have a correlation of −0.15 (there are many similar examples). In certain cases, the patterns can be different owing to differences in mRNA stability – for example, the same amount of mRNA might be transcribed, but differential degradation rates could affect the expression measurements of the microarrays. However, in the many examples of completely different expression pattern pairs, it is more likely that separate regulatory systems control the genes.

Considering the gene pairs YBR189W–YBR191W, YJL190C–YJL189W and YOL040C–YOL039W in more detail, the first pair is transcribed on the same strand, whereas the other two pairs are transcribed away from each other on opposite strands. In the first case, the regulatory system of both genes could be located upstream of YBR189W. In the other cases, it might be located between the genes. The expression patterns from the cell cycle data are shown in Fig. 2b–d. These genes code ribosomal protein, and as the genes are functionally related, it is not surprising that all six patterns are similar. However, the expression patterns for the genes in each pair are almost identical. The average correlation between adjacent genes is 0.95, whereas the average correlation between the groups is 0.85. In the diauxic shift data, the corresponding correlations are 0.96 and 0.93. The values are 0.83 and 0.67 in the sporulation data. The correlation values support the notion that the consecutive genes are controlled by a single regulatory system.

Many genes that are controlled by separate regulatory systems might have highly correlated expression patterns. This could be the case if the genes are functionally related, or it might occur by random chance. If genes are controlled by a single regulatory system, then their expression patterns should be highly correlated in any data set. Each data set that becomes available can be used to generate a list of adjacent genes with highly correlated expression patterns. By intersecting these lists we obtain the most likely candidates for control by a single regulatory system. Table 1 presents some of the best candidates based on the three expression data sets. A complete candidate list can be found at http://www-hto.usc.edu/~tanghx/yeast-gene-pair.html, and code for conducting searches in other data sets is available from the authors. Once strong candidates have been found, the issue can be resolved by experimental analysis of yeast strains having a mutation in the putative site of the single regulatory region. If both genes are affected, then they are controlled by a single regulatory system.

Acknowledgements

We thank L. Kruglyak and E. Hubbell for valuable comments and feedback. We were supported in part by NSF grant DBI 9504393 and NIH grant R01 GM36230. We thank S. Tavaré and M. Waterman for useful discussions and for providing us with the resources for this project. We also thank the referees for their valuable comments.

References

1 Reznikoff, W.S. (1972) The operon revisited. *Annu. Rev. Genet.* 6, 133–156
2 Ptashne, M. and Gann, A. (1997) Transcriptional activation by recruitment. *Nature* 386, 569–577
3 Struhl, K. (1989) Molecular mechanisms of transcriptional regulation in yeast. *Annu. Rev. Biochem.* 58, 1051–1077
4 Struhl, K. (1996) Chromatin structure and RNA polymerase II connection: implications for transcription. *Cell* 84, 179–182
5 Helden, J. *et al.* (1998) Extracting regulatory sites from the upstream region of yeast genes by computational analysis of oligonucleotide frequencies. *J. Mol. Biol.* 281, 827–842
6 Zhang, X. and Smith, T.F. (1998) Yeast 'operons'. *Microb. Comparative Genomics* 3, 133–140
7 Cho, R.J. *et al.* (1998) A genome-wide transcriptional analysis of the mitotic cell cycle. *Mol. Cell* 2, 65–73
8 Tavazoie, S. *et al.* (1999) Systematic determination of genetic network architecture. *Nat. Genet.* 22, 281–285
9 Arnone, M.I. and Davidson, E.H. (1997) The hardwiring of development: organization and function of genomic regulatory systems. *Development* 124, 1851–1864
10 Heinemeyer, T. *et al.* (1999) Expanding the TRANSFAC database towards an expert system of regulatory molecular mechanisms. *Nucleic Acids Res.* 27, 318–322
11 DeRisi, J.L. *et al.* (1997) Exploring the metabolic and genetic control of gene expression on a genomic scale. *Science* 278, 680–686
12 Chu, S. *et al.* (1998) The transcriptional program of sporulation in budding yeast. *Science* 282, 699–705

Complex evolution of the inositol-1-phosphate synthase gene among archaea and eubacteria

