In situ metabolic activities of uncultivated Ferrovum sp. CARN8 evidenced by metatranscriptomic analysis

Frédéric Plewniak, Sandrine Koechler, Denis Le Paslier, Marina Hery, Odile Bruneel, Philippe N. Bertin

In situ metabolic activities of uncultivated *Ferrovum* sp. CARN8 evidenced by metatranscriptomic analysis

Frédéric Plewniak[1*], Sandrine Koechler[1†], Denis Le Paslier[2], Marina Héry[3], Odile Bruneel[3] and Philippe N. Bertin[1]

Affiliations:

[1] Génétique Moléculaire, Génomique et Microbiologie, UMR7156 CNRS Université de Strasbourg, Strasbourg, France,

[2] Génomique Métabolique, Genoscope, Institut François Jacob, CEA, CNRS, Université d'Évry, Université Paris-Saclay, France,

[3]HydroSciences Montpellier, CNRS, IRD, University of Montpellier, Montpellier, France

[†] Present address: Plateforme d'Analyse de l'Expression Génique, Université de Strasbourg, CNRS, IBMP UPR 2357, 67000 Strasbourg, France.

*Correspondence: f.plewniak@unistra.fr

Abstract

Amongst iron-oxidizing bacteria playing a key role in the natural attenuation of arsenic in acid mine drainages (AMDs), members of the *Ferrovum* genus were identified in mine effluent or water treatment plants, and were shown to dominate biogenic precipitates in field pilot experiments. In order to address the question of the *in situ* activity of the uncultivated *Ferrovum* sp. CARN8 strain in the Carnoulès AMD, we assembled its genome using metagenomic and metatranscriptomic sequences and we determined standardized expression values for protein-encoding genes. Our results showed that this microorganism was indeed metabolically active and allowed us to sketch out its metabolic activity in its natural environment. Expression of genes related to the respiratory chain and carbon fixation suggests aerobic energy production coupled to ferrous iron oxidation and chemolithoautotrophic growth. Notwithstanding the presence of nitrogenase genes in its genome, expression data also indicated that *Ferrovum* sp. CARN8 relied on ammonium import rather than nitrogen fixation. The expression of flagellum and chemotaxis genes hints that at least a proportion of this strain population was motile. Finally, apart from some genes related to metal resistance showing surprisingly low expression values, genes involved in stress response were well expressed as expected in AMDs.

Keywords

gene expression; in situ metabolism; metagenome assembled genome;

1 Introduction

Acid mine drainages (AMDs) are the largest direct environmental impact of mining activities [1]. They are formed when sulphide minerals such as pyrite FeS2 and arsenopyrite FeAsS are exposed to water and oxygen, gradually producing sulphuric acid and dissolving metals and metalloids. This weathering process is considerably hastened by naturally occurring microorganisms, rapidly developing acidic metal-rich run-offs which in turn affect the structure of microbial communities due to low pH, high concentrations of metals and poor availability of organic matter.

Iron-oxidizing bacteria are dominant members of microbial communities dwelling in AMDs and arguably play a key role in the natural arsenic attenuation observed therein [2-7]. Metagenomic sequencing of sediment at the station COWG (Carnoulès Oxydizing Wetland, G) in Carnoulès AMD yielded seven bacterial genomes, amongst which two belonged to the iron-oxidizing genera *Gallionella* and *Acidithiobacillus* [8]. Further exploration of the remaining unassembled sequences from this metagenomic dataset allowed the identification of a complete 16S rRNA gene sequence sharing 99.9% identity with the 16S gene of iron-oxidizing bacterium *Ferrovum myxofaciens* P3G. The *Ferrovum* genus has actually been detected in abundance in different mining sites presenting a wide variety of physicochemical properties [2, 5, 9-17]. The genome of type strain *F. myxofaciens* P3G was sequenced using pure cultures isolated from acidic effluent of an abandoned mine in Wales [18] and the genomes of three more *Ferrovum* strains (JA12, Z-31 and PN-J185) were recently sequenced from mixed cultures. On the basis of sequence similarity and predicted metabolic potentials, strains JA12 and PN-185 have been proposed to constitute a group distinct from *F. myxofaciens* P3G and *Ferrovum* sp. Z-31 [19, 20].

Members of the *Ferrovum* genus are able to grow autotrophically by oxidizing ferrous iron and strains of the group 1, represented by the type strain *F. myxofaciens* P3G, also harbour nitrogen fixation genes. *F. myxofaciens* P3G was reported to produce large amounts of exopolysaccharides (EPS), causing cells to adhere to each other as well as to metal, glass or plastic surfaces. This property strongly reduces washout of cell in continuous flow systems without the need for a growth support matrix, making relevant the use of this strain for treatment of acidic mine waters in bioreactors [21]. On the other hand, isolation in pure culture of strains other than *F. myxofaciens* P3G has not been successful yet, presumably because of their susceptibility to organic compounds and the frequent contamination by heterotrophic iron-oxidizing strains

such as *Acidiphilium* [21-23]. Their physiology and the roles they play in the environment still remain to be fully characterized and only two gene expression studies have been performed to this day, both concerning exclusively members of the JA12 group [15, 23].

Here, we address the question of the *in situ* activity of uncultivated *Ferrovum* sp. CARN8 bacterium at the COWG station in Carnoulès AMD. The genome of this strain, which belongs to group 1 but differs from the type strain, was assembled from a combination of metagenomic [8] and metatranscriptomic sequences. The metagenomic study published in [8] - had yielded seven metagenome-assembled genomes (MAGs) but had left 70851 contigs totalling more than 73 Mbp that had not been incorporated into bins due to low sequencing depth, therefore likely corresponding to less abundant strains. These low-depth contigs were combined with metatranscriptomic sequences obtained from the same samples to assemble the *Ferrovum* sp. CARN8 MAG. *In situ* gene expression for this strain could then be assessed from the metatranscriptomic data: the average number of mRNA copies per cell was estimated for each protein-encoding gene using a method derived from [24] to provide standardized gene expression levels.

2 Materials and methods

2.1 RNA extraction and sequencing

In order to separate and recover microbial cells from the sediment particles, a Nycodenz gradient density treatment was applied to triplicates of the same sediment samples stored at -80°C as those used for the metagenomic analysis of the COWG (Carnoulès Oxydizing Wetland, G) microbial community [8]. This step ensured removal of mineral particles, heavy metals and other contaminants present in the sediments that may impede subsequent RNA extraction while at the same time ensuring consistency of treatment for both metatranscriptomic and metagenomic samples. Total RNA extraction was subsequently performed as described in [25]. The whole process was undertaken at a temperature of 4°C.

