Construction and analysis of a metagenomic library from an enhanced biological phosphorus removal biomass

C. Yeates and L.L. Blackall

Advanced Wastewater Management Centre, The University of Queensland, Brisbane, Queensland 4072, Australia (E-mail: cyeates@awmc.uq.edu.au; blackall@awmc.uq.edu.au)

Abstract The enhanced biological phosphorus removal (EBPR) process is regularly used for the treatment of wastewater, but suffers from erratic performance. Successful EBPR relies on the growth of bacteria called polyphosphate-accumulating organisms (PAOs), which store phosphorus intracellularly as polyphosphate, thus removing it from wastewater. Metabolic models have been proposed which describe the measured chemical transformations, however genetic evidence is lacking to confirm these hypotheses. The aim of this research was to generate a metagenomic library from biomass enriched in PAOs as determined by phenotypic data and fluorescence in situ hybridisation (FISH) using probes specific for the only described PAO to date, "Candidatus Accumulibacter phosphatis". DNA extraction methods were optimised and two fosmid libraries were constructed which contained 93 million base pairs of metagenomic data. Initial screening of the library for 16S rRNA genes revealed fosmids originating from a range of non-pure-cultured wastewater bacteria. The metagenomic libraries constructed will provide the ability to link phylogenetic and metabolic data for bacteria involved in nutrient removal from wastewater.

Keywords DNA extraction; EBPR; metagenomic library; 16S rRNA gene

Introduction

The removal of phosphorus (P) is a crucial step in wastewater treatment and can be achieved through the microbial-mediated process of enhanced biological phosphorus removal (EBPR). EBPR involves the cycling of activated sludge through anaerobic and aerobic stages with volatile fatty acid (VFA) addition during the anaerobic stage. Key to success of this process is the presence of bacteria called polyphosphate (polyP) accumulating organisms (or PAOs) that store polyP in excess of their metabolic requirements during the aerobic period thus removing P from the system. The unique capacity of PAOs to sequester VFAs and generate energy from polyP degradation under anaerobic conditions is thought to be responsible for their selection. Other organisms have been discovered in lab-scale systems that can also sequester VFAs under anaerobic conditions but they utilise intracellular glycogen as their sole energy source for anaerobic processes (Liu et al., 1994). These glycogen-accumulating organisms (GAOs) carry out similar carbon transformations as PAOs but do not participate in the P transformations.

Molecular techniques of cloning and sequencing of 16S rRNA genes from laboratory-scale EBPR biomass have identified a PAO called "Candidatus Accumulibacter phosphatis" (henceforth called Accumulibacter) (Hesselmann et al., 1999; Crocetti et al., 2000) and a GAO called "Candidatus Competibacter phosphatis" (henceforth called Competibacter) (Crocetti et al., 2002). In the meantime, Alphaproteobacteria GAOs have also been described (Beer et al., 2004; Wong et al., 2004). Oligonucleotide probes specific for Accumulibacter, Competibacter and the Alphaproteobacteria have been designed based on the 16S rRNA genes sequences. The use of these probes in fluorescence in situ hybridisation (FISH) has enabled visualisation, quantification and analysis of the spatial...
distribution of these bacteria in full-scale and lab-scale EBPR systems (Nielsen et al., 1999; Crocetti et al., 2000; Crocetti et al., 2002; Wong et al., 2004). Post-FISH chemical staining for polyP and PHA has allowed the linkage of Accumulibacter and Competibacter with aspects of their hypothesised anaerobic and aerobic phenotypes. FISH-microautoradiography has also been used to link specific organisms with uptake of orthophosphate and acetate (for example Kong et al. (2004)). However, there is no genetic information available to assess the metabolic potential of these bacteria. Further genetic information would increase the understanding of the biochemical processes that lead to the PAO and GAO phenotypes, and assist in biochemical modelling of the process.

To date attempts to obtain pure cultures of PAOs or GAOs have proved unsuccessful. The generation of large-insert gene libraries from environmental samples (or metagenomic libraries) has resulted in the discovery of physiological traits of as-yet uncultured bacteria from a variety of environmental niches (for review see Handelsman, 2004). The metagenomic approach was utilised in this study in an attempt to capture genomic information on the physiology of uncultured bacteria from wastewater.

