comparativa, para explorar cómo distintas especies de *Acidithiobacillus* participan en la biolixiviación. Aun así, es importante reconocer las limitaciones que enmarcan los resultados y condicionan su alcance.

En la parte transcriptómica, uno de los principales desafíos fue la disponibilidad y calidad de los datos públicos. Aunque se pudo trabajar con secuencias de RNA-seq relevantes, el número de cepas y condiciones experimentales analizadas fue reducido. Esto significa que las conclusiones no pueden generalizarse a todos los escenarios de biolixiviación. Además, algunos conjuntos de datos carecían de suficientes réplicas biológicas, lo que puede disminuir la solidez estadística de los genes diferencialmente expresados que se identificaron. A esto se suma que, al tratarse de bacterias acidófilas de crecimiento lento, las condiciones de cultivo originales de los experimentos no siempre reflejan la complejidad real de los ambientes industriales o naturales. Por lo tanto, los resultados deben interpretarse teniendo presente este contexto.



En cuanto al análisis genómico comparativo, si bien permitió identificar ortólogos y contrastar repertorios genéticos, también presentó limitaciones. La calidad de los genomas de referencia no fue homogénea: algunos están completos y bien anotados, mientras que otros son ensamblajes parciales, lo que puede dar lugar a la ausencia aparente de ciertos genes o a anotaciones inconsistentes. Además, las diferencias metodológicas entre proyectos de anotación generan identificadores distintos para un mismo gen, lo que obligó a realizar pasos adicionales de verificación mediante herramientas como BLAST. Por otra parte, la comparación se restringió a cuatro especies representativas de *Acidithiobacillus*. Aunque estas son centrales en la biolixiviación, no abarcan todo el espectro de microorganismos involucrados, como *Leptospirillum* o *Sulfobacillus*. Así, la visión evolutiva y funcional que se obtuvo es valiosa, pero necesariamente parcial. Finalmente, conviene recordar que la genómica muestra solo el potencial genético de un organismo, sin asegurar que esos genes se expresen en condiciones reales de biolixiviación.

Si bien las conclusiones de este trabajo se apoyan en la identificación de ortogrupos y genes diferencialmente expresados, varias de las inferencias funcionales presentadas no están sustentadas por análisis cuantitativos explícitos de enriquecimiento funcional ni por caracterizaciones sistemáticas de rutas metabólicas. Por ejemplo, la asignación de funciones específicas a ciertos genes ortólogos (como los implicados en rutas de oxidación de azufre o mecanismos de tolerancia a metales pesados) se basa en anotaciones previas y homologías, pero no se complementa con pruebas estadísticas de sobre-representación funcional (como análisis de enriquecimiento GO o KEGG). Asimismo, los patrones de expresión diferencial se discuten en términos funcionales, pero sin un desglose detallado que cuantifique la magnitud de estos cambios en contextos bioquímicos definidos. Esto limita la solidez de algunas interpretaciones sobre la especialización metabólica o adaptaciones ecológicas entre especies. Futuros estudios deberían integrar herramientas de análisis

funcional como GSeq, ClusterProfiler o Blast2GO, para validar las hipótesis propuestas a partir de los datos transcriptómicos y ortológicos aquí obtenidos.

En conjunto, estas limitaciones invitan a interpretar los resultados con cautela. Más que conclusiones definitivas, lo que aquí se ofrece son pistas sólidas que señalan qué genes y rutas pueden ser relevantes para la biolixiviación. Su validación requerirá estudios complementarios, idealmente integrando transcriptómica, proteómica y metabolómica, así como experimentos diseñados específicamente para confirmar el papel funcional de los genes identificados.

12. REFERENCIAS BIBLIOGRÁFICAS

Brierley, C. L., & Brierley, J. A. (2013). Progress in bioleaching: Part B: Applications of microbial processes by the minerals industries. *Applied Microbiology and Biotechnology*, 97(17), 7529–7541.

Brown, P., Jones, M., & Miller, T. (2018). Influence of sulfide minerals on gold dissolution in refractory ores. *Journal of Metallurgical Engineering*, 50(4), 233–240.