Ribosomal RNAs were removed using the Illumina Bacteria and Plant Ribo-Zero- rRNA removal kits prior to 2x100bp paired-end libraries preparation. The three replicate libraries were then sequenced each on one full line of Illumina Hiseq 2500 following Illumina's instructions. Image analysis and base calling were performed using RTA 1.17.21.3 and CASAVA 1.8.2.

2.2 Metatranscriptome de novo assembly

Reads having a global quality score less than 30 were filtered out with the BBDuk programme from JGI's BBTools (https://jgi.doe.gov/data-and-tools/bbtools/bb-tools-user-guide). Adapters, artefacts and lower-quality read extremities (threshold=10) were subsequently trimmed with the same programme.

The complete metatranscriptomic data set comprising high-quality trimmed reads from all triplicates was assembled *de novo* with Spades version 3.9 in RNA mode [26].

2.3 Genome assembly and annotation

The Carnoulès metagenomic contigs published in [8] were searched using cd-hit-est-2d version 4.6 [27] for those having at least 97% identity over 30% of their length with the sequence of the draft genome of *F. myxofaciens* P3G available in the NCBI Refseq database. A total of 933 contigs were thus identified as potentially belonging to group 1 *Ferrovum* strains present in Carnoulès that will be referred to below as *Ferrovum* sp. CARN8.

Metatranscriptomic paired-end reads were mapped on the *F. myxofaciens* P3G genome with Bowtie2 version 2.3.4.1 [28]. All read pairs mapping with at least 97% identity were subsequently assembled with the 933 metagenomic contigs using Spades version 3.0.9 [26] in single-cell mode to accommodate the highly variable coverage. We thus obtained a pre-draft metagenome of 798 contigs having a mean length of 2.5 kb for a total of 1.99 Mb. The largest scaffold was 30.4 kb and the N50 for the assembly is 5.4 kb.

The whole set of protein sequences from *F. myxofaciens* P3G was then mapped onto the 798 pre-draft contigs using scipio version 1.4.1 [29]. Proteins unambiguously mapping two contigs over at least 30 bases on each one were considered as evidence in favour of the contiguity of those contigs. An in-house Python 3.5 script was used to assemble them, adding the appropriate number of Ns when needed to keep the coding sequence consistent with the reference protein across the ends of both contigs.

In order to fill gaps with sequences that might have been missed in the metagenome but were present in the metatranscriptome, the *de novo*-assembled metatranscripts were subsequently compared with the pre-draft genomic contigs and predicted transcripts using blastn version 2.8.1+ [30]. We could thus select 2270 metatranscripts, showing at least a partial blastn hit with an expect-value of less than 10^{-100} and a minimum identity of 99%, that were considered as potentially belonging to *Ferrovum* sp. CARN8. Mapping back the full set of metatranscriptomic read pairs on those *Ferrovum* sp. CARN8 metatranscripts using Bowtie2 [28]

with a 99% identity threshold allowed us to recruit read pairs corresponding to *Ferrovum* sp. CARN8 but that may have not passed the threshold of 97% identity with the *F. myxofaciens* P3G reference genome sequence in the initial recruitment step.

In order to avoid any bias due to a stepwise inclusion of metatranscriptomic reads, both sets of metatranscriptomic read pairs recruited as described above (97% identity with *F. myxofaciens* P3G genome, and 99% with CARN8 metatranscripts) were pooled together and assembled with Spades in single cell mode, using, to guide the assembly, the original metagenomic contigs as trusted contigs as well as the pre-draft metagenome and the 2270 *Ferrovum* sp. CARN8 metatranscripts sequences as untrusted contigs. Adjacent contigs were determined and joined together using scipio as already described above for the pre-draft assembly.

The resulting *Ferrovum* sp. CARN8 draft metagenome was subsequently annotated using the MicroScope platform [31] .

Average Nucleotide Identity (ANI) between *Ferrovum* sp. CARN8 and other available *Ferrovum* genomes (*F. myxofaciens* P3G, *Ferrovum* sp. Z-31, *Ferrovum* sp. JA12, and *Ferrovum* sp. PN-J185) was determined on-line with OrthoANIu [32].

2.4 Gene expression

The 181 portions of coding sequences (fCDSs) identified by the MicroScope platform as fragmented CDSs (fCDSs) and overlapping contig extremities were merged to reconstitute 88 complete CDSs, for a total number of *Ferrovum* sp. CARN8 CDSs of 2674.

The metatranscriptomic read-pairs were subsequently mapped onto all CDSs, using Bowtie2 version 2.3.4.1 [28] with the following parameters: --sensitive -I 1 -X 500 --no-mixed --no-discordant –fr -k 50.

Read counts, TPM (transcripts per million) and effective length of each CDS were determined with Salmon version 0.9.1 in align mode with variational Bayesian optimization and 1000 replicate Gibbs sampling bootstrapping on metatranscriptomic data [33] .

Assuming an average of 1 mRNA copy per cell per gene [24], the mRNA copy number was estimated for each CDS from their TPM value: $CN_i = n \times TPM_i / 10^6$, where CN_i is the average copy number of transcript i, n is the number of CDSs in the genome, and TPM_i is the estimated TPM of transcript i. In the case of *Ferrovum* sp. CARN8, $CN_i \approx TPM_i / 374$.

In order to indicate the level of confidence that they are expressed indeed in the whole population, genes were classified according to [24] in i) low-abundance mRNA (less than 0.5 copies/cell); ii) medium-abundance mRNA (between 0.5 and 2 copies/cell); iii) robust expression (more than 2 copies/cell). In this study, we did not consider gene expression lower than 0.2 mRNA copy/cell.

3 Results and Discussion

The draft metagenome we obtained had 452 scaffolds longer than 200bp, with an average length of 4.7 kbp for a total of 2.15 Mbp, including the 48458 Ns used to fill the 273 gaps between contigs. The largest scaffold was 51.4 kbp, and 90% of the whole sequence was represented by 204 scaffolds longer than 2.5 kbp. The assembly N50 was 12.2 kbp.

CheckM analysis using *betaproteobacteria* assignation with 235 reference genomes and 419 lineage-specific markers suggested that near completeness was achieved at 91.4% and that contamination was low at 3.4% [34].

