Phylogenetic and metagenomic analysis of Verrucomicrobia in former agricultural grassland soil

Anna Kielak1, Jorge L.M. Rodrigues2, Eiko E. Kuramae1,3, Patrick S.G. Chain4,5, Johannes A. van Veen1,6 & George A. Kowalchuk1,7

1Department of Microbial Ecology, Netherlands Institute of Ecology (NIOO-KNAW), Heteren, The Netherlands; 2Arlington Department of Biology, University of Texas, Arlington, TX, USA; 3Department of Animal Ecology, Institute of Ecological Science, Free University Amsterdam, The Netherlands; 4Microbial Program, Joint Genome Institute, Walnut Creek, CA, USA; 5Center for Microbial Ecology, Michigan State University, East Lansing, MI, USA; 6Institute of Biology, Leiden University, Leiden, The Netherlands; and 7Institute of Ecological Science, Free University Amsterdam, Amsterdam, The Netherlands

Correspondence: George A. Kowalchuk, Department of Microbial Ecology, Netherlands Institute of Ecology (NIOO-KNAW), PO Box 40, 6666 ZG, Heteren, The Netherlands. Tel.: +31 026 479 1314; fax: +31 026 472 3227; e-mail: g.kowalchuk@nioo.knaw.nl

Received 7 April 2009; revised 18 August 2009; accepted 24 August 2009. Final version published online 7 October 2009.

DOI:10.1111/j.1574-6941.2009.00785.x

Editor: Michael Wagner

Keywords
Verrucomicrobia; phylogeny; environmental genomics; bacterial communities; rhizosphere; metagenomics.

Abstract
The bacterial phylum Verrucomicrobia has a widespread distribution, and is known to be one of the most common and diverse phyla in soil habitats. However, members of this phylum have typically been recalcitrant to cultivation methods, hampering the study of this presumably important bacterial group. In this study, we examine the phylogenetic diversity of the Verrucomicrobia in a former agricultural field and gain access to genomic information via a metagenomic approach. We examined Verrucomicrobia-like 16S rRNA gene sequences recovered from general bacterial and phylum-specific libraries, revealing a dominance of subdivisions 1 and 2. A PCR-based screening method was developed to identify inserts containing verrucomicrobial 16S rRNA genes within a large-insert metagenomic library, and on screening of 28 800 clones, four fosmids were identified as containing verrucomicrobial genomic DNA. Full-length sequencing of fosmid inserts and gene annotation identified a total of 98 ORFs, representing a range of functions. No conservation of gene order was observed adjacent to the ribosomal operons. Fosmid inserts were further analyzed for tetranucleotide frequencies to identify remnants of past horizontal gene transfer events. The metagenomic approach utilized proved to be suitable for the recovery of verrucomicrobial genomic DNA, thereby providing a window into the genomes of members of this important, yet poorly characterized, bacterial phylum.

Introduction
Members of the phylum Verrucomicrobia are abundant in diverse habitats, including soil, water and sediments, as well as extreme environments such as in the Antarctic sample (Pearce et al., 2003), hot springs (60 °C) and low pH 2.0–2.5 habitats (Hou et al., 2008). The lifestyles of some Verrucomicrobia have also been linked to eukaryotic gut environments, including termites (Shinzato et al., 2005), sea cucumbers (Sakai et al., 2003) and humans (Wang et al., 2005), or as obligate endosymbionts of nematodes from the genus Xiphinema (Vandekerckhove et al., 2000). In addition, methanotrophs have been identified among the Verrucomicrobia (Dunfield et al., 2007; Pol et al., 2007). Approximately 5% of all characterized 16S rRNA genes are affiliated with the Verrucomicrobia, and this phylum represents 1.2–11% of the total soil bacteria (Sangwan et al., 2004; Kielak et al., 2008). Although only recognized as a separate phylum in the past decade (Garrity & Holt, 2001), their large numbers and broad distribution in soils suggest that they are important members of terrestrial ecosystems.

Although recent advances in cultivation methods have increased the number and diversity of available Verrucomicrobia isolates (Janssen et al., 1997; Janssen et al., 2002; Sangwan et al., 2004, 2005; Stevenson et al., 2004; Yoon et al., 2007a,b), the general poor cultivation of Verrucomicrobia has impaired progress in understanding their ecology. The Verrucomicrobia has been organized into seven
subdivisions based on 16S rRNA gene analyses, although not all subdivisions contain culture representatives (Schlesner et al., 2006). Characterized genera include Verrucomicrobiurn (subdivision 1 – Verrucomicrobiae) (Schlesner, 1987), Prosthecobacter (subdivision 1) (Staley et al., 1976), Akkermansia (subdivision 1) (Derrien et al., 2004), Xiphinema-bacter (subdivision 2 – Spartobacteria) (Vandekerckhove et al., 2000), Chthoniobacter (subdivision 2) (Sangwan et al., 2004), Opitutus (subdivision 4 – Opitutaceae) (Chin et al., 2001), Rubribacteria (subdivision 4) (Scheuermayer et al., 2006) and Victivallis (subdivision 7) (Zoetendal et al., 2003). Complete or draft sequences have recently become available for eight Verrucomicrobiurn strains (bacterium Ellin514, Verrucomicrobiurn bacterium DG1235, Akkermansia muciniphila ATCC BAA-835, Opitutaceae bacterium TAV2, Opitutus terrae PB90-1, Methylacidiphilum infernorum V4, Chthoniobacter flavus Ellin428 and Verrucomicrobiurn spinosum DSM 4136), providing a first insight into the genetic makeup of cultivated members of the Verrucomicrobia. As with other bacteria, it is thought that cultivated members of the Verrucomicrobia represent a very biased sampling of the phylum due to the inherent selection of isolates capable of growth under (novel) laboratory conditions. To date, of the 89 verrucomicrobial 16S rRNA gene sequences deposited in the RDPII database (Cole et al., 2005) derived from isolates, 40 are affiliated to subdivision 4, followed by 18, 14 and six sequences from subdivisions 1, 2 and 3, respectively. Only 17 of these sequences have come from soil isolates, 11 of which are affiliated to subdivision 4 and six to subdivision 6.

