Diversity and spatiotemporal dynamics of bacterial communities: Physicochemical and other drivers along an acid mine drainage

Aurelie Volant, O. Bruneel, A. Desoeuvre, M. Héry, C. Casiot, Noëlle Bru, S. Delpoux, Anne Fahy, Fabien Javerliat, Olivier Bouchez, et al.

Diversity and spatiotemporal dynamics of bacterial communities: physicochemical and other drivers along an acid mine drainage

Aurélie Volant[1], Odile Bruneel[1], Angélique Desoeuvre[1], Marina Héry[1], Corinne Casiot[1], Noëlle Bru[2], Sophie Delpoux[1], Anne Fahy[3], Fabien Javerliat[3], Olivier Bouchez[4], Robert Duran[3], Philippe N. Bertin[5], Françoise Elbaz-Poulichet[1] & Béatrice Lauga[3]

[1]Laboratoire HydroSciences Montpellier, HSM, UMR 5569 (IRD, CNRS, Universités Montpellier 1 et 2), Université Montpellier 2, Montpellier, France; [2]Laboratoire de Mathématiques et de leurs Applications, UMR 5142 (CNRS), Université de Pau et des Pays de l'Adour, Pau, France; [3]Équipe Environnement et Microbiologie, EEM, UMR 5254 (IPREM, CNRS), Université de Pau et des Pays de l'Adour, Pau, France; [4]INRA Auzeville, Plateforme Génomique Chemin de Borde Rouge, Castanet-Tolosan, France; and [5]Département Microorganismes, Génomes, Environnement, Laboratoire de Génétique Moléculaire, Génomique, Microbiologie, GMGM, UMR 7156 (Université de Strasbourg, CNRS), Strasbourg, France

Correspondence: Odile Bruneel, Laboratoire HydroSciences Montpellier, UMR5569, Université Montpellier 2, Place E. Bataillon, CC MSE, 34095 Montpellier, France.
Tel.: (+33)4 67 14 36 59;
fax: (+33)4 67 14 47 74;
e-mail: odile.bruneel@msem.univ-montp2.fr

Keywords
spatial and temporal dynamics; bacterial diversity; acid mine drainage; arsenic.

Abstract

Deciphering the biotic and abiotic factors that control microbial community structure over time and along an environmental gradient is a pivotal question in microbial ecology. Carnoulès mine (France), which is characterized by acid waters and very high concentrations of arsenic, iron, and sulfate, provides an excellent opportunity to study these factors along the pollution gradient of Reigous Creek. To this end, biodiversity and spatiotemporal distribution of bacterial communities were characterized using T-RFLP fingerprinting and high-throughput sequencing. Patterns of spatial and temporal variations in bacterial community composition linked to changes in the physicochemical conditions suggested that species-sorting processes were at work in the acid mine drainage. Arsenic, temperature, and sulfate appeared to be the most important factors that drove the composition of bacterial communities along this continuum. Time series investigation along the pollution gradient also highlighted habitat specialization for some major members of the community (*Acidithiobacillus* and *Thiomonas*), dispersal for *Acidithiobacillus,* and evidence of extinction/re-thriving processes for *Gallionella*. Finally, pyrosequencing revealed a broader phylogenetic range of taxa than previous clone library-based diversity. Overall, our findings suggest that in addition to environmental filtering processes, additional forces (dispersal, birth/death events) could operate in AMD community.

Introduction

Acid mine drainage (AMD) is one of the most pernicious forms of pollution in the world and is widely recognized as having costly environmental and socioeconomic impacts (Hallberg, 2010). AMD occurs when waste from the extraction and processing of sulfide ore comes into contact with oxygenated water. Drainages are typically acidic and usually contain high concentrations of sulfate, metals and metalloids including arsenic. Although perceived as extreme environments hostile to life, a variety of microorganisms are able to thrive in it. For some of them, their role in the oxidation of sulfide minerals, which leads to bioleaching, is well known, as is their role in natural attenuation of such polluted waters (Edwards *et al.*, 2000; Hallberg, 2010; Johnson, 2012). Despite the central role of microorganisms in such ecosystem functioning, our understanding of the mechanisms shaping microbial community structure and diversity in AMD remains limited. As pointed out by Miller *et al.* (2009), deciphering how microbial communities are patterned along environmental gradients is a pivotal question in microbial ecology. Although AMD is characterized by changing conditions over time and space, few studies were interested in comparing the microbial communities along such environmental gradients. The AMD of

Carnoulès in southern France provides an excellent opportunity to investigate these fundamental questions of microbial ecology. This site is characterized by an acid pH (2–4) and high levels of metal and metalloids, in particular As (up to 10 g L^{-1} in the tailings stock pore water, and 100–350 mg L^{-1} at the source of Reigous Creek), and natural attenuation processes result in a strong spatial pollution gradient along the drainage. Indeed, nearly 95% of the arsenic in solution is removed between the source of Reigous Creek, which emerges from the mine tailings, and its confluence with the Amous River, 1.5 km downstream. To our knowledge, this AMD is one of the most As-rich AMDs reported to date (Morin & Calas, 2006). It is an outstanding example of adaptation to life in an extreme environment.

In this study, we used a combination of molecular approaches to investigate the spatial and temporal dynamics of bacterial communities in relation to the physicochemical parameters in Carnoulès acid mine drainage. Using 16S rRNA gene pyrosequencing and terminal restriction fragment length polymorphism (T-RFLP), our aim was (1) to characterize the spatial dynamics of the structure and composition of the bacterial communities along an environmental gradient, (2) to evaluate the temporal changes in the composition of the communities, and (3) to determine whether their dynamics could be linked to variations of environmental conditions.

Materials and methods

Description of the study site

The Pb–Zn Carnoulès mine, located in southern France, produced 1.2 Mt of solid wastes that are stored behind a dam and contain 0.7% Pb, 10% Fe, and 0.2% As. The aquifer is not fed by vertical percolation of rainwater through the tailings, but rather originates from natural springs that were buried under the tailings (Koffi *et al.*, 2003). The water table is 1–10 m below the surface of the tailings stock, depending on the season and location. With the exception of temperature, which is almost constant with average values around 15 °C, the physicochemical parameters of the groundwater vary as a function of the hydrological conditions (Casiot *et al.*, 2003b). In 2001, the groundwater below the tailings contained extremely high levels of arsenic: up to 10 000 mg L^{-1} (Casiot *et al.*, 2003b). The water emerges at the bottom of the dam, forming the source of the Reigous Creek. This AMD is acid ($pH \leq 3$), with high concentrations of sulfate (2000–7700 mg L^{-1}), iron (500–1000 mg L^{-1}), and arsenic (50–350 mg L^{-1}). Iron and arsenic are mainly present in their reduced forms Fe(II) and As(III) (Casiot *et al.*, 2003a). The natural attenuation of As is the result of microbiologically mediated As–Fe coprecipitation (Morin *et al.*, 2003; Bruneel *et al.*, 2006). 10–47% of Fe, and 20–60% of As are removed from the aqueous phase within the first 30 m of the creek. Beyond this point (COWG sampling site, located 30 m downstream from the spring, Fig. 1), the Reigous receives water from quarries and mine galleries, especially after rainfall events, which strongly influence its acidity and metal content (Egal *et al.*, 2010).

Sampling procedure and measurement of physicochemical properties

Six sampling campaigns were carried out in November 2007, February 2008, October 2008, March 2009, November 2009, and January 2010 at five sampling sites, resulting in a set of 30 samples. Groundwater was collected from a borehole (S5, between 10 and 12 m deep) located within the tailings. Water samples were also taken at four sites along Reigous Creek (collecting downstream seepage water from the surroundings) at the spring (S1), 30 m downstream from the spring (COWG), 150 m downstream (GAL), and 1500 m downstream (CONF), just before the confluence between Reigous Creek and the Amous River (Fig. 1). Water samples (300 mL) were immediately filtered through sterile 0.22 μm Nuclepore filters, which were transferred to a collection tube (Nunc), frozen in liquid nitrogen, and stored at −80 °C until DNA extraction. This sampling was carried out in triplicate. Measurements of water conductivity, temperature, redox potential, pH, and dissolved oxygen concentration were carried out as previously described (Bruneel *et al.*, 2011). For chemical analyses, 500 mL water samples were immediately filtered through 0.22 μm Millipore membranes fitted on Sartorius polycarbonate filter holders. For total Fe and As determination, the filtered water was acidified to pH 1 with HNO_3 (14.5 M) and stored at 4 °C in polyethylene bottles until analysis. A 20 μL aliquot of the filtered water sample was added either to a mixture of acetic acid and EDTA (Samanta & Clifford, 2005) for As speciation or to a mixture of 0.5 mL acetate buffer (pH 4.5) and 1 mL of 1,10-phenanthrolinium chloride solution (Rodier *et al.*, 1996) for Fe(II) determination. The vials were filled to 10 mL with deionized water. Samples destined for Fe and As speciation and sulfate determination were stored in the dark and analyzed within 24 h. Chemical analyses were carried out as previously described (Bruneel *et al.*, 2011).

DNA isolation

Genomic DNA was extracted in triplicate from filtered water using the UltraClean Soil DNA Isolation Kit

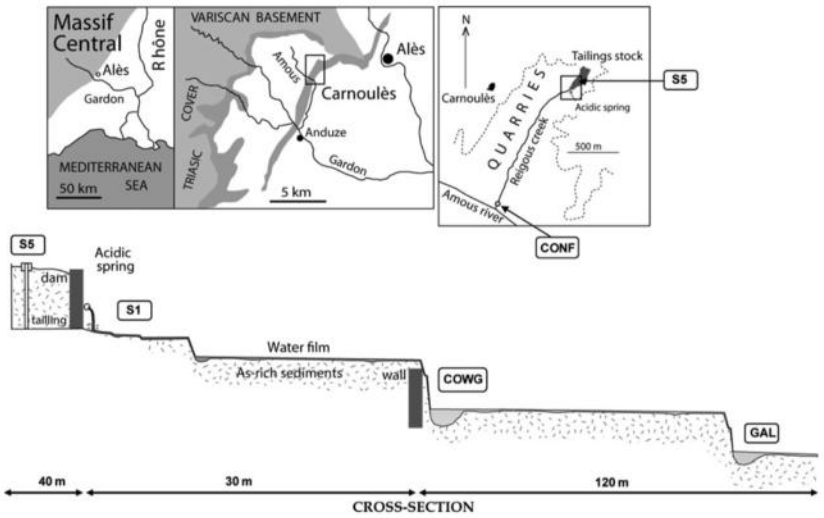

Fig. 1. Map of Carnoulès mine and location of sampling sites.