A full-length 16S rRNA gene (1504 bp) was predicted, having 99.9% identity with the 16S rRNA gene of *F. myxofaciens* P3G and only 96% with those of the group 2 (*Ferrovum* sp. JA12 and PN-185). Average Nucleotide Identity (ANI) was 99.5% between *Ferrovum* sp. CARN8 and both group 1 strains, suggesting that these two strains belong to the same species. In contrast, ANI was less than 70% between *Ferrovum* sp. CARN8 and each strain of group 2. Additionally, with a number of 2674 coding sequences, the gene content of *Ferrovum* sp. CARN8 is comparable to those of group 1 strains *F. myxofaciens* P3G and Z-31 with 2859 and 2492 coding sequences, respectively. It is larger though than the gene content of group 2 strains with 1970 coding sequences for JA12 and 1888 for PN-185. In addition, the 55.01% GC content of *Ferrovum* sp. CARN8 is similar to the group 1 strains' 54.3%, and much higher than the values of 44.5 and 39.9 % for group 2 strains (Table 1).

The synteny statistics provided by the MicroScope platform are consistent with the classification of *Ferrovum* sp. CARN8 into the group 1: 80.7% of CARN8 CDSs are in 381 syntons shared with *F. myxofaciens* P3G, with an average length of 6.21 genes per synton, and a maximum synton length of 46 genes. In contrast, *Ferrovum* sp. CARN8 shares only 45.4% genes in syntons with *Ferrovum* sp. JA12. Putative orthologues as identified by bi-directional best hit represent 72.1% CDSs with *F. myxofaciens* P3G, and only 44.85% with JA12 (Table 1).

Gene expression *in situ* was determined by metatranscriptomic data analysis and genes were grouped by expression values according to [24]: 175 genes were expressed at a robust level (more than 2 mRNA copies/cell), 520 had medium expression values (between 0.5 and 2 mRNA copies/cell) and 822 were expressed at a low level (between 0.2 and 0.5 mRNA copy/cell). In this study, we did not consider as significantly expressed genes with an estimated expression value lower than 0.2 mRNA copy number/cell. The distribution of these expression values is similar to the one experimentally observed in [24] and [35] for *Escherichia coli*, with less than 1 mRNA copy/cell for the vast majority of genes (median value = 0.24 mRNA copy/cell) and a range of two degrees of magnitude.These results, summarized in Fig. 1 and provided in detail in Supplementary Table S1, suggest that *Ferrovum* sp. CARN8 was active at the COWG station in Carnoulès AMD. Amongst 21 predicted genes coding for proteins of the small ribosomal subunit and 31 genes of the large subunit, all but one showed a robust or medium expression level. In this regard, the medium expression of genes coding the ribosome maturation factors RimM and RimP suggests an active assembly of new ribosomes. An indication of an active translation process was also given by the low to robust expression of translation initiation and protein elongation factor genes as well as by the robust expression of the *rrf* gene coding for the ribosome-recycling factor involved in the release of the ribosome at the termination codon. Finally, the robust expression of both *dksA* gene, responsible for rRNA promoters' responsiveness to amino acid availability [36], and *hpf* gene encoding the ribosome hibernation-promoting factor, known to be involved in the response to some stresses, also suggests that *Ferrovum* sp. CARN8 was using these mechanisms to tune its translational capacity to the environmental conditions [37].

A complete set of 32 genes was predicted to encode proteins of the respiratory chain and iron oxidation, consistent with an active energy metabolism based on ferrous iron oxidation in aerobic conditions. Evidence of robust or medium expression was observed for genes encoding: the cytochrome c552, the Cyc2-like protein responsible for ferrous iron oxidation [23, 38], the cytochrome C oxidase, and the ATP synthase complex. Levels of expression were low or medium for genes coding for the cytochrome b/c1 complex, succinate dehydrogenase and NADH dehydrogenase genes. A gene coding for a high-potential iron-sulfur proteins (HiPIP) was also predicted and expressed at a very high level *in situ* by *Ferrovum* sp. CARN8. This gene has 64.6% identity with a *Ferrovum* sp. JA12 HiPIP gene whose transcript levels in a mixed culture has been shown recently to be very high [23]. HiPIP are known to drive a variety of

oxidizing reactions including iron oxidation [39], so, although it has not been demonstrated that iron oxidation can be HiPIP-driven in the *Ferrovum* genus, an iron-oxidation mechanism involving this protein in *Ferrovum* sp. CARN8 may not be excluded.

Two different ribulose-bisphosphate carboxylase (RubisCO) operons were identified in *Ferrovum* sp. CARN8 with the same structure as those found in the *F. myxofaciens* P3G strain. The expression of these genes, along with the medium expression of phosphoribulokinase *prkB* gene, is indicative of carbon fixation activity by *Ferrovum* sp. CARN8. Although both RubisCOs were of the Form IA, one was associated with a carboxysome cluster (Form IAc) and the other one was associated with the *cbbQ* and *cbbO* genes (Form IAq). Since these two forms are expected to respond to different levels of CO_2, their co-presence might be suggestive of the capacity of *Ferrovum* sp. CARN8 to adapt to large fluctuations in CO_2 and O_2 concentrations in the environment [40]. In the present study, the robust expression of Form IAq genes and the lower expression of Form IAc RubisCO genes with low or no expression of carboxysome genes hint at *in situ* conditions more suitable to Form IAq, *i.e.* medium to high CO_2 in the presence of O_2 [40], which is consistent with the clearly aerobic conditions encountered in COWG (Table 2).

The presence of nitrogenase-encoding genes in its genome suggests that *Ferrovum* sp. CARN8 has the capacity to fix dinitrogen. However, no significant expression was observed for these genes, while genes coding for the ammonium transporter AmtB and for the ammonia-assimilatory glutamine synthetase GlnA were expressed at low and medium level, respectively, suggesting ammonium-based nitrogen assimilation. In support of this hypothesis, we observed the robust expression of the *glnK* gene involved in the negative regulation of the Ntr regulon and no clear-cut expression of genes involved in amino-acid intake or catabolism. Since oxygen is known to have a direct inhibitory effect on nitrogen fixation [41], it may not be surprising that this process could be inactive in COWG with a concentration of 8.3 $mg.l^{-1}$ of dissolved oxygen (Table 2), close to the saturation concentration of 10 $mg.l^{-1}$ in freshwater at 15°C. Other parameters indirectly affecting nitrogen fixation such as a high concentration of sulphate and a low dissolved organic carbon (DOC) content [41] may also contribute to the inhibition of nitrogen fixation in *Ferrovum* sp. CARN8 in COWG with concentrations of 3200 $mg.l^{-1}$ sulphate (Table 2) and only 3 $mg.l^{-1}$ DOC [42].