Although Verrucomicrobia have commonly been recovered from diverse soil environments, relatively little is known about their diversity and distribution with respect to the environmental conditions present in different soilborne habitats. Soil management and nutritional regimes appear to influence Verrucomicrobia community structure (Buckley & Schmidt, 2001), and bacterial 16S rRNA gene library and quantitative PCR analyses have shown that Verrucomicrobia tend to be more dominant in the rhizosphere (Kielak et al., 2008). However, differences between the rhizospheres of different plants tended not to be significant (Kielak et al., 2008). Thus, it appears that Verrucomicrobia represent important rhizosphere colonizers, and yet virtually no information is available to date on the genotypic or the phenotypic characteristics of these bacteria. Furthermore, their importance in terrestrial ecosystems and potential as sources of natural products remains unexplored.

In this study, we therefore sought to gain an insight into the phylogenetic diversity of Verrucomicrobia in a former agricultural field and to develop and apply a cultivation-independent approach towards gaining access to genomic information from this phylum. To address the first of these goals, Verrucomicrobia-like 16S rRNA gene sequences were examined from the libraries recovered using general bacterial and Verrucomicrobia-specific PCR primers. In addition, a large insert fosmid library (28 800 clones) was constructed from high-molecular-weight DNA isolated from the rhizosphere of Festuca ovina grown at the same field site. A PCR-based screening procedure was developed and implemented to identify clones containing inserts with Verrucomicrobia-like 16S rRNA genes. Four inserts, positively determined to be of verrucomicrobial origin, were subsequently subjected to full-length DNA sequence analysis, followed by identification and annotation of ORFs. The genomic properties of verrucomicrobial genomic inserts were studied and patterns of oligonucleotide usage were examined to identify sites of potential horizontal gene transfer (HGT).

**Materials and methods**

**Field site and sampling**

All soil samples were collected from a former arable field site near Ede (52°04′N, 5°45′E), the Netherlands, where an ecosystem restoration experiment was initiated after the 1995 harvest season within the European project ‘Changing land usage: enhancement of biodiversity and ecosystem development’. Detailed characteristics of the soil and plant diversity treatment plots are provided in Van der Putten et al. (2000) and Bezemer & Van der Putten (2007).

Samples for the construction of general bacterial 16S rRNA gene libraries were derived from collective rhizosphere and bulk soil from each of the plant diversity treatments in the above experiment as described by Kielak et al. (2008). For the construction of Verrucomicrobia-specific 16S rRNA gene libraries, rhizosphere soil samples were collected in August 2005 from *F. ovina* L. plants from within monoculture plots of this species, adjacent to the main plant diversity treatment experiment, as well as from *Lotus corniculatus* L. plants from various locations within the borders of the plots described above (Bezemer et al., 2006; Kielak et al., 2008).

Rhizosphere and bulk soil samples were taken as follows: roots were first separated and shaken to remove loosely adhering soil. Pieces of roots with the remaining adhering soil were cut and taken as a combined sample of rhizosphere and root surface (rhizoplane), further referred to as rhizosphere samples. Soil remaining behind after the removal of plant material was defined as bulk soil. All soil samples were frozen at 20°C for approximately 2 weeks before DNA extraction.

**DNA isolation**

For the construction of all 16S rRNA gene libraries, DNA was isolated from 250 mg wet weight soil using the Ultra Clean™ DNA Isolation Kit, according to the manufacturer's
PCr-based cloning strategies

Bacterial 16S rRNA gene libraries were constructed as described by Kielak et al. (2008). For Verrucomicrobia-specific libraries, a phylum-specific forward primer, VerS3F, was combined with a bacterial reverse primer, 1378r (Stevenson et al., 2004). The specificity of the primer was confirmed by probe match check against the Silva 16S rRNA gene database (Fruess et al., 2007). Thermocycling consisted one 2-min denaturation step at 96°C, followed by 30 cycles of 30 s at 92°C; 60 s at 65°C and 45 s at 68°C. The terminal elongation step was extended by 10 min, and reactions were cooled to 10°C upon completion.