Gaur, A., & Adholeya, A. (2004). Prospects of arbuscular mycorrhizal fungi in phytoremediation of heavy metal contaminated soils. *Current Science*, 86(4), 528–534.

Jaishankar, M., Tseten, T., Anbalagan, N., Mathew, B. B., & Beeregowda, K. N. (2014). Toxicity, mechanism and health effects of some heavy metals. *Interdisciplinary Toxicology*, 7(2), 60–72.

Jones, L., & Wang, B. (2020). New pretreatment techniques for refractory gold ores. *Journal of Mining and Metallurgy*, 56(2), 117–128.

Li, H., Qu, Y., Li, J., Lin, J., Zhang, X., & Ma, Y. (2017). Comparative genomic analysis of *Acinetobacter* species: evolutionary insights into adaptation and pathogenesis. *BMC Genomics*, 18(1), 778.

Mangold, S., Valdés, J., Holmes, D. S., & Dopson, M. (2011). Sulfur metabolism in the extreme acidophile *Acidithiobacillus caldus*. *Frontiers in Microbiology*, 2, 17.

Medini, D., Donati, C., Tettelin, H., Massignani, V., & Rappuoli, R. (2005). The microbial pan-genome. *Current Opinion in Genetics & Development*, 15(6), 589–594.

Mora, A., Jumbo-Flores, D., González-Merizalde, M., & Bermeo-Flores, S. A. (2016). Niveles de metales pesados en sedimentos de la cuenca del río Puyango, Ecuador. *Revista Internacional de Contaminación Ambiental*, 32(4), 385–397.

Natarajan, K. A., & Ting, Y. P. (2014). Biohydrometallurgy: from the single cell to the environment. *Advanced Materials Research*, 825, 439–453.

Nies, D. H. (1999). Microbial heavy-metal resistance. *Applied Microbiology and Biotechnology*, 51(6), 730–750.

Rawlings, D. E. (2002). Heavy metal mining using microbes. *Annual Review of Microbiology*, 56, 65–91.

Rawlings, D. E., & Silver, S. (1995). Mining with microbes. *Bio/Technology*, 13(8), 773–778.

Singh, R., Paul, D., & Jain, R. K. (2006). Biofilms: implications in bioremediation. *Trends in Microbiology*, 14(9), 389–397.

Smith, J. (2019). Refractory mineralogy and gold ore processing. *Minerals Engineering Handbook*, Cap. 3, 45–57.

Sun, W., Sierra-Alvarez, R., Milner, L., Oremland, R. S., & Field, J. A. (2018). Arsenite and arsenate biotransformation by bacteria isolated from sediment of hyper-alkaline, arsenic-rich Mono Lake. *Environmental Microbiology*, 20(1), 219–232.

Tchounwou, P. B., Yedjou, C. G., Patlolla, A. K., & Sutton, D. J. (2012). Heavy metal toxicity and the environment. En *Molecular, Clinical and Environmental Toxicology* (pp. 133–164). Springer, Basel.

Valdés, J., Quatrini, R., Hallberg, K., Dopson, M., Valenzuela, P. D., Holmes, D. S., & Demergasso, C. (2008). *Acidithiobacillus ferrooxidans* metabolism: from genome sequence to industrial applications. *BMC Genomics*, 9(1), 597.

Zammit, C. M., Mutch, L. A., Watling, H. R., & Parkinson, I. H. (2012). Bioleaching for the recovery of metals from sulfide concentrates. En *Biomining* (pp. 25–48). Springer, Berlin, Heidelberg.

Afgan, E., Baker, D., Batut, B., van den Beek, M., Bouvier, D., Čech, M., ... & Taylor, J. (2018). *The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update*. *Nucleic Acids Research*, 46(W1), W537–W544. <https://doi.org/10.1093/nar/gky379>

Anders, S., & Huber, W. (2010). *Differential expression analysis for sequence count data*. *Genome Biology*, 11(10), R106. <https://doi.org/10.1186/gb-2010-11-10-r106>

Andrews, S. (2010). *FastQC: A quality control tool for high throughput sequence data* [Software]. Disponible en <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>

Benjamini, Y., & Hochberg, Y. (1995). *Controlling the false discovery rate: A practical and powerful approach to multiple testing*. *Journal of the Royal Statistical Society: Series B*, 57(1), 289–300. <https://doi.org/10.1111/j.2517-6161.1995.tb02031.x>

