

RESEARCH ARTICLE

Codon usage bias reveals genomic adaptations to environmental conditions in an acidophilic consortium

Andrew Hart¹, María Paz Cortés^{2,3}, Mauricio Latorre^{2,3,4,5*}, Servet Martínez^{6*}

1 UMI 2071 CNRS-UCHILE, Facultad de Ciencias Físicas y Matemáticas, Centro de Modelamiento Matemático, Universidad de Chile, Casilla 170, Correo 3, Santiago, Chile, **2** Mathomics, Centro de Modelamiento Matemático, Universidad de Chile, Santiago, Chile, **3** Fondap-Center of Genome Regulation, Facultad de Ciencias, Universidad de Chile, Santiago, Chile, **4** Laboratorio de Bioinformática y Expresión Génica, INTA, Universidad de Chile, Macul, Santiago, Chile, **5** Universidad de O'Higgins, Instituto de Ciencias de la Ingeniería, Rancagua, Chile, **6** Departamento de Ingeniería Matemática, UMI 2071 CNRS-UCHILE, Facultad de Ciencias Físicas y Matemáticas, Centro de Modelamiento Matemático, Universidad de Chile, Casilla 170, Correo 3, Santiago, Chile

* mauricio.latorre@uoh.cl (ML); smartine@dim.uchile.cl (SM)



OPEN ACCESS

Citation: Hart A, Cortés MP, Latorre M, Martínez S (2018) Codon usage bias reveals genomic adaptations to environmental conditions in an acidophilic consortium. PLoS ONE 13(5): e0195869. <https://doi.org/10.1371/journal.pone.0195869>

Editor: Ulrich Melcher, Oklahoma State University, UNITED STATES

Received: December 21, 2017

Accepted: March 30, 2018

Published: May 9, 2018

Copyright: © 2018 Hart et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Funding: This work was supported by FONDAP Grant 15090007 of the Center for Genome Regulation (CGR), Apoyo a la Formación de Redes Internacionales para Investigadores REDI170193 and FONDECYT Grant N° 11150679 to Mauricio Latorre. CONICYT Basal Grant PFB 03 of the Center for Mathematical Modeling (CMM) UMI2807 UCHILE-CNRS to Servet Martínez and Mauricio

Abstract

The analysis of codon usage bias has been widely used to characterize different communities of microorganisms. In this context, the aim of this work was to study the codon usage bias in a natural consortium of five acidophilic bacteria used for biomining. The codon usage bias of the consortium was contrasted with genes from an alternative collection of acidophilic reference strains and metagenome samples. Results indicate that acidophilic bacteria preferentially have low codon usage bias, consistent with both their capacity to live in a wide range of habitats and their slow growth rate, a characteristic probably acquired independently from their phylogenetic relationships. In addition, the analysis showed significant differences in the unique sets of genes from the autotrophic species of the consortium in relation to other acidophilic organisms, principally in genes which code for proteins involved in metal and oxidative stress resistance. The lower values of codon usage bias obtained in this unique set of genes suggest higher transcriptional adaptation to living in extreme conditions, which was probably acquired as a measure for resisting the elevated metal conditions present in the mine.

Introduction

A total of 61 sense codons translate into 20 different amino acids, which is known as the redundancy of the genetic code or degeneracy of codons. Codon usage bias (CUB) refers to differences in the relative frequencies of synonymous codons within a coding sequence, differences which have been correlated with functional and adaptive properties [1–3]. The absence of CUB means that synonymous codons are used randomly without preference to code for their corresponding amino acids. A coding sequence is said to have low or weak CUB when synonymous codons are employed in a mostly random way. In contrast, high or strong CUB

Latorre. All the authors also acknowledge the support of the National Laboratory of High Performance Computing (NLHPC) at the CMM (PIA ECM-02.-CONICYT).

Competing interests: The authors have declared that no competing interests exist.

ensues when synonymous codons are used in a preferential manner to code for amino acids, the most extreme case being when exactly one codon is used to represent each amino acid.

The analysis of CUB has been used to characterize both specific and general properties of genes from communities of microorganisms [4]. Botzman *et al* determined an association between the lifestyles of several prokaryotic organisms and variations in their CUB [5]. Their results indicated that species living in a wide range of habitats have low CUB, which is consistent with the need to adapt to different environments. In addition, results also suggest that species may more readily adjust to metabolic variability by maintaining low CUB.

Bacteria which use a small subset of optimal codons (high CUB) also present fast growing rates [6], supporting the idea that optimization of the translation machinery is correlated with the maximization of growth rate. Complementing these studies, the analysis of 11 sequenced microbial samples showed that organisms living in the same ecological niche share a common preference for CUB, regardless of their phylogenetic diversity [7]. Such evidence highlights the importance of analyzing CUB in order to characterize bacterial communities, studies not hitherto addressed in acidophilic species.

Acidophilic bacteria are characterized by their survival under low pH and high concentration of metal cations. They are some of the most studied microorganisms living in extreme environments and are widely employed for the recovery of precious metals from mineral ores. During the process of extracting metal ions from different ores or concentrates, several microbial species work in concert in order to convert insoluble metal sulfides into water-soluble metal sulfates [8,9]. Currently, it is known that biomining communities of extremophile microorganisms that act in a coordinated manner are able to achieve higher levels of performance in metal extraction processes [10–12].

Several efforts have been made to isolate and characterize bacterial species and communities from differing extreme environmental sites [13–18]. At the molecular level, most criteria have been focused on identifying and quantifying particular components of each bacterium, such as proteins involved in iron/sulfur oxidation, metal resistance and biofilm formation [19–21]. While this strategy is able to suggest direct correlations between some of these components and a greater capacity for mineral bioleaching, only a few global-scale studies with the objective of imputing genomic advantages or common properties to such communities have been undertaken.

The first study to shed light on community gene structures in a mine environment was presented in 2004 by Tyson *et al*. [22]. Performing a metagenomic analysis, it was determined that a microbial community inhabiting acid mine drainage combines carbon and nitrogen fixation pathways in order to survive in such an extreme environment. With the aim of investigating genomic properties of a bacterial community from an industrially bioleached mine, a metagenome analysis of a surface layer of low grade copper tailings was recently undertaken at the Dexing Copper Mine in China [23,24]. The results illustrated that metal cation transport and DNA repair are highly represented processes inside the community, highlighting the presence of *Acidithiobacillus* and *Acidiphilium* species. In addition, the afore-mentioned studies provide a complete dataset of genes from acidophilic bacterial species, opening the possibility to study, characterize and classify extreme communities of microorganisms according to their CUB.

Recently, a consortium of five natural copper-bioleaching acidophilic bacteria was presented [25]. The consortium is made up of the bacteria *Acidithiobacillus thiooxidans* Licanantay, *Acidiphilium multivorum* Yenapatur, *Leptospirillum ferriphilum* Pañiwe, *Acidithiobacillus ferrooxidans* Wenelen and *Sulfobacillus thermosulfidooxidans* Cutipay, which were directly isolated from copper mines and selected based on their high capacity to solubilize copper and resist high concentrations of metal cations. In addition, this consortium is currently employed in a fully operational biotechnology system at CODELCO, Radomiro Tomic Division (Patent Registration No. CL 48319, Antofagasta, Chile).

In order to determine if this natural consortium of extreme acidophilic bacteria exhibits any particular genomic advantages, the CUB of genes belonging to the consortium were contrasted with: i) an acidophilic biomining consortium (metagenomic data) from a surface layer of low grade copper tailings (Dexing Copper Mine, China), ii) an alternative (non-consortium) collection of reference acidophilic bacterial strains (which were independently isolated from different geographic locations around the world) and iii) a global bacterial CUB profile generated from a set of reference genes compiled from the 2014 COG database. Considering the particular niches they inhabit, the consortium, non-consortium and metagenomic data were compared with a view towards determining if discrepancies in patterns of CUB are correlated with the extreme environments they inhabit.

