Environmental Microbiology (2004) **6**(9), 990–998 doi:10.1111/j.1462-2920.2004.00673.x

Phylogeny–function analysis of (meta)genomic libraries: screening for expression of ribosomal RNA genes by large-insert library fluorescent *in situ* hybridization (LIL-FISH)

Johan H. J. Leveau,* Saskia Gerards, Wietse de Boer and Johannes A. van Veen
Netherlands Institute of Ecology (NIOO-KNAW), Centre for Terrestrial Ecology, Boterhoeksestraat 48, 6666 GA Heteren, the Netherlands.

Summary

We assessed the utility of fluorescent *in situ* hybridization (FISH) in the screening of clone libraries of (meta)genomic or environmental DNA for the presence and expression of bacterial ribosomal RNA (rRNA) genes. To establish proof-of-principle, we constructed a fosmid-based library in *Escherichia coli* of large-sized genomic DNA fragments of the mycophagous soil bacterium *Collimonas fungivorans*, and hybridized 768 library clones with the *Collimonas*-specific fluorescent probe CTE998-1015. Critical to the success of this approach (which we refer to as large-insert library FISH or LIL-FISH) was the ability to induce fosmid copy number, the exponential growth status of library clones in the FISH assay and the use of a simple pooling strategy to reduce the number of hybridizations. Twelve out of 768 *E. coli* clones were suspected to harbour and express *Collimonas* 16S rRNA genes based on their hybridization to CTE998-1015. This was confirmed by the finding that all 12 clones were also identified in an independent polymerase chain reaction-based screening of the same 768 clones using a primer set for the specific detection of *Collimonas* 16S ribosomal DNA (rDNA). Fosmids isolated from these clones were grouped by restriction analysis into two distinct contigs, confirming that *C. fungivorans* harbours at least two 16S rRNA genes. For one contig, representing 1–2% of the genome, the nucleotide sequence was determined, providing us with a narrow but informative view of *Collimonas* genome structure and content.

Received 27 February, 2004; revised 25 May, 2004; accepted 25 May, 2004. *For correspondence. E-mail j.leveau@nioo.knaw.nl; Tel. (+31) 26 479 1316; Fax (+31) 26 472 3227.

Introduction

Metagenomic libraries are databases of bacterial clones, usually *Escherichia coli*, carrying DNA fragments that originate from the collective genomes of all organisms present in a particular environment, habitat or assemblage, i.e. the so-called metagenome (Handelsman *et al*., 2002). Metagenomic libraries (also referred to as environmental or community DNA libraries) are a valuable resource for microbial ecologists for a number of reasons. First, they contain genetic information that has been obtained in a culture-independent manner. This permits the study of DNA from all microorganisms in a sample, including those that remain unculturable and from which no DNA could have been isolated otherwise. Secondly, each DNA fragment can potentially be placed and interpreted in the context of all other DNA fragments in the same library. Thirdly, a metagenomic library contains a vast amount of unexplored DNA, which offers possibilities for the discovery of novel metabolic pathways and enzymatic functions.

To uncover the vast amount of information that lies contained within a metagenomic library, several established approaches are available. One is screening for gene-encoded activities. Some examples are the searches for expression of chitinases, lipases, proteinases or esterases (Henne *et al*., 2000; Rondon *et al*., 2000), antibiotic production (Wang *et al*., 2000; MacNeil *et al*., 2001; Gillespie *et al*., 2002), biocatalyst activity (Lorenz *et al*., 2002), metabolic pathways (Henne *et al*., 1999) and antiporter activity (Majerník *et al*., 2001). An absolute requirement for such activity screenings is the expression of foreign DNA in the host strain. A problem arises if the host strain cannot provide the proper transcription factors for a particular gene, as it will not be able to express the corresponding activity. Proven solutions to this problem are the use of alternative host strains (Wang *et al*., 2000) or expression vectors that stimulate the transcription of cloned genes independent of their native promoter (Henne *et al*., 1999; 2000). A second type of screening approach is aimed at establishing the phylogenetic diversity of the metagenome represented in the library. This is accomplished by screening for clones that

carry DNA inserts with a phylogenetic marker such as the 16S rRNA gene. The most common method for 16S rDNA screening is the polymerase chain reaction (PCR) in combination with primers that are specific for individual species, genera or higher taxa (Vergin *et al*., 1998; Béjà *et al*., 2000a; 2002; Quaiser *et al*., 2002; Liles *et al*., 2003; Lopez-Garcia *et al*., 2004). A bonus feature of phylogenetic screening is the unique opportunity to mine the DNA flanking the phylogenetic marker gene for functions that could disclose details on the physiology or ecology of the organism from which the DNA originated (Stein *et al*., 1996; Béjà *et al*., 2000a; 2002; Quaiser *et al*., 2002; Liles *et al*., 2003; Lopez-Garcia *et al*., 2004). One of the most striking illustrations of this is the discovery of a bacteriorhodopsin in an uncultured marine bacterium, suggesting a lifestyle that depends on energy from sunlight (Béjà *et al*., 2000b).

Here, we have explored the possibility of merging a gene activity screening and a phylogenetic screening into one using fluorescent *in situ* hybridization (FISH) for the detection of 16S rRNA gene expression in (meta)genomic library clones (large-insert library FISH or LIL-FISH). We reasoned that, of all the genes in a given bacterial genome, the rRNA genes are probably among the most likely ones to be expressed in an *E. coli* background. This is based on the observation that, at least in *E. coli*, the expression of ribosomal RNA (rRNA) genes requires only a minimal number of ubiquitous transcription factors (Schneider *et al*., 2003). Furthermore, rRNA genes are in general very highly expressed (Schneider *et al*., 2003), which would facilitate detection of the product of 16S rRNA gene expression, i.e. rRNA. We successfully performed a proof-of-principle experiment, which showed that a fluorescently labelled probe targeting 16S rRNA of the mycophagous bacterium *Collimonas fungivorans* (de Boer *et al*., 2004) could be used to identify library clones in a *Collimonas* genomic library that expressed the *Collimonas* 16S rRNA gene from its native promoter.