(MoBio Laboratories Inc., Carlsbad, CA) according to the manufacturer's recommendations. These triplicate extractions were pooled before PCR amplification. All genomic DNA extracts were stored at −20 °C until further analysis.

Terminal restriction fragment length polymorphism

The 16S rRNA genes were amplified by PCR, and the bacterial community structure was identified by T-RFLP. The fluorescent labeled primers HEX 357F (5′-hexachloro-fluorescein-phosphoramidite-CCTACGGGAGGCAGCAG-3′) (Lane, 1991) and 926R (5′-CCGTCAATTCMTTTRAGT-3′) (Muyzer & Ramsing, 1995), described as universal within the bacterial domain, were used. Triplicate PCR amplifications were performed on each sample. The reaction mixture contained 1 μL of DNA template, 1 μL of both primers (10 μM), and 12.5 μL of PCR Master Mix Ampli Taq Gold 360 (Applied Biosystems, Foster City, CA). Sterile distilled water was added to obtain a final volume of 25 μL. PCR conditions were as follows: one cycle at 95 °C for 10 min, 35 cycles at 95 °C for 45 s, 55 °C for 45 s, and 72 °C for 45 s, followed by 10 min at 72 °C. The 90 PCR products were purified with Illustra GFX™ PCR DNA and the Gel Band Purification Kit (GE Healthcare, Munich, Germany). The concentration of PCR product was determined by comparison with molecular markers (Smartlader, Eurogentec) after migration on agarose gel. Approximately 100 ng of purified amplicon was digested in 10 μL reaction with 0.3 U of enzyme HpaII or AluI (New England Biolabs Inc., Ipswich, MA) at 37 °C for 3 h. Terminal restriction fragment (T-RF) profiles were obtained from the digested amplicons by suspending 1 μL aliquots in 8.75 μL formamide with 0.25 μL of Genescan ROX 500 size standard (Applied Biosystems). T-RFs were separated on an ABI PRISM 3130xl Genetic Analyser (Applied Biosystems). Data were collected and analyzed using GENEMAPPER software (version 1.4, Applied Biosystem). To increase stringency for the T-RF profiles of 16S rRNA genes, T-RFs outside the range of the size standard (35–500 bp) were discarded, and the background noise level was set at 30 fluorescence units. T-ALIGN software (Smith *et al.*, 2005) was used to compare replicate profiles and to generate consensus profiles containing only T-RFs that occurred in replicate reactions. Consensus profiles were then aligned on the basis of the length of the T-RFs and individual peak areas as previously described by Smith *et al.* (2005) with the confidence interval set at 0.5, resulting in the generation of data sets of aligned T-RFs that gave individual relative peak areas as a percentage of the overall profile. T-RFs were included in the subsequent analysis if they represented > 1% of the cumulative peak height for the sample.

Construction of the libraries, 454-pyrosequencing, and sequence quality control

The 16S rRNA genes were also amplified by PCR for multiplex pyrosequencing using barcoded primers. The primer pairs used, targeting the V3 to V5 variable regions of the 16S rRNA gene, were 357F (5′-AxxxCCTACGGGAGGCAGCAG-3′) and 926R (5′-BxxxCCGTCAATTCMTTTRAGT-3′). A and B represent the two FLX Titanium adapters (A adapter sequence: 5′-CGTATCGCCTCCCTCGCGCCATCAG-3′; B adapter sequence: 5′-CTAT

GCGCCTTGCCAGCCCGCTCAG-3′), and xxx represent the sample-specific barcode sequence. PCR was performed using 30–35 cycles under conditions identical to those described above for T-RFLP. The number of cycles was varied with the samples to obtain a strong band with a minimum number of cycles to respect the initial abundances of bacterial communities. The 90 PCR products with a proximal length of 569 bp were excised from 1% agarose gel and purified with the QIAquick Gel Extraction Kit (QIAGEN Inc., Valencia, CA). To minimize random PCR bias, triplicates were pooled in equimolar ratios prior to pyrosequencing. Pyrosequencing of the 30 amplicon libraries was performed on a GS-FLX-Titanium sequencer (Roche 454 Life Sciences) at the GenoToul genomic platform in Toulouse (France) using four separate 1/8 region of a plate.

Processing of pyrosequencing data and taxonomic classification

Preliminary quality checks, sorting, and trimming of the 454 reads were carried out using the NG6 pipeline (http://vm-bioinfo.toulouse.inra.fr/ng6/). Tags were extracted from the 454 reads using the sff file (Roche software), and three kinds of analysis were performed as described by Ueno *et al.* (2010): (1) BLAST search for *E. coli*, phage, and yeast contaminants, (2) read quality analysis, and (3) removal of sequences that were too long or too short (sequences with more or less than two standard deviations from the mean), sequences containing more than 4% of N, low-complexity sequences and duplicated reads, using Pyrocleaner. The sequences were then analyzed with the software Mothur version 1.30 (Schloss *et al.*, 2009). Preprocessing of unaligned sequences included removing sequences < 450 bp, all sequences containing ambiguous characters, and sequences with more than eight homopolymers. We also removed sequences that did not align over the same span of nucleotide positions. Identical sequences were grouped, and representative sequences were aligned against the SILVA bacterial and archaeal reference database using the Needleman–Wunsch algorithm (Needleman & Wunsch, 1970). Chimeric sequences were detected and removed using the implementation of Chimera Uchime. A further screening step (precluster) was carried out to reduce sequencing noise by clustering reads differing by only one base every 100 bases (Huse *et al.*, 2010). The remaining high-quality reads were used to generate a distance matrix and were clustered into operational taxonomic units (OTUs) defined at 97% cutoff using the average neighbor algorithm. Next, the OTUs were phylogenetically classified to genus level using the naive Bayesian classifier (80% confidence threshold) trained on the RDP taxonomic outline implemented in Mothur and a modified bacterial database. *In silico* T-RF prediction of the 16S rRNA gene sequences obtained in this study was performed using the program TRiFLE (Junier *et al.*, 2008), and predicted T-RFs were linked to measured T-RFs from the microbial community profiles.

Estimation of diversity and statistical analysis

Diversity indices

Nonparametric Chao1 and Shannon alpha diversity estimates, as well as coverage and rarefaction curves, were calculated with MOTHUR v.1.30 for each sample. Analysis of variance (ANOVA) was performed with Tukey's tests to identify differences between sampling sites.

Cluster analysis

To compare community composition based on T-RFLP and 454-pyrosequencing data, normalized OTUs abundances were square-root-transformed and pairwise dissimilarities among samples were calculated using the relative abundance-based Bray–Curtis index (BC). Nonmetric multidimensional scaling (nMDS) analysis was performed on the dissimilarity matrices to visualize patterns of community composition. Using the 454-pyrosequencing data, we carried out a random sampling procedure to make equal the number of sequences per sample (486 sequences) and we removed singleton OTUs (sequences that only occurred once) to reduce the influence of rare OTUs. One-way analysis of similarity (ANOSIM) and multiple pairwise comparisons were used to test whether there were significant differences in community composition in space. R-values > 0.75 are commonly interpreted as well-separated bacterial compositions; $R > 0.5$ as overlapping, but clearly different; and $R < 0.25$ as practically not separable.

CCA

Canonical correspondence analyses (CCAs) were used to explore variations in the bacterial communities under the constraint of our set of environmental variables. Explanatory variables were $\log(x + 1)$-transformed where necessary to approximate normal distribution. This model was tested with Monte Carlo permutation tests (499 randomized runs) to determine significance, and each environmental parameter was tested by stepwise analysis to detect significant predictors. All statistical analyses were performed with R 3.0.1 (R Development Core Team, 2012) including the *VEGAN* package.

Results

Spatial and temporal analyses of the environmental data set

The physicochemical characteristics of the water samples were determined for six sampling dates and at five different sites in a borehole and along Reigous Creek (Fig. 1; Supporting Information, Table S1). With the exception of pH, the environmental variables measured differed significantly between sites (ANOVA, $P < 0.05$). A significant decrease (Tukey's test, $P < 0.01$) in temperature was observed in Reigous Creek, where the influence of air temperature causes a larger range of values (Fig. 2a) in contrast to the temperature of the water in the borehole (S5) and at the source of the Reigous (S1), which did not differ significantly (13.3 ± 1.4 °C and 13.6 ± 1 °C, respectively, and Tukey's test, $P = 0.99$). All samples were characterized by low pH (≤ 3.7). No significant variations in pH (ANOVA, $P = 0.49$) were observed between the sampling sites located downstream from the source (COWG, GAL and CONF; Fig. 2b). Dissolved oxygen (DO) concentrations in the water at the upstream sites presented a mean of 1.0 ± 1.0 mg L^{-1} for S5 and 0.8 ± 0.6 mg L^{-1} for S1, denoting generally suboxic conditions at these sites. DO increased sharply between S1 and COWG (mean of 6.1 ± 1.8 mg L^{-1}) (Fig. 2c) and continued to increase slightly all along the creek, to reach a mean of 10.5 ± 1.3 mg L^{-1} at CONF. The redox potential (Eh) showed average values of 558 ± 85 mV at S5. Eh increased along Reigous Creek from 512 ± 39 mV at S1 to 635 ± 99 mV at CONF (Fig. 2d). In contrast, average conductivity decreased from 7588 ± 5203 μS cm^{-1} at S5 to 5121 ± 631 μS cm^{-1} at S1 and reached minimum (1612 ± 97 μS cm^{-1}) at CONF (Fig. 2e). Sulfate (SO_4^{2-}) concentrations were maximum in the groundwater at S5 with a mean of $14\,080 \pm 10\,630$ mg L^{-1}. After a sharp decrease at S1 (average values of 2682 ± 1180 mg L^{-1}), concentrations gradually decreased along Reigous Creek (Fig. 2f). Dissolved Fe concentrations in the groundwater at S5 exhibited average values of 4474 ± 2855 mg L^{-1}. Fe concentrations decreased from the source (S1, average values of 1317 ± 383 mg L^{-1}) to CONF, where Fe remained below 82 mg L^{-1} (Fig. 2g). The proportion of Fe(III) (difference between Fe(total) and Fe(II)) was generally negligible except at some sampling dates at S5 and CONF (Table S1). At S5, concentrations of dissolved arsenic (As) exhibited an average value of 440 ± 184 mg L^{-1} (Fig. 2h). Along Reigous Creek, As concentrations decreased with increasing distance from the source (average value of 175 ± 71 mg L^{-1}), to values below 6 mg L^{-1} at CONF (Fig. 2h), with predominance of the reduced form As(III). An average of 65% of sulfate, 96% of iron, and 99% of arsenic were removed from the aqueous phase between S1 and CONF sampling sites.

