IN SILICO ANALYSIS OF REGULATORY ELEMENTS OF ANTIBIOTIC PRODUCING GENES IN STREPTOMYCES SPECIES

Pramod Katara*, Neha Sanghi, Vinay Sharma, India.
Department of Bioscience and Biotechnology, Banasthali University, India-304022
***Email:** pmkatara@gmail.com

ABSTRACT

One of the great challenges currently facing biologists is to understand the varied and complex mechanisms that regulate gene expression. Recent advances in genome sequence availability and in high-throughput gene expression analysis technologies i.e. EST and cDNA Microarray provide us lot of significant information, which can be used for various biological purposes. The gene expression data observed over a range of conditions and time points has become widely used in modern biology to discover relationships between different genes specifically in functional and regulatory aspects i.e. to develop gene regulatory network and TFBS. Microarray data analysis for such purpose involves clustering approach which can be based on different concepts and grouped genes on the basis of their similar gene expression patterns. In present work we focus on regulatory genomics of antibiotics producing genes belong to *Streptomyces* species whose medicinal importance is already well reported. For our study we used DNA microarray data from SMD and available software to analyze the gene expression patterns and to predict *cis*-regulatory elements (TFBS) of genes in *Streptomyces* species involved in antibiotic biosynthesis. As a result of our study we observe that there are some genes which share similar gene expression pattern with antibiotic producing genes, we can assume that such genes are somehow related with the antibiotic production, may be by taking part in antibiotic producing pathway. Further analysis of TFBS suggests that mostly all TFBS tend to be 5 to 20 nucleotides long and shows a significant conserverdness.

Keywords: Co-expression, Gene expression pattern, K-mean clustering, Transcription factor binding sites.

INTRODUCTION

The fame of Streptomycetes as versatile producers of secondary metabolites started with the discovery of Actinomycin in 1940, followed by streptomycin in 1943. Two-thirds of the marketed microbial drugs are produced by Streptomycetes. The total number of antimicrobial compounds, which this genus is capable of producing, to be of the order of a 100,000 - a tiny fraction of which has been unearthed so far [1]. The reason why *Streptomyces* produce so many kinds of antibiotics and bioactive compounds is that *Streptomyces* strains have many gene clusters, which encode enzymes for many secondary pathways. The genus *Streptomyces* belongs to the order Actinomycetales (high-G+C Gram-positive bacteria)[2], which include unicellular and filamentous microbes, the latter having complex life cycles and forming hyphae, spores and secondary metabolites Recent estimates indicate that nearly 50% of the 20,000 bioactive secondary metabolites described from 1900 onwards are produced by filamentous actinomycetes that originated in the soil. Among them, the easiest to isolate from soils are *Streptomyces* species[2].

Streptomyces sp. has capacity to synthesize secondary metabolites and they are attractive model for the study of microbial differentiation. *Streptomyces coelicolor* is a representative of the group of soil-dwelling, filamentous bacteria responsible for producing most natural antibiotics used in human and veterinary medicine. The genome contains 8,667,507 base pair linear chromosome, containing the largest number of genes so far discovered in a bacterium. The 7,825 predicted genes include more than 23 clusters coding for known or predicted secondary metabolites. The genome contains an unprecedented proportion of regulatory genes, predominantly those likely to be involved in responses to external stimuli and stresses, and many duplicated gene sets that may represent 'tissue-specific' isoforms operating in different phases of colonial development, a unique situation for a bacterium[3].

An understanding of the organization and regulation of biosynthetic genes for antibiotic pathways will help in the search for new antibiotics in a bacterial group[4].However, unraveling the mechanisms that regulate gene expression is a major challenge in biology. An important task in this is to identify regulatory elements, especially the binding sites in deoxyribonucleic acid (DNA) for transcription factors (TF). These binding sites are short DNA segments that are called motifs[5]. The motifs can be located far away from genes, although most of times a good chunk appears upstream of the genes a few hundred bases apart. Key components of transcriptional regulatory machinery in prokaryotes are TF and Transcription factor binding sites (TFBS), sigma factor and promoters[6]. A simple general mechanism in prokaryotes for coordinating the regulation of genes encoding products that participate in a set of related processes these genes are clustered on the chromosome and are transcribed together. The gene cluster and promoter, plus additional sequences that function together in regulation, are called operon[7].