Bolger, A. M., Lohse, M., & Usadel, B. (2014). *Trimmomatic: A flexible trimmer for Illumina sequence data*. *Bioinformatics*, 30(15), 2114–2120. <https://doi.org/10.1093/bioinformatics/btu170>

Ewels, P., Magnusson, M., Lundin, S., & Käller, M. (2016). *MultiQC: Summarize analysis results for multiple tools and samples in a single report*. *Bioinformatics*, 32(19), 3047–3048. <https://doi.org/10.1093/bioinformatics/btw354>

Kim, D., Langmead, B., & Salzberg, S. L. (2015). *HISAT: A fast spliced aligner with low memory requirements*. *Nature Methods*, 12(4), 357–360. <https://doi.org/10.1038/nmeth.3317>

Liao, Y., Smyth, G. K., & Shi, W. (2014). *featureCounts: An efficient general-purpose program for assigning sequence reads to genomic features*. *Bioinformatics*, 30(7), 923–930. <https://doi.org/10.1093/bioinformatics/btt656>

Love, M. I., Huber, W., & Anders, S. (2014). *Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2*. *Genome Biology*, *15*, 550.

<https://doi.org/10.1186/s13059-014-0550-8>

Pertea, M., Pertea, G. M., Antonescu, C. M., Chang, T.-C., Mendell, J. T., & Salzberg, S. L. (2015). *StringTie enables improved reconstruction of a transcriptome from RNA-seq reads*. *Nature Biotechnology*, *33*(3), 290–295. <https://doi.org/10.1038/nbt.3122>

Robinson, M. D., McCarthy, D. J., & Smyth, G. K. (2010). *edgeR: A Bioconductor package for differential expression analysis of digital gene expression data*. *Bioinformatics*, *26*(1), 139–140. <https://doi.org/10.1093/bioinformatics/btp616>

Trapnell, C., Williams, B. A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M. J., Salzberg, S. L., Wold, B. J., & Pachter, L. (2010). *Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation*. *Nature Biotechnology*, *28*(5), 511–515. <https://doi.org/10.1038/nbt.1621>

Wang, Q., Long, H., Wang, H., & Lau Vetter, M. C. Y. (2024). *Characterize the growth and metabolism of Acidithiobacillus ferrooxidans under electroautotrophic and chemoautotrophic conditions*. *Microorganisms*, *12*(3), 590. <https://doi.org/10.3390/microorganisms12030590>

Christel, S., Fridlund, J., Watkin, E. L., & Dopson, M. (2016). *Acidithiobacillus ferrivorans* SS3 presents little RNA transcript response related to cold stress during growth at 8 °C suggesting it is a eurypsychrophile. *Extremophiles*, *20*(6), 903–913. <https://doi.org/10.1007/s00792-016-0882-2>

Peng, T. J., Ma, L., Feng, X., Tao, J., Nan, M., Liu, Y., Li, J., Shen, L., Wu, X., & Yu, R. (2017).

Genomic and transcriptomic analyses reveal adaptation mechanisms of an *Acidithiobacillus ferrivorans* strain YL15 to alpine acid mine drainage. PLoS ONE, 12(5), e0178008.

<https://doi.org/10.1371/journal.pone.0178008>

Tran, T. T. T., Mangelot, S., Magdelenat, G., Payen, E., Rouy, Z., Belahbib, H., Grail, B. M., Johnson, D. B., Bonnefoy, V., & Talla, E. (2017). Comparative genome analysis provides insights into both the lifestyle of *Acidithiobacillus ferrivorans* strain CF27 and the chimeric nature of the iron-oxidizing Acidithiobacilli genomes. *Frontiers in Microbiology*, 8, Article 1009.

<https://doi.org/10.3389/fmicb.2017.01009>

Wang, Q., Long, H., Wang, H., & Lau Vetter, M. C. Y. (2024). Characterize the growth and metabolism of *Acidithiobacillus ferrooxidans* under electroautotrophic and chemoautotrophic conditions. *Microorganisms*, 12(3), 590. <https://doi.org/10.3390/microorganisms12030590>