Diversity and species richness estimators of bacterial communities

Hex-labeled PCR products were digested separately with two restriction enzymes. HpaII that produced the largest numbers of T-RFs (data not shown) was used to assess the differences in the microbial communities. T-RFLP profiles generated showed a total of 43 different T-RFs for the five sites, and the number of T-RFs detected in each sample varied from 2 to 17 (Fig. 3). Average T-RF richness (number of T-RFs) and average Shannon diversity indices calculated from relative peak intensity data were highest at S1 and COWG ($H = 2.03 \pm 0.3$ and $H = 1.96 \pm 0.3$, respectively), while the lowest values were observed at GAL ($H = 1.26 \pm 0.4$; Fig. 4a). Values at CONF were intermediate ($H = 1.67 \pm 0.7$). Although bacterial community diversity varied among the sampling sites, the differences were not significant (ANOVA, $F = 2.46$, $P = 0.071$). For each site, the bacterial community showed variations over time, but no particular trend could be identified. Some T-RFs were found in the majority of the profiles (e.g. T-RF 150), where they usually accounted for a high proportion of the total T-RFs (Fig. 3). Between one and three site-specific T-RFs were identified in all the sites (in red in Fig. 3).

A total of 99 441 sequence reads were generated in a single run of 454-pyrosequencing from 30 independent 16S rRNA gene libraries. Note that pyrosequencing of two samples taken in February 2008 (S5 and S1) failed and were thus excluded from analysis. After trimming and processing with Mothur, 63 442 reads remained with an average length of 530 bp. Clustering of the remaining sequences led to the identification of 6613 OTUs (including 4510 singletons) defined at 97% identity. Although singletons represented 68% of the total number of OTUs, they only accounted for 7% of the total DNA sequences. The results of rarefaction analysis along with the Chao1 and the Shannon indexes and coverage values are listed in Table 1. In the resampled data set, Good's coverage ranged from 69% to 97% with an average value of 85%, indicating that the majority of bacterial phylotypes were recovered. Species richness (Chao1 index) of the bacterial communities presented significant variations along Reigous Creek (ANOVA, $P = 0.001$, $F = 6.66$) (Fig. 4b). The nonparametric estimators Chao1 ranged between 52 and 495 estimated OTUs for all the sites considered (Table 1). The highest average OTU richness was found at CONF and S1 (Chao1 = 364 ± 145 and 296 ± 32, respectively), suggesting that an important number of OTUs were not revealed by the analysis of these two sites. Situated

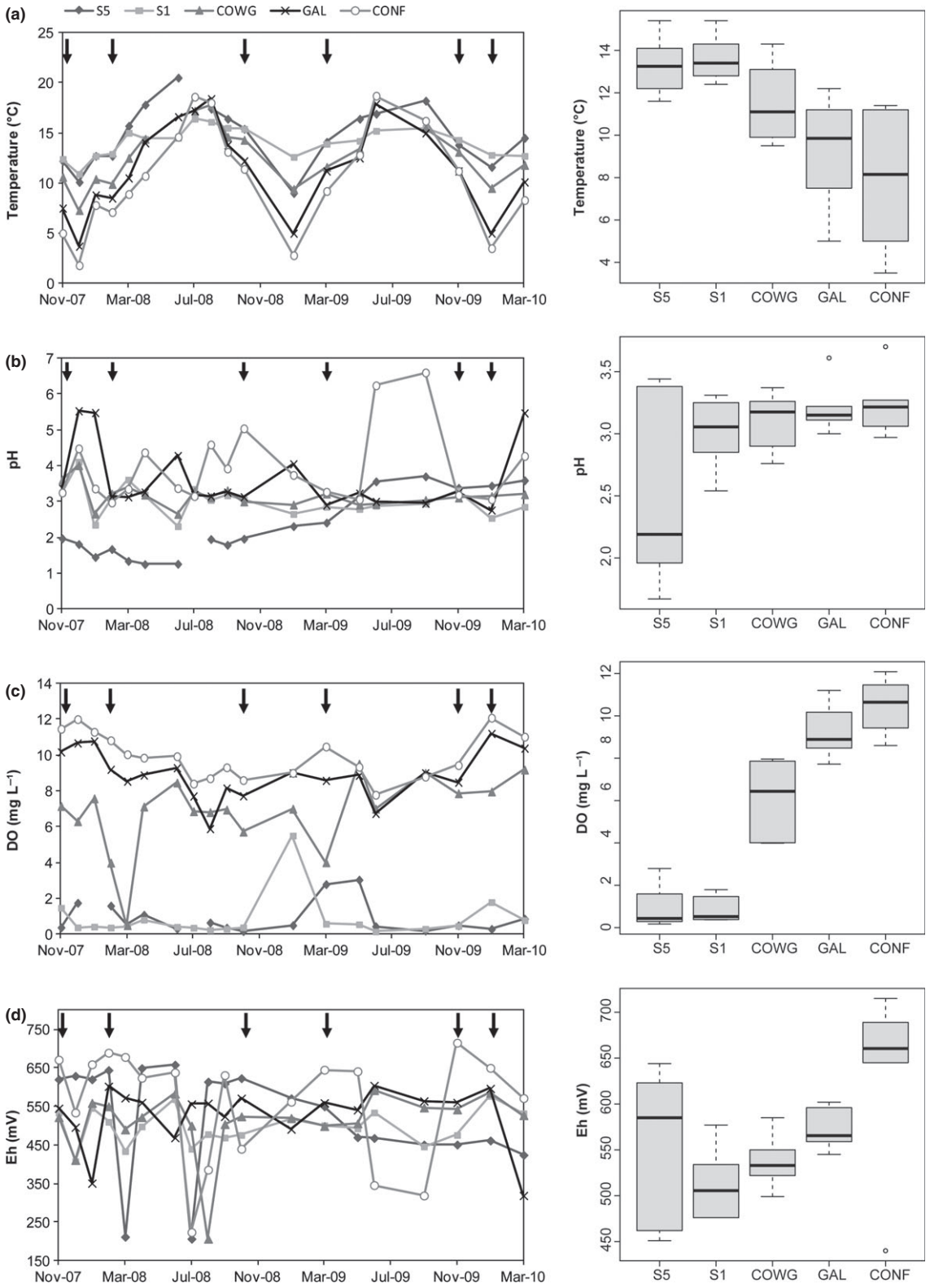

Fig. 2. Variations in the main physicochemical parameters over the course of the study and boxplot of each variable per sampling site. Arrows indicate sampling dates for T-RFLP and pyrosequencing analysis. Note that some data are missing, as shown by the gaps in the curves. DO, dissolved oxygen; Eh, redox potential.