Expression was robust or medium for 12 flagellum genes out of 39, low for 19 and below the threshold of 0.2 mRNA copy/cell for 8 only. With 15 genes expressed at medium or low level and 7 below the defined

minimum threshold, a similar situation was observed for chemotaxis, including genes coding for methyl-accepting chemotaxis proteins and *aer* involved in aerotaxis. The observation that the majority of those genes had an expression above the fixed threshold suggests that *Ferrovum* sp. CARN8 was motile *in situ* . The observed heterogeneity of expression levels might reflect the heterogeneity of the studied population with cells at different stages of the flagellum assembly cascade or showing different behaviours. As a matter of fact, medium expression was observed for a gene encoding the diguanylate cyclase enzyme that catalyses the formation of cyclic di-guanosine monophosphate, a key molecular messenger controlling bacterial exopolysaccharide biofilm formation [43]. This could be explained by the fact that, though a part of the *Ferrovum* sp. CARN8 population expressing flagellum genes might be motile, a proportion of the same population might nonetheless be in a sessile biofilm-forming state. Furthermore, the flagellum can also play an important role in cell adhesion and in the initial steps of biofilm formation [44]. Flagellar motility helps bacteria to reach the most favourable environments and to successfully compete with other micro-organisms. These complex organelles also play an important role in adhesion to substrates, biofilm formation and virulence process. In addition, because their synthesis and functioning are very expensive for the cell (about 2% of biosynthetic energy expenditure in *Escherichia coli*) and may induce a strong immune response in the host organism, the expression of flagellar genes is highly regulated by environmental conditions. In the past few years, many data have been published about the regulation of motility in polarly and laterally flagellated bacteria. However, the mechanism of motility control by environmental factors and by some regulatory proteins remains largely unknown. In this respect, recent experimental data suggest that the master regulatory protein-encoding genes at the first level of the cascade are the main target for many environmental factors. This mechanism might require DNA topology alterations of their regulatory regions. Finally, despite some differences the polar and lateral flagellar cascades share many functional similarities, including a similar hierarchical organisation of flagellar systems. The remarkable parallelism in the functional organisation of flagellar systems suggests an evolutionary conservation of regulatory mechanisms in Gram-negative bacteria. Our study thus suggests the coexistence in Carnoulès AMD of both *Ferrovum* sp. CARN8 streamers and planktonic-phase cells, as previously observed for *F. myxofaciens* P3G in shake flask cultures [21]. Contrary to what our study suggests though this lab experiment showed only non motile free cells in culture.

Expression of many genes involved in stress response was observed at medium or robust level, as expected in an acid-mine drainage environment. This included genes involved in oxidative stress (super-oxide dismutase, rubrerythrin, thioredoxin), genes involved in chaperoning (*groEL*, *groES*, *dnaK*,…), general stress response (*yhdN*, *nhaX*), and arsenic resistance (*arsC*, *arsA, arsD, arsR*). Remarkably, the gene coding for GlbN, which belongs to the subfamily of truncated haemoglobins involved in the response to oxidative stress or oxygen sensing, showed a very robust expression level at an average 17.91 mRNA copies/cell.

However, except *copZ*, *copA and cusF*, genes involved in heavy metal resistance like *cusABC* and *czcABC* had a surprisingly low level of expression. The determination of gene expression levels for the dominant strains in Carnoulès whose genomes were published in [8] showed a similar trend for metal resistance genes (Table 3). Only *copA* genes showed robust expression in CARN2, CARN6 and CARN7 while medium levels of expression were observed only for *copA* in CARN1, CARN4 and CARN5, for *cusABC* in CARN2 and for *czcABC* in CARN3. This might suggest that different systems of heavy metal resistance were employed by different strains, possibly in response to differences in their local environment.

We have presented here a metatranscriptomic analysis of *Ferrovum* sp. CARN8, an uncultivated bacterium of the *Ferrovum* genus, which allowed us to get a glimpse of the ecophysiology of this strain by identifying genes expressed *in situ* (Fig. 2), thus emphasizing the importance of metatranscriptomics to highlight the metabolic capacities and functions that are actually triggered or not in natural conditions. It is often difficult to extrapolate results from lab studies under controlled conditions to the much more complex conditions encountered in the natural environment of microorganisms. Reckoning with knowledge gained from previous lab experiments and *in situ* physico-chemical parameters, meta-omic approaches can provide indirect evidence of the activity of microorganisms in their natural environment where direct experimental observations are extremely difficult if possible at all. Notably, a substantial fraction representing 33.5% (889 genes) of the total gene set in the *Ferrovum* sp. CARN8 genome could not be assigned any known function. Medium or robust expression was observed in this study for 227 of them, confirming that a portion of these unidentified genes are not only real functional genes but that they are actually expressed in the particular conditions encountered by *Ferrovum* sp. CARN8 cells in their natural environment. This also confirms that new functions or mechanisms remain to be discovered in the large reservoir of genes represented by environmental microorganisms. In this regard, machine learning and large-scale meta-

analyses focusing on conserved genes of unknown function in metatranscriptomic data from different environments could help expand our knowledge of microbial ecophysiology and discover new activities that might be harnessed for biotechnology applications or bioremediation [45] .

Acknowledgements and data availability

The *Ferrovum* sp. CARN8 metagenome complete sequence was deposited in DDBJ/EMBL/GenBank database as under the bioproject number PRJEB28721. The assembly received the accession number GCA_900568365, contigs were numbered UOYP01000001 to UOYP01000725, and scaffolds from LR025218 to LR025669.

Metatranscriptomic data were made available in the SRA database under the bioproject number PRJEB31731. Replicate samples received the accession numbers ERS3230182 (replicate SKR15), ERS3230183 (replicate SKR16) and ERS3230184 (replicate SKR19).

Metatranscriptomic sequencing was performed by the GenomEast platform, a member of the 'France Génomique' consortium (ANR-10-INBS-0009)

We are grateful to the Génoscope (Évry, France) for providing on-line access to MicroScope, the Microbial Genome Annotation and Analysis Platform, (http://www.genoscope.cns.fr/agc/microscope).

We thank the OSU OREME (SO POLLUMINE Observatory) for long term funding of physicochemical and microbiological monitoring since 2009.

This work benefited from the financial support of the CoMMERCE project subsidized by the CNRS national program EC2CO-MicrobiEn.

Conflict of interest

The authors declare no conflict of interest.

References

[1] Sarkar A, Paul B, The global menace of arsenic and its conventional remediation - A critical review. Chemosphere 2016;158:37-49.

[2] Hallberg K, New perspectives in acid mine drainage microbiology. Hydrometallurgy 2010;104:448-453.

[3] Kay C, Rowe O, Rocchetti L, Coupland K, Hallberg K, Johnson D, et al., Evolution of microbial "streamer" growths in an acidic, metal-contaminated stream draining an abandoned underground copper mine. Life 2013;3:189-210.