Results

Construction of a large-insert DNA library from Collimonas *genomic DNA*

We successfully used a CopyControl fosmid library production kit (Epicentre) to construct a large-insert DNA library from *C. fungivorans* strain Ter331 genomic DNA. The first two steps of the standard protocol called for isolation of high-molecular-weight genomic DNA and subsequent shearing into ≈40 kb fragments. We combined isolation and shearing into a single step by isolation of genomic DNA from *Collimonas* using an UltraClean soil DNA isolation kit (Mo Bio Laboratories). We found that, by subjecting different amounts of an overnight culture of *Collimonas* to the UltraClean procedure, differently sized DNA fragments could be obtained (Fig. 1A). For example, 0.5 ml of an overnight culture resulted in >20% of the DNA sheared to the desired size of ≈40 kb (Fig. 1A, lane 3). We used this DNA preparation to create a large-insert *Collimonas* DNA library in fosmid pCC1FOS in host strain *E. coli* EPI300. It would have been possible to produce a library of about 4×10^6 clones from 0.5 ml of overnight *Collimonas* culture, but we randomly picked $n = 3200$ clones, representing a 26-fold genome coverage (assuming a genome size G of 5 Mbp and an insert size i of 40 kb). Given the formula $n = \ln(1-P)/\ln(1-i/G)$ (Sambrook *et al*., 1989), this guaranteed a probability P (expressed as a fraction) of nearly 1 that a given DNA sequence of the *Collimonas* genome was present in the library. A first quality check of the library for DNA insert size and clone diversity was done by restriction analysis of fosmid DNA isolated from eight randomly selected library clones. This revealed eight different restriction patterns and an average insert size of 33.8 ± 3.7 kb (Fig. 1B).

LIL-FISH screening of the genomic DNA library for expression of Collimonas *16S rRNA genes*

We used FISH and *Collimonas*-specific probe CTE998-1015 (de Boer *et al*., 2004) to screen 768 clones from the

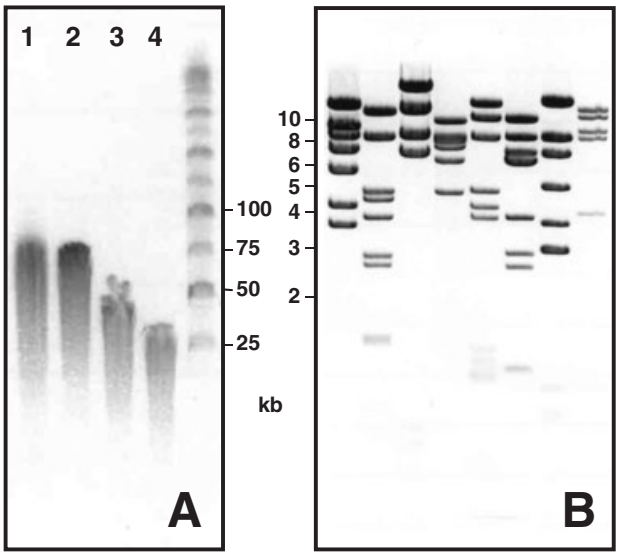

Fig. 1. Construction of the *Collimonas fungivorans* genomic DNA fosmid library.
A. One-step isolation and shearing of total DNA for ligation into vector pCC1FOS. Decreasing volumes (2, 1, 0.5 or 0.25 ml) of an overnight culture of *C. fungivorans* strain Ter331 produced increasingly sheared DNA (lanes 1, 2, 3 or 4 respectively). Details of the procedure are given in *Experimental procedures*. The fifth lane shows a 25 kb size marker (MidRange II PFG; New England Biolabs).
B. *Bam*HI restriction analysis of fosmid DNA isolated from eight randomly picked library clones. Marker sizes are given to the left. In all lanes, the 8.1 kb band represents vector pCC1FOS.

E. coli EPI300 library for the expression of *Collimonas* 16S rRNA genes. Three factors were critical to the success of this approach. The first was the ability to increase 10- to 20-fold the copy number of a vector carrying *Collimonas* genomic DNA by the CopyControl feature of fosmid pCC1FOS. Increased copy number of the *Collimonas* 16S rRNA gene resulted in intensified fluorescence of positive clones and thus greatly facilitated our ability to distinguish them from clones that did not express *Collimonas* 16S rRNA. Secondly, cell cultures were deliberately harvested in the exponential phase of growth to maximize fluorescence signal in positive clones. Thirdly, using a simple pooling scheme, the number of FISH reactions was reduced dramatically. Instead of having to perform 768 individual hybridizations, we first hybridized 64 pools of 12 clones each, then split positive pools into individual clones for a second round of FISH (Fig. 2). By this strategy, we identified a total of 12 *E. coli* clones that hybridized with the *Collimonas*-specific probe, i.e. 1B5, 1D2, 2C10, 4H12, 5B7, 5F10, 6A3, 6D9, 7C9, 7E9, 8E7 and 8G3. This result indicates the utility of LIL-FISH for searching large-insert DNA libraries for the expression of heterologous 16S rRNA genes.