Transcription factors, as these regulators are called, bind to short (5-15 bp), highly specific DNA sequences and can regulate large networks of functionally related genes. Due to the importance of transcriptional regulation, one of the main goals in the post-genomic era is to predict how a

gene's expression is regulated based on the presence of TFBS in the adjacent genomic regions[8]. Genome wide knowledge of TFBS could be used to build models of transcriptional regulatory networks that operate in cell fate specification during development. Currently two computational approaches are available to predict regulatory motif, first approach involves the use of microarray technology[9, 10], to find genes that are co-expressed or that respond similarly to a stimulus and then to search the promoter/enhancer regions of these genes for conserved motifs that could be TFBS[11, 12], second approach is based on the sequence comparative approach[13], and known as phylogenetic footprinting[14, 15]. In present work we focus on regulatory genomics of antibiotics producing genes belong to *Streptomyces* species. We concentrate on the prediction of *cis*-regulatory elements by the help of gene expression data and available data analysis tools.

Methodology: In gene expression based approach we downloaded the gene expression data files of *S. coelicolor* (486.19 Mb) available at SMD database[16]; from these files raw data sheets were extracted. We then prepared the working files by taking gene ID and $\log_2$ mean values of gene expression. All ratio values are log transformed (base 2 for simplicity) to treat inductions or repressions of identical magnitude as numerically equal but with opposite sign[17]. Normalization of data was done using flagging method[18]. All the genes present in Actinorhodin biosynthesis gene cluster and red gene cluster were searched in these prepared files and those showing more than one occurrence were averaged to reduce the redundancy[19]. The resulting files having averaged values were used as input for K-mean clustering process. After K-mean clustering, we used Regulatory Sequence Analysis Tool[20, 21] for upstream sequence retrieval which takes gene ID and gene name as the input and find their upstream region along with the distance. Further to predict regulatory motif we used Multiple EM for Motif Elicitation, which takes as input a set of upstream sequences[22].

Figure 1: Antibiotic genes showing co-expression with other *S. coelicolor* genes.