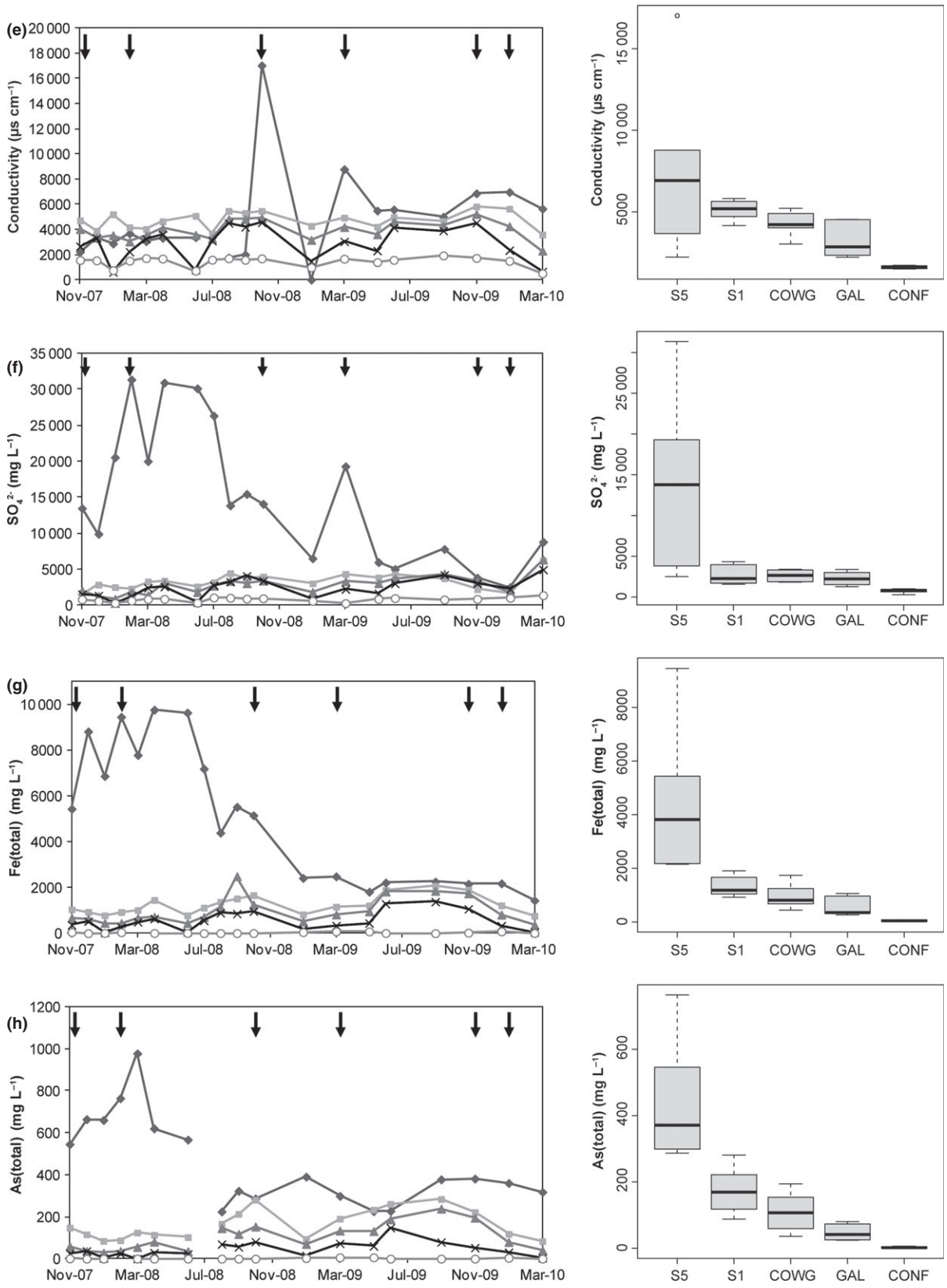

Fig. 2. Continued

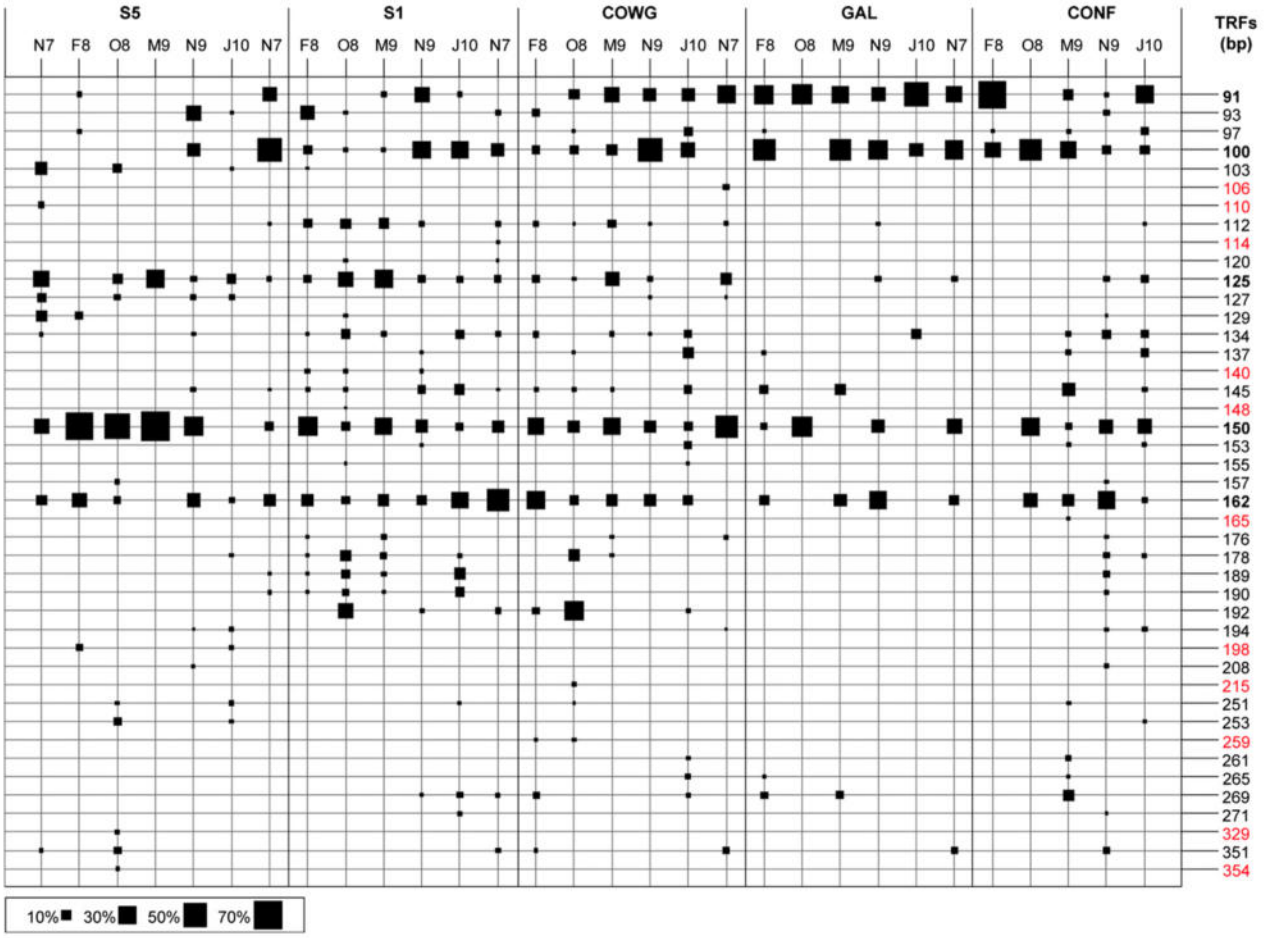

Fig. 3. Relative abundance of terminal restriction fragments (T-RFs) derived from bacterial communities. Single T-RFs per sampling site are in red. Dominant T-RFs are in bold. Taxonomic affiliation of T-RFs was carried out by *in silico* T-RFLP analysis. T-RF 91 and 100 could not be assigned to any phylogenetic group; T-RF 125 represented *Armatimonadetes* gp4 and *Chlorobi*; T-RF 150 was mainly related to *Deinococcus-Thermus*, *Spirochaetes* and *Actinobacteria*; T-RF 162 was mainly related to *A. ferrooxidans* but could be assigned to other proteobacterial phylotypes detected in the AMD. N7: November 2007; O8: October 2008; M9: March 2009; N9: November 2009; J10: January 2010.

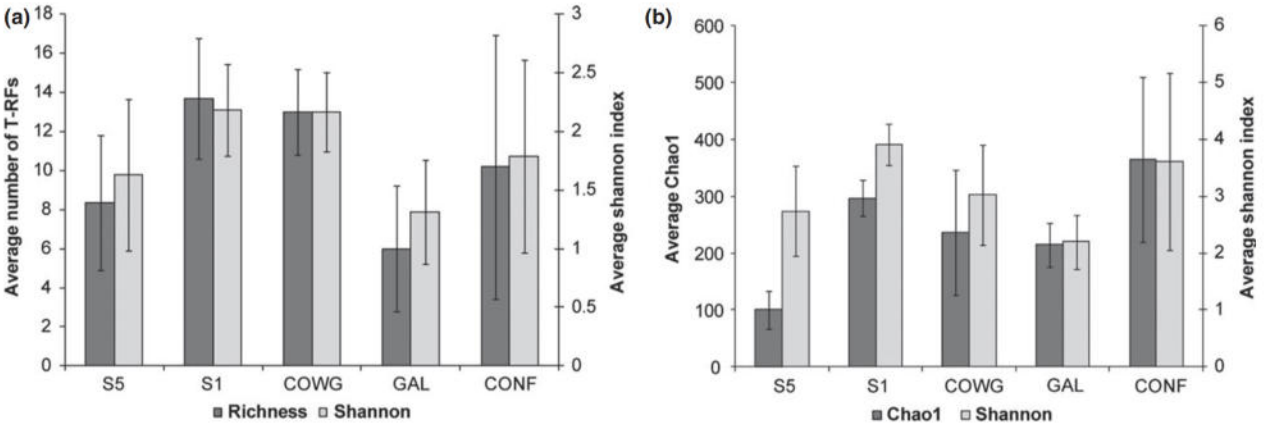

Fig. 4. Average diversity and species richness index per group ± standard deviation calculated based on (a) T-RFLP profiles and (b) 454 pyrosequencing reads of the reduced data set based on the 16S rRNA genes.

Table 1. Estimated OTU richness, diversity indices, and estimated sample coverage for each 16S rRNA gene library. Results are presented for full data set reads (full data set) and for reduced data sets without singletons and randomly resampled to make the sample size equal (reduced data set)

	Reduced data set					Full data set				
Sampling sites	No. of reads	Obs. OTUs*	Chao1	Shannon†	Coverage‡	No. of reads	Obs. OTUs*	Chao1	Shannon†	Coverage‡
S5										
S5N7	486	79	116 (95; 163)	3.18 (3.04; 3.32)	92	2089	216	576 (435; 810)	3.16 (3.07; 3.24)	93
S5O8	486	32	52 (38; 98)	1.66 (1.52; 1.80)	97	2523	87	258 (165; 465)	1.49 (1.42; 1.56)	98
S5M9	486	52	89 (67; 146)	2.14 (1.99; 2.29)	94	2838	177	381 (298; 520)	2.20 (2.13; 2.28)	96
S5N9	486	92	142 (115; 199)	3.31 (3.17; 3.46)	91	1086	159	481 (334; 751)	3.22 (3.11; 3.34)	91
S5J10	486	70	102 (82; 150)	3.40 (3.29; 3.51)	94	2354	195	598 (423; 908)	3.41 (3.35; 3.48)	95
S1										
S1N7	486	138	266 (209; 368)	3.86 (3.71; 4.01)	83	2719	436	1329 (1051; 1773)	3.88 (3.79; 3.96)	90
S1O8	486	131	289 (217; 422)	3.93 (3.80; 4.05)	83	3422	727	2396 (1994; 2925)	4.59 (4.52; 4.66)	85
S1M9	486	127	270 (204; 392)	3.82 (3.68; 3.95)	84	2057	365	1068 (844; 1397)	3.97 (3.89; 4.06)	88
S1N9	486	137	315 (237; 455)	3.48 (3.30; 3.65)	81	2021	392	1040 (848; 1313)	3.53 (3.41; 3.64)	86
S1J10	486	181	342 (277; 449)	4.48 (4.35; 4.61)	78	4573	845	2158 (1859; 2544)	4.81 (4.75; 4.88)	88
COWG										
CGN7	486	121	193 (159; 256)	3.75 (3.61; 3.89)	87	688	194	367 (301; 474)	4.06 (3.93; 4.19)	82
CGF8	486	105	201 (154; 295)	3.16 (2.99; 3.33)	87	2160	305	835 (659; 1099)	3.21 (3.11; 3.30)	90
CGO8	486	98	206 (152; 312)	2.90 (2.72; 3.07)	87	2119	255	1025 (715; 1544)	2.82 (2.72; 2.92)	92
CGM9	486	146	382 (279; 565)	3.87 (3.72; 4.01)	79	2756	511	1710 (1364; 2196)	4.11 (4.02; 4.19)	87
CGN9	486	112	349 (232; 578)	3.00 (2.81; 3.18)	84	1317	201	624 (455; 907)	2.80 (2.67; 2.93)	89
CGJ10	486	50	88 (65; 148)	1.44 (1.26, 1.62)	94	1827	136	472 (311; 781)	1.51 (1.41; 1.62)	95
GAL										
GLN7	486	91	144 (117; 200)	2.28 (2.08; 2.49)	90	1638	223	543 (423; 734)	2.23 (2.10; 2.35)	91
GLF8	486	65	224 (135; 430)	1.33 (1.14; 1.52)	90	1679	177	665 (456; 1030)	1.56 (1.44; 1.68)	92
GLO8	486	93	253 (172; 418)	2.57 (2.39; 2.75)	87	2093	236	614 (473; 839)	2.64 (2.54; 2.74)	93
GLM9	486	107	239 (178; 354)	2.36 (2.14; 2.57)	85	2489	356	1070 (843; 1403)	2.49 (2.37; 2.60)	90
GLN9	486	110	227 (171; 334)	2.60 (2.39; 2.82)	86	1682	285	715 (570; 934)	2.65 (2.51; 2.78)	88
GLJ10	486	83	201 (140; 331)	2.05 (1.85; 2.25)	88	2246	271	862 (654; 1183)	2.12 (2.01; 2.23)	91
CONF										
CFN7	486	239	492 (400; 635)	5.07 (4.97; 5.18)	70	3256	933	1995 (1764; 2290)	5.92 (5.87; 5.98)	84
CFF8	486	62	186 (115; 352)	1.46 (1.27; 1.65)	91	1768	240	972 (692; 1428)	1.94 (1.82; 2.06)	89
CFO8	486	191	335 (279; 426)	4.40 (4.25; 4.55)	77	1731	548	1163 (1004; 1377)	5.02 (4.93; 5.12)	80
CFM9	486	84	202 (142; 323)	1.87 (1.66; 2.07)	88	527	125	462 (306; 751)	2.28 (2.07; 2.50)	81
CFN9	486	233	477 (390; 612)	4.82 (4.69; 4.95)	69	1994	699	1708 (1467; 2026)	5.45 (5.37; 5.53)	77
CFJ10	486	180	495 (367; 711)	4.06 (3.89; 4.23)	74	5718	1343	3763 (3330; 4289)	4.79 (4.71; 4.86)	84

*OTUs were defined at 97% cutoff.

