Pakistan J. Zool., vol. 42(1), pp. 33-36, 2010.

Metagenomic Accessing of Genes From Environmental Clone Library From an Oxidation Tank of a Waste Treatment Plant Using 59-be Conserved Sequence Specific Primers

Muhammad Afzal Ghauri*, William Duncan Grant and Shaun Heaphy

Industrial Biotechnology Division, National Institute for Biotechnology and Genetic Engineering, P.O. Box 577, Jhang Road, Faisalabad, Pakistan (MAG) and Department of Infection, Immunity & Inflammation, Maurice Shock Building, University of Leicester, P.O. Box 138, Leicester, LE1 9HN, United Kingdom (WDG, SH)

Abstract.- Environmental samples were collected from oxidation tank of a waste treatment plant in Leicester, United Kingdom. Genomic DNA was recovered from the environmental samples and amplified by PCR using 59-be conserved sequence specific primers HS286 and HS287. Different 59-be conserved sequence containing genes were recovered from the environmental DNA clone library. Most of the recovered genes were novel with no identifiable ORF homologues in the databases. This technique demonstrates the usefulness of a culture-independent gene mining approach as an alternative to traditional gene recovery technologies.

Key Words: Environmental samples, 59-be conserved sequence, gene mining,

INTRODUCTION

The recognition that most of the micro-organisms in the environment cannot be cultured by standard methods stimulated the development of the area of "Metagenomics", specifically leading to the development and use of new molecular methods for detecting taxon specific genes without the need for culture (Lane, 1991). "Metagenomics" describes the functional and sequence-based analysis of the collective microbial genomes contained in an environmental sample. Two types of analysis have been used to obtain information from metagenomic libraries: a function-driven approach, in which metagenomic libraries are initially screened for an expressed trait, and a sequence-driven approach, in which libraries are initially screened for particular DNA sequences.

Vertical and horizontal gene transfers have greatly affected the composition of bacterial gene pools. It is well documented that a significant proportion of the bacterial gene pool is extrachromosomal in origin and can vary between different species and strains of the same species. Moreover, most of bacterial genes have been acquired by horizontal gene transfer (Ochman *et al*., 2000).

Integrons have been reported to be quite widespread amongst genomes of bacteria. They have been emerged as very efficient gene capturing and expressing systems. Integrase enzyme being coded by integrons can integrate gene cassettes into the attachment site *att*I (Bennett, 1999). Gene cassettes are promoterless and can exist as free, circular, non-replicating DNA entities. Integron associated gene cassettes make use of promoter of integron for their expression and are mobile in nature (Hall and Collis, 1995). Another very striking feature of such gene cassettes is that they contain a very specialised sequence called 59 base element, which functions as a specific recombination site for the entry of additional genes (Hall *et al*., 1991). Bacterial genera now known to contain integrons are highly diversified and are extending as more investigations are being made (Ghauri *et al*., 2006).

The current study reports the screening of culture-independent recovery of 59-be conserved sequence containing (integron-linked) genes from the oxidation tank of a waste treatment plant in Leicester, United Kingdom. Moreover, the usefulness of integron gene capture technique as a tool for gene mining and as an alternative to traditional gene recovery methodologies is highlighted.

* Corresponding author: maghauri@nibge.org
0030-9923/2010/0001-0033 $ 8.00/0

MATERIALS AND METHODS

Acquisition of environmental samples

Environmental samples were collected from oxidation tank of a waste treatment plant in Leicester, UK in sterile screw capped bottles and were transferred into refrigerator till further use.

Isolation of DNA

Genomic DNA was isolated from 10ml of environmental sample using the GenomicPrep™ Cells and Tissue DNA Isolation Kit (Amersham Pharmacia Biotech.) and stored at -20°C till further use.

Recovery, cloning and sequencing of genes

Two primers to 59 base element conserved sequences of integrons were HS286 (5` GGGATCCTC (GC)GCT(GT)GA(GA)CGA(AC) TTGTTAG(GCA)C 3`) and HS287 (5` GGGATCCGC(GC)GCT(GT)A(AGCT)CTC (GCA)(GA)(GA)CGTTAG(GC)C 3`) (Stokes *et al*., 2001). The underlined sequence encodes a *Bam*HI site to help cloning. Reaction mixtures consisted of approximately 5ng of template DNA, 100 pmol of each of the primer, 200nM deoxynucleoside triphosphate (dNTP) mix, 2mM $MgCl_2$, and 1U of *Taq* DNA polymerase. The PCR was carried out by standard techniques with the following cycling programme: 94°C for 3 minutes for 1 cycle, 94°C for 30 seconds, 55°C for 30 seconds, 72°C for 2 minutes and 30 seconds for 35 cycles, and 72°C for 5 minutes for 1 cycle. Cloning and sequencing of PCR products was carried out as reported earlier by us (Ghauri *et al*., 2003). A brief description is that PCR products were cloned into the pGEM-T Easy vector (Promega) followed by transfer of ligation mixture into *Escherichia coli* JM109 competent cells (Promega). Plasmids containing inserts were isolated from 3ml overnight cultures by using the Wizard Plus Miniprep DNA purification system (Promega). Selection of clones was carried out after digesting plasmid DNA with *Eco*RI restriction enzyme. Dried plasmid DNA samples, approximately containing 750ng of DNA, were sequenced with *M13* forward and reverse primers.