Methods
Anaerobic-aerobic sequencing batch reactors (SBRs) were operated to enrich for bacteria exhibiting the PAO phenotype. The biomass was enriched in Accumulibacter as determined by FISH (Crocetti et al., 2000). DNA was extracted using the Qiagen Genomic Tip 100/G kit from 4.5 or 6 ml of freshly collected, homogenised biomass for the construction of two large-insert fosmid libraries (F00 or P00, respectively).

The fosmid libraries were constructed using the CopyControl™ fosmid library production kit (Epigen) following the manufacturers instructions. Briefly, extracted genomic DNA was end-repaired according to manufacturers instructions and size-fractionated using agarose gel electrophoresis. For library F00, DNA greater than 40 kb was size-fractionated by pulsed field gel electrophoresis (PFGE) on a CHEF system (BioRad) using 1% Seaplaque Agarose (Biowhitaker) under the following conditions: 12 °C, 6 V cm⁻¹ for 15 h and 5–15 s pulse time in 1 X TAE (40 mM Tris-acetate, 1 mM EDTA, pH 8.0) buffer. For library P00, DNA was size-fractionated using a 25 cm 0.8% Seaplaque agarose gel run at 200 V for 40 min, 60 V for 16 h and 200 V for 30 min in 0.5 X TBE (45 mM Tris-borate, 1 mM EDTA). The part of the gel containing the extracted DNA was removed from the rest of the gel that contained the size markers, and this latter section was stained with ethidium bromide and examined by UV illumination. The gel region containing the extracted genomic DNA corresponding in size with 35–50 kb was excised and the DNA isolated by treatment with GELase™ (Epicentre). The DNA was ligated into the pCC1FOS™ vector, packaged in vitro using MaxPlax™ Lambda packaging extracts and transduced into E. coli EPI300 cells. The library was stored as −80°C glycerol stocks in 96-well microtitre plate format.

To confirm the size and assess the diversity of inserts, minipreps from randomly selected clones were digested overnight with the restriction enzyme NorI (New England Biolabs). The recognition site for this enzyme is GCGGCGGC. The presence of rRNA gene-containing clones was determined using a PCR-based screening method based upon that described by Suzuki et al. (2004) and utilising the ability to induce the fosmids to 10–50 copies per cell using methods described in Levaue et al. (2004). For the PCR-based screening, each 96-well plate of fosmid clones was inoculated into another microtitre plate containing 100 μl of LB with 12.5 μg ml⁻¹ chloramphenicol. After overnight incubation at 37 °C, 40 μl from all wells in each row were combined, resulting in 8 pools for each plate. LB with 12.5 μg ml⁻¹ chloramphenicol was added to the pooled cell
suspensions to a total volume of 5 ml and the fosmids were induced to high copy number with 1 × induction solution (Epicentre) and incubation for 5 h at 37 °C with shaking. The cells were pelleted by centrifugation and the fosmids were extracted using standard mini-prep procedures followed by isopropanol precipitation (Sambrook and Russell, 2001).

Fosmid DNA from pooled induced clones was used as a PCR template to screen for the presence of 16S-23S rRNA spacer region genes. The forward primer was from the 16S rRNA gene (1406f 5’ TGYACACACCTCCCGT) and the reverse primer was from the 23S rRNA gene (242r 5’ KTTGCCTCGCCRCRTAC) (Lane, 1991). In a final volume of 50 µl, PCRs contained 1 × Platinum Taq buffer (Invitrogen), 200 µM dNTPs, 1.5 mM MgCl2, 0.2 µM of each primer and 1U of Platinum Taq DNA Polymerase with 1 µl of fosmid DNA from induced and pooled clones. Reactions were run under the following conditions: 94 °C for 2 min followed by 40 cycles of 94 °C for 30 s, 55 °C for 30 s, and 72 °C for 1 min followed by a final extension at 72 °C for 5 min.