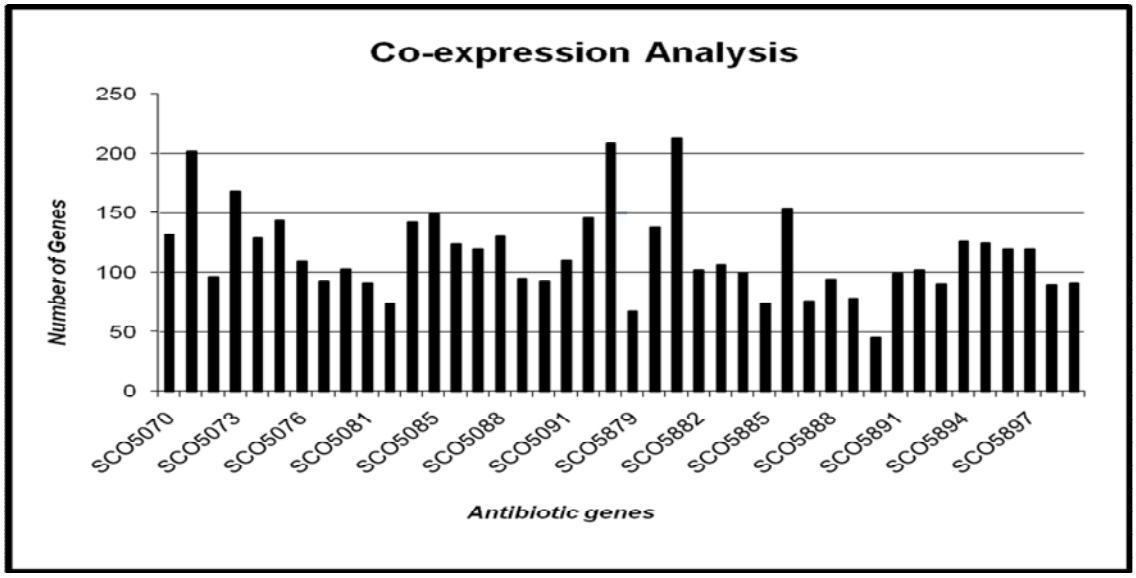


Figure 2: Graph depicting the number of nucleotide bases contained by each antibiotic gene in their upstream region.

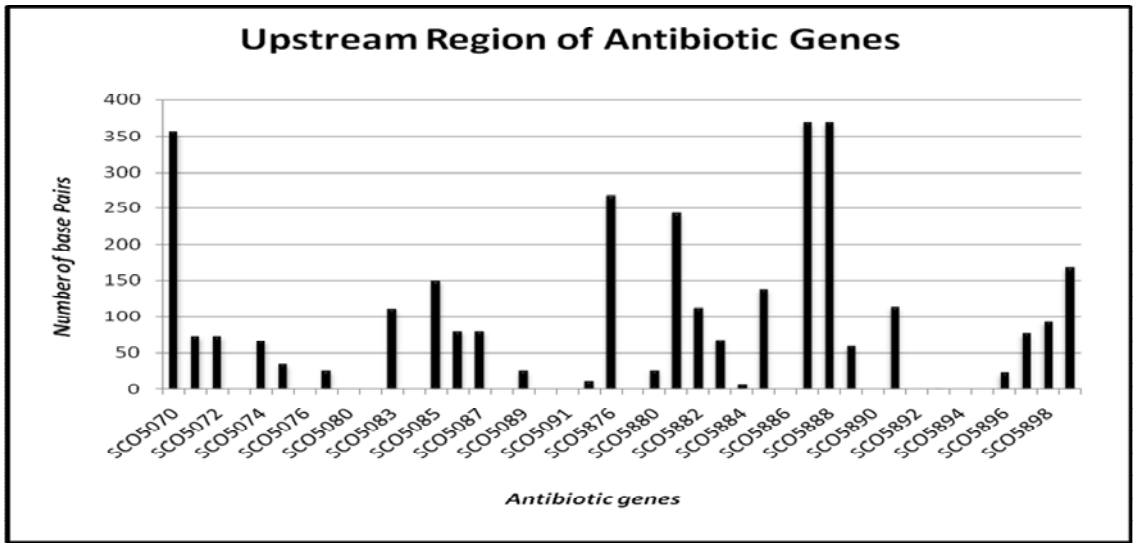
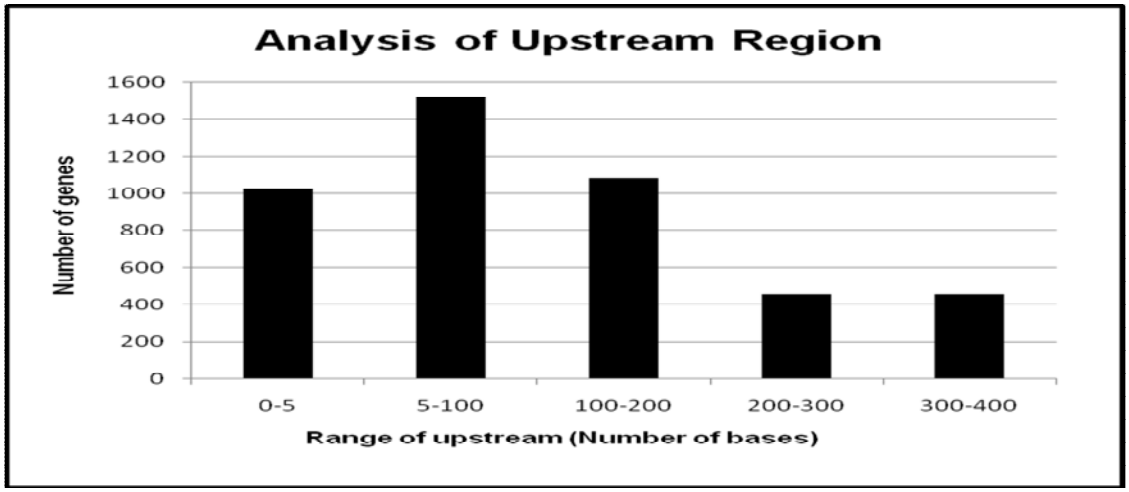