†Takes into account the number and evenness of species.

‡Coverage: sum of probabilities of observed classes calculated as $(1 - (n/N))$, where n is the number of singleton sequences and N is the total number of sequences.

Values in brackets are 95% confidence intervals.

between these two sites, COWG and GAL exhibited intermediate richness estimates (Chao1 = 236 ± 110 and 215 ± 39, respectively). The lowest richness was observed in the tailing groundwater at S5 (Chao1 = 100 ± 33). Bacterial OTU diversity, estimated by the Shannon index, also differed significantly between sites (ANOVA, $F = 3.01$, $P = 0.039$), with values ranging from 1.33 to 5.07 (Table 1). In agreement with T-RFLP data analysis (Fig. 4a), the highest average diversity value was found at S1 (H = 3.91 ± 0.36) and the lowest value at GAL (H = 2.20 ± 0.47) (Fig. 4b). As predictable, average T-RF diversity is lower than OTU diversity (*c.* 50%); indeed, taxon-specific resolution of pyrosequencing is much higher than fingerprinting (Pilloni *et al.*, 2012). Again, no seasonal trend was observed. The same richness and diversity patterns were observed in both the full and resampled data sets, although the richness estimator and Shannon index were higher in the full data set, due to the larger number of sequences (data not shown).

Taxonomic assignment of bacterial pyrosequencing reads and T-RFs

At a confidence threshold of 80%, we were able to assign 56 426 of 63 442 qualified reads (that is, 89%) to a known phylum (Table S2) and 76% to a known order (Supporting Information, Fig. S1). Most of the unclassified reads (55% representing 9.6% to 37% of the qualified reads of each sample) were associated with samples collected at S1. Altogether, 23 bacterial phyla were recovered from our samples, with 4–8 different phyla found in samples collected at S5, 12–13 at S1, 9–14 at COWG, 10–13 at GAL, and 9–20 at CONF (Table S2). Most of the bacterial sequences (86%) belonged to phyla that are most often encountered in acid mine drainages worldwide (*Proteobacteria*, *Actinobacteria*, *Firmicutes, Acidobacteria*, *Bacteroidetes*, and *Nitrospirae*). In addition, microorganisms representing 0.5% of the total sequences were related to CARN1, '*Candidatus* Fodinabacter communificans'. *Proteobacteria* was the most abundant phylum in all the samples, accounting for 69.6% of all sequences retrieved. This phylum was represented by bacteria belonging to the *Alphaproteobacteria*, *Betaproteobacteria*, *Gammaproteobacteria*, *Deltaproteobacteria,* and *Epsilonproteobacteria*. The most abundant classes in nearly all the samples were *Betaproteobacteria* and *Gammaproteobacteria* (average values of 63.4% and 30.4% of the pyrosequencing reads, respectively). There were dominated by *Gallionellales* and *Acidithiobacillales*, respectively (Fig. S1a). Three other phyla, *Actinobacteria* represented mainly by *Acidimicrobiales* and *Actinomycetales*, *Firmicutes* (principally *Clostridiales* and *Bacilliales)*, and *Acidobacteria,* were also abundant but their proportion varied depending on the sample analyzed (Fig. S1b–d). Most of the sequences associated with *Acidobacteria* could not be classified to the order level except for *Acidobacteriales* and *Holophagales*. As can be seen in Fig. S2, a relatively small number of OTUs dominated at all sites (> 1% in total abundance per sample). The most abundant OTUs were phylogenetically related to *Gallionella ferruginea (Gallionellales)*, *Acidithiobacillus ferrooxidans* (*Acidithiobacillales*), and *Thiobacillus* sp. (*Hydrogenophilales*), collectively accounting for 41% of all the sequences.

When possible, T-RFs were assigned to a taxon or a group of taxon by *in silico* restriction of 16S rRNA gene sequences. Among the five dominant T-RFs (91, 100, 125, 150, and 162 bp in size), T-RFs 91 and 100 could not be assigned to any specific phylogenetic group. *Armatimonadetes* gp4 and *Chlorobi* were represented by T-RF 125, and T-RF 150 was mainly related to *Deinococcus-Thermus*, *Spirochaetes,* and *Actinobacteria*. While T-RF 162 was mainly related to *A. ferrooxidans*, it could be assigned to other proteobacterial phylotype detected in this AMD.

Spatial and temporal variations in bacterial community structure

Spatiotemporal dynamics of bacterial populations were identified by T-RFLP analysis and 454-pyrosequencing of 16S rRNA genes (Fig. S3).

Although samples formed overlapping clusters on the nMDS plot of the T-RFLP profiles, weak but significantly different bacterial communities at the five sites were revealed (ANOSIM Global $R = 0.2819$, $P < 0.001$). S5 differed significantly from the other sites, with some overlapping communities (pairwise tests: *r*-values ranging from 0.45 to 0.61, $P < 0.05$). These results highlighted changes in the structure of the bacterial communities between the tailing groundwater (S5) and the water in Reigous Creek. The high dissimilarity observed within each site revealed variations in community structure over time, especially at S5, GAL, and CONF (Fig. S3a). These variations may have masked a spatial pattern.

nMDS analyses of 454-pyrosequencing data also showed that the composition of the bacterial communities differed significantly along the spatial gradient from the sterile (S5) to the confluence (CONF) (Fig. S3b). Furthermore, an ANOSIM test corroborated the nMDS plot data, revealing significantly different bacterial compositions in water as a function of the spatial location (Global $R = 0.6192$, $P < 0.001$), except at GAL and COWG which did not differ significantly (ANOSIM pairwise comparison $r = 0.206$, $P = 0.37$). Higher temporal variation at CONF was highlighted by the large cluster on the nMDS plot. The marked temporal variations in the bacterial community at S5 and CONF highlighted by the two data sets

may be linked to the stronger seasonal fluctuation of some physicochemical parameters at these sites, particularly temperature at CONF (Fig. 2a) and pH, Eh, sulfate, Fe, and As at S5 (Fig. 2b, d, f, g and h).

We investigated the four most abundant phyla to get a global view of the variations of bacterial communities along the creek (Fig. S1). *Proteobacteria* distribution varied between samples, with a relative predominance of *Acidithiobacillales* in samples from S5 and S1, followed by the dominance of *Gallionellales* in the majority of other samples (Fig. S1a). A general increase in *Betaproteobacteria* was observed in the downstream direction of Reigous Creek. Among *Actinobacteria*, the *Acidimicrobiales* were present in all samples except those collected at CONF where almost equal proportions of *Acidimicrobiales* and *Actinomycetales* were retrieved (Fig. S1b). The *Firmicutes* phylum revealed the dominance of *Clostridiales* in samples from S5, whereas *Bacillales* were dominant at the other sites, again except CONF. Different orders were dominant at CONF over time, including *Lactobacillales* and *Selenomonadales* (Fig. S1c). No *Acidobacteria* were retrieved at S5 in October 2008 (Fig. S1d).

We also assessed the dynamics of the dominant genera (> 5% in total abundance per sample) (Fig. 5). The relative abundance of genera at each site varied over the sampling period. While the relative abundance of *Gallionella* was almost constant in COWG and GAL, there was an important temporal variation in the other sites. This was evident in S5 where this genus was extinct and re-thrived over time (Fig. 5a). Although *Gallionella* was present in almost all sites for a sampling date, there was no clear relationship between sites. Indeed, GAL and COWG exhibited a relatively high proportion of *Gallionella* at almost all the sampling dates (as much as 85% of all pyrosequencing reads at GAL) without any link with the upstream sites S5 and S1 (Fig. 5b). In contrast, *Acidithiobacillus* represented a minor fraction of the bacterial community, except at S5 where this OTU was dominant (24–72% of the pyrosequencing reads). The relative abundance of *Acidithiobacillus* showed a decreasing trend along the continuum for each sampling date (Fig. 5b). This genus also exhibited an increase in October 2008 and March 2009 for S5 and S1 (Fig. 5a). Members of the *Thiobacillus* genus showed higher proportion in COWG at each sampling date independently of the other sites (Fig. 5b). The temporal variation of this genus was minor except an important increase at GAL in October 2008 (Fig. 5a).

T-RFLP profiles from the 30 samples were investigated to assess the dynamics of T-RFs (Fig. 3). Among the five dominant T-RFs, T-RFs 91 and 100 represented a large proportion of the T-RFLP profiles in almost all samples except those from S5. T-RF 125 related to *Armatimonadetes* gp4 and *Chlorobi* was more abundant in samples from S5 and S1 than in downstream samples. T-RF 150 (*Deinococcus-Thermus*, *Spirochaetes,* and *Actinobacteria* – related) was the most abundant phylotype in the tailings groundwater (S5), accounting for up to 70% of the T-RFLP profiles. It was relatively abundant along the creek especially at GAL. T-RF 162 (related to *A. ferrooxidans* but also to other proteobacterial phylotype) was not dominant at S5 and represented a minor fraction of the bacterial community along the creek.

