Journal of Industrial Microbiology (1996) 17, 179–184

Genetic diversity of soil microorganisms assessed by analysis of *hsp70* (*dnaK*) sequences

WH Yap[1,3], X Li[1], TW Soong[1,3] and JE Davies[1,2]

[1]*West-East Centre for Microbial Diversity, BC Research Building, 3650 Wesbrook Mall, Vancouver, BC V6S 2L2;* [2]*Department of Microbiology, University of British Columbia, 300–6174 University Boulevard, Vancouver, BC V6T 1Z3, Canada;* [3]*Institute of Molecular and Cell Biology, National University of Singapore, Kent Ridge, Singapore 0511, Republic of Singapore*

The genetic diversity of a soil microbial community was assessed by analysis of cloned *hsp70* sequences. A clone library was generated by polymerase chain reaction-mediated amplification of a 650-base pair fragment of the *hsp70* gene, using DNA extracted from soil, without culturing the microorganisms. Fifty-five random clones were sequenced and their amino acid sequences deduced. Analysis of the amino acid sequence of the clones revealed the presence of signature sequences in common with known prokaryotic and lower eukaryotic HSP70 homologs. None of the 55 analyzed sequences were identical to each other or to a published sequence. These results confirm the presence of considerable genetic diversity within soil microbial communities, the major proportion of which remains uncharacterized.

Keywords: *hsp70*; soil microorganisms; soil DNA; microbial diversity

Introduction

The evaluation of diversity in naturally-occurring microbial communities has been limited by the non-culturable nature of a large percentage of the microbial biota [18]. However, the recent development of techniques to isolate DNA from the environment [10,17,19], together with the application of polymerase chain reaction (PCR) to amplify 16S rRNA genes from environmental DNA has provided methods to obtain information about the diversity of microbial communities [1,7,14,16]. The results of these studies carried out on terrestrial [1,14,16] as well as on a marine environment [7], confirm the view that culturable microbial species constitute a small proportion of the entire microbial population in an ecosystem.

Apart from the use of the ribosomal rRNA genes as tools for molecular ecological studies, other functional genes such as the nitrogenase gene and the hydrogenase gene have been used to evaluate genetic diversity in an environment [20,21,22]. Members of the 70-kDa heat shock family of proteins (HSP70) are ubiquitous in both eukaryotic and prokaryotic organisms and are one of the most conserved classes of proteins found in all species [2,8]. Because of their large size, ubiquity and high degree of sequence conservation, HSP70 homologs have been useful for studying evolutionary relationships [2,4,8]. In this report, we evaluate the genetic diversity of a soil environment by analyzing *hsp70* sequences derived by PCR amplification.

Correspondence: JE Davies, Department of Microbiology, University of British Columbia, 300–6174 University Boulevard, Vancouver, BC V6T 1Z3, Canada
Received 20 February 1996; accepted 12 July 1996

Materials and methods

DNA extraction

A soil sample was collected from a site at the Botanical Gardens of the University of British Columbia, Canada. Total DNA was extracted from the soil sample by a direct lysis procedure described by Barns *et al* [1]. The high molecular weight (>20 kb) extracted DNA was subsequently purified by chromatography on Sephadex G200 columns (Pharmacia, Uppsala, Sweden) [19].

Polymerase chain reaction amplification of hsp70 genes

Degenerate oligonucleotide primers for two conserved regions of the HSP70 family of proteins (amino acid sequence: ^{152}QATKDAG^{158} and ^{366}NPDEAVA^{372}) [6] were synthesized using a Beckman Oligo 1000 DNA synthesizer (Beckman Instruments, Fullerton, CA, USA). The sequences of the amplification primers were 5′ CAR GCI ACI AAR GAY GCI GG 3′ and 5′ GC IAC IGC YTC RTC IGG RTT 3′ (where I represents inosine, R represents A or G, Y represents T or C) [6]. PCR amplification was performed in a total volume of 100 μl containing 10–50 ng of template DNA, 20 mM Tris-HCl (pH 8.4), 50 mM KCl, 2 mM $MgCl_2$, 200 μM of each deoxyribonucleotide triphosphate, 25 pmol of each primer and 5.0 units of *Taq* polymerase (BRL Life Technologies, Gaithersburg, MD, USA). The thermal cycling conditions were as follows: denaturation at 93°C for 1.0 min, annealing at 58°C for 30 s and extension at 72°C for 1.0 min, repeated for a total of 30 cycles. PCR products were electrophoresed in 1% agarose gels and stained with ethidium bromide using standard techniques. The expected size of the PCR product was 650 base pairs. No amplification was observed for negative controls (water used instead of template DNA). The band containing the PCR product was excised from the gel and purified using a Qiaex gel extraction kit (Diagen