PCr product size (approximately 1300 bp) and integrity were examined by standard agarose gel electrophoresis and ethidium bromide staining. PCr products were purified using the Qiagen PCR Purification Kit (Qiagen, Venlo, the Netherlands) and cloned into the pGEM-T vector, as per the manufacturer’s protocol (Promega, Madison, WI), for transformation into E. coli JM109 competent cells using the pGEM-T Vector System. Fifteen positive transformants were selected randomly for each rhizosphere soil library, screened for potential chimera structures using CHIMERA-CHECK (http://rdp.cme.msu.edu) and BELLEROPHON (Huber et al., 2004) software. After the screening of libraries, 26 clones from the Verrucomicrobia-specific libraries and nine clones previously identified from the general bacterial libraries (Kielak et al., 2008) were selected for further analysis. All sequences were also examined by BLAST (Altschul et al., 1997) to identify the most closely related database sequences.

Screening of the metagenomic library for verrucomicrobial inserts

A total of 28 800 clones in the metagenome library were subjected to a PCr-based screening strategy using a semi-nested clone pooling strategy (Kielak et al., 2008, 2009).

Briefly, clones were cultured overnight in 96-well plates, with the contents of four plates (384 clones) combined in a single plasmid extraction (QIAprep Spin Miniprep Kit; Qiagen). The resulting mixed templates were subjected to PCr with the pAF/1378R primers (Edwards et al., 1989), followed by primers VerS3F/1378R (Stevenson et al., 2004) as described above. If a PCr product was detected, clone pooling was reduced to a single plate. Pooling was further reduced to a single row of a culture plate and eventually a single clone for subsequent reactions that produced positive results. Four clones were identified of putative verrucomicrobial origin.

Fosmid clone sequencing and assembly

Four fosmid DNA preparations for each positive clone (designated clones 106, 118, 184 and 286) were performed using the Qiagen Midi Kit, and DNA concentrations were quantified using a Nanodrop (Thermo Scientific) instrument. We then pooled 2.5 μg from each fosmid DNA sample for a total of 10 μg DNA. The pooled fosmid DNA was subjected to pyrosequencing using a 1/2 FLX run on a Roche/454 Life Sciences Genome Sequencer. The resulting data were assembled using NEWBLER ASSEMBLY software v1.1.02.15 (Roche/454 Life Sciences). A total of 32 Mb of sequence data were assembled into 52 (500 kb) contigs. All large contigs were ‘shredded’ into 1000-bp fragments, overlapping one another by 100 bp. The resulting 192 ‘shreds’ were reassembled with PHRED/PHRAP software, yielding 21 contigs, totalling 137 016 bp. The four fosmid clones were also end-sequenced using the vector-targeting primers T7 and T7 RV primer (5′-ATGACCATGATTACGCAAG-3′) to facilitate orientation of the fosmid assignment. Gap closure by primer walking was completed with the aid of 32 additional custom primers.

ORF characterization

ORFs were assigned using the GLIMMER (http://www.ncbi.nlm.nih.gov/geneBank/MICROBES/glimmer_3.cgi) and FGENESB (http://linux1.softberry.com) software tools. Annotation of the identified ORFs was based on a similarity search against nonredundant protein databases using BLASTX.
Phylogenetic analyses

All recovered Verrucomicrobiium-like 16S rRNA gene sequences (four from metagenomic, nine from general bacterial and 26 from phylum-specific libraries) were aligned using the ARB software package (Ludwig et al., 2004) along with 68 sequences obtained from public databases, and the alignment was manually edited based on the secondary structure. Phylogenetic reconstruction was performed by neighbor-joining (NJ), maximum likelihood (ML) and Bayesian inference (BI). MRMODELTEST 2.2 (Nylander, 2004) was used to determine the most appropriate nucleotide substitution model. The likelihood ratio test, as well as the Akaike information criterion, identified the general time-reversible model with invariant sites and the Γ-shaped distribution of substitution rates as the best-fitting model. The NJ and ML trees were constructed using the PHYLIP 3.67 package (Felsenstein, 2005) with the model parameters calculated by MRMODELTEST. The NJ tree was bootstrapped 1000 times, and the ML tree 100 times. The Bayesian tree was constructed using the MRBAYES 3.1.2 program (Ronquist & Huelsenbeck, 2003). Four independent computations, with random starting trees and Markov chains, were run for 3 000 000 generations with a sampling frequency of 100 generations. The first 600 000 generations were discarded as burn-in.

Comparative genomics and identification of potential sites of HGT

The gene orders within the recovered verrucomicrobial fosmid inserts and ORF similarities were compared with the eight publicly available verrucomicrobial draft genome sequences by pairwise TBLASTX using BLAST v2.2.18 (ftp://ftp.ncbi.nih.gov; Altschul et al., 1990) and ACT software for gene/ORF visualization (Teeling et al., 2004; Carver et al., 2005). The e-value threshold for ORFs compared with these genomes was 10^{-30}, because the diversity within the Verrucomicrobiium phylum and the low number of available Verrucomicrobiium genomes limit the chance for perfect match detection.