Linking bacterial community structure to environmental variables

Canonical correspondence analysis (CCA) was performed to elucidate the main relationships between physicochemical variables and bacterial community structure and composition (Fig. 6). Samples were plotted in different areas of the diagram depending on their environmental characteristics. The resolution of 454-pyrosequencing allowed to account for more variation than T-RFLP (36.4% and 20.5%, respectively) in the species–environment relationship across the first two canonical axes. CCA axis 1 based on T-RFLP data only separated the samples into two clusters, one containing the tailings site (S5) and the other grouping the sites along the creek (S1, COWG, GAL, and CONF), with sulfate, DO, and pH being the strongest determinants of bacterial community structure (Fig. 6a). In contrast, a higher resolution was observed with CCA axis 1 based on 454-pyrosequencing data, which was most closely correlated with iron, arsenic, and conductivity and separated the sampling sites into three clusters (CONF, GAL+COWG, and S1+S5) as a function of the pollution gradient (Fig. 6b). The upstream site S5 was highly polluted and little oxygenated, whereas the downstream site CONF was less polluted and characterized by a higher redox potential (Eh). CCA axis 2 separated samples according to water temperature. After the Monte Carlo permutation test, the environmental variables significantly correlated with the canonical axes based on 454-pyrosequencing data were arsenic (F-ratio = 1.9, P = 0.01), temperature (F-ratio = 1.4, P = 0.01), and sulfate (F-ratio = 1.4, P = 0.01). The differences between the two data sets were probably due to the power of 454-pyrosequencing over T-RFLP for taxon resolution. Focusing on 454-pyrosequencing data, the influence of environmental variables on dominant OTUs (> 5% of total abundance per sample) was also investigated (Fig. S4). Nineteen of the 23 dominant OTUs showed a strong correlation with the physicochemical parameters. Five OTUs (15, 16, 28, 32, and 52) were strongly correlated with elevated DO and Eh and 14 with high temperatures and high concentrations of As, Fe, and sulfate. As indicated by the position of OTUs 1, 3, 4, and 11 on the graph, near the origin of the

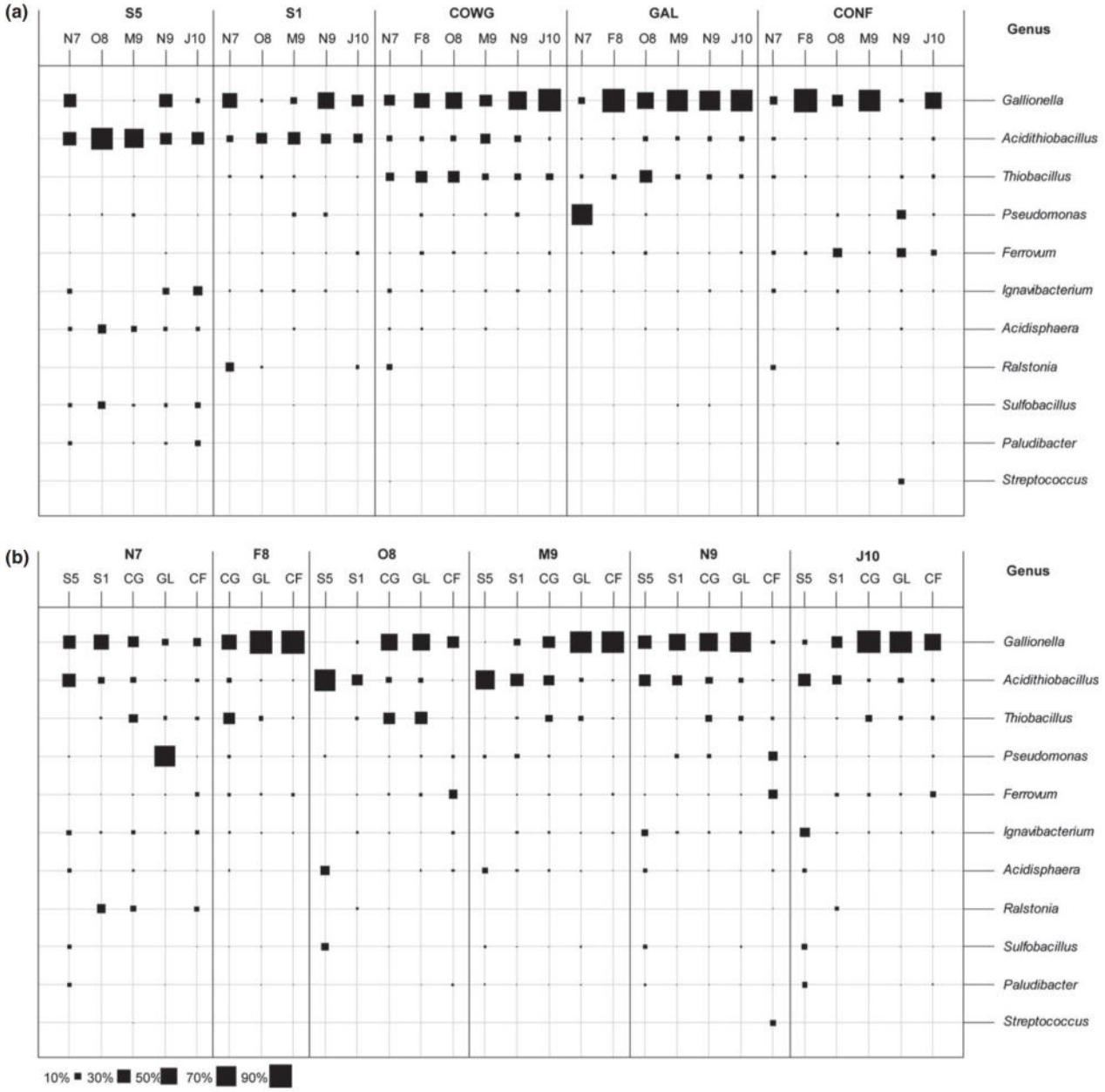

Fig. 5. Relative abundance of the dominant genera (> 5% in total abundance per sample) presented by (a) sampling sites and (b) sampling date. N7: November 2007; O8: October 2008; M9: March 2009; N9: November 2009; J10: January 2010.

axes, none of the environmental variables measured in the study could explain their distribution and thus their niche. At the least polluted site (CONF), *Gallionella*, *Ferrovum*, and *Acidiferrobacter* were the main genera detected, whereas, at the most polluted sites (S5 and S1), a higher number of genera were codominant (*Acidithiobacillus*, *Ignavibacterium*, *Ralstonia*, *Leptospirillum*, *Gallionella*, *Ferrovum*, etc.).

Discussion

This study combined a classical fingerprinting method (T-RFLP) and a high-throughput barcoded pyrosequencing of 16S rRNA genes to investigate the diversity, spatial distribution, and seasonal variation of bacterial communities in Carnoulès AMD (France), which is heavily contaminated with As.

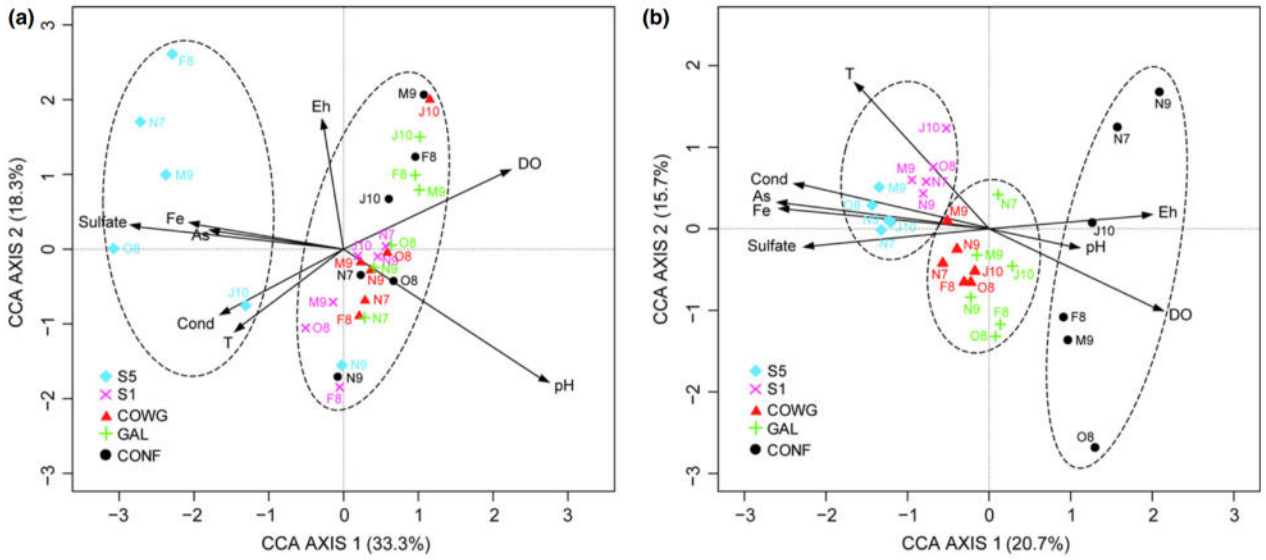

Fig. 6. Canonical correspondence analysis (CCA) correlating the bacterial community structure at each sampling site with arsenic (As), iron (Fe), conductivity (Cond), temperature (T), dissolved oxygen (DO), redox (Eh), pH and sulfate. The bacterial community structures correspond to OTU abundances from (a) T-RFLP data and (b) pyrosequencing data. The main clusters are highlighted.

Spatial and temporal variations in the environmental data set and in the bacterial community

Monitoring the physicochemical parameters of Reigous Creek confirmed previous results (Casiot *et al.*, 2003a; Egal *et al.*, 2010), showing a significant decrease in concentrations of dissolved As, Fe, and sulfate with increasing distance from the source: 72% of sulfate, 96% of iron, and 99% of arsenic had been removed by the time Reigous Creek flowed into the River Amous (Table S1). In addition, the concentrations of As and Fe in the water from the tailings stock were much lower in 2007–2010 (average values of 440 ± 184 mg L^{-1} and 4474 ± 2855 mg L^{-1}, respectively), than those measured in 2001 (up to 10 g L^{-1} for As and around 20 g L^{-1} for Fe, Casiot *et al.*, 2003b), although these concentrations are still very high compared to other AMDs.

Both molecular methods highlighted a higher bacterial diversity than expected in this extreme habitat. T-RFLP profiles showed for the five sites a total of 43 T-RFs ranging from 2 to 17 T-RFs per sampling site (Fig. 3). For pyrosequencing data, a total of 63 442 reads led to the identification of 6613 OTUs, including 4510 singletons representing 68% of the total number of OTUs. As expected, a larger number of phylotypes were identified using the pyrosequencing method leading to a significant increase in resolution. Average Good's coverage was over 89%, suggesting that the 16S rRNA gene sequences into each sample represented the majority of the bacterial phylotypes present. Nonetheless, additional sequencing effort would be required to exhaustively characterize the bacterial community, particularly for samples from the least polluted site CONF, as shown by the lower coverage values and the lack of asymptote in the rarefaction curves (data not shown).

nMDS analyses revealed significant differences in the composition of the bacterial communities in the five sites along the AMD (Fig. S3). However, different clustering patterns were obtained based on T-RFLP or pyrosequencing data. With pyrosequencing, individual sequences can be classified at the genus level. In contrast, one T-RF can correspond to several different bacterial phylotypes (belonging to different genera or even different higher taxonomic levels). Such differences in the resolution of the two methods may explain the differences obtained in the cluster analyses (Hwang *et al.*, 2012). Nevertheless, changes in bacterial community structure between the tailings groundwater (S5) and the Reigous Creek were revealed by the two sets of data, reflecting important differences in ecological conditions between the two habitats. According to both methods, S1 was the most diverse bacterial community, while GAL was the least diverse. Therefore, bacterial diversity varied independently of the sampling site, suggesting that globally upstream communities do not influence downstream communities. The apparent minimum effect of immigration suggests that species-sorting processes best describe bacterial community structure in these connected environments, with local environmental factors driving the composition of the

bacterial community at each site. Temporal variations of the bacterial communities could also be observed at each site although no particular trend could be identified. However, the important temporal variation of the bacterial community observed at S5 and CONF may be due to a higher seasonal fluctuation in physicochemical parameters at these two sites, especially temperature at CONF, and pH, Eh, sulfate, Fe, and As at S5.