Table 1 Phylogenetic parameters based on the comparisons of partial amino acid (lower left triangle) and related DNA (upper right triangle) sequences of representative clones. Distance matrices were generated using the PHYLIP phylogeny inference package

	H12	H35	H54	H33	H13	H5	H22	H10	H43	H53	H51	H24	H28	H36	H3	H25	H47	H30	H6	H14
H12	–	0.5730	0.5653	0.4663	0.8091	0.7596	0.6815	0.6606	0.7903	0.7078	0.7040	0.6262	0.6193	0.6131	0.6650	0.7411	0.8340	0.7726	0.6991	0.6748
H35	0.44494	–	0.4773	0.4852	0.8995	0.7292	0.7812	0.6874	0.8313	0.8119	0.6968	0.7630	0.6983	0.6511	0.6642	0.6945	0.8314	0.8652	0.7972	0.7444
H54	0.43414	0.49663	–	0.5493	0.9483	0.8339	0.7489	0.6636	0.7738	0.7879	0.7604	0.7333	0.7032	0.6530	0.6730	0.7712	0.8809	0.8980	0.8265	0.7237
H33	0.37043	0.34298	0.41885	–	0.7849	0.8545	0.7668	0.6798	0.7353	0.7421	0.7646	0.6721	0.6990	0.6735	0.6926	0.8443	0.8042	0.8214	0.7382	0.6868
H13	0.84697	0.97882	0.96797	0.90004	–	0.6108	0.8070	0.6893	0.8094	0.7857	0.9426	0.7189	0.7339	0.8666	0.7235	0.6059	0.3915	0.2939	0.6757	0.5937
H5	0.77491	0.88292	0.94595	0.84208	0.30950	–	0.5442	0.4852	0.5247	0.5702	0.5066	0.6021	0.4526	0.5695	0.5078	0.3069	0.5343	0.5164	0.5029	0.5124
H22	0.85181	0.94476	1.02844	0.83229	0.58855	0.54088	–	0.5291	0.6297	0.6019	0.5811	0.5648	0.4952	0.5993	0.5655	0.5951	0.8094	0.8044	0.6257	0.5723
H10	0.67744	0.84922	0.88634	0.73241	0.53537	0.52783	0.60177	–	0.4601	0.5142	0.4342	0.3726	0.3456	0.4761	0.3992	0.5089	0.6360	0.6073	0.4774	0.4730
H43	0.83890	1.01555	1.02269	0.85227	0.67531	0.55077	0.69273	0.54788	–	0.4992	0.5081	0.5242	0.4728	0.5847	0.4997	0.5807	0.7475	0.7160	0.6113	0.5626
H53	0.78865	1.02017	1.03494	0.84038	0.64615	0.50018	0.65521	0.46910	0.48537	–	0.5641	0.4949	0.4887	0.5167	0.5170	0.5448	0.7429	0.7439	0.5344	0.5065
H51	0.79989	0.93403	1.02991	0.83550	0.67983	0.59728	0.67926	0.48084	0.53614	0.61866	–	0.5270	0.4249	0.4980	0.4455	0.5342	0.8191	0.8697	0.5605	0.5654
H24	0.72632	0.88370	0.88314	0.76872	0.51592	0.47995	0.56573	0.25600	0.53952	0.42637	0.44159	–	0.3571	0.5203	0.5187	0.5972	0.6559	0.6305	0.4960	0.5203
H28	0.69202	0.82073	0.86776	0.76375	0.53042	0.47743	0.57359	0.33310	0.54247	0.44015	0.49623	0.20077	–	0.5108	0.4010	0.4969	0.6775	0.6676	0.5068	0.5117
H36	0.68977	0.76852	0.82098	0.70825	0.62353	0.52899	0.56499	0.49672	0.55636	0.60124	0.53126	0.44167	0.47647	–	0.4267	0.5082	0.7542	0.7973	0.5871	0.5504
H3	0.77879	0.79159	0.86915	0.71533	0.58797	0.49413	0.63472	0.39712	0.63524	0.52547	0.50947	0.41292	0.40405	0.52840	–	0.4748	0.7131	0.6703	0.5393	0.5050
H25	0.74277	0.91545	0.95751	0.77455	0.34160	0.24396	0.54275	0.51785	0.54027	0.49882	0.53661	0.45650	0.49495	0.57516	0.47772	–	0.5531	0.5144	0.4872	0.4341
H47	0.84636	0.93794	0.96783	0.79613	0.31038	0.24106	0.55277	0.48556	0.60033	0.59076	0.57836	0.45433	0.44850	0.59139	0.49410	0.28448	–	0.3122	0.6678	0.6260
H30	0.75355	0.89468	0.89280	0.77035	0.25760	0.22911	0.49500	0.46954	0.52202	0.47622	0.56963	0.44594	0.43596	0.58008	0.42569	0.22841	0.22764	–	0.7192	0.6019
H6	0.76351	0.99357	1.06819	0.88080	0.53713	0.45348	0.63414	0.53498	0.63567	0.53378	0.54663	0.55143	0.51289	0.60277	0.53209	0.47522	0.51860	0.46383	–	0.4260
H14	0.76887	0.91073	0.97039	0.73397	0.46612	0.49085	0.50494	0.45774	0.58130	0.50049	0.51827	0.42888	0.42550	0.53351	0.55350	0.43582	0.48966	0.42365	0.40599	–