[4] Andres J, Bertin PN, The microbial genomics of arsenic. FEMS Microbiol Rev 2016;40:299-322.

[5] Arce-Rodríguez A, Puente-Sánchez F, Avendaño R, Libby E, Rojas L, Cambronero JC, et al., Pristine but metal-rich Río Sucio (Dirty River) is dominated by *Gallionella* and other iron-sulfur oxidizing microbes. Extremophiles 2017;21:235-243.

[6] Bruneel O, Mghazli N, Hakkou R, Dahmani I, Filali Maltouf A, Sbabou L, In-depth characterization of bacterial and archaeal communities present in the abandoned Kettara pyrrhotite mine tailings (Morocco). Extremophiles 2017;21:671-685.

[7] Laroche E, Casiot C, Fernandez-Rojo L, Desoeuvre A, Tardy V, Bruneel O, et al., Dynamics of bacterial communities mediating the treatment of an As-rich acid mine drainage in a field pilot. Front Microbiol 2018;9:3169.

[8] Bertin PN, Heinrich-Salmeron A, Pelletier E, Goulhen-Chollet F, Arsène-Ploetze F, Gallien S, et al., Metabolic diversity among main microorganisms inside an arsenic-rich ecosystem revealed by meta- and proteo-genomics. ISME J 2011;5:1735-1747.

[9] Heinzel E, Janneck E, Glombitza F, Schlömann M, Seifert J, Population dynamics of iron-oxidizing communities in pilot plants for the treatment of acid mine waters. Environmental Science & Technology 2009;43:6138-6144.

[10] González-Toril E, Aguilera A, Souza-Egipsy V, López Pamo E, Sánchez España J, Amils R, Geomicrobiology of La Zarza-Perrunal acid mine effluent (Iberian Pyritic Belt, Spain).. Appl Environ Microbiol 2011;77:2685-94.

[11] Fabisch M, Beulig F, Akob DM, Küsel K, Surprising abundance of *Gallionella*-related iron oxidizers in creek sediments at pH 4.4 or at high heavy metal concentrations. Front Microbiol 2013;4:390.

[12] Ziegler S, Dolch K, Geiger K, Krause S, Asskamp M, Eusterhues K, et al., Oxygen-dependent niche formation of a pyrite-dependent acidophilic consortium built by archaea and bacteria. ISME J 2013;7:1725-1737.

[13] Kay CM, Haanela A, Johnson DB, Microorganisms in subterranean acidic waters within Europe's deepest metal mine. Res Microbiol 2014;165:705-712.

[14] Volant A, Bruneel O, Desoeuvre A, Héry M, Casiot C, Bru N, et al., Diversity and spatiotemporal dynamics of bacterial communities: physicochemical and other drivers along an acid mine drainage. FEMS Microbiol Ecol 2014;90:247-63.

[15] Hua Z-S, Han Y-J, Chen L-X, Liu J, Hu M, Li S-J, et al., Ecological roles of dominant and rare prokaryotes in acid mine drainage revealed by metagenomics and metatranscriptomics. ISME J 2015;9:1280-1294.

[16] Jones DS, Kohl C, Grettenberger C, Larson LN, Burgos WD, Macalady JL, Geochemical niches of iron-oxidizing acidophiles in acidic coal mine drainage. Appl Environ Microbiol 2015;81:1242-1250.

[17] Hao C, Wei P, Pei L, Du Z, Zhang Y, Lu Y, Dong H, Significant seasonal variations of microbial community in an acid mine drainage lake in Anhui Province, China. Environmental Pollution 2017;223:507-516.

[18] Moya-Beltrán A, Cárdenas JP, Covarrubias PC, Issotta F, Ossandon FJ, Grail BM, et al., Draft genome sequence of the nominated type strain of *Ferrovum myxofaciens*, an acidophilic, iron-oxidizing Betaproteobacterium. Genome Announc 2014;2.

[19] Ullrich SR, González C, Poehlein A, Tischler JS, Daniel R, Schlömann M, et al., Gene loss and horizontal gene transfer contributed to the genome evolution of the extreme acidophile *Ferrovum*. Front Microbiol 2016;7:797.

[20] Ullrich SR, Poehlein A, Tischler JS, González C, Ossandon FJ, Daniel R, et al., Genome analysis of the biotechnologically relevant acidophilic iron oxidising strain JA12 indicates phylogenetic and metabolic diversity within the novel genus *Ferrovum*. PLOS ONE 2016;11:e0146832.

[21] Johnson DB, Hallberg KB, Hedrich S, Uncovering a microbial enigma: isolation and characterization of the streamer-generating, iron-oxidizing, acidophilic bacterium *Ferrovum myxofaciens*. Appl Environ Microbiol 2014;80:672-680.

[22] Tischler JS, Jwair RJ, Gelhaar N, Drechsel A, Skirl A-M, Wiacek C, et al., New cultivation medium for *Ferrovum* and Gallionella-related strains. J Microbiol Methods 2013;95:138-144.

[23] Ullrich SR, Poehlein A, Levicán G, Mühling M, Schlömann M, Iron targeted transcriptome study draws attention to novel redox protein candidates involved in ferrous iron oxidation in *Ferrovum* sp. JA12. Res Microbiol 2018;169:618-627.

[24] Bartholomäus A, Fedyunin I, Feist P, Sin C, Zhang G, Valleriani A, Ignatova Z, Bacteria differently regulate mRNA abundance to specifically respond to various stresses. Philosophical transactions Series A, Mathematical, physical, and engineering sciences 2016;374:20150069.

[25] Farasin J, Andres J, Casiot C, Barbe V, Faerber J, Halter D, et al., *Thiomonas* sp. CB2 is able to degrade urea and promote toxic metal precipitation in acid mine drainage waters supplemented with urea. Front Microbiol 2015;6.

[26] Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al., SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. Journal of computational biology : a journal of computational molecular cell biology 2012;19:455-477.

[27] Li W, Godzik A, Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics 2006;22:1658-1659.

[28] Langmead B, Salzberg SL, Fast gapped-read alignment with Bowtie 2. Nat Methods 2012;9:357-359.

[29] Keller O, Odronitz F, Stanke M, Kollmar M, Waack S, Scipio: Using protein sequences to determine the precise exon/intron structures of genes and their orthologs in closely related species. BMC Bioinformatics 2008;9:278.

[30] Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL, BLAST+: architecture and applications. BMC Bioinformatics 2009;10:421.

[31] Médigue C, Calteau A, Cruveiller S, Gachet M, Gautreau G, Josso A, et al., MicroScope—an integrated resource for community expertise of gene functions and comparative analysis of microbial genomic and metabolic data. Brief Bioinform 2017;:bbx113.