However, our fine-scale investigation at the genus level of the bacterial communities along the Reigous Creek over time provided some important data and allowed to establish some hypothesis about community composition. Indeed, *Gallionella* in contrast to *Acidithiobacillus* do not seem to benefit from the seed bank provided by the most upstream sites (S5 and S1). This suggests that *Gallionella*, under a process that still need to be elucidated, extinct/re-thrived at each site over time. In contrast, *Acidithiobacillus* that is preferentially encountered upstream of the Reigous Creek or *Thiobacillus* that thrived at COWG could be found at these sites, under conditions that reflect their preferential habitats. The presence of these organisms downstream of the sites would instead reflect dispersal from upstream sites.

Physicochemical parameters shape the composition of the bacterial community

This work highlighted a spatial gradient of physicochemical conditions linked to a significant shift in bacterial community composition along the continuum. Indeed, canonical correspondence analysis of the whole pyrosequencing data set indicated that arsenic, temperature, and sulfate were the factors that most influence the composition of the bacterial communities (Fig. 6b). The level of pollution affects also some dominant bacterial populations (> 5% of relative abundance). *Gallionella*, *Ferrovum,* and *Acidiferrobacter* were the dominant genera in water sampled at the least polluted site (CONF) and were correlated with high DO and Eh, whereas in water from the most polluted sites (S5 and S1), a larger number of dominant genera were detected (*Acidithiobacillus*, *Ignavibacterium*, *Ralstonia*, *Leptospirillum*, *Gallionella,* and *Ferrovum*) whose relative abundance was correlated with higher temperature and high concentrations of As, Fe, and sulfate (Fig. S4). Thus, different members of a given genus such as *Gallionella* (OTUs 19, 15 and 1) or *Ferrovum* (OTUs 25 and 28) were correlated with different environmental parameters, suggesting that these OTUs correspond to bacterial phylotypes with some specificity explaining these different behaviors. Furthermore, the high abundance of *Gallionella*-related sequences in these acidic ecosystems characterized by contrasted levels of pollution is consistent with results of a previous study suggesting that *Gallionella*-like organisms may be more tolerant to acid and metal than currently thought (Fabisch *et al.*, 2013). In accordance with our results, temperature has been previously suggested as a primary factor controlling the structure and dynamics of microbial communities in AMD (Edwards *et al.*, 1999) and in various natural environments like hot springs (Ward *et al.*, 1998; Miller *et al.*, 2009) or marine environments (Fuhrman *et al.*, 2008). Nevertheless, sulfate and arsenic concentrations have not previously been shown to be significantly correlated with bacterial diversity in AMDs. Earlier studies identified different environmental predictors of microbial populations in AMD including conductivity and rainfall (Edwards *et al.*, 1999), pH (Lear *et al.*, 2009), oxygen gradient (González-Toril *et al.*, 2011), and season (Streten-Joyce *et al.*, 2013), which may result in site-specific physicochemical and geochemical characteristics (Kuang *et al.*, 2012). Furthermore, while many studies highlighted pH as the most important factor structuring AMD communities (Kuang *et al.*, 2012; Chen *et al.*, 2013), our study produced no evidence of the influence of this parameter, probably due to the limited variation in pH among our samples (average values of 2.5 $\pm$ 0.8–3.2 $\pm$ 0.3).

Composition of the bacterial communities

In this study, we were able to identify a wider phylogenetic range of taxa than in any previous clone library-based diversity survey of the Carnoulès AMD, including sequences of several previously undetected taxa. These new taxa include members of the *Bacteroidetes*, *Chlorobi*, *Chloroflexi*, *Elusimicrobia*, *Chlamydiae*, *Cyanobacteria*, *Deinococcus-Thermus*, *Spirochaetes*, *Fibrobacteres*, *Fusobacteria*, *Gemmatimonadetes*, *Plantctomycetes*, *Verrumicrobia,* and of the uncultured *OD1-PO11-TM7* clade. The majority of phyla that were not previously detected on clone libraries accounted only for < 1% of the pyrosequencing data, explaining why they were missed with the clone library approach. The high rate of low-abundance populations (68% of singletons) increased the phylogenetic bacterial diversity. However, despite the preponderance of this rare biosphere in most studies, its ecological and functional roles remain largely unexplored (Galand *et al.*, 2009). Recent studies indicated that such organisms may be at a dormant or a spore stage, but in favorable conditions they may become active and even dominant (Delavat *et al.*, 2012). Thus, these taxa may play an important role in extreme habitats like AMD, buffering the effects of important environmental shifts (Sogin *et al.*, 2006; Monchy *et al.*, 2011). However, further investigations will be needed to determine whether they play a role in this ecosystem and/or whether they reflect allochthonous input from surrounding environments. Moreover, the high-

throughput sequencing also questions the accuracy of OTU richness estimates, as sequencing errors and inadequate clustering algorithms can lead to overestimates of community richness (Huse *et al.*, 2010). The majority of the most abundant taxa detected in this study were related to orders commonly encountered in AMD, most of which are known to be involved in Fe, As, and S cycles: namely *Gallionellales* (*Betaproteobacteria*), *Acidithiobacilliales* (*Gammaproteobacteria*), *Acidimicrobiales* (*Actinobacteria*), *Hydrogenophilales* (*Betaproteobacteria*), *Burkholderiales* (*Betaproteobacteria*), *Nitrospiralles* (*Nitrospirae*), *Desulforomonadales* (*Deltaproteobacteria*), and *Desulfobacterales* (*Deltaproteobacteria*). The ecological role of previously detected taxa has been widely characterized (Bruneel *et al.*, 2005, 2006, 2011; Bertin *et al.*, 2011) in this ecosystem. A relatively small number of OTUs dominated at each sampling site (Fig. S2) and the majority of them were phylogenetically related to taxa previously found in AMD (*Gallionella ferruginea*, *Acidithiobacillus ferrooxidans*, and *Thiobacillus* sp.), as well in Carnoulès revealing their persistence in such ecosystems (Baker & Banfield, 2003; Bruneel *et al.*, 2006, 2011; Hallberg *et al.*, 2006; Heinzel *et al.*, 2009; Hallberg, 2010). These three genera varied in their relative abundance over the sampling period. *Gallionella* was present in high proportions in almost all samples, mainly at GAL and COWG. In contrast, except at S5 where this genus was dominant, *Acidithiobacillus* accounted for a minor fraction of the bacterial community (Fig. 5). Furthermore, our study confirmed the presence of relatives of a novel bacterial phylum, '*Candidatus* Fodinabacter communificans' detected by a recent metagenomic investigation of Carnoulès AMD and prominent in the active COWG community (Bertin *et al.*, 2011; Fahy *et al.*, unpublished data). The relatively high number of unclassified bacteria per sample (0.3–37%) supports the fact that many bacteria remain to be cultured. These results thus corroborate the main observations made in previous studies, except for the predominance of organisms related to sulfate reducing bacteria (SRB) identified in water from the tailings by Bruneel *et al.* (2005) using a cloning–sequencing approach. Instead, our results revealed the dominance of *Acidithiobacillales* over SRB in these samples. The very low proportion of SRB populations in our study (on average 0.1% of total abundance per sample) could be partly due to differences in the physicochemical variables of the water, to the choice of a stringent similarity cutoff but also to the different primers used for PCR amplification. Furthermore, relatives of *Thiomonas* belonging to the *Burkholderiales* order were retrieved and accounted for < 1% of the total sequences (Fig. S1a). Despite their low abundance, several strains of *Thiomonas* sp. have been previously isolated and shown to be active in the oxidation of As (Bruneel *et al.* 2003). A metaproteomic approach also showed that *Gallionella*, *Thiomonas,* and *A. ferrooxidans* actively express proteins *in situ,* thus probably playing a functional role in this AMD (Bruneel *et al.*, 2011). These populations could play an important role in the efficient remediation process observed along this creek by favoring the oxidation of Fe(II) and the co-precipitation of As (Casiot *et al.*, 2003a; Bruneel *et al.*, 2011).

This work has increased our knowledge of bacterial diversity and dynamics in acid mine drainage. Bacterial diversity in Carnoulès AMD was revealed to be much higher than previously evidenced using clone library techniques (Bruneel *et al.*, 2011), as suggested by culture-dependent methods (Delavat *et al.*, 2012). Our study revealed complex patterns of spatial and temporal variations in bacterial community composition, suggesting that community composition reflects changes in physicochemical conditions. This investigation provided a first step to the study of spatial and temporal structure of bacterial communities and the factors that control it. To improve our understanding of the functioning of this ecosystem, future efforts should be oriented toward active communities and how they fluctuate in response to environmental changes. Such knowledge will help to determine their roles in the functioning of the AMD ecosystems and explain important assembly processes in microbial ecology.

Acknowledgements

This study was financed by the FRB (*Fondation pour la recherche sur la Biodiversité*) *program blanc* AAP-IN-2009-039, the « *Observatoire de Recherche Méditerranéen de l' Environnement* » (OSU-OREME). A.V. was supported by a grant from the French Ministry of Education and Research and F.J. by a grant from the Direction Générale de l'Armement (DGA). This work was performed within the framework of the *Groupement de recherche: Métabolisme de l'Arsenic chez les Microorganismes* (GDR2909-CNRS).

References

Baker BJ & Banfield JF (2003) Microbial communities in acid mine drainage. *FEMS Microbiol Ecol* **44**: 139–152.

Bertin PN, Heinrich-Salmeron A, Pelletier E *et al.* (2011) Metabolic diversity among main microorganisms inside an arsenic-rich ecosystem revealed by meta- and proteo-genomics. *ISME J* **5**: 1735–1747.

Bruneel O, Personné JC, Casiot C, Leblanc M, Elbaz-Poulichet F, Mahler BJ, Le Flèche A & Grimont P AD (2003) Mediation of arsenic oxidation by *Thiomonas* sp. in acid-mine drainage (Carnoulès, France). *J Appl Microbiol* **95**: 492–499.