Searches for GC islands were performed using CPGFINDER (SoftBerry, http://linux1.softberry.com). For identification of potential HGT regions, the method described by Tamames & Moya (2008) was applied to examine tetrancleotide frequencies. Fosmid sequences were edited to include all ORFs after removal of rRNA operons, and the resulting sequences were cut into 300-bp fragments and tiled with an overlap of 290 bp for each subsequent fragment. Pearson’s correlation coefficients of tetrancleotide frequencies for all fragments were calculated using TETRA software (Teeling et al., 2004), and correlation matrix visualization was performed in R (R statistical programming environment) using the ‘gplots’ package.

Sequence accession numbers

The 16S rRNA gene sequences described in this study have been deposited in GenBank under accession numbers FJ822616–FJ822655. The accession numbers of fosmid clone sequences are as follows: clone 106, FJ872572; clone 118, FJ872573; clone 184, FJ872574; and clone 286, FJ872575.

Results and discussion

Several studies based on sequencing of 16S rRNA genes have suggested that Verrucomicrobiium represent one of the more dominant phyla in soil environments (Janssen, 2006; Kielak et al., 2008). They typically represent between 2% and 8% of the total bacterial community, with the highest densities observed in the rhizosphere (Kielak et al., 2008). This relatively high abundance suggests that members of this phylum may play significant roles in the soil environment and have important interactions with plants. This study was therefore geared towards recovering information regarding the phylogenetic diversity of Verrucomicrobiium in soil, as well as providing tools for gaining access to genomic information from this potentially important phylum.

Phylogenetic analyses (NJ, ML and BI) of 16S rRNA gene sequences recovered from three independent cloning approaches showed strong phylogenetic support for the established clustering system within the Verrucomicrobiium (Hugenholtz et al., 1998; Schlesner et al., 2006) (NJ tree in Fig. 1; ML and BI trees not shown). Subdivision 4 was found to be the basal group in all phylogenetic trees, followed by subdivision 3. Subdivisions 1 and 2 appear to be sister groups (Fig. 1).

In agreement with other studies (Janssen, 2006), we identified subdivision 2 (Spartobacteria) to be the most dominantly detected cluster within all libraries, representing 18 of 27, six of eight and three of four clones for the Verrucomicrobiium phylum-specific, general bacterial and metagenomic libraries, respectively. Subdivision 1 (Verrucomicrobiia) was also represented in all libraries with one identified metagenomic clone, two clones in the bacterial-specific library and four in the Verrucomicrobiium-specific library. Subdivisions 3 and 4 were relatively rare and were only identified in the Verrucomicrobiium phylum-specific libraries (three clones and one clone for subdivisions 3 and 4, respectively). Although all three cloning strategies yielded highly similar results, it is premature to state that the three...
Fig. 1. NJ tree of the phylum Verrucomicrobia, highlighting the sequences recovered in this study. Names in italics indicate sequences retrieved from GenBank, red, general bacterial library; blue, metagenome library; green, phylum specific library; underlined, genomes being sequenced. Numbers near nodes indicate bootstrap values. Sequences from the phylum Planctomycetes were used as an outgroup.
methods are truly similar in their recovery of sequences from across the breadth of the phylum, given the low number of clones examined. No clones closely related to subdivisions 5, 6 or 7 were identified in any of the libraries, and these subdivisions were therefore excluded from the presented phylogenetic analyses.

The number of clones from the metagenomic library determined to contain genomic inserts of verrucomicrobial origin was rather low (four fosmid clones), but within the range (three to nine positive clones) that would be expected, given the library size (28 800 clones), size of the insert (32 kb), estimated average genome size (6000 kb), 16S rRNA gene copies per genome (1–3) and the relative abundance of Verrucomicrobia within the soil community sampled (4.9–5.7%) (Kielak et al., 2008). The average insert size per positive fosmid clone was 32 ± 1.5 kb (Table 1), yielding a total of approximately 130 kb of Verrucomicrobia genetic information. All four genome fragments revealed typical bacterial rrs operon organization, with 16S rRNA gene, followed by tRNAIle, the 23S rRNA gene and then the 5S rRNA gene. In two clones (118 and 286), the 16S rRNA and tRNAIle genes were separated by tRNAAla.

The total G+C content of the fosmid inserts ranged from 60% to 62%, with a range from 61% to 64% for coding regions. These values are comparable to those found for verrucomicrobial genomes for which drafts are available, with G+C contents ranging from 52% for Ellin514 (http://www.jgi.doe.gov) to 65% for O. terrae PB90-1 (NC010571). Methylacidiphilum infernorum V4 (NC010794) is a marked exception, with a G+C content of only 45%. Pearson’s correlation coefficients for the four verrucomicrobial fosmid inserts were also calculated based on tetranucleotide usage patterns. Using this measure, clones 118 and 106 were most similar ($r = 0.78$), followed by the clone pair 118 and 286 ($r = 0.75$). The lowest correlation was between clones 106 and 286 ($r = 0.62$), while clone 184 was similarly distant to the other three clones ($r = 0.64$). The divergence in 16S rRNA genes followed a similar pattern, with the lowest dissimilarity (12%) between clones 118 and 106 and the highest dissimilarity (20%) between clones 184 and 286 (Table 2).