Validation by PCR screening and analysis of fosmids from positive clones

To verify the results obtained by LIL-FISH, we performed a PCR-based search on the same 768 clones using *Collimonas*-specific primer set 6F/1015R. We found 13 clones that produced the expected ≈ 1 kb fragment, i.e. 1B5, 1D2, 2C10, 4H12, 5B7, 5F10, 6A3, 6C3, 6D9, 7C9, 7E9, 8E7 and 8G3. All but one had also been identified by the LIL-FISH method. The failure of clone 6C3 to be expressed was probable due to the absence of a functional promoter to direct expression of the 16S rRNA gene (see *Discussion*). From the 13 *E. coli* library clones carrying *Collimonas* 16S rDNA, fosmid DNA was isolated, digested with *Bam*HI, analysed by gel electrophoresis and hybridized in a Southern blot to a *Collimonas* 16S rRNA gene probe (Fig. 3). First of all, this confirmed that all 13 fosmids contained *Collimonas* 16S rDNA sequences. Secondly, restriction patterns revealed that the insert DNAs could be organized into two contigs, i.e. A and B. For both, we were able to draw a *Bam*HI restriction map (Fig. 4). Finding two distinct contigs suggests that *Collimonas* harbours two copies (A and B) of the 16S rDNA gene. Hybridization of *Collimonas* total DNA actually revealed three *Bam*HI fragments (Fig. 3, lane C). The largest fragment corresponded to copy B found on inserts from contig B, and the second largest to copy A on contig A. The putative third copy of 16S rDNA appears to be absent from the *Collimonas* large-insert library, possibly because of its association with genes that are toxic in *E. coli*. Interestingly, the under-representation of copy A in the fosmid library (five positive clones compared with eight for copy B) may mean that this copy too is associated with genes that are toxic to *E. coli*. The presence of such genes ≈ 20 kb upstream of 16S rDNA copy A could explain the skewed recovery of copy A from the fosmid library (Fig. 4A).

Sequence analysis of an ≈ 71 kb DNA fragment containing Collimonas *sp. 16S rRNA gene copy B*

To obtain a first glance into the genome content and structure of *C. fungivorans*, we determined the nucleotide sequence of the largest of the two contigs by sequence analysis of overlapping DNA inserts from fosmids pCFUFOS19 and pCFUFOS26 (Fig. 4B). This contig B was 71 580 bp in size, with a GC content of 59.4%. *In silico Bam*HI restriction analysis confirmed the map shown in Fig. 4B. The nucleotide sequence of contig B allowed us to determine precisely the individual sizes of pCFUFOS inserts that constitute this contig and to calculate an average insert size of 36.6 ± 2.3 kb. The 16S rDNA sequence on contig B is in complete agreement with a previously published (de Boer *et al*., 2004) partial sequence of the *Collimonas* Ter331 16S rRNA gene (accession number AJ310395). Full annotation of the pCFUFOS19 and pCFUFOS26 DNA inserts is available at GenBank as two overlapping entries, AY593479 and AY593480 respectively.

Discussion

We have demonstrated the utility of LIL-FISH to screen a large-insert DNA library for clones that carry and express bacterial 16S rRNA genes. Although we used a single-organism genomic library to establish proof-of-principle, we think that it is feasible to extend the utility of the method to the screening of metagenomic libraries. However, its success will depend on a number of factors. The first and major one is whether and to what degree 16S rRNA genes from other bacteria are expressed in host strain *E. coli*. We found that information on heterologous expression of 16S rDNA in *E. coli* is practically non-existent. One study (Schramm *et al*., 2002) used FISH to confirm that expression and detection of foreign 16S rRNA in *E. coli* is possible; however, the foreign 16S rRNA genes were transcribed from a vector-located, not their native, promoter. From the current study, it might be concluded that β-proteobacteria such as *C. fungivorans* are related enough to *E. coli* to allow for their 16S rRNA genes to be expressed. It is not clear at all whether the same holds true for more distantly related microorganisms. However, as increasing numbers of metagenomic libraries are constructed and screened phylogenetically, it becomes pos-

