
Zohier B. Salah¹, Christopher J. Charles¹, Paul N. Humphreys¹, Andrew. P. Laws² and Simon P. Rout¹*

1. Department of Biological and Geographical Sciences, University of Huddersfield, Queensgate Campus, Huddersfield, United Kingdom, HD1 3DH.
2. Department of Chemical Sciences, University of Huddersfield, Queensgate Campus, Huddersfield, United Kingdom, HD1 3DH.

*Corresponding author*: Simon Rout, Department of Biological Sciences, University of Huddersfield, Huddersfield, UK, E-mail: S.Rout@hud.ac.uk

**Data deposition**: The genome sequence of *Azonexus hydrophilus* strain ZS02 is deposited within BioProject PRJNA345295 under GenBank ID MTHD0000000

**Keywords**: Azonexus, facultative aerobic, diazotroph, alkalitolerant, whole genome sequence, Nitrogen fixation, Nitrogen cycle, nitrogen fixing bacteria.

**Article = Short Research Communication**
Abstract

Alkaline environments represent a significant challenge to the growth of micro-organisms. Despite this, there are a number of alkaline environments which contain active microbial communities. Here we describe the genome of a diazotrophic, alkalitolerant strain of *Azonexus*, which was isolated from a microcosm seeded with hyperalkaline soils resulting from lime depositions. The isolate has a genome size 3.60 Mb with 3431 protein coding genes. The proteome indicated the presence of genes associated with the cycling of nitrogen, in particular the fixation of atmospheric nitrogen. Although closely related to *Azonexus hydrophilus* strain d8-1 by both 16S (97.9%) and *in silico* gDNA (84.1%) relatedness, the isolate demonstrates a pH tolerance above that reported for this strain. The proteome contained genes for the complete Na⁺/H⁺ antiporter (subunits A to G) for cytoplasmic pH regulation; this may account for the phenotypic characteristics of this strain which exhibited optimal growth conditions of pH 9 and 30°C.

Introduction

From a microbiological perspective alkaline environments fall into three broad categories based on their chemistry and origins. By far the most studied alkaline environments are the saline, sodium dominated Soda Lakes [1-3]. There are also calcium and magnesium dominated, alkaline environments generated by serpentinization processes [4] which are associated with groundwater [5, 6] and its subsequent environmental interactions [7-9]. Finally, a range of anthropogenic alkaline systems exist [10]; generated by the waste disposal practices of lime [11], steel [12], soda lime [13] and bauxite [14] manufacture.

Despite the harsh geochemical environment these sites often support extensive and diverse bacterial and archaeal populations [2, 15] capable of a wide range of metabolic and energy generating processes [9, 16]. Such processes have gained interest since they may be exploited
for industrial applications [17]. In the case of nitrogen, the fate of ammonia, nitrate and nitrite has received considerable attention [2, 18]. Nitrogen fixation in alkaline saline systems appears to be associated with free living diazotrophs from the Firmicute division [19-21], although anoxygenic photoautotrophs have also been implicated [22]. In serpentinization associated systems, some authors have reported a lack of nitrogen fixation [8] whilst it has been inferred from community data in others [23].

Our understanding of the nitrogen cycle within anthropogenic lime based environments is currently limited to that of denitrification processes through the measurement of nitrate/nitrite within the soil core [24]. It has also been observed through the amendment of microcosms, reaction vessels with constrained ecological parameters [25], with nitrate as a terminal electron acceptor [26, 27]. Nitrate enrichment of these soils resulted in the detection of *Azoarcus sp.* [26], whilst *Azonexus sp* and *Azovibrio sp* were detected in nitrate free systems [28]. This suggests that these free living diazotrophs play an important role in the nitrogen flow through these systems, potentially replacing the nitrogen fixing *Clostridium sp* [29]. When the studies of both Rizoulis, Milodowski [7] and Kyeremeh, Charles [28] are considered, *Clostridium sp* were below detection limits in the community analyses of these alkaline soil microcosms.

In this paper we provide the whole genome sequence, discussion of the associated proteome and a description of the metabolic capabilities of isolate ZS02, an alkaliphilic nitrogen fixing bacteria. ZS02 was isolated from an anaerobic microcosm.

**Materials and Methods**

**Microcosm operation and characterisation**

Hyperalkaline soil was obtained from Harpur Hill, Derbyshire and diluted in a 10 % mix of alkaline hydrolysed cellulose (prepared as per Rout, Radford [30]) and anoxic mineral media
The microcosm was kept under a nitrogen gas headspace and operated under a 10% waste feed ratio every 30 days at pH 9.5 for 12 months. Genomic DNA was then extracted and 16S rDNA library prepared as per Kyeremeh et al [28].