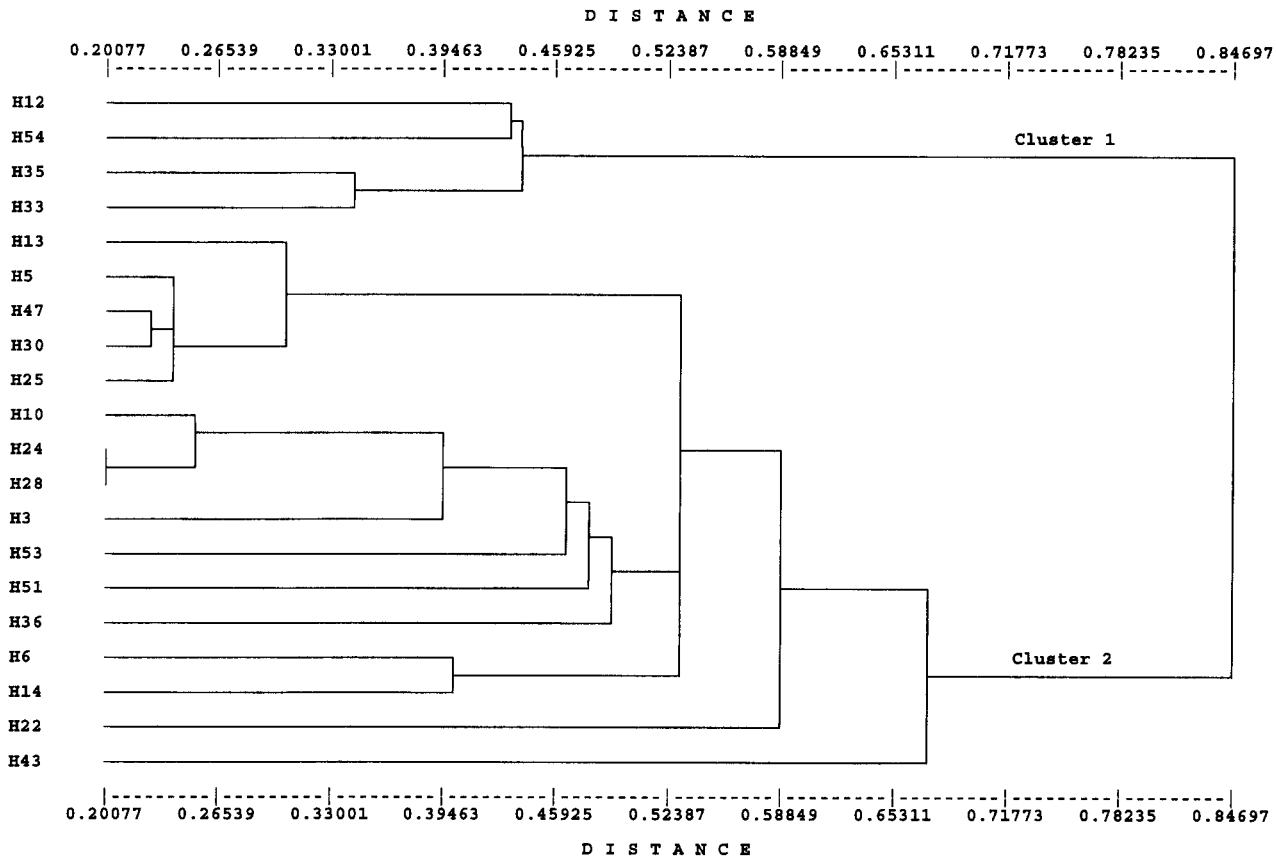

Figure 1 Detailed dendrogram for HSP70 clones. The distance matrix based on partial amino acid sequences is presented in Table 1.