[32] Yoon S-H, Ha S-m, Lim J, Kwon S, Chun J, A large-scale evaluation of algorithms to calculate average nucleotide identity. Antonie van Leeuwenhoek 2017;110:1281-1286.

[33] Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C, Salmon provides fast and bias-aware quantification of transcript expression. Nat Methods 2017;14:417-419.

[34] Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW, CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res 2015;25:1043-1055.

[35] Gorochowski TE, Chelysheva I, Eriksen M, Nair P, Pedersen S, Ignatova Z, Absolute quantification of translational regulation and burden using combined sequencing approaches. Mol Syst Biol 2019;15.

[36] Paul BJ, Barker MM, Ross W, Schneider DA, Webb C, Foster JW, Gourse RL, DksA: a critical component of the transcription initiation machinery that potentiates the regulation of rRNA promoters by ppGpp and the initiating NTP. Cell 2004;118:311-322.

[37] McKay SL, Portnoy DA, Ribosome hibernation facilitates tolerance of stationary-phase bacteria to aminoglycosides.. Antimicrob Agents Chemother 2015;59:6992-6999.

[38] Castelle C, Guiral M, Malarte G, Ledgham F, Leroy G, Brugna M, Giudici-Orticoni M-T, A new iron-oxidizing/O2-reducing supercomplex spanning both inner and outer membranes, isolated from the extreme acidophile *Acidithiobacillus ferrooxidans*. The Journal of biological chemistry 2008;283:25803-25811.

[39] Kusano T, Takeshima T, Sugawara K, Inoue C, Shiratori T, Yano T, et al., Molecular cloning of the gene encoding *Thiobacillus ferrooxidans* Fe(II) oxidase. High homology of the gene product with HiPIP. The Journal of biological chemistry 1992;267:11242-11247.

[40] Badger MR, Bek EJ, Multiple Rubisco forms in proteobacteria: their functional significance in relation to CO2 acquisition by the CBB cycle. J Exp Bot 2008;59:1525-1541.

[41] Howarth RW, Marino R, Cole JJ, Nitrogen fixation in freshwater, estuarine, and marine ecosystems. 2. Biogeochemical controls. Limnol Oceanogr 1988;33:688-701.

[42] Giloteaux L, Duran R, Casiot C, Bruneel O, Elbaz-Poulichet F, Goñi-Urriza M, Three-year survey of sulfate-reducing bacteria community structure in Carnoulès acid mine drainage (France), highly contaminated by arsenic. FEMS Microbiol Ecol 2013;83:724-737.

[43] Whiteley CG, Lee D-J, Bacterial diguanylate cyclases: Structure, function and mechanism in exopolysaccharide biofilm development. Biotechnol Adv 2015;33:124-141.

[44] Soutourina OA, Bertin PN, Regulation cascade of flagellar expression in Gram-negative bacteria. FEMS Microbiol Rev 2003;27:505-523.

[45] Plewniak F, Crognale S, Rossetti S, Bertin PN, A genomic outlook on bioremediation: the case of arsenic removal. Front Microbiol 2018;9:820.

Legends to figures

Figure 1

Boxplot of *Ferrovum* sp. CARN8 *in situ* gene expression values for protein-coding genes related to energy and nitrogen metabolism, carbon fixation, stress response, cell mobility and the ribosome. Dotted circles represent the median of the data points in each row. Thick lines represent the interquartile range (IQR) of expression values, covering at least 50% of the data for the corresponding genes. Thin lines extend to the most extreme data points within a distance of 1.5 x IQR from the lower and upper quartiles, respectively. Diamonds are genes whose expression values fall outside this range and are considered outliers. Values are represented on a log scale for better readability. The plot was produced with Matlab R2018b.

The standardized expression values were determined as follows: metatranscriptomic read-pairs were mapped onto all CDSs with Bowtie2 version 2.3.4.1 [28] and the number of transcripts per million (TPM) were quantified using Salmon version 0.9.1 [33] in align mode. Then, assuming an average of 1 mRNA copy per cell per gene [24], the mRNA copy number was estimated for each CDS from their TPM value: $CN_i = n \times TPM_i / 10^6$, where CN_i is the average copy number of transcript i, n is the number of CDSs in the genome, and TPM_i is the estimated TPM of transcript i.

According to [24], expression can be defined as i) low (less than 0.5 copies/cell); ii) medium (between 0.5 and 2 copies/cell); iii) robust (more than 2 copies/cell). In this study, we did not consider as significantly expressed genes with less than 0.2 mRNA copy/cell.

Figure 2

In situ metabolic activity of *Ferrovum* sp. CARN8 deduced from metatranscriptomic data analysis. Processes and systems with robust and medium expression are represented by solid filled boxes and disks. Low expression proteins and complexes are represented by dashed filled boxes. White boxes represent proteins and complexes not showing in this study any strong evidence in favour of their expression.

NADH dh: NADH dehydrogenase; sdh: succinate dehydrogenase; b/c1: cytochrome b/c1 complex; Cyc2 like: cytochrome C2 like; HiPIP like: high potential iron-sulphur protein; PRPP: phosphoribosyl pyrophosphate synthetase; Amt: ammonium transporter; GlnA: glutamine synthetase; GlbN: Group 1

truncated haemoglobin GlbN; Tpx: reduced lipid hydroperoxide peroxidase; Tpx*: oxidized lipid hydroperoxide peroxidase; TrxA: reduced thioredoxin; TrxA*: oxidized thioredoxin; TrxB: thioredoxin reductase; Sod: superoxide dismutase; ArsC: arsenate reductase; ArsAB: arsenite efflux pump; Mcp: methyl-accepting proteins; Aer: aerotaxis receptor; Che: chemotaxis proteins; CopA: copper exporting ATPase; ActP: ActP/CopZ copper transporting P-type ATPase; CusA, CusB, CusC and CusF: copper/silver efflux system.