Materials and methods

Genome sequences

The gene and protein sequences for *Sulfobacillus thermosulfidooxidans* strain Cutipay, *Acidithiobacillus thiooxidans* strain Licanantay, *Acidiphilium multivorum* strain Yenapatur, *Leptospirillum ferriphilum* strain Pañiwue and *Acidithiobacillus ferrooxidans* strain Wenelen, which constitute the bacterial species belonging to a Chilean biomining consortium, were previously described [25–27]. Their sequences are available at <http://biominingdb.cmm.uchile.cl/genomes/>. Sequences for the group of non-consortium species were downloaded from the NCBI. This group consists of *Acidimicrobium ferrooxidans* DSM 10331, *Acidiphilium cryptum* JF-5, *Acidiphilium multivorum* AIU301, *Acidiphilium* sp. PM, *Acidithiobacillus caldus* ATCC 51756, *At. caldus* SM-1, *At. ferri- vorans* SS3, *At. ferrooxidans* ATCC 23270, *At. ferrooxidans* ATCC 53993, *At. thiooxidans* ATCC19377, *Desulfosporosinus acidiphilus* DSM 22704, *Leptospirillum ferriphilum* ML-04, *L. ferrooxidans* C2-3, *Sulfobacillus acidophilus* TPY, *Sb. thermosulfidooxidans* DSM 9293, *Sb. thermo- sulfidooxidans* CBAR-13, *Thiomonas intermedia* K12 and *Thiomonas* sp 3As (Accessions NC_013124.1; NC_009484.1, NC_009467.1-NC_009474.1; NC_015186.1; NZ_AFPR01000001.1- NZ_AFPR01000627.1; NZ_CP005986.1-NZ_CP005989.1; NC_015850.1-NC_015854.1; NC_015942.1; NC_011761.1; NC_011206.1; NZ_AFOH0000000.1; NC_018066.1-NC_018068.1; NC_018649.1; NC_017094.1; NC_015757.1; FWY01000001.1-FWY01000002.1; NZ_LGRO0000000.1; NC_014153.1-NC_014155.1 and NC_014144.1-NC_014145.1 respectively) [28–40]. Metagenomics data from a bioleaching heap sample presented in Zhang, X. *et al.* [23] were downloaded from the MG-RAST repository (Accession 4664533.3) [41]. Putative gene sequences in this sample and their taxonomic categories were also obtained from this repository (328571 sequences in total with taxonomic assignments made through similarity searches against RefSeq proteins with cutoffs: 15bp alignment length; e^{-5} e-value; 60% identity).

Orthologous genes

Each bacterium in the consortium group was paired with one of the same species in the non-consortium group: *Sb. thermosulfidooxidans* Cutipay and CBAR-13; *At. thiooxidans* Licanantay and ATCC19377; *A. multivorum* Yenapatur and AIU301; *L. ferriphilum* Pañiwue and ML-04; and *At. ferrooxidans* Wenelen and ATCC23270. For each pair, orthologous genes were calculated using both ORTHOMCL v1.4 [42] and Inparanoid v4.1 [43]. Only gene pairs predicted as orthologous by both tools were kept.

COG category assignment

COG categories for all groups were assigned based on a protein BLAST search against the 2014 COG database with e-value and identity cutoffs of $1e^{-5}$ and 40% respectively. In the case

of protein sequences from the metagenomic sample, only those with a length of at least 90% of the hit length were considered (35000 sequences in total). This set was taken as the metagenome group.

COG database to gene database

A gene sequence database based on the 2014 COG protein database was generated for use as a reference set of non-lifestyle-specific genes. GenBank accessions for proteins in the COG database were retrieved from NCBI (<ftp://ftp.ncbi.nih.gov/pub/COG/COG2014/data>). Using those accession codes the associated genome gbk files were downloaded from NCBI. Finally, gene sequences corresponding to COG proteins were retrieved from these files and a gene sequence database was constructed containing a total of 1,737,559 DNA sequences.

Kullback–Leibler codon information bias (CIB)

We use the Kullback–Leibler codon information bias (CIB) defined in [44] as a way of quantifying the use of synonymous codons in genes relative to the reference scenario in which each synonymous codon is used equally often to code for its corresponding amino acid (see [45], for an examination of various measures of CUB based on other principles). More explicitly, CIB is a measure of codon usage bias based on information theoretic concepts, namely entropy, which takes account of how amino acids are distributed. As such, CIB is a natural and intuitively appealing quantity for measuring the departure of a coding sequence from equal usage of synonymous codons (details in [S1 Appendix](#)). CIB is zero if and only if the codons that code for each amino acid are used equally often to represent that amino acid, that is, there is unbiased synonymous codon usage. It attains its maximum value, which is determined by the relative frequencies of all the amino acids, precisely when each amino acid is represented by exactly one codon. Small values of CIB correspond to low (less selective or weak) codon usage bias while larger values of CIB correspond to a greater concentration of the codon relative frequencies on fewer codons (stronger or more selective codon usage bias). For this study, CIB was rescaled to have a value in the range 0–1.

Data analysis and statistical tools

The value of CIB was computed for every gene annotated for all bacterial species under consideration and for every putative gene belonging to the metagenomic sample. In addition, CIB was calculated for the 1,737,559 genes in the gene database derived from the COG database. This constitutes 97.3% of the 1,785,722 genes listed in the 2014 COG database. The remaining 2.7% of genes in the COG database were excluded from this study as it was not technically possible to recover the coding sequences needed for calculating the codon relative frequencies; the computation of CIB requires both the amino acid relative frequencies and the codon relative frequencies.

Differences in the pattern of CIB were analyzed between strains of the same organism, as well as between individual organisms and the gene database generated from the COG database. This was accomplished as follows. Consider two groups of CIB values, for instance, genes of *A. multivorum* Yenapatur with COG category P and genes from *A. multivorum* AIU301 that also have COG category P. Firstly, the distributions of CIB were tested for equivalence using the two-sample Anderson–Darling test [46]. The Anderson–Darling test is similar to the more familiar two-sample Kolmogorov–Smirnov test, but is generally more powerful with greater sensitivity to discrepancies in the tails of the distributions. It has null hypothesis “the two groups have the same distribution” and alternative hypothesis “the two groups have different distributions”. Secondly, if a difference was detected by the Anderson–Darling test, a further

test was performed to see if the values of CIB in one group stochastically dominate those in the other group.

Stochastic dominance, also known as simple stochastic ordering or strong stochastic ordering [47], means that the probability of observing a value of CIB greater than a specified threshold in one group is always greater than the probability of seeing a value greater than the same threshold in the other group. Equivalently, one group will stochastically dominate the other if graphs of their cumulative distribution functions do not cross, though they may touch. When it applies, stochastic dominance establishes a strong relationship between two statistical samples and provides a method of comparison, in which case it can be said that one sample is stochastically smaller or larger than the other. Two groups satisfying this relationship can be ranked, say, according to their mean values, without the need to consider measures of dispersion. For the analysis in this paper, a permutation test for stochastic dominance using Monte Carlo estimation to compute the p -value was implemented using version 11 of the C++ programming language in conjunction with the R statistical computing software V3.3.2 (refer to [S1 Appendix](#)).

Unless otherwise indicated, the computation of CIB and all statistical analyses were carried out using the R statistical computing software V3.3.2. The `kSamples` package was used for the Anderson-Darling test and the Bioconductor `Biostrings` package was used to process DNA sequence data.

Hierarchical clustering of CIB using average linkage was carried out by means of the [TM4 MeV v4.9.0 stand-alone local client](#) using the Pearson product correlation coefficient as the distance metric [48].

Results and discussion

Codon usage bias in biomining organisms

Mining sites are characterized by the presence of low-pH and the prevalence of aerobic environments. These extreme conditions induce selective pressures which have an impact on indigenous organisms, for example, the principal acidophilic organisms are autotrophic, able to use ferrous iron and reduced sulfur compounds as electron donors which are released from sulfide minerals during oxidative dissolution [8]. Apart from nutritional selection, it is plausible to hypothesize that genes from such organisms have also been genetically selected in order to improve the ability of the organism to survive under extreme conditions [49].

To assess putative gene sequence differences between copper-bioleaching acidophilic species and other organisms, we considered a set of five such bacteria which inhabit the same niche (the consortium group) [25] and compared these with three specifically selected groups of bacteria ([Table 1](#)). The first group (non-consortium) includes a total of 18 previously sequenced acidophilic bacteria, which were isolated from different mining sites. The second comparison group (metagenome) is made up of 35000 sequences from a copper mine metagenomic sample including 274 bacterial families. This group is that subset of the complete metagenomic sample to which a COG category could be assigned and which had a length of at least 90% of the size of the match in the COG database. The third and final comparison group (COG), was constructed from almost all genes present in the 2014 edition of the COG database [50]. In order to make these comparisons, we used the codon information bias (CIB) as a measure of codon usage bias (see [Materials and Methods](#)).

First, all the genes in each bacterial group were assigned COG categories. Then, the distribution of CIB values calculated for genes in each category were compared with the distribution of CIB values computed for genes in the corresponding category in the COG database. Results indicate that the consortium and non-consortium groups of acidophilic species showed significant differences in the distribution of CIB relative to the COG database in almost every

Table 1. Origins of biomining strains under study.