Gmbh, Hilden, Germany). The purified DNA was ligated to pCR™II (Invitrogen, San Diego, CA, USA) to generate a *hsp70* library using *E. coli* INVαF′ competent cells (Invitrogen).

Sequencing and phylogenetic analysis

Fifty-five cloned *hsp70* sequences (referred to as HSP70 clones H1 to H55) were chosen at random from the library and sequences in both directions using the *Taq* Dye Primer Cycle Sequencing kit and the Applied Biosystem DNA sequencer model 373 (Applied Biosystem, Foster City, CA, USA). The sequencing primers used included the universal M13 (–20) forward primer, the M13 reverse primer and primers designed from the sequence data obtained; DNA sequences were translated into partial amino acid sequences. Partial amino acid sequences of known HSP70 homologs were obtained from Genbank (GB), Swiss-Prot (SP) and PIR (PI) databases. Multiple alignment of partial amino acid sequences and related DNA sequences was carried out in a pairwise manner using a software program package from Geneworks (Intelligenetics Inc, CA, USA), with further manual adjustments. Steps were taken to ensure that the highly conserved regions were properly aligned in all sequences and the number of gaps was kept to a minimum. Phylogenetic analysis was performed on the entire 650-base pair fragment, with omission of the primer sequences used to amplify the *hsp70* genes. Phylogenetic distances based on DNA and protein sequences were calculated according to the algorithms of Jukes and Cantor [12] and Dayhoff *et al* [3], respectively, using the PHYLIP phylogeny inference package version 3.5 [5]. The distance matrix was analyzed using the Taxan program (University of Maryland, Baltimore, MD, USA) with unweighted, average linkage clustering to generate phylogenetic dendrograms [13].

Nucleotide sequence accession numbers

Sequences from the clones reported have been submitted to Genbank under accession numbers U49131 to U49150.

Results

Analysis of cloned hsp70 sequences

A total of 55 randomly selected HSP70 clones (H1 to H55) from a PCR-generated clone library of a soil DNA preparation were sequenced. None of the sequences analyzed were identical to each other or to a published sequence. A distance phylogenetic dendrogram of deduced partial amino acid sequence of the clones was obtained using the Taxan program. At a phylogenetic distance of 0.20, the clones could be divided into 20 different groups (data not shown). A representative clone of each group was then selected and subjected to further phylogenetic analysis. Table 1 shows the DNA and protein distance matrix of the representative clones. Comparison of the sequence data from various representative clones shows that phylogenetic distance based on amino acid sequence ranges from 0.20 to 0.85 (Figure 1 and Table 1), and distance based on nucleic acid sequences

from 0.30 to 0.95 (Table 1). A phylogenetic dendrogram constructed on the basis of the protein distance matrix is shown in Figure 1. At a phylogenetic distance of 0.80, the clones could be divided into two major clusters (Clusters 1 and 2; Figure 1). It was also observed that representative clones in Cluster 2 could be further divided into a number of subclusters (Figure 1).

Amino acid sequence comparison and signature sequences

Based on the global alignment of known HSP70 sequences, a number of signature sequences have been identified that provide clear distinctions between eukaryotic cytosolic homologs and prokaryotic homologs and between Gram-positive and Gram-negative bacteria [8,9]. In addition, Falah and Gupta [4] also identified a number of signature sequences that distinguish mitochondrial homologs and eubacteria from chloroplast homologs and cyanobacteria, as well as a four-amino acid insert specifically present in β- and γ-subdivisions of proteobacteria.