Figure 3: Graph showing the range of upstream region of *S. coelicolor* genes.



RESULTS AND DISCUSSION

Clustering of co-expressed genes: Analysis of co-expressed genes performed by cluster using loose K-mean clustering, resulting cluster showed groups of co-expressed genes associated with our known antibiotic producing genes. We observed that lot of genes share similar expression pattern with antibiotic genes but their numbers are vary gene to gene and it ranged from 40-210. It suggest that every antibiotic producing genes share similar gene expression pattern with different number of genes i.e genes like SCO5072, SCO5876, and SCO5881 show co-expression with more number of genes than compared to all other genes and some other genes like SCO5879 and SCO5890 shows co-expression with very low number of genes (Fig 1).

Result suggest that though very less number of genes are known to produce antibiotics, but lot of other genes i.e which share similar gene expression pattern are also play a crucial role in antibiotic production. At present their role in antibiotic production might be an enigmatic, but from current study it is now clear that somehow they act as an important factor for antibiotic production.

Identification of upstream regions: The key components of transcriptional regulatory machinery in prokaryotic genome mainly include TF, TFBS, Sigma factors and promoters which are present in upstream region. Thus, the identification of upstream region can shed light on further analysis of regulatory machinery.

RSAT output gives the upstream sequences of various lengths. Some genes like SCO5070, SCO5887 and SCO5888 shows high length of upstream region, while some of the antibiotic genes like SCO5073, SCO5080, SCO5081, SCO5084, SCO5090, SCO5091, SCO5879, SCO5884, SCO5886, SCO5890, SCO5892, SCO5893, SCO5894 and SCO5895 has zero upstream size. Here the variation in upstream regions provide a very significant information it indicating that those genes whose upstream shows zero size, are lacking regulatory region, thus their regulation may be controlled by some other genes (fig 2). Thus we can infer that these genes may be a part of a gene cluster[23], and act as operon and participate in a set of related processes[24].

Above figure-3 can be summarized as, 1027 genes fall in the range of 0-5 bases, 1520 genes are in the range of 5-100 bases, 1086 genes have bases in the range of 100-200,459 genes fall in the range of 200-300, while 460 genes are in the range of 300-400 bases. It can be inferred that most of the upstream region of genes in prokaryotes lie in the range of 5-100 and genes in the range of 0-5 does not provide any background information for the prediction of conserved motifs.

Prediction of conserved motifs: To understand the regulatory mechanism involved in antibiotic biosynthesis we need to have knowledge about the conserved patterns present in the upstream region of gene sequences. These patterns can serve as putative regulatory elements. The identified motifs are depicted in the form of regular expressions along with their information content and E-value. The motif with the highest information content can be considered as highly conserved for their respective gene. The search by MEME for potential conserved motifs of 43 antibiotic genes of actinorhodin and red gene cluster of *S. coelicolor* resulted in putative motifs showing the conserved patterns among the upstream region of gene sequences and might be work as *cis*-regulatory elements. Observed length of most of the motif is ranged between 5-20 nucleotides. The predicted motifs for a given set of genes in which some genes have zero upstream region shows that their biosynthesis is controlled by the action of specific regulatory genes located within the particular biosynthetic clusters or several pleiotropic genes outside the biosynthetic clusters may have been implication in the regulation of the multiple antibiotic pathways[25]. It can also be inferred that some of these motifs have binding sites for TFs or serve as TFBSs. Furthermore, evidence suggests that transcriptional regulatory regions often occur in modules, so TFBS adjacent to genes will be clustered into regulatory modules that can be distinguished from non-regulatory areas by their high base conservation[26]. Finally, some of these conserved