The numbers of putative protein-encoding genes assigned for each clone varied between 22 ORFs (clone 184) and 28 ORFs (clone 118) (Table 1). All ORFs were compared with the sequences available in public databases with special emphasis on the sequences available for the eighth Verrucomicrobia genome project currently underway. A total of 10% of the predicted genes, found predominantly within clones 118 and 286, showed low or no similarity to sequences deposited in public databases (Supporting Information, Tables S1 and S2). The number of ORFs with best matches to verrucomicrobial sequences ranged between 11 out of 22

### Table 1. Verrucomicrobial genome fragment features

| Fosmid clone ID | 286 | 118 | 106 | 184 |
|----------------|-----|-----|-----|-----|
| Number of bases (kb) | 32 226 | 32 436 | 33 568 | 29 919 |
| rRNA | 3 (16S, 5S, 23S) | 3 (16S, 5S, 23S) | 3 (16S, 5S, 23S) | 3 (16S, 5S, 23S) |
| tRNA | 2 | 2 | 1 | 1 |
| Number of ORFs | 22 | 28 | 26 | 22 |
| Coding percentage | 82% | 90% | 85% | 95% |
| GC percentage in coding regions | 63.51% | 63.69% | 61.29% | 63.84% |
| Overall GC percentage | 60.42% | 61.62% | 59.65% | 61.54% |

### Table 2. Correlation coefficients between Verrucomicrobia 106, 118, 184 and 286 genome fragments (including and excluding the ribosomal operon) based on tetranucleotide usage patterns and distance matrix of 16S rRNA gene sequences

| Fosmid clone ID | 106 (With RNA operon) | 184 (Without RNA operon) | 118 (Without RNA operon) | 286 (With RNA operon) |
|----------------|------------------------|--------------------------|--------------------------|------------------------|
| 118 Pearson’s correlation | 0.89 | 0.78 | 0.12 |
| 184 Pearson’s correlation | 0.81 | 0.64 | 0.18 | 0.82 | 0.64 | 0.17 |
| 286 Pearson’s correlation | 0.86 | 0.62 | 0.14 | 0.88 | 0.75 | 0.12 | 0.80 | 0.64 | 0.20 |

© 2009 NIDO-KNAW
Journal compilation © 2009 Federation of European Microbiological Societies
Published by Blackwell Publishing Ltd.
Fig. 2. Tetranucleotide frequency distributions of fosmid clones of verrucomicrobial origin. (a) Fosmid 106; (b) fosmid 184; (c) fosmid 286 and (d) fosmid 118. Color indicates the tetranucleotide distribution relative to the full fragment lengths, with red representing typical (high correlation) and green atypical (low correlation) regions. The positions of ORFs in each of the fosmids are indicated by arrows above the diagrams. Numbers in arrows correspond to the CDS number of fosmids 118 (Table S1); 284 (Table S2); 184 (Table S3) and 106 (Table S4). The ribosomal operon illustrated in the figure in each fosmid was excluded from the analysis.
for clone 184 (Table S3) and 23 out of 26 for clone 106 (Table S4). For clone 184, the genome of *C. flavus* Ellin428 provided the best match for four of the 11 ORFs with verrucomicrobial best matches, whereas all 23 of the ORFs from clone 106 matched this genome. However, the 16S rRNA gene from clone 106 showed lower similarity to the Ellin428 strain compared with clone 118 (96% identity), which also showed several best BLAST hit matches with *Proteobacteria*, other verrucomicrobial strains and even *Planctomycetes* (Table S1).

Assignment of ORFs to COG functional categories revealed that all four clones encoded proteins involved in information storage/processing, metabolism and cellular processes/signalling, with clone 106 having the greatest proportion of COG hits belonging to metabolism or cellular processes and signalling. A large number of ORFs (between nine and 15 per clone) could not be classified into any COG category (Table S5).

The three fosmid clones affiliated with subdivision 2 (106, 118 and 286) were compared with the draft genome of Ellin428 by TBLASTX to examine patterns of gene synteny and homology. Homologous protein-encoding regions identified in these comparisons were no longer than 5 kb and contained a maximum of four ORFs (Fig. S1). Clone 184, affiliated with subgroup 1, contained no regions of conserved gene order compared with the other clones or available genomes. Thus, even though some of the recovered clones were very close to sequenced verrucomicrobial strains, based on 16S rRNA gene comparisons, large levels of divergence appear at the level of gene order and origin. Together, these observations highlight the fact that available sequence information represents but a small fraction of the depth of genomic information across the phylum as a whole.

Oligonucleotide frequencies within DNA sequences follow species-specific patterns (Nei & Gojobori, 1986; Sueoka, 1988; Karlin *et al.*, 1998; Bellgard *et al.*, 2001; Bentley & Parkhill, 2004). Such patterns carry an innate phylogenetic signal, and aberrant patterns of oligonucleotide usage may thus be indicative of regions recently acquired from foreign genetic sources (Pride *et al.*, 2003; Abe *et al.*, 2005). Upon analysis of tetranucleotide usage patterns (Tamames & Moya, 2008), clones 106 (Fig. 2a) and 184 (Fig. 2b) each appeared to contain one region of potential horizontal transfer. Several regions of aberrant tetranucleotide usage were observed in fosmid 286 (Fig. 2c), while no such regions were found in fosmid 118 (Fig. 2d). The ORFs identified within the regions of aberrant tetranucleotide usage typically exhibited only low levels of similarity to available database entries (29% identity, e-value = $7 \times 10^{-4}$ for clone 106; 43% identity, e-value = 0.002 for clone 184; and 32% identity, e-value = 7.9 for clone 286). Interestingly, with the exception of fosmid 106, these regions corresponded with ORFs that did not yield verrucomicrobial BLAST matches. The largest potential HGT region detected (in clone 286) contained several genes (UvrB/UvrC protein, putative peptidoglycan peptidase and ankyrin) flanked by insertion sequence elements, an organization typical of mobile DNA regions. It is noteworthy that the putative ankyrin repeat belongs to a functionally diverse group of eukaryotic proteins, for which HGT has been implicated in its transfer to prokaryotic genomes (Bork, 1993).