Supporting information

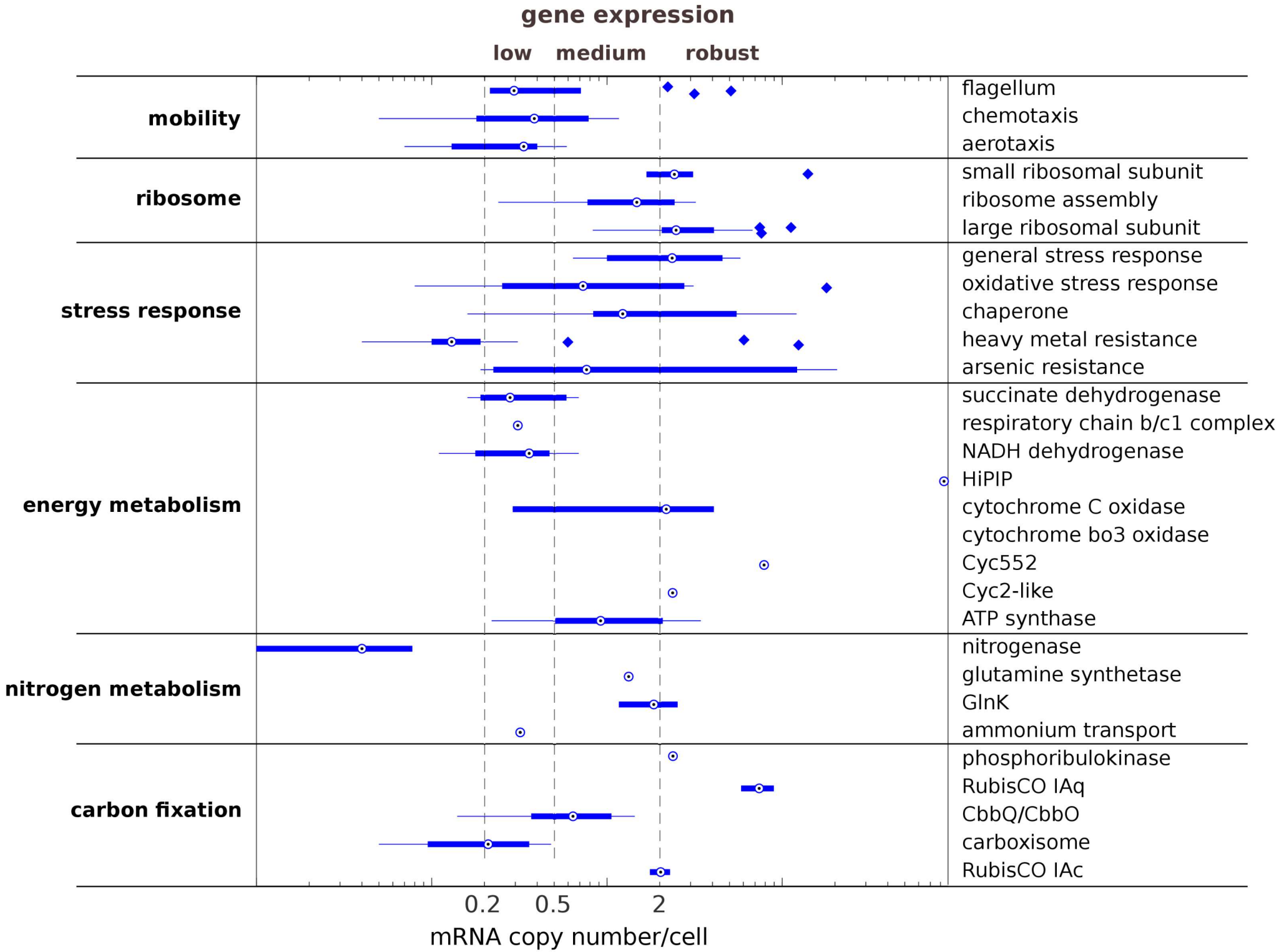
Table S1:

file: SupportingInformation_Table_S1.xls

Gene expression values determined as described in Fig. 1 caption and experimental procedures for all CDSs of *Ferrovum* sp. CARN8.

'Name' is the name of the CDS, 'Mean' is the average of mRNA copy number/cell values for the three replicates 'SKR15', 'SKR16' and 'SKR19'. The columns 'Gene', 'Synonyms', 'Product', 'Biological process', and 'Roles' contain MicroScope annotations. Columns 'start' and 'end' give the positions of the CDS in the metagenome.

A dark green background indicates robust expression (over 2 mRNA copies/cell) of the corresponding gene. Medium green denotes medium expression (between 0.5 and 2 mRNA copies/cell) and light green low expression (between 0.2 and 0.5 mRNA copy/cell). Genes with less than 0.2 mRNA copy per cell and not considered as significantly expressed in this study are shown here against a white background.


gene expression
low
medium
robust
mobility
flagellum
chemotaxis
aerotaxis
ribosome
small ribosomal subunit
ribosome assembly
large ribosomal subunit
stress response
general stress response
oxidative stress response
chaperone
heavy metal resistance
arsenic resistance
energy metabolism
succinate dehydrogenase
respiratory chain b/c1 complex
NADH dehydrogenase
HiPIP
cytochrome C oxidase
cytochrome bo3 oxidase
Cyc552
Cyc2-like
ATP synthase
nitrogen metabolism
nitrogenase
glutamine synthetase
GlnK
ammonium transport
carbon fixation
phosphoribulokinase
RubisCO IAq
CbbQ/CbbO
carboxisome
RubisCO IAc
0.2
0.5
2
mRNA copy number/cell