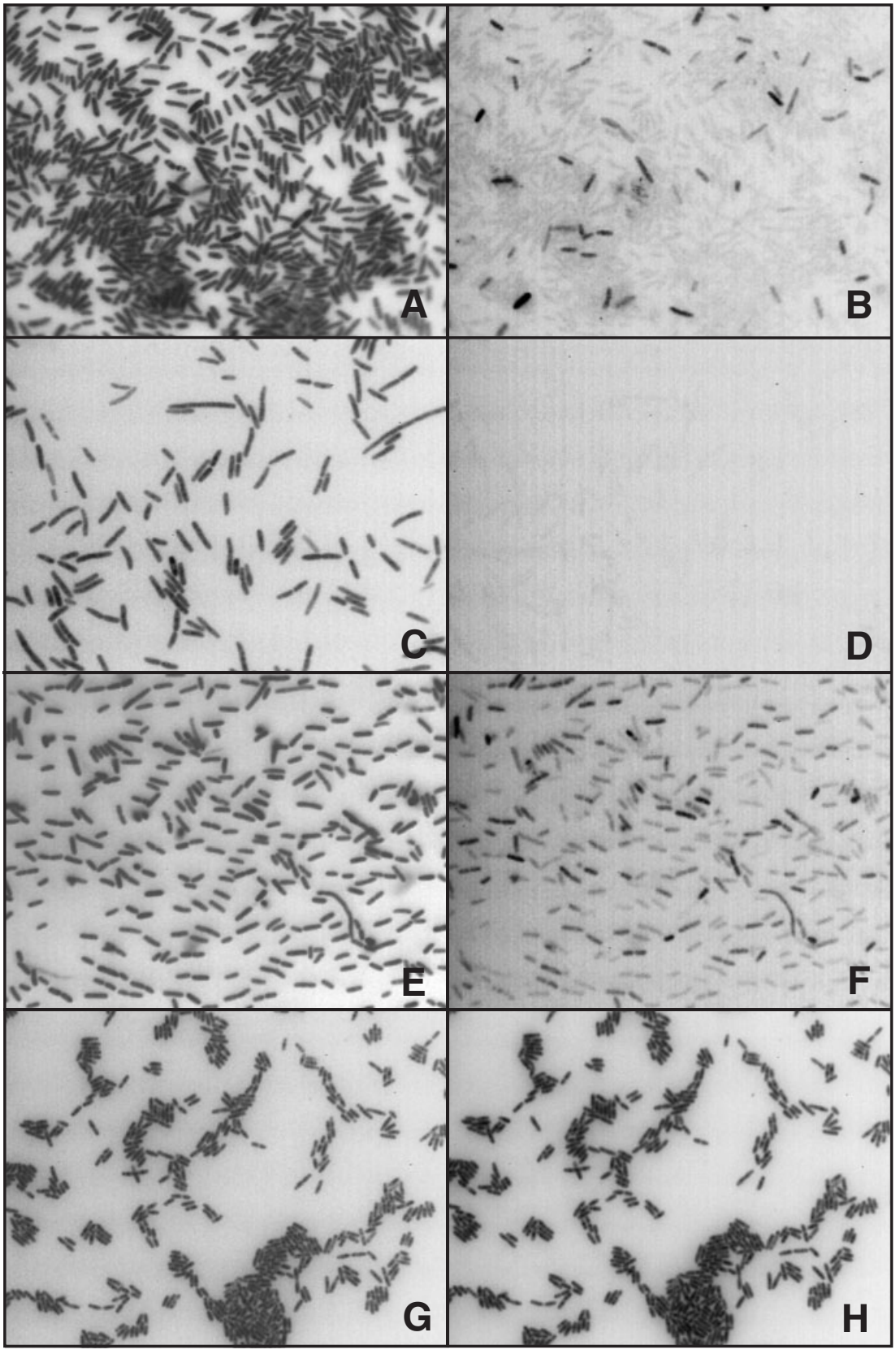
A
B
C
D
E
F
G
H

Fig. 2. Fluorescent *in situ* hybridization (FISH) of paraformaldehyde-fixed fosmid library clones (A–F) with eubacterial probe EUB338 (A, C and E) and *Collimonas*-specific probe CTE998-1015 (B, D and F).
A and B. A pool of 12 library clones which was scored as positive for *Collimonas* 16S rRNA expression.
C and D. A negative clone.
E and F. A positive clone.
G and H. A control hybridization of EUB338 (G) or CTE998-1015 (H) to *Collimonas* sp. Ter331 cells.

sible to select individual library clones that have been confirmed by PCR to contain 16S rDNA sequences and subject them to the FISH procedure as described here. If done in a systematic fashion for a wide range of phylogenetic groups, including unculturable representatives, a clearer picture will arise on the restrictions of LIL-FISH for the screening of metagenomic libraries for the expression of 16S rRNA.

A second determinant in the success of the FISH approach is the choice of probe. A universal probe or set of probes (Daims *et al*., 1999) would theoretically allow the identification of all 16S rDNA-containing fragments in a metagenomic library. However, such probes can obviously not be used because hybridization to 16S rRNA of the host strain would result in a positive score for all library clones. This impediment of host strain background is not unique to FISH: PCR-based methods suffer the same problem. There, solutions include the use of species-, genus- or other taxa-specific primer sets (Béjà *et al*., 2000a; 2002; Quaiser *et al*., 2002), or modification of the PCR conditions to inhibit amplification of *E. coli* sequences, e.g. by template-specific termination (Liles *et al*., 2003). Similar adaptations to the FISH approach are likely to improve the method for use on metagenomic libraries. For example, many taxa-specific FISH probes are already available (Loy *et al*., 2003). An as yet untested alternative is the use of unlabelled *E. coli*-specific competitor or quenching probes (Wagner *et al*., 2003) in combination with a universal probe. This could reduce the hybridization signal from the host strain, either by hindering the hybridization of universal probe to *E. coli* rRNA or by preventing the emission from universal probe bound to *E. coli* rRNA.