Bacterial species	Isolation site	Reference
Consortium		
<i>Ac. multivorum</i> Yenapatur	Copper mine in the north of Chile.*	[14]
<i>At. ferrooxidans</i> Wenelen	Copper mine in the north of Chile.*	[14]
<i>At. thiooxidans</i> Licanantay	Copper mine in the north of Chile.*	[27]
<i>L. ferriphilum</i> Pañiwe	Copper mine in the north of Chile.*	[14]
<i>Sb. thermosulfidooxidans</i> Cutipay	Copper mine in the north of Chile.*	[26]
Non-consortium		
<i>A. ferrooxidans</i> ICP, DSM 10331	Hot spring run-off in Iceland.	[62]
<i>Ac. cryptum</i> JF-5	Coal mine lake sediment.	[32]
<i>Ac. multivorum</i> AIU301	pyritic acid mine drainage in the Matsuo mines.	[63]
<i>Ac. sp.</i> PM	Rio Tinto's acidic, heavy metal-rich waters.	[33]
<i>At. caldus</i> ATCC 51756	Coal spoils, United Kingdom.	[34]
<i>At. caldus</i> SM-1	-	[64]
<i>At. ferrivorans</i> SS3	Norilsk mining area, Russia.	[65]
<i>At. ferrooxidans</i> ATCC 23270	Coal mine acid, bituminous effluent.	[66]
<i>At. ferrooxidans</i> ATCC 53993	-	[35]
<i>At. thiooxidans</i> ATCC 19377	Kimmeridge clay. Dorset, England, UK.	[67]
<i>D. acidiphilus</i> SJ4 DSM 22704	Sediment from acid mining effluent decantation pond; France, Chessy les Mines.	[36]
<i>L. ferriphilum</i> ML-04	Acidic water near a hot spring in Tengchong, Yunnan, China.	[68]
<i>L. ferrooxidans</i> C2-3	Volcanic ash deposit on Mount Oyama in the island of Miyake, Japan.	[37]
<i>Sb. acidophilus</i> TPY	Hydrothermal vent in the Pacific Ocean.	[69]
<i>Sb. thermosulfidooxidans</i> AT-1	Spontaneous ore deposit in Eastern Kazakhstan.	Unpublished
<i>Sb. thermosulfidooxidans</i> CBAR-13	Percolate solution of the bioleaching heap at Escondida copper mine, Chile.	Unpublished
<i>Thiomonas intermedia</i> K12	Sewage pipe in Hamburg Germany.	Unpublished
<i>Thiomonas</i> sp 3As	Acidic waters draining the Carnoulès mine tailings, France.	[40]
Metagenome		
Microbial community	Surface layer of low grade copper tailings bioleached at Dexing Copper Mine, China.	[23]

*Patent Registration No. CL 48319, Antofagasta, Chile

<https://doi.org/10.1371/journal.pone.0195869.t001>

category (see Fig 1 and S1 Table), with the largest differences observed in processes related to protein and nucleotide metabolism, cell motility and inorganic ion transport (COG categories E, F, N, and P). In particular, autotrophic species shown in Fig 2 (*At. thiooxidans*, *At. ferrooxidans*, *L. ferriphilum* and *Sb. thermosulfidooxidans*) exhibited smaller values of CIB on average compared to the COG database, independently of gene length.

As indicated, this is a typical characteristic found in organisms that are able to live in a wide range of habitats and which require the ability to efficiently adapt their metabolisms to different environments [51]. Also, acidophilic organisms show wide and versatile metabolic diversity, coupled with an extraordinary physiological capacity to live under extreme conditions [52]. The lower CIB seen in both acidophilic groups coincides with their capacity to adjust their metabolic variability, which correlates with previous analyses of codon usage made in other communities of microorganisms [5]. In addition, the acidophilic strains studied here are characterized by low growth rates [25], supporting the hypothesis that bacterial species with low codon usage experience slow growth [6].

All three groups of genes, (consortium, non-consortium and metagenome) exhibit distributions of CIB that differ from genes in the COG database (S1 Table). It is remarkable that genes in the first two groups have CIB values that are stochastically smaller than genes in the COG

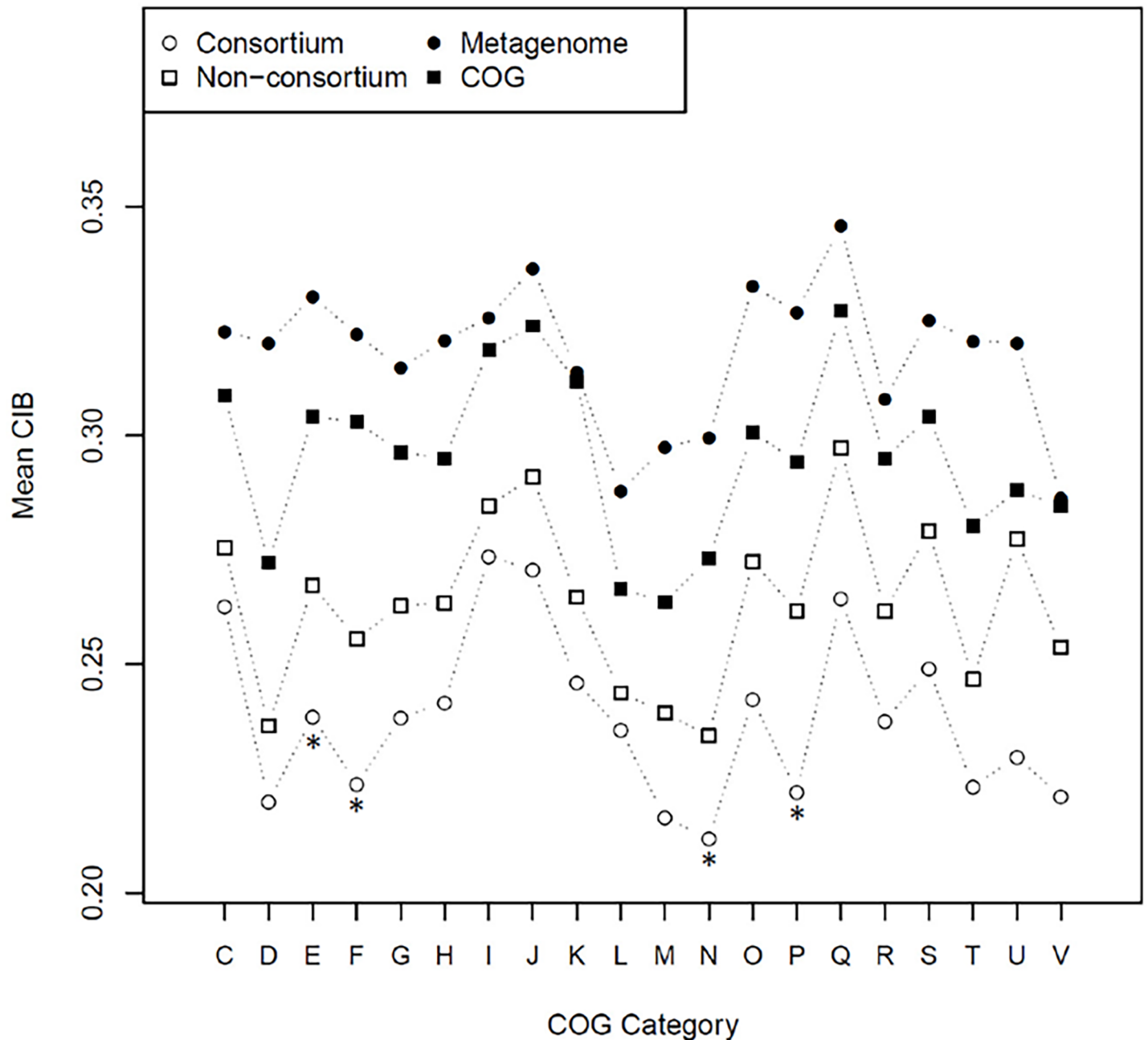


Fig 1. Average value of CIB for genes belonging to the consortium biomining species and the selected comparison groups under study. Each value is the average CIB calculated over all the species from each independent group classified according to COG category. The asterisks mark the four COG categories for which the greatest difference was observed between the mean CIB for the consortium and the mean CIB for the 2014 COG database.

<https://doi.org/10.1371/journal.pone.0195869.g001>

database in almost all COG categories: this is much stronger than merely saying that they have smaller CIB on average.