A comparison of partial amino acid sequences of representative HSP70 clones and a number of known eubacterial and eukaryotic homologs was carried out. Figure 2 shows the multiple alignment of the partial amino acid sequences. As observed, the sequences showed extensive sequence similarity throughout the entire length (Figure 2). The representative clones also showed the presence of a number of signature sequences (Figure 2). Out of 20 clones examined, four clones (H12, H54, H35 and H33) had signatures typical of eukaryotic cytosolic HSP70 homologs (Figure 2). These clones were found in Cluster 1 of the phylogenetic dendrogram (Figure 1). The remaining 16 clones had signatures typical of eukaryotic organellar and prokaryotic homologs (Figure 2) and constituted Cluster 2 of the dendrogram (Figure 1). In addition, clone H36 had a number of signatures in common with chloroplast homologs and cyanobacteria and five other clones (H25, H30, H47, H13 and H5) had a four-amino acid insert characteristic of β- and γ-subdivisions of proteobacteria (Figure 2). Interestingly, clones H25, H30, H47, H13 and H5 formed a subcluster within Cluster 2 (Figure 1). Apart from the 20 representative clones described here, the remaining 35 HSP70 clones of the 55 randomly selected clones had signatures typical of eukaryotic organellar and prokaryotic homologs (data not shown). Among these 35 clones, four clones also had signatures typical of chloroplast homologs