The cultivation-independent approach used in this study allowed for genomic information to be recovered specifically from the *Verrucomicrobia*, but a large number of clones was required for the identification of a relatively small number of verrucomicrobial clones. Despite the recent advances in shotgun metagenomic analysis, it remains clear that metagenomes derived from highly diverse soil environments will not yield large genomic scaffolds (Tringe *et al.*, 2005; Kowalchuk *et al.*, 2007). Thus, directed approaches, such as the one used here, will continue to be necessary for the recovery of extended regions of verrucomicrobial genomic information from soil. Although this study has provided some insight into the genome content and genomic properties of some members of the phylum *Verrucomicrobia*, defining the ecological and evolutionary properties from relatively small genomic regions (32 kb) remains problematic. Much larger sample sizes will be necessary in order to retrieve sufficient information to make ecological inferences, and novel high-throughput screening methods will need to be used to achieve this goal (Park *et al.*, 2008). Also, the approach utilized here is currently restricted to regions adjacent to rRNA operons. Now that verrucomicrobial genomic sequences are becoming available, detecting verrucomicrobial genomic fragments could be achieved by end-sequencing of large insert libraries followed by sequence comparison and binning. Particularly exciting are methods of single-cell sequencing and microcultivation (Abulencia *et al.*, 2006; Ingham *et al.*, 2007; Podar *et al.*, 2007), which could provide novel windows into the genomic diversity of this important bacterial phylum.

**Acknowledgements**

We thank Stephanie Malfatti and Mari Christensen from Joint Genome Institute (JGI) for help with DNA sequencing and assembling. Ellin514 sequence data were produced by the US Department of Energy JGI (http://www.jgi.doe.gov/). This work was funded by the BSIK programme 'Ecogenomics' (http://www.ecogenomics.nl/). This is Netherlands Institute of Ecology (NIOO-KNAW) publication 4606.

**References**

Abe T, Sugawara H, Kinouchi M, Kanaya S & Ikemura T (2005)
Novel phylogenetic studies of genomic sequence fragments
Verrucomicrobia: phylogenetic and metagenomic analysis