In contrast, while the metagenome group and COG database have different distributions of CIB, only COG categories D (cell division) and T (signal transduction mechanisms) from the metagenomics sample display a clear stochastic relationship to genes in the COG database (S1 Table), despite the mean CIB for genes in the metagenome group exceeding the mean CIB of

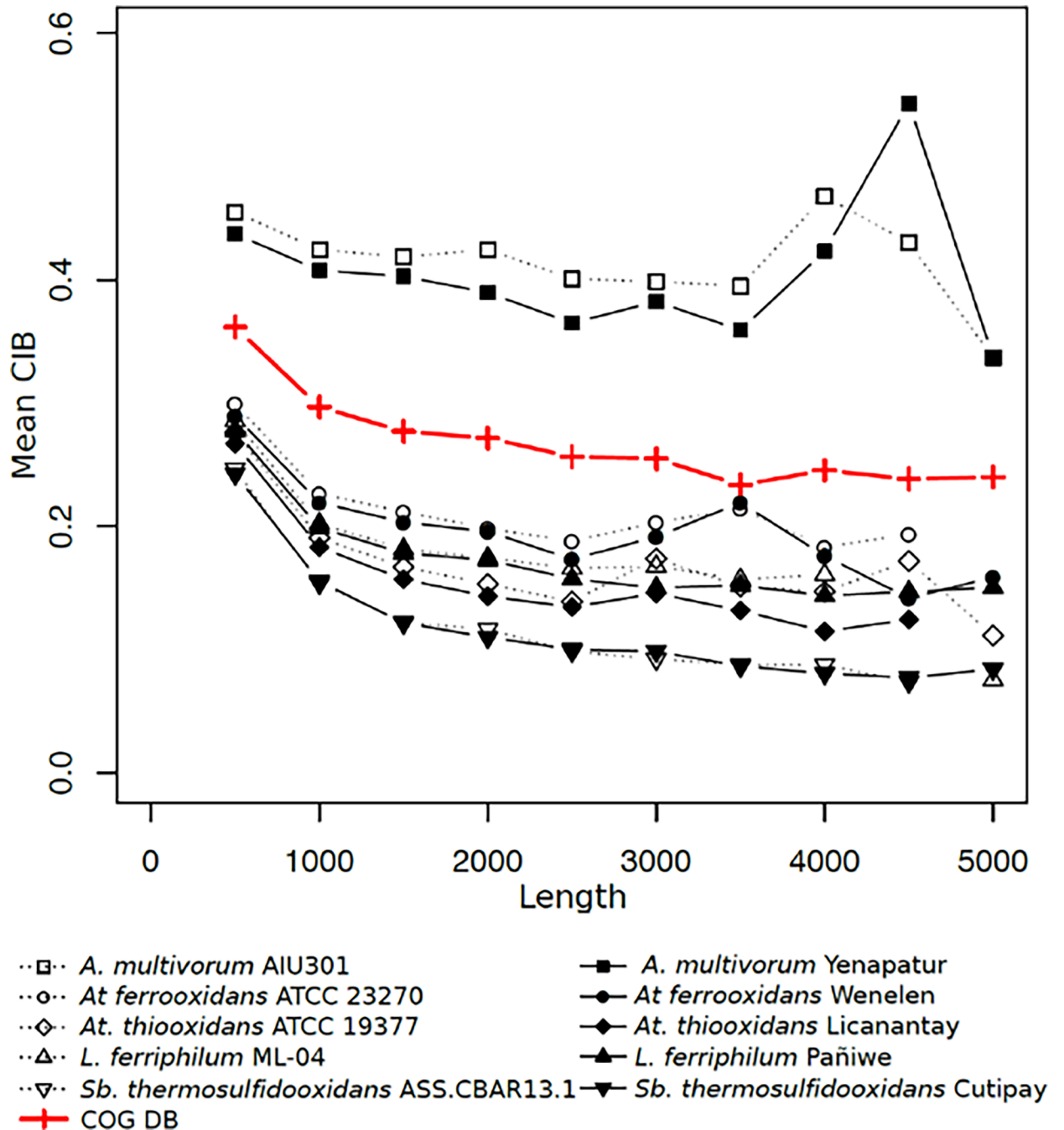


Fig 2. Average value of CIB for genes belonging to ten strains of bacteria (consortium and non-consortium) and the COG database binned by gene length in bases. Each bin contains genes from $\ell-499$ to ℓ bases in length, where ℓ can be read off the x-axis. The y-axis indicates the mean value of CIB for genes belonging to the bin indicated on the x-axis. Bacterial strains belonging to the same species are plotted using the same point shape. Strains belonging to the biomining consortium are distinguished by filled points linked by solid lines while non-consortium strains are hollow points linked by dotted lines. The average CIB values for the 2014 COG database are plotted as crosses linked by a red solid line.

<https://doi.org/10.1371/journal.pone.0195869.g002>

COG database genes (Fig 1). Considering that this metagenomics sample covers a total of 274 bacterial families, it is plausible to argue that the high diversity of microorganisms in this community covers an extensive variety of patterns of CIB, which is reflected in the statistical results obtained.

Specific codon usage bias in the acidophilic-bacteria consortium used for biomining

In order to study whether or not a higher capacity to survive in extreme environments is correlated with a particular pattern of codon usage, a fourth group of species were selected from the non-consortium group as counterpart strains to the acidophilic bacteria consortium used for biomining. This new group was composed by: *Sb. thermosulfidooxidans* CBAR-13, *At. thiooxidans* ATCC19377, *At. ferrooxidans* ATCC23270, *L. ferriphilum* ML-04 and *A. multivorum* AIU301, all of which were isolated from different mining sites.

In general, the observed differences in CIB are similar in both groups (Fig 2 and S2 Table), indicating that these acidophilic organisms probably share some aspects of codon usage bias independently of the place where they were isolated or their phylogenetic relationship. This is suggestive of co-evolution of the genetic code in these species [53].

In particular, the heterotrophic *A. multivorum* strains showed larger values of CIB on average compared to the remaining species, which are autotrophic. This was seen mainly in COG categories related to translation, transcription, signaling and general metabolism. Unlike the other members of the consortium, *A. multivorum* has the specific role of degrading organic metabolites highly toxic to autotrophic organisms [54]. High CIB is associated with high functional specialization and faster translational rates [55], which in this case probably improves the ability of *A. multivorum* to sense, metabolize and degrade organic compounds.

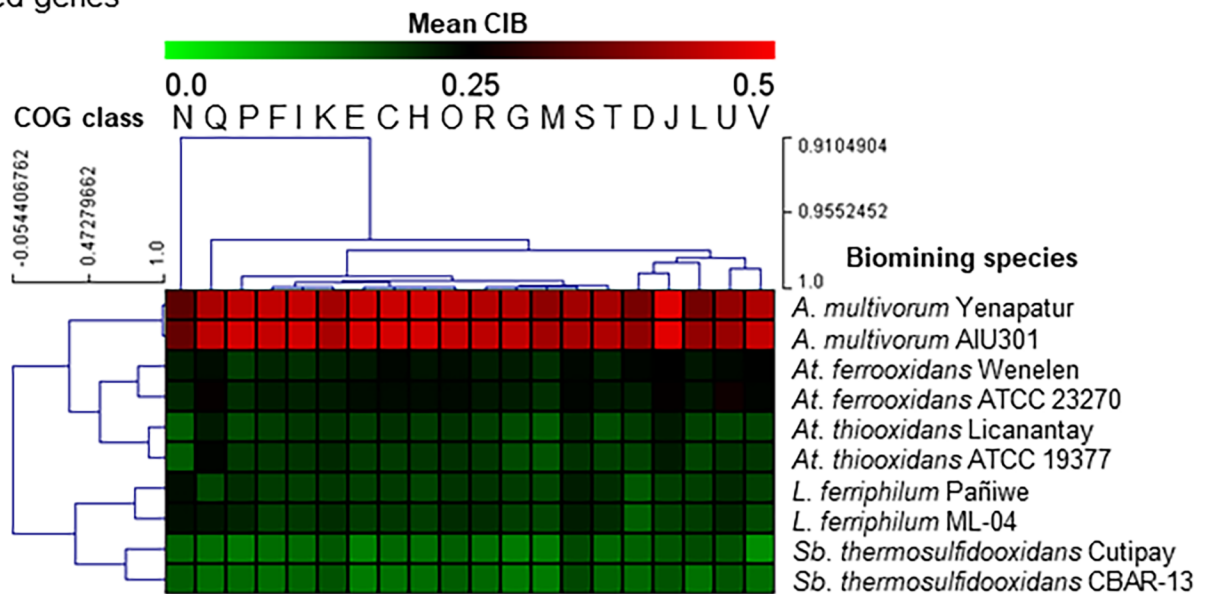
Unexpectedly, the low growth rate of the two *A. multivorum* strains [30] does not correspond to their higher CIB. However, in their extreme environmental niche (mining site), the growth of *A. multivorum* depends on the presence of other members of the community to produce the organic sources the bacterium consumes and to oxidize the thiosulfate compounds toxic to it [54,56]. This establishes mutual dependence within the consortium which is reflected by the similar growth rates observed in these species [56].

The next step was to divide all the genes belonging to each species into two sets, those that are conserved in the two strains of the species and those that are unique to one of the strains. For each of these sets, the two-sample Anderson-Darling test was used to decide whether or not the two strains of each species had the same distribution of CIB among the genes in each COG category.

The results reveal that the conserved genes in all the pairs of biomining strains studied exhibit essentially the same distribution of CIB (S3 Table), supporting the previous observation that the conserved genes in biomining lifestyle organisms apparently co-evolved in order to survive in extreme environments. In addition, the clustering of conserved genes in Fig 3A shows that, biomining strains from the same species fall naturally into the same clade and hence the distribution of CIB for conserved genes among the various COG categories clusters the bacteria in a way similar to phylogenetic distance [25]. This indicates that the CIB of the conserved genes conforms to a phylogenetic relationship and has not been significantly affected by the specific niche they inhabit.