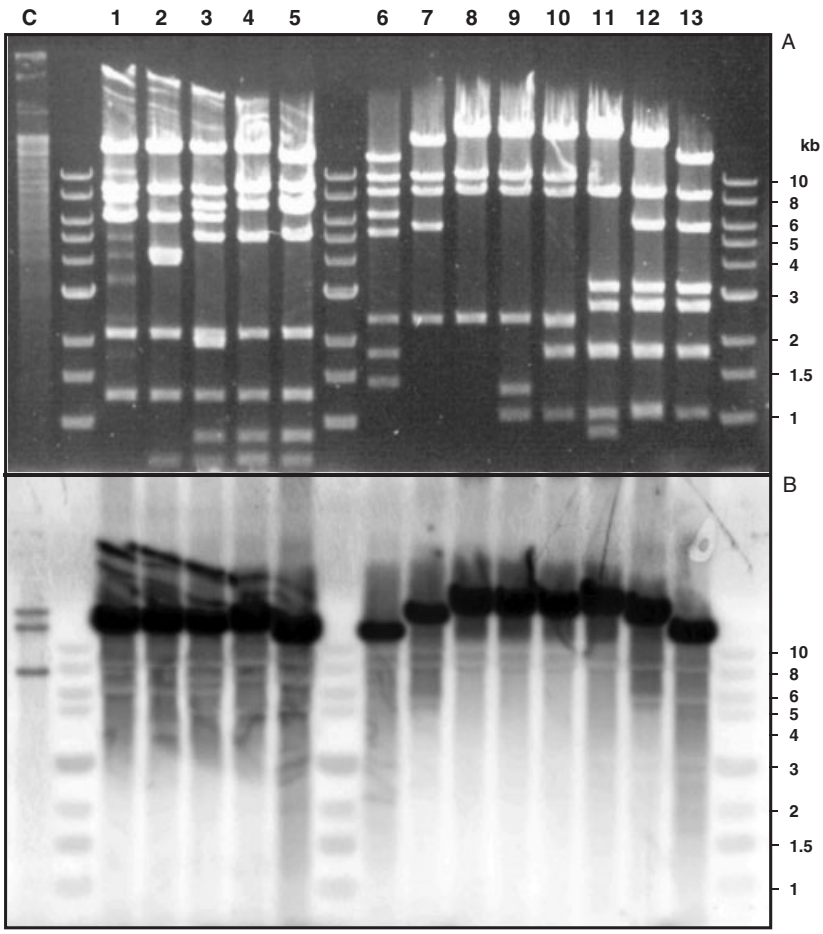

Fig. 3. Analysis of fosmid DNA isolated from library clones that scored positive in the screening for *Collimonas* 16S rDNA.
A. Gel electrophoresis of *Bam*HI-digested fosmid DNA isolated from clones 6A3, 8E7, 8G3, 6D9, 6C3, 1D2, 7E9, 5F10, 5B7, 2C10, 1B5, 7C9 and 4H12 (lanes 1–13 respectively). Lane C contains total DNA isolated from *C. fungivorans* strain Ter331.
B. Southern blot of the gel in (A), using as a probe a DNA fragment containing the *Collimonas* 16S rDNA gene.

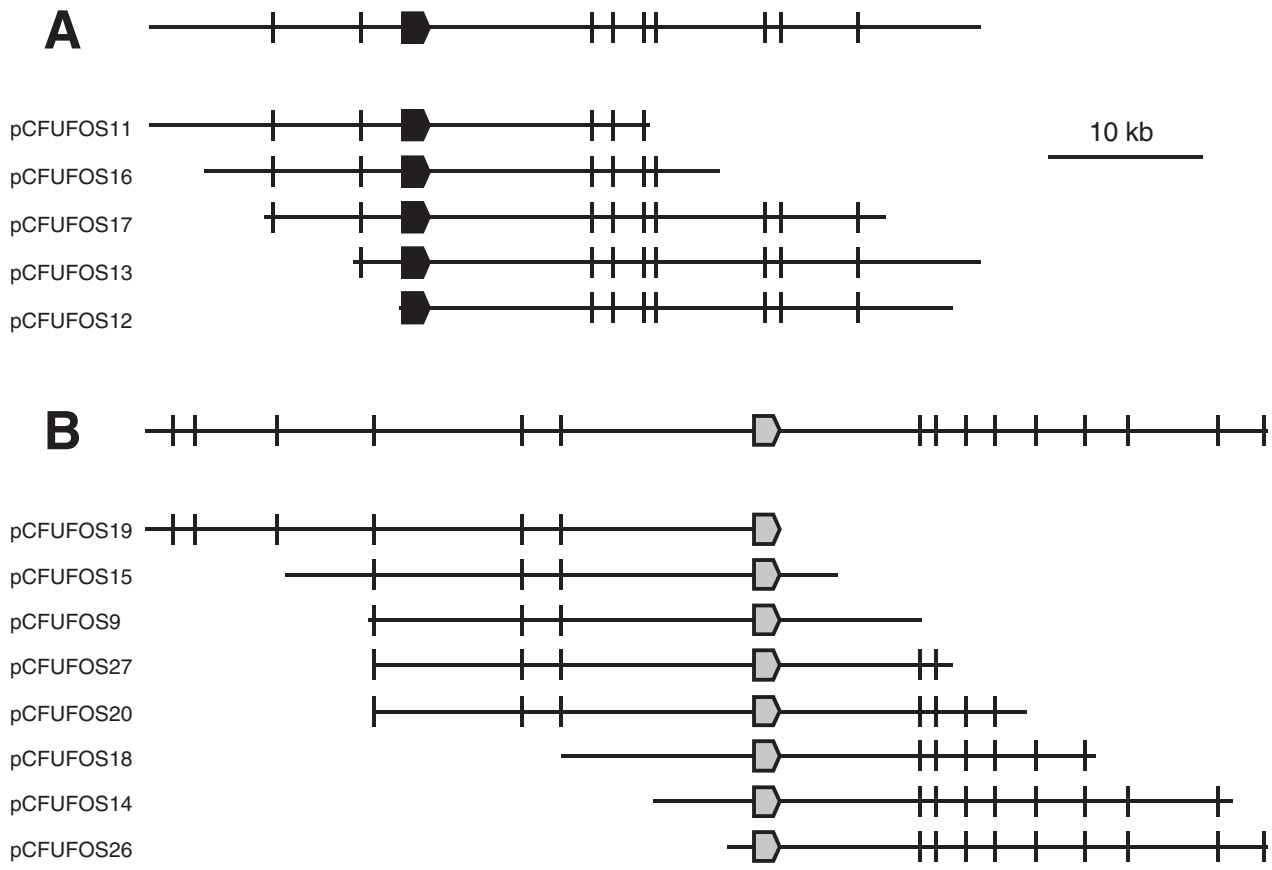

Fig. 4. *Bam*HI restriction map of fosmids isolated from 13 positive clones and their projection on to contig A (top) or B (bottom). Vertical lines represent *Bam*HI restriction sites. The block arrows show the location and orientation of 16S rDNA genes A and B. Fosmid pCFUFOS11 was isolated from clone 6A3, pCFUFOS16 from 8E7, pCFUFOS17 from 8G3, pCFUFOS13 from 6D9, pCFUFOS12 from 6C3, pCFUFOS19 from 1D2, pCFUFOS15 from 7E9, pCFUFOS9 from 5F10, pCFUFOS27 from 5B7, pCFUFOS20 from 2C10, pCFUFOS18 from 1B5, pCFUFOS14 from 7C9, and pCFUFOS26 from 4H12.