**Isolation of pure cultures and phenotypic characterisation.**

10 μL of the microcosm fluid was used to inoculate pH 9.5 fastidious anaerobe agar (LabM, UK) with growth observed following 5 days incubation at 25°C in anaerobic conditions ((10% H₂:10% CO₂: 80% N₂; DW Scientific). Single colonies were further purified by sub-culture and cell morphology was determined via Gram staining and SEM (Quanta FEG 250, FEI). A pH and temperature profile was obtained using Bioscreen C technology (Oy Growth Curves). Briefly, fastidious anaerobe broth was prepared and pH values adjusted to 4-12 using 4M NaOH, each pH broth (180μl) was then transferred to duplicate wells and inoculated with 20μl of the isolate prepared to an OD of ~0.2. Manipulations were carried out in an anaerobic chamber and the plate sealed with clear, gas proof tape. Replica plates were prepared and incubated at a range of temperatures (10, 20, 25, 30, 35 and 40 °C), with OD₆₀₀ measured every hour for 24 hours. The pH range and optimal pH for growth were then determined by calculating the fractional area of the growth curves compared with uninoculated control data.

**Genome Sequencing, Assembly and Annotation**

gDNA from the isolate was extracted using an UltraClean Microbial Isolation Kit. The concentration and purity of the extracted DNA was assessed using a NanoDrop ND1000 spectrophotometer. Sanger sequencing (Eurofins, Germany) of a partial fragment of the 16S rRNA gene coupled to BLASTn search allowed for initial identification of the isolate. For whole genome sequence generation, paired end sequence reads were prepared using an Illumina HiSeq 2500 system (Illumina, US) by Microbes NG (Birmingham, UK). A standard analysis pipeline assigned reads to the closest available reference genome using Kraken [32].
and BWA mem [33]. Further de novo assembly of the reads using SPAdes [34] was also performed and the reads mapped back to the resultant contigs and quality checked using BWA mem. Gaps within the contigs were identified and filled using GapFiller v1.10 [35].

The average nucleotide and average amino acid identities of the genome to near neighbours were calculated using the methods of Rodriguez and Konstantinidis [36]. Annotation of the genome was carried out through RAST [37] and Pfam domain predictions carried out using the NCBI conserved domain database [38]. For the prediction of signal peptides and transmembrane domains, SignalP 4.1 server [39] and the TMHMM server v. 2.0 [40] were used respectively. CRISPRs were found online using the CRISPRfinder tool [41] and metabolic pathways were reconstructed from the annotated proteins using KAAS [42]. Proteins encoding carbohydrate active enzymes identified using the CAZy analysis toolkit [43].

**Results and discussions**

Community analysis on the microcosm suggested that it was dominated (94.5%) by sequences showing homology to *Azonexus hydrophilus* strain d8-1 [44], indicating that the microcosm was a source of nitrogen fixing bacteria by sequencing of the V4 region of the 16S rRNA gene via MiSeq platform. A pure culture of a facultative anaerobe was obtained, designated ZS02, with morphology of Gram negative rods ~1µm in length (Figure S1). Sanger sequencing generated a 934bp fragment of the 16S rRNA gene which showed 97.9% sequence homology to that of *Azonexus hydrophilus* strain d8-1 (DSM23864) and clustered with the other 16S sequences of this genus (Figure S2). The final assembly of the genome of ZS02 consisted of 3,596,517 bp within 17 scaffolds, representing a coverage of 116x and a GC content of 61.80% containing no gaps or ambiguous nucleotide positions (Figure 1). In silico comparisons of available whole genome sequences of nitrogen fixing bacteria indicated that ZS-02 still showed greatest sequence identity to type strain d8-1 (84.1%, Table 1).
The genome contained 3,494 genes, of which 3,431 were protein coding, 63 were RNA coding, 43 were pseudogenes and 1 CRISPR repeat. 75.3% of the protein-coding genes were assigned a putative function, 2,443 were found to have Pfam domains, 819 transmembrane helicases and 309 signal peptides. When amino acid identities were compared with type strain nitrogen fixing bacteria, again ZS02 showed greatest homology to *Azonexus hydrophilus* d8-1 (84.2%, Table S1).