On the other hand, significant differences were identified within the group of unique genes (S4 Table). The smallest values of CIB were observed in the autotrophic species in the consortium, corresponding to low codon usage bias. Clustering based on unique genes revealed a different organization (species and COG categories) compared to clustering in terms of conserved genes (Fig 3B). Interestingly, unlike the situation observed for the conserved group of genes, consortium species *At. thiooxidans* Licanantay and *L. ferriphilum* Pañiwe (both of which stand out due to their high copper-bioleaching performance) [25], cluster together in the same clade. This supports the idea that the CIB of unique genes is most likely affected by the common niche (copper mine site) rather than any phylogenetic relationship. Note that

A. Conserved genes



B. Unique genes

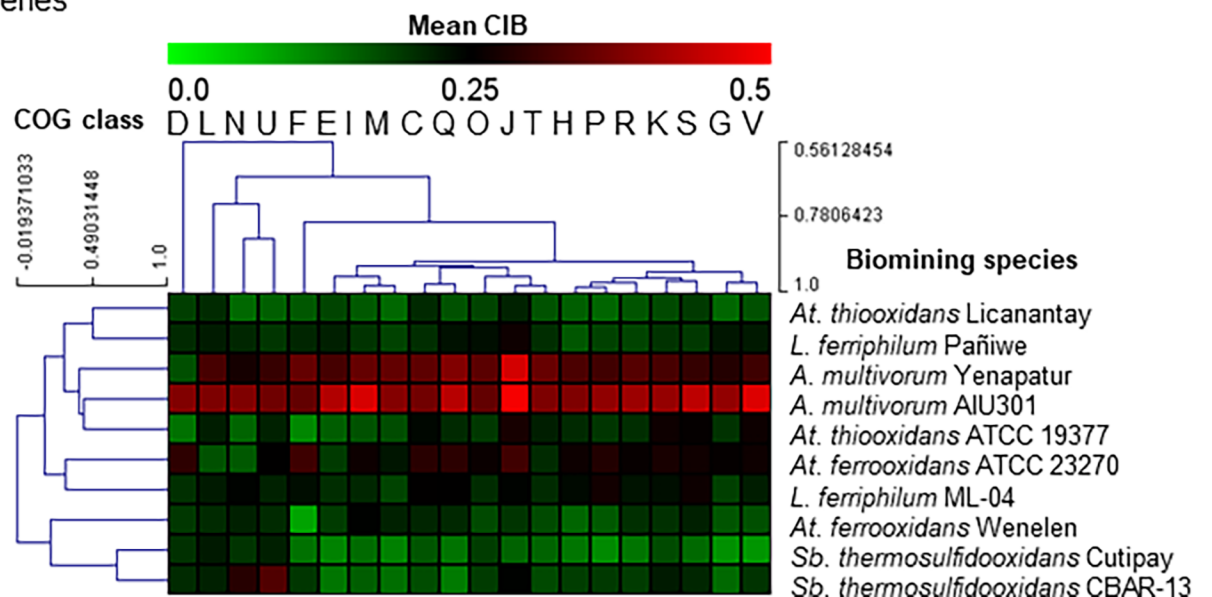


Fig 3. Hierarchical clustering of 10 biomining bacteria according to CIB values grouped by COG category. A. Conserved genes, B. Unique genes. For each strain, the mean value of CIB was calculated for genes (both conserved and unique) within each COG category. The strains were then hierarchically clustered using average linkage with the Pearson product correlation coefficient measuring the distance between the vectors of mean CIB per COG category. The color bar ranges from green (low CIB, 0.0) to red (high CIB, 0.5).

<https://doi.org/10.1371/journal.pone.0195869.g003>

although the two strains of *Sb. thermosulfidooxidans* share the same clade, both of these strains were isolated from Chilean copper deposits, which is also consistent with unique genes causing these species to be clustered according to geographic/environmental effects.

As mentioned, low CIB may suggest better environmental adaptation as a product of higher metabolic variability, which probably (directly or indirectly) affects the adaptation of the consortium to the extreme environmental conditions it inhabits. In this context, mining sites are also characterized by a high concentration of metal cations. The COG groups directly related

Table 2. Unique genes from the consortium involved in copper bioleaching with the smallest CIB values.

Gene ID	Annotation	COG	Consortium species	CIB
AFWEN_837	TonB-dependent receptor	P	<i>At. ferrooxidans</i> Wenelen	0.045
AFWEN_932	Iron complex outermembrane receptor protein	P	<i>At. ferrooxidans</i> Wenelen	0.069
ATLIC_1742	Cation efflux system protein	P	<i>At. thiooxidans</i> Licanantay	0.076
ATLIC_1133	TonB-dependent receptor	P	<i>At. thiooxidans</i> Licanantay	0.078
ATLIC_2149	Iron complex outermembrane receptor protein	P	<i>At. thiooxidans</i> Licanantay	0.085
ATLIC_3279	ATPase-like protein	L	<i>At. thiooxidans</i> Licanantay	0.085
ATLIC_1945	superfamily I DNA helicase	L	<i>At. thiooxidans</i> Licanantay	0.085
ATLIC_223	DNA mismatch repair protein MutS domain protein	L	<i>At. thiooxidans</i> Licanantay	0.089
ATLIC_3543	TonB-dependent receptor	P	<i>At. thiooxidans</i> Licanantay	0.089
ATLIC_3440	Iron complex outermembrane receptor protein	P	<i>At. thiooxidans</i> Licanantay	0.104
ATLIC_3544	Iron complex outermembrane receptor protein	P	<i>At. thiooxidans</i> Licanantay	0.105
ATLIC_3112	Phosphate transporter	P	<i>At. thiooxidans</i> Licanantay	0.113
ATLIC_3827	Cu ²⁺ -exporting ATPase	P	<i>At. thiooxidans</i> Licanantay	0.118
ATLIC_3828	Cu ²⁺ -exporting ATPase	P	<i>At. thiooxidans</i> Licanantay	0.143
ATLIC_3862	Cu(I)/Ag(I) efflux system membrane protein CusA	V	<i>At. thiooxidans</i> Licanantay	0.160
LFPA_2449	Heavy-metal exporter, HME family	P	<i>L. ferriphilum</i> Pañiwe	0.120
LFPA_613	Cu ²⁺ -exporting ATPase	P	<i>L. ferriphilum</i> Pañiwe	0.169
STCUT_2973	Major facilitator superfamily MFS_1	P	<i>Sb. thermosulfidooxidans</i> Cutipay	0.062
STCUT_3247	Putative efflux protein	P	<i>Sb. thermosulfidooxidans</i> Cutipay	0.077
STCUT_3508	Putative efflux protein	P	<i>Sb. thermosulfidooxidans</i> Cutipay	0.077
STCUT_2348	DNA repair protein RecN (Recombination protein N)	L	<i>Sb. thermosulfidooxidans</i> Cutipay	0.093
STCUT_1513	ATP-dependent DNA helicase RecQ	L	<i>Sb. thermosulfidooxidans</i> Cutipay	0.115
STCUT_1001	Cation efflux system protein, CDF family	P	<i>Sb. thermosulfidooxidans</i> Cutipay	0.158

<https://doi.org/10.1371/journal.pone.0195869.t002>

to the adaptation to these conditions [25], such as categories L, P and V, which involve metal resistance, and iron and sulfur oxidation, presented the greatest differences between species in the consortium and their non-consortium counterparts.

Within this group of unique genes (Table 2), the following genes presented some of the lowest CIB values recorded: the protein complex TonB, which participates in iron acquisition through siderophore mechanisms [57]; the cation efflux systems, phosphate transporter and Cu-ATPase, principal proteins related to copper resistance [58,59], and the enzymes RecN and MutS which are involved in double-strand break and mismatch DNA repair respectively [60,61].

One of the principal characteristics of the acidophilic consortium used for bioleaching in comparison to other acidophilic organisms, is the ability of its members to resist elevated concentrations of heavy metal cations [25]. Most of the consortium species resist at least twice the external copper concentration in relation to their counterpart biomining species (S5 Table). Next, a collection of genes previously classified in copper resistance and oxidative stress protection were selected from the acidophilic consortium [25], together with homologs from their non-consortium counterparts. Inside this group, copper efflux proteins CopA and Cus, antioxidant defense enzymes thioredoxin reductase (NadpH), superoxide dismutase (sodA) and peroxidoreoxin (bcp) components exhibit lower CIB values in relation to the homologous genes of their acidophilic counterpart species (Table 3). The species belonging to the consortium not only contain a higher number of these components [25], but these genes also present with a low value of CIB, suggesting higher transcriptional adaptation to living in a wide range of habitats, which was probably acquired as a measure for resisting the extreme, elevated copper conditions present in the mine.

Table 3. List of genes from the consortium with lower CIB values compared to its non-consortium counterpart.