motifs reside upstream of genes with similar functional annotations or similar expression pattern or those bound by same TF and are thus may serve as good candidates for functional regulatory sequences[27].

CONCLUSION

Current study concludes that antibiotic producing genes of *streptomyces sp.* Share similar gene expression pattern with lot of other genes whose role till now is no known for antibiotic production, but analysis suggest that somehow they are involve in antibiotic production. Numbers of predicted co-expressed genes are varying gene to gene. Analyzed length of upstream regions indicate that TFBS for most of the genes are located within 100 base upstream, and lot of genes are arranged in cluster and form Operon or operon-like structure and regulate by common regulatory factors. Further analysis suggest that the size of most of the motif is ranged between 5-20 nucleotide with high G/C content, and because of this high G/C content they shows more conservedness. The predicted motifs can be further analyzed to check palidromicity, because it is a known fact that many microbial TFBSs are palindromic. All the overrepresented motifs can be screened for their overlap with RBS (ribosome binding sites).

REFERENCES

1. Watve M.G., Tickoo R., Jog M.M., Bhole B.D. (2001) How many antibiotics are produced by the genus Streptomyces? *Arch Microbiol* 5: 386-390
2. Marinelli F. (2009) Antibiotics and Streptomyces: the future of antibiotic discovery. *Microbiology today* 20-23
3. Bentley S.D., Chater K.F., Cerdeño-Tárraga A.M., Challis G.L., Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H., Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M., Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S., Huang C.H., Kieser T., Larke L., Murphy L., Oliver K., O'Neil S., Rabbinowitsch E., Rajandream M.A., Rutherford K., Rutter S., Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K., Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J., Hopwood D.A. (2002) Complete genome sequence of the model actinomycete *Streptomyces coelicolor A3(2)*. *Nature* 417:141-147
4. Malpartida F., Hopwood D.A. (1986) Physical and genetic characterization of the gene cluster for antibiotic actinorhodin in *Streptomyces coelicolor A3(2)*. *Mol gen Genet* 205:66-73
5. Das M.K., Dai H.K. (2007) A survey of DNA motif finding algorithms. *BMC Bioinformatics*. 8 Suppl 7:1-13
6. Lee E.J., Karoonuthaisiri N., Kim H.S., Park J.H., Cha C.J., Kao C.M., Roe J.H. (2005) A master regulator sigmaB governs osmotic and oxidative response as well as differentiation via a network of sigma factors in *Streptomyces coelicolor*. *Mol Microbiol* 5:1252-1264
7. Yanai K., Murakami T. (2004) The Kanamycin Biosynthetic Gene Cluster from *Streptomyces kanamyceticus*. *J. Antibiotics* 57:351-354
8. Li H., Rhodius V., Gross C., Siggia E.D. (2002) Identification of the binding sites of regulatory proteins in bacterial genomes. *Proc Natl Acad Sci USA* 99:11772- 1777
9. Schena M., Shalon D., Davis R.W., Brown P.O. (1995) Quantitative Monitoring of Gene Expression Pattern with a Complementary DNA Microarray. *Science* 270: 467-470
10. DeRisi J.L., Iyer V.R., Brown P.O. (1997) Exploring the Metabolic and Genetic Control of Gene Expression on a Genomic Scale. *Science* 278: 680-686
11. Huang J., Lih C.J., Pan K.H., Cohen S.N. (2001) Global analysis of growth phase responsive gene expression and regulation of antibiotic biosynthetic pathways in *Streptomycin coelicolor* using DNA microarrays. *Genes & Dev*.15:3183–3192
12. Katara P., Singh A., Ragav D., Sharma V. (2010) Analysis of Regulatory Genomics and Gene Expression Pattern of Medicinal Importance Genes of *Helicobacter Pylori*. *J Comput Sci Syst Biol* 3: 010-015. doi: 10.4172/jcsb.1000049
13. Tompa M. (2001) Identifying functional elements by comparative DNA sequence analysis. *Genome Res*.11:1143-1144
14. Blanchette M., Tompa M. (2002) Discovery of Regulatory Elements by a Computational Method for Phylogenetic Footprinting. *Genome Res.* 12:739-748
15. Studholme D.J., Bentley S.D., Kormanec J. (2004) Bioinformatics identification of novel regulatory DNA sequence motifs in *Streptomyces coelicolor*. *BMC Microbiology* 4:14
16. Sherlock G., Hernandez-B.T., Kasarskis A., Binkley G., Matese J.C., Dwight S.S., Kaloper M., Weng S., Jin H., Ball C.A., Eisen M.B., Spellman P.T., Brown P.O., Botstein D., Cherry J.M. (2001) The Stanford Microarray Database. *Nucleic Acid Res.* 29:152-155
17. Eisen M.B., Spellman P.T., Brown P.O., Botstein D. (1998) Cluster analysis and display of genome-wide expression pattern. *Proc Natl Acad Sci USA*.95:14863-14868
18. Yang M.C., Ruan Q.G., Yang J.J., Eckenrode S., Wu S., Mcindoe R.A., She J.X. (2001) A statistical method for flagging weak spots improved normalization and ratio estimate in microarrays. *Physiol Genomics*.7 (1):45-53
19. Yeung K.Y., Medvedovic M., Bumgarner R.E. (2003) Clustering gene-expression data with repeated measurements. *Genome Biol.* 4:R34.1-17