derived from uncultured microbe mixtures in environmental and clinical samples. DNA Res 12: 281–290.
Abulencia CB, Wyborski DL, Garcia JA et al. (2006) Environmental whole-genome amplification to access microbial populations in contaminated sediments. Appl Environ Microb 72: 3291–3301.
Altschul SF, Gish W, Miller W, Myers EW & Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215: 403–410.
Altschul SF, Madden TL, Schaffer AA, Zhang JJ, Zhang Z, Miller W & Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25: 3389–3402.
Bellgard M, Schibeci D, Trifonov E & Gojobori T (2001) Early detection of GIC differences in bacterial species inferred from the comparative analysis of the two completely sequenced Helicobacter pylori strains. J Mol Evol 53: 465–468.
Bentley SD & Parkhill J (2004) Comparative genomic structure of prokaryotes. Annu Rev Genet 38: 771–792.
Bezemer TM & van der Putten WH (2007) Ecology – diversity and stability in plant communities. Nature 446: E6–E7.
Bezemer TM, Lawson CS, Hedlund K, Edwards AR, Brook AJ, Igual JM, Mortimer SR & Van der Putten WH (2006) Plant species and functional group effects on abiotic and microbial soil properties and plant-soil feedback responses in two grasslands. J Ecol 94: 893–904.
Bork P (1993) Hundreds of Ankyrin-like repeats in functionally diverse proteins – mobile modules that cross phyla horizontally. Proteins 17: 363–374.
Buckley DH & Schmidt TM (2001) Environmental factors influencing the distribution of rRNA from Verrucomicrobia in soil. FEMS Microbiol Ecol 35: 105–112.
Carver TJ, Rutherford KM, Berrieman M, Rajandream MA, Barrell BG & Parkhill J (2005) ACT: the Artemis comparison tool. Bioinformatics 21: 3422–3423.
Chin KJ, Liesack W & Janssen PH (2001) Opitutus terrae gen. nov., sp. nov., to accommodate novel strains of the division Verrucomicrobia isolated from rice paddy soil. Int J Syst Evol Micr 51: 1965–1968.
Cole JR, Chai B, Farris RJ, Wang Q, Kulam SA, McGarrell DM, Garrity GM & Tiedje JM (2005) The Ribosomal Database Project (RDP-II): sequences and tools for high-throughput rRNA analysis. Nucleic Acids Res 33: D294–D296.
Derrien M, Vaughan EE, Plugge CM & de Vos WM (2004) Akkermansia muciniphila gen. nov., sp. nov., a human intestinal mucin-degrading bacterium. Int J Syst Evol Micr 54: 1469–1476.
Dunfield PF, Yuriev A, Senin P et al. (2007) Methane oxidation by an extremely acidophilic bacterium of the phylum Verrucomicrobia. Nature 450: 879–870.
Edwards U, Rogall T, Blocker H, Emde M & Bottger EC (1989) Isolation and direct complete nucleotide determination of entire genes – characterization of a gene coding for 16S-ribosomal RNA. Nucleic Acids Res 17: 7843–7853.
Felsenstein J (2005) PHYLP (Phylogeny Inference Package) Version 3.6. Distributed by the Author, Department of Genome Sciences, University of Washington, Seattle, WA.
Garrity GM & Holt JG (2001) The Road Map to the Manual, Vol. 1. pp. 119–166. Springer-Verlag, New York, NY.
Hou SB, Makarova KS, Saw JHW et al. (2008) Complete genome sequence of the extremely acidophilic methanotroph isolate V4, Methylocacidiphilum infernorum, a representative of the bacterial phylum Verrucomicrobia. Biol Direct 3: 26.
Huber T, Faulkner G & Hugenholtz P (2004) Bellerophon: a program to detect chimeric sequences in multiple sequence alignments. Bioinformatics 20: 2317–2319.
Hugenholtz P, Goebel BM & Pace NR (1998) Impact of culture-independent studies on the emerging phylogenetic view of bacterial diversity. J Bacteriol 180: 4765–4774.
Ingham CJ, Sprenkels A et al. (2007) The micro-Petri dish, a million-well growth chip for the culture and high-throughput screening of microorganisms. P Natl Acad Sci USA 104: 18217–18222.
Janssen PH (2006) Identifying the dominant soil bacterial taxa in libraries of 16S rRNA and 16S rRNA genes. Appl Environ Microb 72: 1719–1728.
Janssen PH, Schuhmann A, Morschel E & Rainey FA (1997) Novel anaerobic ultramicrobacteria belonging to the Verrucomicrobiales lineage of bacterial descent isolated by dilution culture from anoxic rice paddy soil. Appl Environ Microb 63: 1382–1388.
Janssen PH, Yates PS, Grinton BE, Taylor PM & Sait M (2002) Improved culturability of soil bacteria and isolation in pure culture of novel members of the divisions Acidobacteria, Actinobacteria, Proteobacteria, and Verrucomicrobia. Appl Environ Microb 68: 2391–2396.
Karlin S, Campbell AM & Mrazek J (1998) Comparative DNA analysis across diverse genomes. Annu Rev Genet 32: 185–225.
Kielak A, Pijl AS, van Veen JA & Kowalchuk GA (2008) Differences in vegetation composition and plant species identity lead to only minor changes in soil-borne microbial communities in a former arable field. FEMS Microbiol Ecol 63: 372–382.
Kielak A, Pijl AS, van Veen JA & Kowalchuk GA (2009) Phylogenetic diversity of Acidobacteria in a former agricultural soil. ISME J 3: 378–382.
Kowalchuk GA, Speksnijder AGCL, Zhang K, Goodman RM & van Veen JA (2007) Finding the needles in the meta-genome haystack. Microb Ecol 53: 475–485.
Ludwig W, Strunk O, Westram R et al. (2004) ARB: a software environment for sequence data. Nucleic Acids Res 32: 1363–1371.
Nei M & Gojobori T (1986) Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Mol Biol Evol 3: 418–426.
Nylander JAA (2004) MrModeltest v2. Program. Distributed by the Author, Evolutionary Biology Centre, Uppsala University, Uppsala, Sweden.
Park SJ, Kang CH et al. (2008) Metagenome microarray for screening of fosmid clones containing specific genes. FEMS Microbiol Lett 284: 28–34.

Pearce DA, van der Gast CJ, Lawley B & Ellis-Evans JC (2003) Bacterioplankton community diversity in a maritime Antarctic lake, determined by culture-dependent and culture-independent techniques. FEMS Microbiol Ecol 45: 59–70.

Podar M, Abulencia CB, Walcher M, Hutchison D, Zengler K, Garcia JA, Holland T, Cotton D, Hauser L & Keller M (2007) Targeted access to the genomes of low-abundance organisms in complex microbial communities. Appl Environ Microb 73: 3205–3214.

Pol A, Heijmans K, Harhangi HR, Tedesco D, Jetten MSM & den Camp H (2007) Methanotrophy below pH1 by a new Verrucomicrobia species. Nature 450: 874–870.

Pride DT, Meinersmann RJ, Wassenaa TM & Blaser MJ (2003) Evolutionary implications of microbial genome tetraneucleotide frequency biases. Genome Res 13: 145–158.

Pruesse E, Quast C, Knittel K, Fuchs B, Ludwig W, Peplies J & Glöckner FO (2007) SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. Nucleic Acids Res 35: 7188–7196.

Ronquist F & Huelsenbeck JP (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19: 1572–1573.