Id consortium	CIB consortium	Id counterpart	CIB counterpart	name	Gene annotation	COG
Copper resistance						
LFP_A_558	0.254	LFML04_RS04475	0.257	cutA	periplasmic divalent cation tolerance protein	P
ATLIC_3492	0.142	ATHIO_RS0103355	0.145		Cu ²⁺ -exporting ATPase	P
STCUT_215	0.069	AOA63_RS11310	0.074	copA	heavy metal translocating P-type ATPase	P
STCUT_220	0.093	AOA63_RS11335	0.100		Cu ²⁺ -exporting ATPase	P
LFP_A_1158	0.208	LFML04_RS01790	0.217		Cu ²⁺ -exporting ATPase	P
ATLIC_2000	0.245	ATHIO_RS0110590	0.254		lipoprotein	M,P
ATLIC_1679	0.288	ATHIO_RS0112045	0.299	hmrR	transcriptional regulator, MerR family	K
STCUT_1350	0.158	AOA63_RS16975	0.196		NIF3-related protein	S
AFWEN_2071	0.231	AFE_RS11165	0.278	cusA	Cu(I)/Ag(I) efflux system membrane protein CusA	V
LFP_A_601	0.148	LFML04_RS03050	0.216	cusA	Cu(I)/Ag(I) efflux system membrane protein CusA	V
LFP_A_601	0.148	LFML04_RS08675	0.224	cusA	Cu(I)/Ag(I) efflux system membrane protein CusA	V
AMYEN_2455	0.250	ACMV_RS13835	0.521		Cu ²⁺ -exporting ATPase	P
Average	0.186		0.232			
ROS						
ATLIC_1147	0.332	ATHIO_RS0106580	0.333	sodA	superoxide dismutase, Fe-Mn family	P
ATLIC_1534	0.188	ATHIO_RS0115785	0.196		peroxidase (EC:1.11.1.7)	O
LFP_A_122	0.167	LFML04_RS12970	0.176	resA	thiol-disulfide oxidoreductase	O,C
LFP_A_382	0.254	LFML04_RS05360	0.264	trxA	thioredoxin	O,C
ATLIC_2542	0.304	ATHIO_RS0105865	0.314		alkylhydroperoxidase like protein, AhpD family	S
LFP_A_1163	0.227	LFML04_RS01815	0.237		thioredoxin 2	O,C
STCUT_347	0.284	AOA63_RS12055	0.295		alkylhydroperoxidase like protein, AhpD family	S
AFWEN_3145	0.259	AFE_RS01735	0.271		glutaredoxin family protein	O
AMYEN_133	0.392	ACMV_RS07445	0.406	trxA	thioredoxin 1	O,C
ATLIC_1449	0.140	ATHIO_RS0107230	0.155	msrA	peptide-methionine (S)-S-oxide reductase	O
AMYEN_129	0.305	ACMV_RS07425	0.323	trxB	thioredoxin reductase (NADPH)	O
ATLIC_1450	0.183	ATHIO_RS0107225	0.204	msrB	peptide-methionine (R)-S-oxide reductase	O
AFWEN_2550	0.162	AFE_RS06860	0.184	bcp	peroxiredoxin Q/BCP	O
ATLIC_2479	0.201	ATHIO_RS0106370	0.227	bcp	peroxiredoxin Q/BCP	O
ATLIC_130	0.240	ATHIO_RS0108025	0.272	ahpC	peroxiredoxin (alkyl hydroperoxide reductase subunit C)	O
STCUT_838	0.130	AOA63_RS14385	0.163		glutaredoxin-like domain-containing protein	O
AFWEN_1580	0.267	AFE_RS13515	0.300	msrA	peptide-methionine (S)-S-oxide reductase	O
STCUT_1306	0.178	AOA63_RS16755	0.223	sodA	superoxide dismutase, Fe-Mn family	P
LFP_A_1179	0.304	LFML04_RS01895	0.351	bcp	peroxiredoxin Q/BCP	O
STCUT_518	0.252	AOA63_RS12875	0.317	bcp	peroxiredoxin Q/BCP	O
LFP_A_1314	0.244	LFML04_RS02930	0.316		monothiol glutaredoxin	O
AFWEN_3137	0.162	AFE_RS01775	0.234	trxB	thioredoxin reductase (NADPH)	O
ATLIC_589	0.246	ATHIO_RS0110035	0.319		monothiol glutaredoxin	O
Average	0.236		0.264			

<https://doi.org/10.1371/journal.pone.0195869.t003>

Conclusions

In general, acidophilic organisms present similar patterns of CIB independently of the place where they were isolated or their phylogenetic relationships and this is mainly characterized by low CIB in autotrophic species. However, the particular copper mining environment influences the CIB in unique genes and genes for copper-resistance, probably conferring the acidophilic consortium with a greater capacity to resist high concentrations of metal cations.

Finally, studies of CIB in acidophilic organisms provide an alternative application for identifying and characterizing new strains with higher capacities for bioleaching metal ores.

Supporting information

S1 Table. Summary of statistical relationships between the CIB distribution in each group under study and genes from the COG2014 database for each COG category. Each table entry displays the p-value from the permutation test for stochastic dominance between genes from the indicated group and COG category and genes assigned the same COG category in the COG database.

(PDF)

S2 Table. Summary of statistically significant differences in CIB distribution between bio-mining strains (the consortium species and their non-consortium counterparts) and genes from the COG2014 database for each COG category. Each table entry displays the difference in the mean value of CIB for genes of the indicated strain and COG category, and the mean value of CIB for genes assigned the same COG category in the COG database.

(PDF)

S3 Table. Statistically significant differences in CIB distribution of conserved genes between the consortium strain and its counterpart for each species and COG category. Each table entry displays the p-value for the two-sample Anderson-Darling test applied to genes conserved in the consortium and non-consortium strains for the indicated species and COG category.

(PDF)

S4 Table. Statistically significant differences in CIB distribution of unique genes between the consortium strain and its non-consortium counterpart for each species and COG category. Each table entry displays the p-value for the two-sample Anderson-Darling test applied to unique genes in the consortium and non-consortium strains for the indicated species and COG category.

(PDF)

S5 Table. Minimum inhibitory concentration (MIC) of copper between each consortium strain and its non-consortium counterpart. *data from [25].

(PDF)

S1 Appendix. Kullback–Leibler codon information bias (CIB) and stochastic dominance.

(PDF)

Acknowledgments

All the authors also acknowledge the support of the National Laboratory of High Performance Computing (NLHPC) at the CMM (PIA ECM-02.-CONICYT).

Author Contributions

Data curation: Mauricio Latorre.

Formal analysis: María Paz Cortés, Mauricio Latorre.

Funding acquisition: Mauricio Latorre.

Investigation: Andrew Hart, María Paz Cortés, Servet Martinez.

Methodology: Andrew Hart.

Software: María Paz Cortés.

Writing – original draft: Andrew Hart, Mauricio Latorre, Servet Martinez.