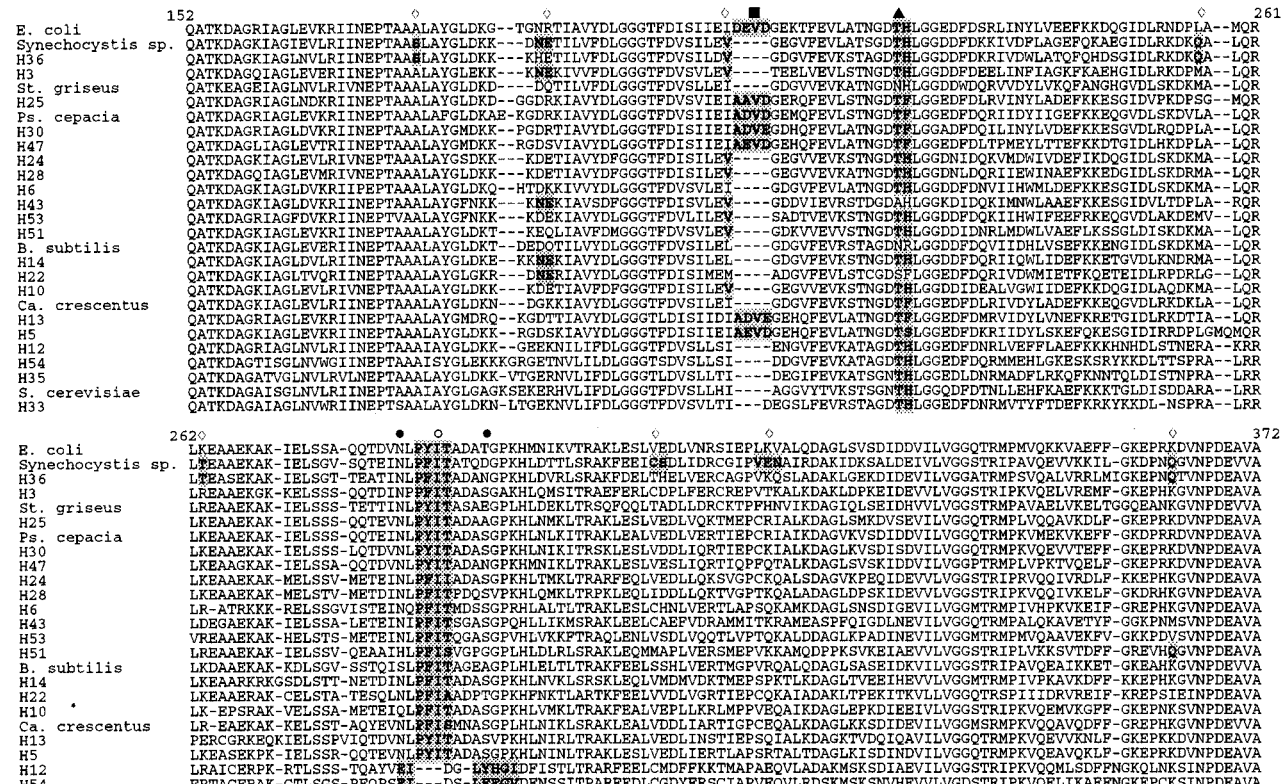

Figure 2 Alignment of representative cloned HSP70 sequences with known eukaryotic and prokaryotic homolog sequences. The numbers at the top refer to positions in the *E. coli* sequence. Signature sequences are presented as boldface letters on a shaded background. The symbols describe different types of signature sequences. ◇, Signatures shared by chloroplast homologs and cyanobacteria; ■, the four-amino acid insert shared by homologs of β- and γ-subdivisions of proteobacteria; ▲, signatures shared by eukaryotic homologs and Gram-negative bacteria; ●, signatures shared by eukaryotic cytosolic homologs; ○, signatures shared by eukaryotic organellar and prokaryotic homologs. Dashes (–) indicated gaps or deletions in the sequence. Partial amino acid sequences of known HSP70 homologs were obtained from Genbank (GB), Swiss-Prot (SP) and PIR (PI) databases. Species (accession number): B. subtilis, *Bacillus subtilis* (SP P17820), Ca. crescentus, *Caulobacter crescentus* (SP P20442), E. coli, *Escherichia coli* (SP P04475), Ps. cepacia, *Pseudomonas cepacia* (GB L36603), S. cerevisiae, *Saccharomyces cerevisiae* (PI S20149), St. griseus, *Streptomyces griseus* (GB D14499), Synechocystis sp, *Synechocystis* sp (SP P22358).

and cyanobacteria and 20 clones had a four-amino acid insert characteristic of β- and γ-subdivisions of proteobacteria (data not shown). The presence of signature sequences within the HSP70 clones suggests a relationship between the clones and known homologs and provides information regarding identity of the clones.

Discussion

Polymerase chain reaction-mediated amplification of 16S rRNA genes obtained directly from DNA extracted from the environment, followed by gene cloning, sequencing and data comparison has provided a way of assessing the composition of a microbial community [1,7,14,16]. We have employed a similar approach to determine the genetic diversity of a soil community by analyzing sequences of a highly conserved gene, *hsp70*, derived by PCR amplification. None of the sequences of 55 randomly selected HSP70 clones were identical to each other or to a published sequence, suggesting the presence of a variety of novel HSP70 types within the soil community. The observed variation cannot be accounted for by errors in PCR in view of the location of sequence changes and the maximum error rate reported for *Taq* polymerase (1 in 400 per 30 amplification cycles) [15].

The presence of a distribution of signature sequences in the HSP70 clones (Figure 2) provides information about the identity of novel HSP70 types. On a broad basis, the 55 HSP70 clones could be divided into two major groups, one comprising eukaryotic cytosolic homologs and the other eukaryotic organellar and prokaryotic homologs, constituting 7% (four clones) and 93% (51 clones) of the clones, respectively. Of the latter group 49% (25 clones) of the clones had signatures distinctive of the β- and γ-proteobacteria group and 10% (five clones) had signatures characteristic of the cyanobacteria and chloroplast group suggesting a close relationship of these clones to the groups mentioned. However, although detailed phylogenetic analysis of the clones and known homologs was carried out, the limitation in length of the partial amino acid sequences did not allow for precise phylogenetic assignment.

Using a strategy involving PCR amplification of 16S rRNA genes, Liesack and Stackebrandt [14] and Stackebrandt *et al* [16], demonstrated the existence of a variety of novel groups of the domain *Bacteria* within a subtropical terrestrial soil sample. In a similar manner, we demonstrated the existence of a number of novel HSP70 types within a soil community. Taken together, these results further support the notion that culturable microorganisms constitute only a small proportion of a microbial community in a terrestrial environment. A corollary of this finding is that such communities possess much unrecognised microbial metabolic diversity that presents an opportunity in terms of industrial and pharmaceutical [11] applications and which can be made accessible by using the molecular techniques outlined here.

In summary, through PCR amplification of *hsp70* genes from DNA extracted from a soil sample, we demonstrate the existence of a large proportion of uncharacterized genetic diversity in a soil community. Efforts are now underway to use the PCR-amplified *hsp70* genes as probes for the screening of libraries constructed from extracted soil DNA. The isolation of full length clones will enable a more precise phylogenetic assignment of novel HSP70 types.

Acknowledgements

We gratefully acknowledge M Berbee and D Carmean for helpful discussions and V Miao and J McDermott for critical reading of the manuscript.