To determine which clone had the 16S rRNA gene from rows that produced positive results in the above PCR, minipreps were performed on 5 ml overnight cultures from individual wells. When the clone that harboured the 16S rRNA gene was identified, further PCRs were carried out from that clone using the forward primer 530f (5’ GTGCCAGCMGCGCGG) (Lane, 1991) to generate additional sequence information. The 16S rRNA gene sequence was analysed by the basic local alignment search tool (BLASTN) (Altschul et al., 1997), aligned in ARB (www.arb-home.de) and data analysis was performed as described in Björnsson et al. (2002). GenBank accession numbers for the sequences reported are AY962315-AY962323, DQ010158 and DQ010159.

Results and discussion

Two large insert fosmid libraries were constructed from EBPR biomass collected from two lab-scale SBRs. The biomass used to construct the F00 library consisted of approximately 70–80% Accumulibacter and 5% Competibacter as determined by FISH. The biomass for the P00x library consisted of approximately 85–90% Accumulibacter and 5% Competibacter. Besides using different biomass, the other difference between the libraries was the electrophoresis method used to separate DNA of 40 kb in size. There was no significant difference found between the methods, with the use of the less expensive agarose gel electrophoresis compared to PFGE proving adequate. These two libraries comprised 2332 clones, each carrying inserts of ~40 kb, totalling 93 million nucleotides. Assuming an average genome size of 3 × 10^6, the libraries contain 31 genome equivalents.

The diversity of inserts was examined by restriction digestion with NorI (Figure 1). A number of the clones produced several bands upon digestion, which is possibly indicative of a high-GC content of the inserts.

Screening for phylogenetically relevant marker genes like the 16S rRNA gene enables linking of the genomic information on the fosmid to specific taxa. The presence of the background of E. coli DNA can make screening for 16S rRNA genes difficult. The PCR screening method used in this study relied on inducing the fosmids to high copy number and using the difference in size of the spacer region between the large and small subunit of the rRNA genes to detect PCR products from non- E. coli DNA. The appearance of an additional amplicon besides the two bands seen with E. coli DNA (lane E) was indicative of the presence of a 16S rRNA gene on the fosmid (Figure 2).

Once a row of clones had been identified that contained a 16S rRNA gene, the individual clones within that row were examined to establish which was the source of the rRNA gene. Sequence analysis of the 16S rRNA gene from the individual fosmid clones revealed the presence of a diverse range of bacteria that had been captured in the metagenomic libraries (Table 1). Whilst some clones show high identities to known cultured
bacteria (P003_C1, P006_B, P007_E3, P008_H3), others showed high identities to uncultured bacteria.

Despite the high enrichment of Accumulibacter within the initial biomass used for DNA extraction, only one clone (F003_A8) containing a 16S rRNA gene from this bacterium was identified using the screening approach described. Some problems were associated with the development of the DNA extraction method, possibly associated with the co-elution of polyP with the DNA. Initially, CsCl density centrifugation was attempted to generate high quality DNA, but the presence of what could have been polyP prevented the clear separation of DNA within the gradient. With the extraction method developed, it was found that the yield of DNA was improved when the lysate was centrifuged before addition to the Qiagen Genomic Tip column, possibly indicative of the presence of excessive proteinaceous substance in the eluate. It is possible that DNA from Accumulibacter may have been associated with any co-extracted polyP, and removal of this material to allow improved extraction may result in a selective reduction in the amount of Accumulibacter DNA in the final extract.

Only 11 clones within the libraries were identified to contain a 16S rRNA gene, representing 0.5% of the library. This is comparable with other fosmid libraries from

![Figure 1](image1.png)  
**Figure 1** Agarose gel electrophoresis of randomly selected fosmid clones digested with NotI. Lane 1, 1kb + ladder (Gibco BRL); lane 2, λ DNA digested with HindIII; lane 3, 36 kb T7 size marker (Epicentre); lanes 4–10, digested fosmids. The arrow indicates the fosmid vector.