Sequence retrieval and analysis

The gene sequences were analysed by comparing with others in the GenBank databases using the NCBI BLAST programme (www.ncbi.nlm.nih.gov). Homologues of nucleotide sequences were found using BLASTN and homologues to inferred proteins were found using BLASTX. Open reading frames (ORFs) were identified using the ORF Finder programme. ORF cognitor compares ORFs with the clusters of orthologous groups (COG) protein database.

Nucleotide sequence accession numbers

Nucleotide sequences from clones were deposited with GenBank of National Center for Biotechnology Information (www.ncbi.nlm.nih.gov) under the accession numbers AY221026, AY221027 and AY221028.

RESULTS

Description of environmental samples

A sample from the oxidation tank of treatment plant of general sewage in Leicester, United Kingdom was collected in a sterile plastic bottle, which was immediately processed to avoid any change in the physico-chemical conditions. The pH of the sample was 6.9. The colour of the sample was black/brown and smell of methane gas was indicative of anaerobic microbial activity in such environments.

Metagenomic recovery of 59-be conserved sequence containing genes

Genomic DNA from sludge sample was amplified through polymerase chain reaction (PCR) with 59-be conserved sequence specific primers HS286 and HS287 yielding fragments ranging from about 300bp to 600bp.

PCR products were cloned into pGEM T-Easy vector. Thirty six clones were picked at random from hundreds of transformed clones. On the basis of restriction fragment length polymorphism analyses following restriction with *EcoRI* three different clones were identified. DNA sequences were analysed using GenBank databases by performing BlastN, BlastX and ORF finder analysis. BlastN hits were due to the presence of the

59 nucleotide recombination sequence in the gene cassettes found in all three amplicons with E value of 10^{-6}. Amplicons S-1, S-2 and S-3 showed BlastX top hits to Rachiplusia ou multiple nucleopolyhedrovirus (E value 0.031), Zn-ribbon protein (E value $2x10^{-4}$) and Ribosomal protein L6 [*Trichodesmium erythraeum* IMS101](E value 3.1), respectively (Table I).

Table I.- Sequence analysis of genes recovered from waste water treatment plant environmental metagenomic DNA sample.

Clone code	Insert (bp)	Top Database Hit (BlastX)
S-1 (AY221026)	360	Unknown [Rachiplusia ou multiple nucleopolyhedrovirus]. (NP_703087) E 10^{-3}
S-2 (AY221027)	558	Zn-ribbon protein [Pelobacter carbinolicus DSM 2380] (ABA90161) E 10^{-3}
S-3 (AY221028)	580	Ribosomal protein L6 [*Trichodesmium erythraeum* IMS101](ZP_00675520) E 2.8

Highly conserved region of 59-be with 7 base pair core site GTTAGGC with inverted repeats CAATCCG was observed in all sequences. Inverted repeats in all sequences were found near 3/ ends. Amplicon S-1 showed imperfect inverted repeats whereas amplicons S-2 and S-3 showed perfect inverted repeats.

Amplicons S-1, S-2 and S-3 displayed diverse profile of ORFs when screened by ORF finder programme of NCBI search tools. Amplicon S-1 had three ORFs, among which two ORFs had complete coding regions with standard start and stop codons. Amplicon S-2 showed a total of four ORFs with three having complete coding regions. Maximum number of ORFs was found in amplicon S-3. Out of five ORFs four had complete coding regions with standard start and stop codons.

Submission of sequences in GenBank

Nucleotide sequences of amplicons S-1, S-2 and S-3 were deposited with GenBank of National Center for Biotechnology Information (www.ncbi.nlm.nih.gov) and accession numbers obtained were AY221026, AY221027 and AY221028, respectively.

DISCUSSION

Gene mining activities involving culture-dependent and/or culture-independent methodologies are becoming very popular day by day in biotechnology and related disciplines because of the flexibility and success rate of such technologies. In many studies, integrons have been shown to work as efficient gene acquisition and expression systems and proving themselves a good alternate option to traditional methods for gene acquisition from metagenomes. The current study has further strengthened the position of integrons as fishing tools by recording the recovery of novel integron-linked genes from metagenomic DNA extracted from oxidation tank of a waste water treatment plant. Moreover, this study has further extended the range of environments known to host integrons.