References

1 Barns SM, RE Fundyga, MW Jefferies and NR Pace. 1994. Remarkable archaeal diversity detected in a Yellowstone National Park hot spring environment. Proc Natl Acad Sci 91: 1609–1613.
2 Boorstein WR, T Zeigelhoffer and EA Craig. 1994. Molecular evolution of the HSP70 multigene family. J Mol Evol 38: 1–17.
3 Dayhoff MO, RM Schwartz and BC Orcutt. A model of evolutionary change in proteins. In: Atlas of Protein Sequence and Structure (Dayhoff MO, ed), vol 5, suppl 3, pp 345–352, National Biomedical Research Foundation, Washington DC.
4 Falah M and RS Gupta. 1994. Cloning of the *hsp70 (dnaK)* genes from *Rhizobium meliloti* and *Pseudomonas cepacia*: phylogenetic analyses of mitochondrial origin based on a highly conserved protein sequence. J Bacteriol 176: 7748–7753.
5 Felsenstein J. 1988. Phylogenies from molecular sequences: inference and reliability. Annu Rev Genet 22: 525–565.
6 Galley KA, B Singh and RS Gupta. 1992. Cloning of HSP70 (*dnaK*) gene from *Clostridium perfringens* using a general polymerase chain reaction based approach. Biochim Biophys Acta 1130: 203–208.
7 Giovannoni SJ, TB Britischgi, CL Moyer and KG Field. 1990. Genetic diversity in Sargasso Sea bacterioplankton. Nature 345: 60–63.
8 Gupta RS and GB Golding. 1993. Evolution of HSP70 gene and its implications regarding relationships between archaebacteria, eubacteria and eukaryotes. J Mol Evol 37: 573–582.
9 Gupta RS and B Singh. 1994. Phylogenetic analysis of 70 kD heat shock protein sequences suggests a chimeric origin for the eukaryotic cell nucleus Curr Biol 4: 1104–1114.
10 Holben WE, JK Jansson, BK Chelm and JM Tiedje. 1988. DNA probe method for the detection of specific microoganisms in the soil bacterial community. Appl Environ Microbiol 54: 703–711.
11 Jindal S. 1996. Heat shock proteins: applications in health and disease. Trends Biotechnol 14: 17–20.
12 Jukes TH and CR Cantor. 1969. Evolution of protein molecules. In: Mammalian Protein Metabolism (Munro HH, ed), vol 3, pp 21–132, Academic Press, New York.
13 Li X and SH DeBoer. 1995. Comparison of 16S ribosomal RNA genes in *Clavibacter michiganensis* subspecies with other coryneform bacteria. Can J Microbiol 41: 925–929.
14 Liesack W and E Stackebrandt. 1992. Occurrence of novel groups of the domain *Bacteria* as revealed by analysis of genetic material isolated from an Australian terrestrial environment. J Bacteriol 174: 5072–5078.
15 Saiki RK, DH Gelfand, S Stoffel, SJ Scharf, R Higuchi, GT Horn, KB Mullis and HA Ehrlich. 1988. Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. Science 239: 487–491.
16 Stakebrandt E, W Liesack and GM Goebel. 1993. Bacterial diversity in a soil sample from a subtropical Australian environment as determined by 16S rDNA analysis. FASEB J 7: 232–236.
17 Steffan RJ, J Goksyor, AK Bej and RM Atlas. 1988. Recovery of DNA from soils and sediment. Appl Environ Microbiol 54: 2908–2915.
18 Torsvik V, J Goksoyr and FL Daae. 1990. High diversity in DNA of soil bacteria. Appl Environ Microbiol 56: 782–787.
19 Tsai Y-L and BH Olson. 1992. Rapid method for separation of bacterial DNA from humic substances in sediments for polymerase chain reaction. Appl Environ Microbiol 58: 2292–2295.
20 Ueda T, Y Suga, N Yahiro and T Matsuguchi. 1995. Remarkable N_2-fixing bacterial diversity detected in rice roots by molecular evolutionary analysis of *nifH* gene sequences. Appl Environ Microbiol 177: 1414–1417.

21 Wawer C and G Muyzer. 1995. Genetic diversity of *Desulfovibrio* spp in environmental samples analyzed by denaturing gradient gel electrophoresis of [NiFe] hydrogenase gene fragments. Appl Environ Microbiol 61: 2203–2210.

22 Zehr JP, M Mellon, S Braun, W Litaker, T Steppe and HW Paerl. 1995. Diversity of heterotrophic nitrogen fixation genes in a marine cyanobacterial mat. Appl Environ Microbiol 61: 2527–2532.