In terms of performance, the FISH method scored as well as the PCR method in our search for *Collimonas* 16S rDNA. The only anomaly was clone 6C3, which was shown to contain but not express a 16S rRNA gene. End-sequencing of pCFUFOS12, the fosmid isolated from clone 6C3, revealed that, although a complete copy A of the 16S rRNA gene was present at one end of the DNA insert (Fig. 4A), the upstream DNA region that we suspect to contain the promoter region was missing (data not shown). This is a significant finding. First, it implies that the 16S rRNA genes on all the 12 other fosmids were being expressed from their native promoters. Furthermore, it provides indirect evidence that, by the FISH method, we detected the expression, not the mere presence, of the *Collimonas* 16S rRNA genes in the *E. coli* library clones.

Searching the *Collimonas* genomic library for 16S rDNA sequences has provided us with valuable information on the genome of this organism. For one, based on the frequency with which copies A and B were found in the large-insert library, we were able to estimate the *C. fungivorans* genome size. Finding eight copies of copy B in the library suggested an 8× genome coverage, so that a low-end estimate for genome size *G* could be calculated as 36.6 (average insert size in kb) × 768 (number of clones)/8 (coverage) = 3.5 Mb. Similarly, a high-end estimate of $36.6 \times 768/5 = 5.6$ Mb could be calculated from the observed frequency of copy A. However, the latter value is probably an overestimate because of the under-representation of copy A in the fosmid library (Fig. 4A), as discussed earlier.

One important question is if and in what instances LIL-FISH might replace or complement the more commonly used PCR approach for phylogenetic screening of (meta)genomic libraries. On a small scale, such as the one presented here, LIL-FISH is somewhat more laborious than the PCR method and, with the current uncertainty as to whether 16S rRNA genes are actually expressed in *E. coli*, the PCR method probably remains first choice as the standard screening procedure. How-

ever, the strength of a FISH-based method such as LIL-FISH lies in the fact that it can be combined with the power of flow cytometry (Amann *et al*., 1990a; Wallner *et al*., 1997). We envisage a LIL-FISH-FACS procedure in which LIL-FISH is combined with fluorescence-activated cell sorting (FACS) for (i) the identification and (ii) the enrichment of fluorescently labelled cells.

(i) Identification of LIL-FISH-positive cells by flow cytometry would take advantage of the fact that a large number of cells can be analysed in a short period of time, often thousands of cells or even more per second. This high throughput, combined with the fact that a LIL-FISH-FACS procedure could be fully automated (from growing the cells to hybridization to flow cytometry) opens up the possibility of deploying pooling strategies that allow for the one-pass identification of LIL-FISH-positive clones in a clone library. Such pooling strategies are based on the 'smart' pooling of library clones in such a way that, by single screening of many pools, it can be resolved which clones are positive (Bruno *et al*., 1995).

(ii) FACS also allows for the separation of fluorescently labelled cells from non-fluorescent ones (Wallner *et al*., 1997), so that, in principle, it would be possible to isolate from a pool of library clones only those cells that are LIL-FISH positive. This feature would negate the need for pooling designs altogether. A major drawback that still needs to be addressed is that LIL-FISH requires fixation of the cells by paraformaldehyde, which renders the cells non-viable and makes impossible direct isolation of fosmid DNA from such cells for subsequent analysis. However, it has been shown that DNA can be amplified, cloned and sequenced from fixed cells by PCR (Wallner *et al*., 1997; Fode-Vaughan *et al*., 2003), and it is therefore conceivable that one could obtain DNA from sorted LIL-FISH cells in a similar way, e.g. by a combination of fosmid-specific and 16S rDNA primers. Possibly, the method could be modified to exploit the unique properties of DNA polymerases such as Φ29 (Blanco *et al*., 1989) for the amplification and (sub)cloning of large DNA fragments from 16S rDNA-containing fosmids.

Experimental procedures

Construction of a large-insert fosmid DNA library from Collimonas *genomic DNA*

We isolated genomic DNA from an overnight King's broth (King *et al*., 1954) culture of *C. fungivorans* strain Ter331 (de Boer *et al*., 2004), using an UltraClean soil DNA isolation kit (Mo Bio Laboratories). Genomic DNA was size checked by pulsed-field gel electrophoresis on a CHEF Mapper XA (Bio-Rad), and ≈3 μg was used as starting material in a Copy-Control fosmid library production kit protocol (Epicentre). This resulted in a library of ≈3200 *E. coli* EPI300 clones, each carrying an ≈40 kb DNA fragment of *Collimonas* in vector pCC1FOS. The library was stored as −80°C glycerol stocks in microtitre plate format.

Isolation of large-insert fosmid DNA from E. coli *EPI300*

Overnight cultures of *E. coli* EPI300 clones were diluted 10× in LB containing 12.5 μg ml^{-1} chloramphenicol and 1× induction solution (Epicentre) and incubated for 5 h at 37°C, 300 r.p.m. This culture (5 ml) was used in a QIAprep spin miniprep kit protocol (Qiagen), applying 50 μl of EB buffer that was heated to 70°C in the final elution step.

FISH-based search for Collimonas *16S rRNA expression in fosmid library clones*

From the large-insert *Collimonas* DNA library, 768 clones were inoculated in eight microtitre plates containing in each well 100 μl of LB with 12.5 μg ml^{-1} chloramphenicol. After overnight incubation at 37°C, 20 μl from all wells in each row was combined, resulting in 64 pools. Pooled cell suspensions were diluted 1000× in 5 ml of LB containing 12.5 μg ml^{-1} chloramphenicol and 1× induction solution (Epicentre) and incubated for 5 h at 37°C, 300 r.p.m. Cells from each pool were then fixed and hybridized as described elsewhere (Egli *et al*., 2003) with a Cy3-labelled (Thermo Electron) probe CTE998-1015 that is specific for *Collimonas* sp. 16S rDNA (de Boer *et al*., 2004). A pool was scored positive if about 10% of the cells appeared to be more fluorescent than the remaining ≈90% during analysis of the sample by epifluorescence microscopy (illumination with a Hg lamp using filter combination BP515-560, FT580, LP 580) on a Leica Diaplan. To visualize all cells, 6-FAM-labelled (Thermo Electron) universal probe Eub338 (Amann *et al*., 1990b) was used in the same FISH reaction. Positive pools were split into individual clones for a second round of FISH to identify *E. coli* library clones carrying and expressing *Collimonas* 16S rDNA.