KAAS assignment indicated that the genome contained a number of nitrogen metabolism genes. Of these, ZS02 proteome includes the most well studied molybdenum dependent nitrogenases encoded by the catalytic nifHDK cluster (Dos Santos, Fang [45]). Within this cluster, nifH is highly conserved and this was reflected in the nucleotide sequence homology of the nifH of ZS02 with that of other nitrogen fixing bacteria, with greatest homology (96.5%) seen *Azonexus hydrophilus* (DSM23864, Figure 2). In contrast, the remaining genes in the cluster showed greater homology to other organisms, with nifD nucleotide sequence showing greatest homology to *Zoogloea* sp LCSB751 (94.2%) and *Azovibrio restrictus* DSM23866 (93.1%). Comparisons of the NifK resulted in homologies >90%, with greatest homology to *Canididatus accumulibacter* (87.9%) and *Azovibrio restrictus* (82.6%).

The proteome of ZS02 contained a complete napAB gene cluster encoding periplasmic nitrate reductase [46], with the presence of nirBD cluster suggesting that nitrite can then be further reduced to ammonia within the cytoplasm [47]. Although genes encoding NasAB were present, the absence of nirA suggests ZS02 is incapable of assimilatory nitrate reduction. The absence of nirK within the proteome also suggests that conversion of nitrite to nitric oxide is not a metabolic capability, however the presence of norBC and nosZ suggests that ZS02 is capable of the conversion of nitric oxide to molecular nitrogen [47]. The proteome also suggest that ZS02 is not capable of ammonium oxidation or nitrification of ammonia.
The presence of a complete Na⁺/H⁺ antiporter (subunits A to G) are a common feature of alkaliphilic microorganisms relating to cytoplasmic pH regulation [48], and were present in the proteome of ZS02. pH/temperature profiles (Figure 3) indicated that there was no growth above a pH of 10 or below 5, with the limit of temperature appearing to be 40°C. Optimum growth was observed at a temperature of 30°C and a pH of 9. Although moderate growth was observed at this temperature between a pH of 6 and 10. CAZy analysis indicated a number of carbohydrate active enzymes within the proteome of ZS02, and a greater number than the type strain (Figure S3). These increased numbers of enzymes may provide an advantage in harsh alkaline environments. Nutrient cycling in natural and anthropogenic alkaline environments requires micro-organisms that are capable of tolerating the harsh pH conditions. Since nitrate is limited in these environments, micro-organisms capable of utilising atmospheric nitrogen may play a pivotal role in delivering nitrogen. Here we present the whole genome sequence of an alcalitolerant Azonexus hydrophilus strain, containing nitrogen cycling abilities within its proteome.

References
<table>
<thead>
<tr>
<th>Organism</th>
<th>Average Nucleotide Identity (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
</tr>
<tr>
<td><strong>Azonexus hydrophilus ZS-02</strong> (1)</td>
<td>100</td>
</tr>
<tr>
<td><strong>Azonexus hydrophilus DSM23864</strong> (2)</td>
<td>100</td>
</tr>
<tr>
<td><strong>Dechloromonas denitrificans ATCC BAA-841</strong> (3)</td>
<td>100</td>
</tr>
<tr>
<td><strong>Azovibrio restrictus DSM23866</strong> (4)</td>
<td>100</td>
</tr>
<tr>
<td><strong>Azoarcus sp BH72</strong> (5)</td>
<td></td>
</tr>
<tr>
<td><strong>Zoogloea sp LCSB751</strong> (6)</td>
<td></td>
</tr>
<tr>
<td><strong>Aquaspirillum sp LM1</strong> (7)</td>
<td></td>
</tr>
<tr>
<td><strong>Candidatus accumulibacter sp SK-02</strong> (8)</td>
<td>100</td>
</tr>
</tbody>
</table>

Table 1: Average nucleotide identity comparisons between nitrogen fixing type strains.
Figure 1: Graphical representation of the features of the genome of isolate ZS-02, Circles (from outside to inside) 1 and 2 (protein) coding sequences on the forward and reverse strand, blue indicates CDS, peach shows tRNAs, pink rRNAs, and grey, other). Circle 3 shows GC content% and Circles 4 and 5 show positive (green) and negative (magenta) GC skew.
Figure 2: Amino acid similarities of *Azonexus hydrophilus* ZS02 NifHDK cluster. Estimate of phylogeny based on whole genome sequences using neighbour joining method is also shown (branch lengths not to scale).
Figure 3: pH/temperature profile of isolate ZS02 indicated the isolate was capable of growth between pH 5 and 10, with an optimal pH for growth of 9 at 30°C