References

1. Ikemura T (1985) Codon usage and tRNA content in unicellular and multicellular organisms. *Mol Biol Evol* 2: 13–34. <https://doi.org/10.1093/oxfordjournals.molbev.a040335> PMID: 3916708
2. Wang H, Liu S, Zhang B, Wei W (2016) Analysis of synonymous codon usage bias of Zika virus and its adaption to the hosts. *PLoS One* 11: e0166260. <https://doi.org/10.1371/journal.pone.0166260> PMID: 27893824
3. Shen X, Huang T, Wang G, Li G (2015) How the sequence of a gene specifies structural symmetry in proteins. *PLoS One* 10: e0144473. <https://doi.org/10.1371/journal.pone.0144473> PMID: 26641668
4. Ran W, Higgs PG (2012) Contributions of speed and accuracy to translational selection in bacteria. *PLoS One* 7: e51652. <https://doi.org/10.1371/journal.pone.0051652> PMID: 23272132
5. Botzman M, Margalit H (2011) Variation in global codon usage bias among prokaryotic organisms is associated with their lifestyles. *Genome Biol* 12: R109. <https://doi.org/10.1186/gb-2011-12-10-r109> PMID: 22032172
6. Rocha EP (2004) Codon usage bias from tRNA's point of view: redundancy, specialization, and efficient decoding for translation optimization. *Genome Res* 14: 2279–2286. <https://doi.org/10.1101/gr.2896904> PMID: 15479947
7. Roller M, Lucic V, Nagy I, Perica T, Vlahovicek K (2013) Environmental shaping of codon usage and functional adaptation across microbial communities. *Nucleic Acids Res* 41: 8842–8852. <https://doi.org/10.1093/nar/gkt673> PMID: 23921637
8. Rawlings DE (2005) Characteristics and adaptability of iron- and sulfur-oxidizing microorganisms used for the recovery of metals from minerals and their concentrates. *Microb Cell Fact* 4: 13. <https://doi.org/10.1186/1475-2859-4-13> PMID: 15877814
9. Panda S, Akcil A, Pradhan N, Deveci H (2015) Current scenario of chalcopyrite bioleaching: a review on the recent advances to its heap-leach technology. *Bioresour Technol* 196: 694–706. <https://doi.org/10.1016/j.biortech.2015.08.064> PMID: 26318845
10. Li S, Zhong H, Hu Y, Zhao J, He Z, Gu G (2014) Bioleaching of a low-grade nickel-copper sulfide by mixture of four thermophiles. *Bioresour Technol* 153: 300–306. <https://doi.org/10.1016/j.biortech.2013.12.018> PMID: 24374030
11. Yang H, Feng S, Xin Y, Wang W (2014) Community dynamics of attached and free cells and the effects of attached cells on chalcopyrite bioleaching by *Acidithiobacillus sp.* *Bioresour Technol* 154: 185–191. <https://doi.org/10.1016/j.biortech.2013.12.036> PMID: 24389460
12. Goebel BM, Stackebrandt E (1994) Cultural and phylogenetic analysis of mixed microbial populations found in natural and commercial bioleaching environments. *Appl Environ Microbiol* 60: 1614–1621. PMID: 7517131
13. Hodar C, Moreno P, di Genova A, Latorre M, Reyes-Jara A, Maass A, et al. (2012) Genome wide identification of *Acidithiobacillus ferrooxidans* (ATCC 23270) transcription factors and comparative analysis of ArsR and MerR metal regulators. *Biomaterials* 25: 75–93. <https://doi.org/10.1007/s10534-011-9484-8> PMID: 21830017
14. Latorre M, Ehrenfeld N, Cortes MP, Travisany D, Budinich M, Aravena A, et al. (2016) Global transcriptional responses of *Acidithiobacillus ferrooxidans* Wenelen under different sulfide minerals. *Bioresour Technol* 200: 29–34. <https://doi.org/10.1016/j.biortech.2015.09.110> PMID: 26476161
15. Liu Y, Yin H, Zeng W, Liang Y, Baba N, Qiu G, et al. (2011) The effect of the introduction of exogenous strain *Acidithiobacillus thiooxidans* A01 on functional gene expression, structure and function of indigenous consortium during pyrite bioleaching. *Bioresour Technol* 102: 8092–8098. <https://doi.org/10.1016/j.biortech.2011.06.012> PMID: 21705214
16. Acuna LG, Cardenas JP, Covarrubias PC, Haristoy JJ, Flores R, Nuñez H, et al. (2013) Architecture and gene repertoire of the flexible genome of the extreme acidophile *Acidithiobacillus caldus*. *PLoS One* 8: e78237. <https://doi.org/10.1371/journal.pone.0078237> PMID: 24250794
17. Peng T, Ma L, Feng X, Tao J, Nan M, Liu Y, et al. (2017) Genomic and transcriptomic analyses reveal adaptation mechanisms of an *Acidithiobacillus ferrivorans* strain YL15 to alpine acid mine drainage. *PLoS One* 12: e0178008. <https://doi.org/10.1371/journal.pone.0178008> PMID: 28542527

18. Chen L, Ren Y, Lin J, Liu X, Pang X, Lin J (2012) *Acidithiobacillus caldus* sulfur oxidation model based on transcriptome analysis between the wild type and sulfur oxygenase reductase defective mutant. PLoS One 7: e39470. <https://doi.org/10.1371/journal.pone.0039470> PMID: 22984393
19. Farah C, Vera M, Morin D, Haras D, Jerez CA, Guilian N (2005) Evidence for a functional quorum-sensing type AI-1 system in the extremophilic bacterium *Acidithiobacillus ferrooxidans*. Appl Environ Microbiol 71: 7033–7040. <https://doi.org/10.1128/AEM.71.11.7033-7040.2005> PMID: 16269739
20. Ramirez P, Guilian N, Valenzuela L, Beard S, Jerez CA (2004) Differential protein expression during growth of *Acidithiobacillus ferrooxidans* on ferrous iron, sulfur compounds, or metal sulfides. Appl Environ Microbiol 70: 4491–4498. <https://doi.org/10.1128/AEM.70.8.4491-4498.2004> PMID: 15294777
21. Frankel ML, Demeter MA, Lemire JA, Turner RJ (2016) Evaluating the metal tolerance capacity of microbial communities isolated from Alberta oil sands process water. PLoS One 11: e0148682. <https://doi.org/10.1371/journal.pone.0148682> PMID: 26849649
22. Tyson GW, Chapman J, Hugenholtz P, Allen EE, Ram RJ, Richardson PM, et al. (2004) Community structure and metabolism through reconstruction of microbial genomes from the environment. Nature 428: 37–43. <https://doi.org/10.1038/nature02340> PMID: 14961025
23. Zhang X, Niu J, Liang Y, Liu X, Yin H (2016) Metagenome-scale analysis yields insights into the structure and function of microbial communities in a copper bioleaching heap. BMC Genet 17: 21. <https://doi.org/10.1186/s12863-016-0330-4> PMID: 26781463
24. Hu Q, Guo X, Liang Y, Hao X, Ma L, Yin H, et al. (2015) Comparative metagenomics reveals microbial community differentiation in a biological heap leaching system. Res Microbiol 166: 525–534. <https://doi.org/10.1016/j.resmic.2015.06.005> PMID: 26117598
25. Latorre M, Cortes MP, Travisany D, Di Genova A, Budinich M, Reyes-Jara A, et al. (2016) The bioleaching potential of a bacterial consortium. Bioresour Technol 218: 659–666. <https://doi.org/10.1016/j.biortech.2016.07.012> PMID: 27416516
26. Travisany D, Cortes MP, Latorre M, Di Genova A, Budinich M, Bobadilla-Fazzini RA, et al. (2014) A new genome of *Acidithiobacillus thiooxidans* provides insights into adaptation to a bioleaching environment. Res Microbiol 165: 743–752. <https://doi.org/10.1016/j.resmic.2014.08.004> PMID: 25148779
27. Travisany D, Di Genova A, Sepulveda A, Bobadilla-Fazzini RA, Parada P, Maass A (2012) Draft genome sequence of the *Sulfobacillus thermosulfidooxidans* Cutipay strain, an indigenous bacterium isolated from a naturally extreme mining environment in Northern Chile. J Bacteriol 194: 6327–6328. <https://doi.org/10.1128/JB.01622-12> PMID: 23105067
28. Mi S, Song J, Lin J, Che Y, Zheng H, Lin J (2011) Complete genome of *Leptospirillum ferriphilum* ML-04 provides insight into its physiology and environmental adaptation. J Microbiol 49: 890–901. <https://doi.org/10.1007/s12275-011-1099-9> PMID: 22203551
29. Kelly DP, Wood AP (2000) Reclassification of some species of *Thiobacillus* to the newly designated genera *Acidithiobacillus* gen. nov., *Halothiobacillus* gen. nov. and *Thermithiobacillus* gen. nov. Int J Syst Evol Microbiol 50 Pt 2: 511–516.
30. Wakao N, Nagasawa N, Matsuura T, Matsukura H, Matsumoto T, Hiraishi A, et al. (1994) *Acidiphilium multivorum* sp. nov., an acidophilic chemoorganotrophic bacterium from pyritic acid mine drainage. Journal of General and Applied Microbiology 40: 143–159.
31. Colmer AR, Temple KL, Hinkle ME (1950) An iron-oxidizing bacterium from the acid drainage of some bituminous coal mines. J Bacteriol 59: 317–328. PMID: 15436401
32. Magnuson TS, Swenson MW, Paszczynski AJ, Deobald LA, Kerk D, Cummings DE (2010) Proteogenomic and functional analysis of chromate reduction in *Acidiphilium cryptum* JF-5, an Fe(III)-respiring acidophile. Biometals 23: 1129–1138. <https://doi.org/10.1007/s10534-010-9360-y> PMID: 20593301
33. San Martin-Uriz P, Gomez MJ, Arcas A, Bargiela R, Amils R (2011) Draft genome sequence of the electricigen *Acidiphilium* sp. strain PM (DSM 24941). J Bacteriol 193: 5585–5586. <https://doi.org/10.1128/JB.05386-11> PMID: 21914891
34. Valdes J, Quatrini R, Hallberg K, Dopson M, Valenzuela PD, Holmes DS (2009) Draft genome sequence of the extremely acidophilic bacterium *Acidithiobacillus caldus* ATCC 51756 reveals metabolic versatility in the genus *Acidithiobacillus*. J Bacteriol 191: 5877–5878. <https://doi.org/10.1128/JB.00843-09> PMID: 19617360
35. Orellana LH, Jerez CA (2011) A genomic island provides *Acidithiobacillus ferrooxidans* ATCC 53993 additional copper resistance: a possible competitive advantage. Appl Microbiol Biotechnol 92: 761–767. <https://doi.org/10.1007/s00253-011-3494-x> PMID: 21789491
36. Pester M, Brambilla E, Alazard D, Rattei T, Weinmaier T, Han J, et al. (2012) Complete genome sequences of *Desulfosporosinus orientis* DSM765T, *Desulfosporosinus youngiae* DSM17734T, *Desulfosporosinus meridiei* DSM13257T, and *Desulfosporosinus acidiphilus* DSM22704T. J Bacteriol 194: 6300–6301. <https://doi.org/10.1128/JB.01392-12> PMID: 23105050