PCR-based search for Collimonas *16S rDNA sequences in the fosmid library*

The 768 *E. coli* library clones that were tested by FISH for expression of *Collimonas* 16S rRNA were also screened by PCR for the presence of DNA fragments containing *Collimonas* 16S rDNA. Twenty microlitres of overnight microtitre plate cultures were combined from all wells in each row, resulting in 64 pools. Pooled cell suspensions were centrifuged, resuspended in 40 μl of deionized H_2O, boiled for 10 min and centrifuged again. From the supernatant, 0.5 μl was used in a 50 μl PCR, containing 1× reaction buffer, 0.2 mM each dNTP, 2.5 mM $MgCl_2$, 12.5 pmol of primer 6F (Thermo Electron), 12.5 pmol of primer 1015R (Thermo Electron) and 2.5 U of r*Taq* DNA polymerase. Primer 6F is a universal 16S rDNA primer (van der Meer *et al*., 1998), whereas primer 1015R is in fact an unlabelled version of *Collimonas*-specific CTE998-1015. PCR conditions were as follows: 10 min at 94°C, followed by 30 cycles of 1 min at 94°C, 1 min at 50°C, 1 min 30 s at 72°C. Primer set 6F/1015R was expected to give an ≈1 kb amplification product.

Southern blot analysis of Collimonas *total DNA and* E. coli *library fosmids carrying* Collimonas *16S rDNA sequences*

Collimonas genomic DNA and fosmids isolated from positive *E. coli* library clones were digested with *Bam*HI, run on a 0.8% agarose gel in 1× TAE, transferred on to nylon membrane (Roche Diagnostics) by capillary transfer under neutral conditions (Sambrook *et al*., 1989) and hybridized with a digoxigenin (DIG)-labelled 1.5 kb DNA fragment containing the *Collimonas* 16S rDNA gene. Labelling, hybridization and probe detection were done according to specifications in the DIG-high prime DNA labelling and detection starter kit (Roche Diagnostics).

DNA sequencing of large-insert Collimonas *DNA plasmids*

Sequencing of overlapping fosmids pCFUFOS19 and pCFUFOS26 was performed by the Joint Genome Institute under the auspices of the US Department of Energy's Office of Science, Biological and Environmental Research Program and by the University of California, Lawrence Livermore National Laboratory under contract no. W-7405-ENG-48, Lawrence Berkeley National Laboratory under contract no. DE-AC03-76SF00098 and Los Alamos National Laboratory under contract no. W-7405-ENG-36.

Acknowledgements

The authors would like to thank Maria Marco, George Kowalchuk and Larissa Folman for their suggestions on the manuscript. This is NIOO publication number 3338.

References

Amann, R.I., Binder, B.J., Olson, R.J., Chisholm, S.W., Devereux, R., and Stahl, D.A. (1990a) Combination of 16S ribosomal-RRA-targeted oligonucleotide probes with flow-cytometry for analyzing mixed microbial populations. *Appl Environ Microbiol* **56:** 1919–1925.

Amann, R.I., Krumholz, L., and Stahl, D.A. (1990b) Fluorescent-oligonucleotide probing of whole cells for determinative, phylogenetic, and environmental studies in microbiology. *J Bacteriol* **172:** 762–770.

Béjà, O., Suzuki, M.T., Koonin, E.V., Aravind, L., Hadd, A., Nguyen, L.P., *et al.* (2000a) Construction and analysis of bacterial artificial chromosome libraries from a marine microbial assemblage. *Environ Microbiol* **2:** 516–529.

Béjà, O., Aravind, L., Koonin, E.V., Suzuki, M.T., Hadd, A., Nguyen, L.P., *et al.* (2000b) Bacterial rhodopsin: evidence for a new type of phototrophy in the sea. *Science* **289:** 1902–1906.

Béjà, O., Koonin, E.V., Aravind, L., Taylor, L.T., Seitz, H., Stein, J.L., *et al.* (2002) Comparative genomic analysis of archaeal genotypic variants in a single population and in two different oceanic provinces. *Appl Environ Microbiol* **68:** 335–345.

Blanco, L., Bernad, A., Lázaro, J.M., Martín, G., Garmendia, C., and Salas, M. (1989) Highly efficient DNA synthesis by the phage Φ29 DNA polymerase: symmetrical mode of DNA replication. *J Biol Chem* **264:** 8935–8940.

de Boer, W., Leveau, J.H.J., Kowalchuk, G.A., Klein Gunnewiek, P.J.A., Abeln, E.C.A., Figge, M.J., *et al.* (2004) *Collimonas fungivorans* gen. nov., sp. nov., a chitinolytic soil bacterium with the ability to grow on living fungal hyphae. *Int J Syst Evol Microbiol* **54:** 857–864.

Bruno, W.J., Knill, E., Balding, D.J., Bruce, D.C., Doggett, N.A., Sawhill, W.W., *et al.* (1995) Efficient pooling designs for library screening. *Genomics* **26:** 21–30.

Daims, H., Bruhl, A., Amann, R., Schleifer, K.H., and Wagner, M. (1999) The domain-specific probe EUB338 is insufficient for the detection of all Bacteria: development and evaluation of a more comprehensive probe set. *Syst Appl Microbiol* **22:** 434–444.

Egli, K., Langer, C., Siegrist, H.R., Zehnder, A.J.B., Wagner, M., and van der Meer, J.R. (2003) Community analysis of ammonia and nitrite oxidizers during start-up of nitrification reactors. *Appl Environ Microbiol* **69:** 3213–3222.