20. van-Helden J. (2003) Regulatory Sequence Analysis Tools. *Nucleic Acid Res.* 31:3593-3596

21. Thomas-Chollier, M., Sand, O., Turatsinze, J. V., Janky, R., Defrance, M., Vervisch, E., Brohee, S. & van-Helden, J. (2008). RSAT: regulatory sequence analysis tools. *Nucleic Acids Res* 36:W119-27

22. Bailey T.L., Boden M., Buske F.A., Frith M., Grant C.E., Clementi L., Ren J., Li W.W., Noble W.S. (2009) MEME Suite: tools for motif discovery and searching. *Nucleic Acid Res*. 37: 202-208

23. Fan Q., Huang F., Leadlay P.F., Spencer J.B. (2008) The neomycin biosynthetic gene cluster of Streptomyces fradiae NCIMB 8233: genetic and biochemical evidence for the roles of two glycosyltransferases and a deacetylase. *Org. Biomol. Chem.* 6:3306–3314

24. Laing E., Sidhu K., Hubbard S.J. (2008) Predicted transcription factor binding sites as predictors of operons in Escherichia coli and *Streptomyces coelicolor*. *BMC Genomics* 9:64-79

25. Martınez-Costa O.H., Arias P., Romero N.M., Parro V., Mellado R.P., Malpartida F. (1996) A relA/spoT Homologous Gene from *Streptomyces coelicolor A3(2)* Controls Antibiotic Biosynthetic Genes. *J. Biol. Chem*. 271:10627-10634

26. Loots G.G., Ovcharenko I., Pachter L., Dubchak I., and Rubin E.M. (2002) rVista for comparative sequence based discovery of functional TFBS. *Genome Res.* 12:832-839

27. Cliften P.F., Sudarsanam P., Desikan A., Fulton L., Fulton B., Majors J., Waterston R.H., Cohen B.A. Johnston M. (2003) Finding functional features in Saccharomyces genomes by Phylogenetic footprinting. *Science* 301:71-76

****