![Figure 2](image2.png)  
**Figure 2** Agarose gel electrophoresis of PCR assay to screen for presence of rRNA genes. The lanes represent amplified products from multiple pooled and induced minipreps from rows of microtitre plates. E, E. coli.
environmental samples which contained 0.2% (Rondon et al., 2000) and 0.9% (Suzuki et al., 2004) rRNA gene-containing clones using DNA extracted from soil and seawater, respectively.

The PCR screening method used in this study relies on the difference in size of the 16S-23S rRNA spacer region from E. coli with that from other bacteria. PCR amplicons from bacteria that possess a spacer region of similar size to E. coli will not be distinguished adequately using agarose gel electrophoresis, and therefore would not be detected. Also, bacteria without linked 16S and 23S ribosomal rRNA genes would not be detected. However, this does not explain the absence of Accumulibacter rRNA genes detected so far, since the size of the Accumulibacter spacer region is approximately 60 nucleotides larger than that of E. coli and the rRNA genes in Accumulibacter are linked (data not shown).

The 16S rRNA genes sequences from the 11 positive clones were analysed using BLASTN (Table 1). The 16S rRNA genes described from the metagenomic libraries are predominantly from bacteria within the Proteobacteria. F001_F8 and P002_C12 showed homology to a Flavobacterium sp. and to a Cytophaga sp., respectively, which are both in the Bacteroidetes phylum.

The 16S rRNA gene from clone F007_D12 and F008_D10 were 95% identical to an uncultured bacterial clone PHOS-HE54, which belongs to the group of Gamma-proteobacteria sequences that encompasses Competibacter. Since bacteria with less than 97% 16S rRNA gene identity do not belong to the same species (Stackebrandt and Goebel, 1994), the source organism of F007_D12 and F008_D10 does not belong to Competibacter phosphatis. However, it could be from another species of Competibacter. Two additional Gammaproteobacterial sequences were found in the libraries. P008_H3 and P003_C1 were 99% identical to the activated sludge bulking bacterium Thiothrix eikelboomii.
Although Betaproteobacteria are common activated sludge organisms, only one 16S rRNA gene from a Betaproteobacteria was discovered (F003_A8), which showed high similarity to clones belonging to the monophyletic group of Accumulibacter (Table 1 and Figure 3).

The 16S rRNA genes from both F002_D12 and F007_B11 were both 98% identical to an uncultured Alphaproteobacterial sludge clone A15b from a nitrifying-denitrifying activated sludge system (Juretschko et al., 2002). As well as being of interest due to their high identity to this uncultured sludge bacterium, some GAOs have recently been identified as members of the Alphaproteobacteria (Wong et al., 2004). One other clone (P007_E3) was also in the Alphaproteobacteria but not close to Alphaproteobacteria GAO sequences and neither to other sludge sequences.

The metagenomic libraries have allowed the identification of fosmid clones that originate from bacteria of possible significance to the process of nutrient removal (F002_D12 and F007_B11) as well as those related to bacteria known to be beneficial (F003_A8) or detrimental to EBPR (F007_D12 and F008_D10). Other clones showed no close affiliation at the species level to any described bacteria (FO01_F8, P002_C12, P003_C1). These clones will provide a source of genomic information on bacteria that are currently relatively unknown.

**Conclusions**

DNA extraction methods were developed to allow the extraction of high molecular weight DNA from EBPR biomass.

A relatively rapid PCR-screening method was described to detect fosmid clones possessing a 16S rRNA gene.

The utilisation of the metagenomic library approach and examination of clones for 16S rRNA genes has lead to the detection of clones containing genetic information from as yet uncultured bacteria.

Fosmid clones were identified that contained DNA originating from Accumulibacter and Competibacter, both bacteria being important in the process of EBPR. These clones...
will be further investigated with additional sequencing to discover genetic information from these bacteria.

**Acknowledgements**

This research was funded by the Australian Research Council (ARC) and Christine Yeates is supported by an ARC Postdoctoral Research Fellowship. The authors would like to thank Ed DeLong and Christina Preston for advice regarding the fosmid library preparation, and Aaron Saunders, Luke Burow and Huabing Lu for providing the SBR biomass.

**References**