Bruneel O, Duran R, Koffi K, Casiot C, Fourçans A, Elbaz-Poulichet F & Personné JC (2005) Microbial diversity in a pyrite-rich tailings impoundment (Carnoulès, France). *Geomicrobiol J* **22**: 249–257.

Bruneel O, Duran R, Casiot C, Elbaz-Poulichet F & Personné JC (2006) Diversity of microorganisms in Fe-As-rich acid mine drainage waters of Carnoulès, France. *Appl Environ Microbiol* **72**: 551–556.

Bruneel O, Volant A, Gallien S *et al.* (2011) Characterization of the active bacterial community involved in natural attenuation processes in arsenic-rich Creek sediments. *Microb Ecol* **61**: 793–810.

Casiot C, Morin G, Juillot F, Bruneel O, Personné JC, Leblanc M, Duquesne K, Bonnefoy V & Elbaz-Poulichet F (2003a) Bacterial immobilization and oxidation of arsenic in acid mine drainage (Carnoulès creek, France). *Water Res* **37**: 2929–2936.

Casiot C, Leblanc M, Bruneel O, Personné JC, Koffi K & Elbaz-Poulichet F (2003b) Geochemical processes controlling the formation of As-rich waters within a tailings impoundment (Carnoulès, France). *Aquat Geochem* **9**: 273–290.

Chen LX, Li JT, Chen YT, Huang LN, Hua ZS, Hu M & Shu WS (2013) Shifts in microbial community composition and function in the acidification of a lead/zinc mine tailings. *Environ Microbiol* **15**: 2431–2444.

Delavat F, Lett MC & Lievremont D (2012) Novel and unexpected bacterial diversity in an arsenic-rich ecosystem revealed by culture-dependent approaches. *Biol Direct* **7**: 28.

Edwards KJ, Gihring TM & Banfield JF (1999) Seasonal variations in microbial populations and environmental conditions in an extreme acid mine drainage environment. *Appl Environ Microbiol* **65**: 3627–3632.

Edwards KJ, Bond PL, Druschel GK, McGuire MM, Hamers RJ & Banfield JF (2000) Geochemical and biological aspects of sulfide mineral dissolution: lessons from Iron Mountain, California. *Chem Geol* **169**: 383–397.

Egal M, Casiot C, Morin G, Elbaz-Poulichet F, Cordier MA & Bruneel O (2010) An updated insight into the natural attenuation of as concentrations in Reigous Creek (southern France). *Appl Geochem* **25**: 1949–1957.

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

Fuhrman JA, Steele JA, Hewson I, Schwalbach MS, Brown MV, Green JL & Brown JH (2008) A latitudinal diversity gradient in planktonic marine bacteria. *P Natl Acad Sci USA* **105**: 7774–7778.

Galand PE, Casamayor EO, Kirchman DL & Lovejoy C (2009) Ecology of the rare microbial biosphere of the Arctic Ocean. *P Natl Acad Sci USA* **106**: 22427–22432.

González-Toril E, Aguilera A, Souza-Egipsy V, López Pamo E, Sánchez España J & Amils R (2011) Geomicrobiology of La Zarza-Perrunal acid mine effluent (Iberian Pyritic Belt, Spain). *Appl Environ Microbiol* **77**: 2685–2694.

Hallberg KB (2010) New perspectives in acid mine drainage microbiology. *Hydrometallurgy* **104**: 448–453.

Hallberg KB, Coupland K, Kimura S & Johnson DB (2006) Macroscopic streamer growths in acidic, metal-rich mine waters in North Wales consist of novel and remarkably simple bacterial communities. *Appl Environ Microbiol* **72**: 2022–2030.

Heinzel E, Hedrich S, Janneck E, Glombitza F, Seifert J & Schlömann M (2009) Bacterial diversity in a mine water treatment plant. *Appl Environ Microbiol* **75**: 858–861.

Huse SM, Welch DM, Morrison HG & Sogin ML (2010) Ironing out the wrinkles in the rare biosphere through improved OTU clustering. *Environ Microbiol* **12**: 1889–1898.

Hwang C, Ling F, Andersen GL, LeChevallier MW & Liu WT (2012) Microbial community dynamics of an urban drinking water distribution system subjected to phases of chloramination and chlorination treatments. *Appl Environ Microbiol* **78**: 7856–7865.

Johnson DB (2012) Geomicrobiology of extremely acidic subsurface environments. *FEMS Microbiol Ecol* **81**: 2–12.

Junier P, Junier T & Witzel KP (2008) TRiFLe, a program for *in silico* terminal restriction fragment length polymorphism analysis with user-defined sequence sets. *Appl Environ Microbiol* **74**: 6452–6456.

Koffi K, Leblanc M, Jourde H, Casiot C, Pistre S, Gouze P & Elbaz-Poulichet F (2003) Reverse oxidation zoning in mine tailings generating arsenic-rich acidic waters (Carnoulès, France). *Mine Water Environ* **22**: 7–14.

Kuang JL, Huang LN, Chen LX, Hua ZS, Li SJ, Hu M, Li JT & Shu WS (2012) Contemporary environmental variation determines microbial diversity patterns in acid mine drainage. *ISME J* **7**: 1038–1050.

Lane DJ (1991) 16S/23S rRNA sequencing. *Nucleic Acid Techniques in Bacterial Systematic* (Stackebrandt E & Goodfellow M, eds), pp. 115–175. Wiley, New York.

Lear G, Niyogi D, Harding J, Dong Y & Lewis G (2009) Biofilm bacterial community structure in streams affected by acid mine drainage. *Appl Environ Microbiol* **75**: 3455–3460.

Miller SR, Strong AL, Jones KL & Ungerer MC (2009) Barcoded pyrosequencing reveals shared bacterial community properties along the temperature gradients of two alkaline hot springs in Yellowstone National Park. *Appl Environ Microbiol* **75**: 4565–4572.

Monchy S, Sanciu G, Jobard M *et al.* (2011) Exploring and quantifying fungal diversity in freshwater lake ecosystems using rDNA cloning/sequencing and SSU tag pyrosequencing. *Environ Microbiol* **13**: 1433–1453.

Morin G & Calas G (2006) Arsenic in soils, mine tailings, and former industrial sites. *Elements* **2**: 97–101.

Morin G, Juillot F, Casiot C, Bruneel O, Personné JC, Elbaz-Poulichet F, Leblanc M, Ildefonse P & Calas G (2003) Bacterial formation of tooeleite and mixed arsenic(III) or arsenic(V)-iron(III) gels in the Carnoulès acid mine drainage, France. A XANES, XRD, and SEM study. *Environ Sci Technol* **37**: 1705–1712.

Muyzer G & Ramsing NB (1995) Molecular methods to study the organization of microbial communities. *Water Sci Technol* **32**: 1–9.

Needleman SB & Wunsch CD (1970) A general method applicable to the search for similarities in the amino acid sequence of two proteins. *J Mol Biol* **48**: 443–453.

Pilloni G, Granitsiotis MS, Engel M & Lueders T (2012) Testing the limits of 454 pyrotag sequencing: reproducibility, quantitative assessment and comparison to T-Rflp fingerprinting of aquifer microbes. *PLoS ONE* **7**: e40467.

R Development Core Team (2012) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org.

Rodier J, Broutin JP, Chambon P, Champsaur H & Rodi L (1996) *L'Analyse des Eaux*. Dunod, Paris, p. 1383.

Samanta G & Clifford DA (2005) Preservation of inorganic arsenic species in groundwater. *Environ Sci Technol* **39**: 8877–8882.

Schloss PD, Westcott SL, Ryabin T *et al.* (2009) Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl Environ Microbiol* **75**: 7537–7541.

Smith CJ, Danilowicz BS, Clear AK, Costello FJ, Wilson B & Meijer WG (2005) T-Align, a web-based tool for comparison of multiple terminal restriction fragment length polymorphism profiles. *FEMS Microbiol Ecol* **54**: 375–380.

Sogin ML, Morrison HG, Huber JA, Mark Welch D, Huse SM, Neal PR, Arrieta JM & Herndl GJ (2006) Microbial diversity in the deep sea and the underexplored "rare biosphere". *P Natl Acad Sci USA* **103**: 12115–12120.

Streten-Joyce C, Manning J, Gibb KS, Neilan BA & Parry DL (2013) The chemical composition and bacteria communities in acid and metalliferous drainage from the wet-dry tropics are dependent on season. *Sci Total Environ* **443**: 65–79.

Ueno S, Le Provost G, Leger V *et al.* (2010) Bioinformatic analysis of ESTs collected by Sanger and pyrosequencing methods for a keystone forest tree species: oak. *BMC Genomics* **11**: 650.

Ward DM, Ferris MJ, Nold SC & Bateson MM (1998) A natural view of microbial biodiversity within hot spring cyanobacterial mat communities. *Microbiol Mol Biol Rev* **62**: 1353–1370.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Fig. S1. Composition of different phyla based on classification of 16S rRNA gene sequences of bacteria from each sample using RDP Classifier: (a) *Proteobacteria* orders, (b) *Actinobacteria* orders, (c) *Firmicutes* orders, and (d) *Acidobacteria* orders.

Fig. S2. Histogram of the relative abundance of dominant OTUs at the Carnoulès sampling sites (*G. ferruginea subsp. capsiferriformans* ES-2: NC_014394; *A. ferrooxidans* strain HL1: JF815535; *Thiobacillus* sp. ML2-16: DQ145970; *Pseudomonas migulae*: AY605698; *A. ferrivorans* SS3: NR_074660; *Actinobacterium* BGR 105: GU168008; *Acidobacteriaceae bacterium* CH1: DQ355184; *Ferrimicrobium* sp. Py-F2: KC208496; *Metallibacterium* sp. ×11: HE858262; *Alicyclobacillaceae bacterium* Feo-D4-16-CH: FN870323; *Acidisphaera* sp. nju-AMDS1: FJ915153; *Betaproteobacterium* OYT1: AB720115.

Fig. S3. Nonmetric multidimensional scaling analysis of the composition of the bacterial community estimated by (a) T-RFLP and (b) 454 pyrosequencing based on 16S rRNA genes.

Fig. S4. (a) Ordination plot of CCA based on pyrosequencing data showing OTUs with relative abundance >5%. (b) Abundant OTUs and their correlation with environmental variables and phylogenetic affiliation determined by BLAST search.

Table S1. Physicochemical characteristics of the water at each sampling site and sampling date.

Table S2. Relative abundance (in %) of total sequences of bacterial 16S rRNA genes from each sample assigned to different phyla.