It is likely that smallness of amplicons and primer set may introduce bias due the competitive nature of PCR technique. The variable part of 59 base element sequence warrants designing of new primers to encompass sequence diversity in this family of recombination sites. Requirement for designing new primer sets is further enhanced by the fact that as new investigations into the biological resources for screening such recombination sites are made novel sequence diversity may emerge. Rowe-Magnus *et al*. (2001) have suggested that 59-base element sites comprise sequence homology groups related to their origin in chromosomal integrons. An element of subjectivity and bias exist in PCR technology which should be taken in to account while analysing such results.

Natural environments can host diverse life forms and exchange of genetic information among such groups is highly likely. It is known that as much as 24% *Escherichia coli* genome has been acquired through lateral transfer and recombination of genetic information (Lawrence and Ochman, 2002). Therefore, the presence of diverse types of gene cassettes in metagenomic DNA and various

possible genetic crosses are highly expected. Although, the localisation and flow of a particular gene or a group of genes cannot be predicted in such cases. Besides gene acquisition and gene expression capability, integrons and associated genes could also be used as bioprobes for monitoring the genetic makeup and exchange between taxa in environments of a varied nature.

Genes recovered from uncultured samples showed significant homologies in databases (Table 1). However, in one of our previous studies we found that integron linked genes from cultured organisms had higher identity to database sequences as compared to uncultured sample (Ghauri *et al*., 2006). The presumptive bias of sequence databases towards cultured organisms may not stand universal when viewed in the light of current findings. Results of Blast N and Blast X showed different top hits representing the diversity of these amplicons. We may say that the origin of these amplicons from waste water treatment plant could be a good reservoir to hunt integron related metagenomes of microbes thriving in such environments. However, All ORFs didn't show any match with COG data base which indicates the novelty of these gene cassettes.

This study further strengthens the position of integron prospecting genes of known and unknown functions from unlimited metagenomic reservoir. Integron gene capture technology makes gene boundaries and location in a sequence fragment highly identifiable. Moreover, the orientation of open reading frames can be predicted and genes can be manipulated by site specific recombination. As a result, we can access huge unexplored metagenomic resource without having pre-hand knowledge of gene sequence. Moreover, integron-driven gene acquisition is likely to be an important factor in the process of horizontal gene transfer and ultimately in the process of evolution of bacterial genomes and proteomes.

ACKNOWLEDGEMENT

M. Afzal Ghauri gratefully acknowledges the provision of a postdoctoral fellowship by the Ministry of Science and Technology, Government of Pakistan.

REFERENCES

BENNETT, P.M., 1999. Integrons and gene cassettes: a genetic construction kit for bacteria. *J. Antimicrob. Chemother*., **43**: 1-4.

GHAURI, M.A., KHALID, A.M., GRANT, S., HEAPHY, S. AND GRANT, W.D., 2003. Phylogenetic analysis of different isolates of sulfobacillus spp. recovered from uranium rich environments and recovery of genes using integron specific primers. *Extremophiles,* **7**: 341-345.

GHAURI, M.A., KHALID, A.M., GRANT, S., GRANT, W.D.AND HEAPHY, S., 2006. Phylogenetic analysis of bacterial isolates from man-made high pH, high salt environments and identification of gene-cassette associated ORFs. *Curr. Microbiol.,* **52**: 487-492.

HALL, R.M. AND COLLIS, C.M., 1995. Mobile gene cassettes and integrons: capture and spread of genes by site specific recombination. *Mol. Microbiol*., **15**: 593-600.

HALL, R.M., BROOKES, D.E. AND STOKES, H.W., 1991. Site-specific insertion of genes into integrons: role of the 59 base-element and determination of the recombination cross-over point. *Mol. Microbiol.,* **5**: 1941-59.

LANE, D.J., 1991. 16S/23S rRNA sequencing. In: *Nucleic acid techniques in bacterial systematics (*eds. E. Stckebrandt and M. Goodfellow), John Wiley and Sons, New York, pp. 115-148.

LAWRENCE, J.G. AND OCHMAN, H., 2002. Reconciling the many faces of lateral gene transfer. *Trends Microbiol*., **10**: 1-4.

OCHMAN, H. LAWRENCE, J.G. AND GROISMAN, E.A., 2000. Lateral gene transfer and the nature of bacterial innovation. *Nature,* **405**: 299-304.

ROWE-MAGNUS, D.A., GUEROT, A.M., PLONCARD, P., DYCHINCO, B., DAVIES, J. AND MAZEL, D., 2001. The evolutionary history of chromosomal super-integrons provides an ancestry for multiresistant integrons. *Proc. natl. Acad. Sci. USA,* **98**: 652-657.

STOKES, H.W., HOLMES, A.J., NIELD, B.S., HOLLEY, M.P., NEVALAINEN, K.M.H., MABBUTT, B.C. AND GILLINGS, M.R., 2001. Gene cassette PCR: sequence-independent recovery of entire genes from environmental DNA. *Appl. environ. Microbiol.,* **67**: 5240-5246.

(Received 11 July 2008, revised 22 July 2009)