37. Fujimura R, Sato Y, Nishizawa T, Oshima K, Kim SW, Hattori M, et al. (2012) Complete genome sequence of *Leptospirillum ferrooxidans* strain C2-3, isolated from a fresh volcanic ash deposit on the island of Miyake, Japan. *J Bacteriol* 194: 4122–4123. <https://doi.org/10.1128/JB.00696-12> PMID: 22815442
38. Anderson I, Chertkov O, Chen A, Saunders E, Lapidus A, Nolan M, et al. (2012) Complete genome sequence of the moderately thermophilic mineral-sulfide-oxidizing firmicute *Sulfobacillus acidophilus* type strain (NAL(T)). *Stand Genomic Sci* 6: 1–13.
39. Li B, Chen Y, Liu Q, Hu S, Chen X (2011) Complete genome analysis of *Sulfobacillus acidophilus* strain TPY, isolated from a hydrothermal vent in the Pacific Ocean. *J Bacteriol* 193: 5555–5556. <https://doi.org/10.1128/JB.05684-11> PMID: 21914875
40. Arsene-Ploetze F, Koehler S, Marchal M, Coppee JY, Chandler M, Bonnefoy V, et al. (2010) Structure, function, and evolution of the *Thiomonas* spp. genome. *PLoS Genet* 6: e1000859. <https://doi.org/10.1371/journal.pgen.1000859> PMID: 20195515
41. Meyer F, Paarmann D, D'Souza M, Olson R, Glass EM, Kubal M, et al. (2008) The metagenomics RAST server—a public resource for the automatic phylogenetic and functional analysis of metagenomes. *BMC Bioinformatics* 9: 386. <https://doi.org/10.1186/1471-2105-9-386> PMID: 18803844
42. Li L, Stoeckert CJ Jr., Roos DS (2003) OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res* 13: 2178–2189. <https://doi.org/10.1101/gr.1224503> PMID: 12952885
43. Remm M, Storm CE, Sonnhammer EL (2001) Automatic clustering of orthologs and in-paralogs from pairwise species comparisons. *J Mol Biol* 314: 1041–1052. <https://doi.org/10.1006/jmbi.2000.5197> PMID: 11743721
44. Comeron JM, Aguade M (1998) An evaluation of measures of synonymous codon usage bias. *J Mol Evol* 47: 268–274. PMID: 9732453
45. Hart A, Martínez S (2016) An entropy-based technique for classifying bacterial chromosomes according to synonymous codon usage. *Journal of Mathematical Biology*: 1–15.
46. Scholz FW, Stephens M. A. (1987) K-Sample Anderson-Darling Tests. *Journal of the American Statistical Association* 82: 6.
47. Szekli R (1995) Stochastic ordering and dependence in applied probability: Springer-Verlag.
48. Saeed AI, Sharov V, White J, Li J, Liang W, Bhagabati N, et al. (2003) TM4: a free, open-source system for microarray data management and analysis. *Biotechniques* 34: 374–378. PMID: 12613259
49. Zhang X, Liu X, Liang Y, Guo X, Xiao Y, Ma L, et al. (2017) Adaptive evolution of extreme acidophile *Sulfobacillus thermosulfidooxidans* potentially driven by horizontal gene transfer and gene loss. *Appl Environ Microbiol* 17:83.
50. Tatusov RL, Galperin MY, Natale DA, Koonin EV (2000) The COG database: a tool for genome-scale analysis of protein functions and evolution. *Nucleic Acids Res* 28: 33–36. PMID: 10592175
51. Carbone A, Kepes F, Zinovyev A (2005) Codon bias signatures, organization of microorganisms in codon space, and lifestyle. *Mol Biol Evol* 22: 547–561. <https://doi.org/10.1093/molbev/msi040> PMID: 15537809
52. Prakash T, Taylor TD (2012) Functional assignment of metagenomic data: challenges and applications. *Brief Bioinform* 13: 711–727. <https://doi.org/10.1093/bib/bbs033> PMID: 22772835
53. Wong JT (1975) A co-evolution theory of the genetic code. *Proc Natl Acad Sci U S A* 72: 1909–1912. PMID: 1057181
54. Okabe S, Odagiri M, Ito T, Satoh H (2007) Succession of sulfur-oxidizing bacteria in the microbial community on corroding concrete in sewer systems. *Appl Environ Microbiol* 73: 971–980. <https://doi.org/10.1128/AEM.02054-06> PMID: 17142362
55. Brandis G, Hughes D (2016) The Selective Advantage of Synonymous Codon Usage Bias in *Salmonella*. *PLoS Genet* 12: e1005926. <https://doi.org/10.1371/journal.pgen.1005926> PMID: 26963725
56. Okibe N, Johnson DB (2004) Biooxidation of pyrite by defined mixed cultures of moderately thermophilic acidophiles in pH-controlled bioreactors: significance of microbial interactions. *Biotechnol Bioeng* 87: 574–583. <https://doi.org/10.1002/bit.20138> PMID: 15352055
57. Moeck GS, Coulton JW (1998) TonB-dependent iron acquisition: mechanisms of siderophore-mediated active transport. *Mol Microbiol* 28: 675–681. PMID: 9643536
58. Navarro CA, Orellana LH, Mauriaca C, Jerez CA (2009) Transcriptional and functional studies of *Acidithiobacillus ferrooxidans* genes related to survival in the presence of copper. *Appl Environ Microbiol* 75: 6102–6109. <https://doi.org/10.1128/AEM.00308-09> PMID: 19666734
59. Orell A, Navarro CA, Arancibia R, Mobarec JC, Jerez CA (2010) Life in blue: copper resistance mechanisms of bacteria and archaea used in industrial biomining of minerals. *Biotechnol Adv* 28: 839–848. <https://doi.org/10.1016/j.biotechadv.2010.07.003> PMID: 20627124

60. Pellegrino S, Radzimanowski J, de Sanctis D, Boeri Erba E, McSweeney S, Timmins J (2012) Structural and functional characterization of an SMC-like protein RecN: new insights into double-strand break repair. *Structure* 20: 2076–2089. <https://doi.org/10.1016/j.str.2012.09.010> PMID: 23085075
61. Kunkel TA, Erie DA (2005) DNA mismatch repair. *Annu Rev Biochem* 74: 681–710. <https://doi.org/10.1146/annurev.biochem.74.082803.133243> PMID: 15952900
62. Clum A, Nolan M, Lang E, Glavina Del Rio T, Tice H, Copeland A, et al. (2009) Complete genome sequence of *Acidimicrobium ferrooxidans* type strain (ICP). *Stand Genomic Sci* 20:38–45.
63. Wakao N, Nagasawa N, Matsuura T, Matsukura H, Matsumoto T, Hiraishi A, et al. (1994) *Acidiphilium multivorum* sp. nov., an acidophilic chemoorganotrophic bacterium from pyritic acid mine drainage. *J Gen Appl Microbiol* 40: 143–159.
64. You XY, Guo X, Zheng HJ, Zhang MJ, Liu LJ, Zhu YQ, et al. (2011) Unraveling the *Acidithiobacillus caldus* complete genome and its central metabolisms for carbon assimilation. *J Genet Genomics* 38:243–52. <https://doi.org/10.1016/j.jgg.2011.04.006> PMID: 21703548
65. Liljeqvist M, Valdes J, Holmes DS, Dopson M (2011) Draft genome of the psychrotolerant acidophile *Acidithiobacillus ferrivorans* SS3. *J Bacteriol* 193:4304–5. <https://doi.org/10.1128/JB.05373-11> PMID: 21705598
66. Valdés J, Pedroso I, Quatrini R, Dodson RJ, Tettelin H, Blake R 2nd, Eisen JA, Holmes DS (2008) *Acidithiobacillus ferrooxidans* metabolism: from genome sequence to industrial applications. *BMC Genomics* 11:597.
67. Valdes J, Ossandon F, Quatrini R, Dopson M, Holmes DS (2011) Draft genome sequence of the extremely acidophilic biomining bacterium *Acidithiobacillus thiooxidans* ATCC 19377 provides insights into the evolution of the *Acidithiobacillus* genus. *J Bacteriol* 193:7003–4. <https://doi.org/10.1128/JB.06281-11> PMID: 22123759
68. Mi S, Song J, Lin J, Che Y, Zheng H, Lin J (2011) Complete genome of *Leptospirillum ferriphilum* ML-04 provides insight into its physiology and environmental adaptation. *J Microbiol* 49:890–901. <https://doi.org/10.1007/s12275-011-1099-9> PMID: 22203551
69. Li B1, Chen Y, Liu Q, Hu S, Chen X (2011) Complete genome analysis of *Sulfobacillus acidophilus* strain TPY, isolated from a hydrothermal vent in the Pacific Ocean. *J Bacteriol* 193:5555–6. <https://doi.org/10.1128/JB.05684-11> PMID: 21914875