Sakai T, Ishizuka K & Kato I (2003) Isolation and characterization of a fucoidan-degrading marine bacterium. Mar Biotechnol 5: 409–416.

Sangwan P, Chen XL, Hugenholtz P & Janssen PH (2004) Chitonibacter flavus gen. nov., sp. nov., the first pure-culture representative of subdivision two, Spartobacteria classis nov., of the phylum Verrucomicrobia. Appl Environ Microb 70: 5875–5881.

Sangwan P, Kovac S, Davis KER, Sait M & Janssen PH (2005) Detection and cultivation of soil Verrucomicrobia. Appl Environ Microb 71: 8402–8410.

Scheuermayer M, Guldner TAM, Bringmann G & Hentschel U (2006) Rubritalea marina gen. nov., sp. nov., a novel marine member of the class Rubritaleaceae gen. nov., sp. nov., a fimbriated prosthecate bacterium. Syst Appl Microbiol 30: 152–158.

Schlesner H (1987) Verrucomicrobium spinosum gen. nov., sp. nov., a marine strain with a fimbriated prosthecate. Syst Appl Microbiol 10: 54–56.

Schlesner H, Jenkins C & Staley JT (2006) The phylum Verrucomicrobia: a phylogenetically heterogeneous bacterial group. Prokaryotes 7: 881–896.

Shinzato N, Muramatsu M, Matsui T & Watanabe Y (2005) Molecular phylogenetic diversity of the bacterial community in the gut of the termite Coptotermes formosanus. Biosci Biotech Bioch 69: 1145–1155.

Staley JT, Bont J & Jonge KD (1976) Prosthecobacter fusiformis gen. et sp., the fusiform caulobacter. Antonie van Leeuwenhoek 42: 333–342.

Stevenson BS, Eichorst SA, Wertz JT, Schmidt TM & Breznak JA (2004) New strategies for cultivation and detection of previously uncultured microbes. Appl Environ Microb 70: 4748–4755.

Sueoka N (1988) Directional mutation pressure and neutral molecular evolution. P Natl Acad Sci USA 85: 2653–2657.

Tamames J & Moya A (2008) Estimating the extent of horizontal gene transfer in metagenomic sequences. BMC Genomics 9: 136.

Teeling H, Waldmann J, Lombardot T, Bauer M & Glöckner FO (2004) TETRA: a web-service and a stand-alone program for the analysis and comparison of tetraneucleotide usage patterns in DNA sequences. BMC Bioinformatics 5: 163.

Tringe SG, von Mering C, Kobayashi A et al. (2005) Comparative metagenomics of microbial communities. Science 22: 554–557.

Vandekerckhove TTM, Willems A, Gillis M & Coomans A (2000) Occurrence of novel verrucomicrobial species, endosymbiotic and associated with parthenogenesis in Xiphinema americanum-group species (Nematoda, Longidoridae). Int J Syst Evol Micr 50: 2197–2205.

Van der Putten WH, Mortimer SR, Hedlund K et al. (2000) Plant species diversity as a driver of early succession in abandoned fields: a multi-site approach. Oecologia 124: 91–99.

Wang M, Ahnre S, Jeppsson B & Molin G (2005) Comparison of bacterial diversity along the human intestinal tract by direct cloning and sequencing of 16S rRNA genes. FEMS Microbiol Ecol 54: 219–231.

Yoon J, Oku N, Matsuda S, Kasai H & Yokota A (2007a) Pelagicoccus croceus sp. nov., a novel marine member of the family Puniciceccoccaceae within the phylum ‘Verrucomicrobia’ isolated from seagrass. Int J Syst Evol Micr 57: 2874–2880.

Yoon J, Yasumoto-Hirose M, Matsuo Y, Nozawa M, Matsuda S, Kasai H & Yokota A (2007b) Pelagicoccus mobilis gen. nov., sp. nov., Pelagicoccus albus sp. nov. and Pelagicoccus litoralis sp. nov., three novel members of subdivision 4 within the phylum Verrucomicrobia, isolated from seawater by in situ cultivation. Int J Syst Evol Micr 57: 1377–1385.

Zoetendal EG, Pluge CM, Akkermans ADL & de Vos WM (2003) Victivallis vadensis gen. nov., sp. nov., a sugar-fermenting anaerobe from human faeces. Int J Syst Evol Micr 53: 211–215.

Supporting Information

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

Fig. S1. Comparative genomic organization of fosmids in relation to Ellin428.

Table S1. Identified ribosomal genes and predicted ORF of the verrucomicrobial genomic fragment 118.

Table S2. Identified ribosomal genes and predicted ORF of the verrucomicrobial genomic fragment 286.

Table S3. Identified ribosomal genes and predicted ORF of the verrucomicrobial genomic fragment 184.

© 2009 NIDO-KNAW
Journal compilation © 2009 Federation of European Microbiological Societies
Published by Blackwell Publishing Ltd.
Table S4. Identified ribosomal genes and predicted ORF of the verrucomicrobial genomic fragment 106.

Table S5. Number of ORFs of Verrucomicrobia genomic fragments (106, 118, 184, and 286) assigned to different COG functional categories.

Please note: Wiley-Blackwell is not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.