Fode-Vaughan, K.A., Maki, J.S., Benson, J.A., and Collins, M.L.P. (2003) Direct PCR detection of *Escherichia coli* O157:H7. *Lett Appl Microbiol* **37:** 239–243.

Gillespie, D.E., Brady, S.F., Bettermann, A.D., Cianciotto, N.P., Liles, M.R., Rondon, M.R., *et al.* (2002) Isolation of antibiotics turbomycin A and B from a metagenomic library of soil microbial DNA. *Appl Environ Microbiol* **68:** 4301–4306.

Handelsman, J., Liles, M., Mann, D., Riesenfeld, C., and Goodman, R.M. (2002) Cloning the metagenome: culture-independent access to the diversity and functions of the uncultivated microbial world. *Funct Microbial Genomics* **33:** 241–255.

Henne, A., Daniel, R., Schmitz, R.A., and Gottschalk, G. (1999) Construction of environmental DNA libraries in *Escherichia coli* and screening for the presence of genes conferring utilization of 4-hydroxybutyrate. *Appl Environ Microbiol* **65:** 3901–3907.

Henne, A., Schmitz, R.A., Bomeke, M., Gottschalk, G., and Daniel, R. (2000) Screening of environmental DNA libraries for the presence of genes conferring lipolytic activity on *Escherichia coli*. *Appl Environ Microbiol* **66:** 3113–3116.

King, E.O., Ward, M.K., and Raney, D.E. (1954) Two simple media for the demonstration of pyocyanin and fluorescein. *J Lab Clin Med* **44:** 301–307.

Liles, M.R., Manske, B.F., Bintrim, S.B., Handelsman, J., and Goodman, R.M. (2003) A census of rRNA genes and linked genomic sequences within a soil metagenomic library. *Appl Environ Microbiol* **69:** 2684–2691.

Lopez-Garcia, P., Brochier, C., Moreira, D., and Rodriguez-Valera, F. (2004) Comparative analysis of a genome fragment of an uncultivated mesopelagic crenarchaeote reveals multiple horizontal gene transfers. *Environ Microbiol* **6:** 19–34.

Lorenz, P., Liebeton, K., Niehaus, F., and Eck, J. (2002) Screening for novel enzymes for biocatalytic processes: accessing the metagenome as a resource of novel functional sequence space. *Curr Opin Biotechnol* **13:** 572–577.

Loy, A., Horn, M., and Wagner, M. (2003) ProbeBase: an online resource for rRNA-targeted oligonucleotide probes. *Nucleic Acids Res* **31:** 514–516.

MacNeil, I.A., Tiong, C.L., Minor, C., August, P.R., Grossman, T.H., Loiacono, K.A., *et al.* (2001) Expression and isolation of antimicrobial small molecules from soil DNA libraries. *J Mol Microbiol Biotechnol* **3:** 301–308.

Majerník, A., Gottschalk, G., and Daniel, R. (2001) Screening of environmental DNA libraries for the presence of genes conferring Na+ (Li+)/H+ antiporter activity on *Escherichia coli*: characterization of the recovered genes and the corresponding gene products. *J Bacteriol* **183:** 6645–6653.

van der Meer, J.R., Werlen, C., Nishino, S.F., and Spain, J.C. (1998) Evolution of a pathway for chlorobenzene metabolism leads to natural attenuation in contaminated groundwater. *Appl Environ Microbiol* **64:** 4185–4193.

Quaiser, A., Ochsenreiter, T., Klenk, H.P., Kletzin, A., Treusch, A.H., Meurer, G., *et al.* (2002) First insight into the genome of an uncultivated crenarchaeote from soil. *Environ Microbiol* **4:** 603–611.

Rondon, M.R., August, P.R., Bettermann, A.D., Brady, S.F., Grossman, T.H., Liles, M.R., *et al.* (2000) Cloning the soil metagenome: a strategy for accessing the genetic and functional diversity of uncultured microorganisms. *Appl Environ Microbiol* **66:** 2541–2547.

Sambrook, J., Maniatis, T., and Fritsch, E.F. (1989) *Molecular Cloning: A Laboratory Manual.* Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

Schneider, D.A., Ross, W., and Gourse, R.L. (2003) Control of rRNA expression in *Escherichia coli*. *Curr Opin Microbiol* **6:** 151–156.

Schramm, A., Fuchs, B.M., Nielsen, J.L., Tonolla, M., and Stahl, D.A. (2002) Fluorescence *in situ* hybridization of 16S rRNA gene clones (Clone-FISH) for probe validation and screening of clone libraries. *Environ Microbiol* **4:** 713–720.

Stein, J.L., Marsh, T.L., Wu, K.Y., Shizuya, H., and DeLong, E.F. (1996) Characterization of uncultivated prokaryotes: isolation and analysis of a 40-kilobase-pair genome fragment from a planktonic marine archaeon. *J Bacteriol* **178:** 591–599.

Vergin, K.L., Urbach, E., Stein, J.L., DeLong, E.F., Lanoil, B.D., and Giovannoni, S.J. (1998) Screening of a fosmid library of marine environmental genomic DNA fragments reveals four clones related to members of the order Planctomycetales. *Appl Environ Microbiol* **64:** 3075–3078.

Wagner, M., Horn, M., and Daims, H. (2003) Fluorescence *in situ* hybridisation for the identification and characterisation of prokaryotes. *Curr Opin Microbiol* **6:** 302–309.

Wallner, G., Fuchs, B., Spring, S., Beisker, W., and Amann, R. (1997) Flow sorting of microorganisms for molecular analysis. *Appl Environ Microbiol* **63:** 4223–4231.

Wang, G.Y.S., Graziani, E., Waters, B., Pan, W.B., Li, X., McDermott, J., *et al.* (2000) Novel natural products from soil DNA libraries in a streptomycete host. *Organic Lett* **2:** 2401–2404.