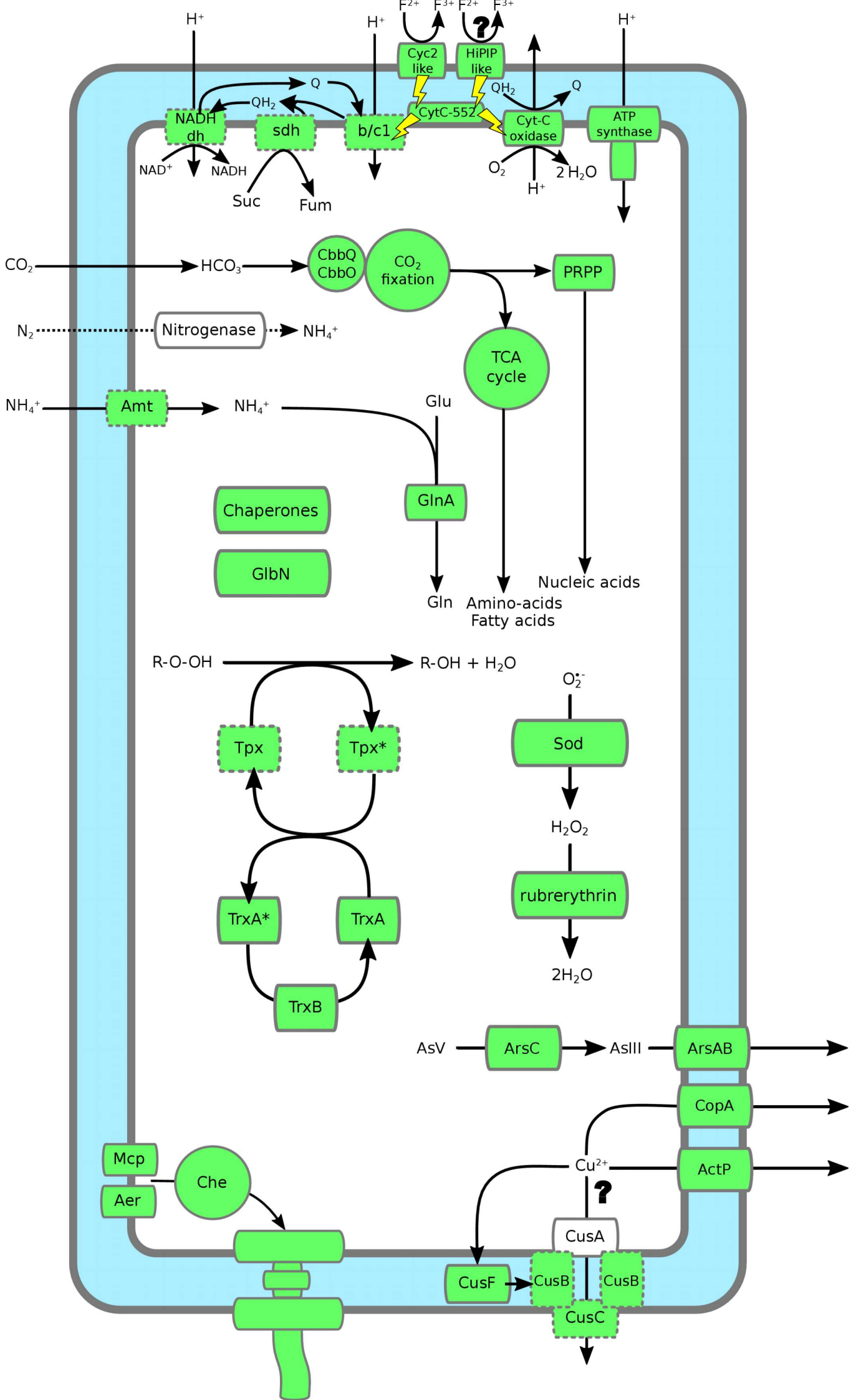

H+
H+
H+
F2+ F3+ F2+ F3+
?
Cyc2 like
HiPIP like
Q
QH2
QH2
Q
NADH dh
sdh
b/c1
CytC-552
Cyt-C oxidase
ATP synthase
NAD+
NADH
Suc
Fum
O2
H+
2 H2O
CO2
HCO3
CbbQ CbbO
CO2 fixation
PRPP
N2
Nitrogenase
NH4+
TCA cycle
NH4+
Amt
NH4+
Glu
Chaperones
GlnA
GlbN
Gln
Amino-acids Fatty acids
Nucleic acids
R-O-OH
R-OH + H2O
O2-
Tpx
Tpx*
Sod
H2O2
rubrerythrin
TrxA*
TrxA
TrxB
2H2O
AsV
ArsC
AsIII
ArsAB
CopA
ActP
Cu2+
?
Mcp
Aer
Che
CusA
CusF
CusB
CusB
CusC

Table 1

***Ferrovum* sp. CARN8 metagenome comparison with other *Ferrovum* public genomes**

	CARN8	**P3G**	**Z-31**	**JA12**	**PN-185**
16S rRNA gene identity	100%	99.9%	n.a.	96.4%	96.3%
Number of genes	2674	2859	2492	1970	1888
%GC	55.0%	54.3%	54.3%	44.5%	39.9%
CheckM completeness	91.4%	94.9%	n.d.	97.7%	n.d.
CheckM contamination	3.40%	0.0%	n.d.	1.0%	n.d.
% CDS in syntons	-	80.7%	n.d.	45.4%	n.d.
Number of syntons	-	381	n.d.	211	n.d.
Average number of genes / synton	-	6.21	n.d.	6.14	n.d.
Max number of genes in a synton	-	46	n.d.	34	n.d.
Bi-directional best blast hits	-	72.1%	n.d.	44.85%	n.d.
Average Nucleotide Identity (ANI)	100%	99.5%	99.5%	67.3%	67.2%
Average Aminoacid Identity (AAI)	100%	98.6%	98.3%	59.7%	60.2%

Ferrovum sp. CARN8 metagenome comparison with public genomes of group 1 strains (*Ferrovum myxofaciens* P3G and *Ferrovum* sp. Z-31) and of group 2 strains (*Ferrovum* sp. JA12 and *Ferrovum* sp. PN-185). CheckM statistics, synteny values and bi-directional best blast hits were determined on-line using the MicroScope platform [31]. ANI and AAI values were computed on-line using the OrthoANIu [32] and the KostasLab AAI Calculator [44].

n.a.: not available (no 16S rRNA gene publicly available for *Ferrovum* Z-31 strain)

n.d.: not determined

Table 2

Physico-chemical characteristics of the COWG water at the time of sampling

pH	Temperature	Dissolved oxygen	FeII	SO_4^{2-}
3.83	15.1 °C	8.3 $mg.l^{-1}$	625 ± 30 $mg.l^{-1}$	3200 ± 300 $mg.l^{-1}$

The complete table can be found in [8].

Table 3

Expression values for metal-resistance genes for the seven dominant strains in Carnoulès

Gene name	CDS	mRNA copies/cell	Expression level
copA	CARN2_3637	4.17	robust
	CARN6_0942	3.98	robust
	CARN7_0874	3.05	robust
	CARN2_3636	1.96	medium
	CARN5_0167	1.73	medium
	CARN1_2059	1.10	medium
	CARN4_2512	0.66	medium
	CARN7_1019	0.58	medium
	CARN2_2558	0.29	low
	CARN2_2568	0.17	
	CARN5_0531	0.17	
cusA	CARN2_3944	0.78	medium
	CARN5_0162	0.49	low
	CARN6_0569	0.42	low
	CARN3_0713	0.06	
cusB	CARN2_3943	0.52	medium
	CARN5_0163	0.22	low
	CARN3_0714	0.06	
cusC	CARN2_3942	0.49	low
czcA	CARN3_0736	0.88	medium
	CARN6_2331	0.24	low
	CARN7_1940	0.10	
	CARN6_2568	0.08	
	CARN2_0914	0.02	
czcB	CARN3_0737	1.13	medium
czcC	CARN3_0738	1.57	medium

Expression values for metal-resistance genes *copA*, *cusABC* and *czcABC* for the seven dominant strains in Carnoulès metagenome published in [8]. These values were determined from metatranscriptomic sequences using the same methodology as in the present study for *Ferrovum* sp. CARN8. Groups of expression levels were assigned depending on the computed mRNA copies/cell values according to [24].