The presence of inositol and its metabolites is widely observed in eukaryotes. However, their occurrence in prokaryotes, with few exceptions, is uncommon. The origin of inositol metabolism in prokaryotes thus remains uncertain[1,2]. The first known enzymatic step in the *de novo* biosynthesis of inositol is mediated by the *INO1* gene, which encodes 1L-*myo*-inositol-1-phosphate synthase (I1-P synthase)[2]. In the presence of nicotinamide adenine dinucleotide (NAD^+), this enzyme catalyzes the conversion of glucose-6-phosphate (G-6-P) into inositol-1-phosphate, which is subsequently dephosphorylated to *myo*-inositol by a specific I1-P phosphatase. In yeast and fungi, genetic evidence indicates that I1-P synthase is essential for viability[3]. We recently identified I1-P synthase in *Mycobacterium tuberculosis H37Rv*, leading to the first evidence for the presence of this enzyme in a prokaryote[2]. To obtain insights into the origin of inositol metabolism among archaea and eubacteria, we have analyzed the evolution of this metabolically important enzyme.

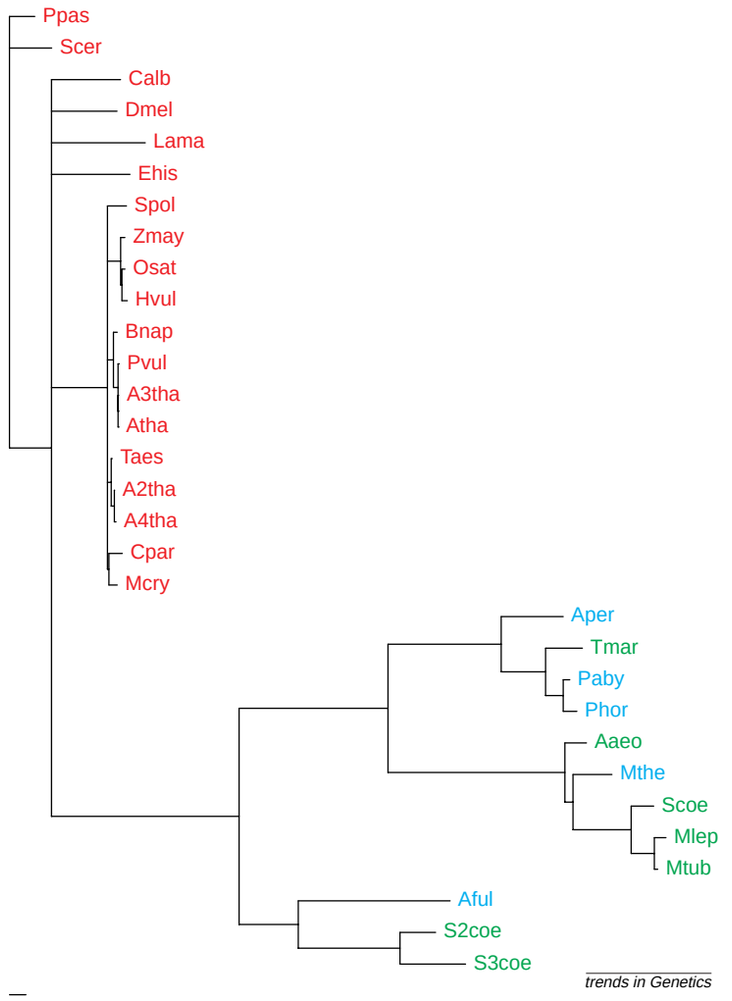
Nandita Bachhawat
nandita@lion.imtech.ernet.in

Shekkar C. Mande
shekkar@bragg.imtech.ernet.in

Institute of Microbial Technology, Sector 39.A, Chandigarh 160036, India.

Both BLAST[4] and PSI-BLAST[4] searches, using yeast INO1p as a search probe in non-redundant protein sequence databases, revealed a mycobacterial protein with insignificant homology (score 10^{-1}–10^{-2}). However, scores for a few other prokaryotic homologues, including two sequences belonging to *Streptomyces coelicolor* (score 10^{-6}–10^{-7}), were high (see http://bragg.imtech.ernet.in/shekhar/tig99/index.html). Similar searches using *Mycobacterial* INO1p as a probe indicated a remarkably low homology to the eukaryotic I1-P synthases but, surprisingly, yielded high homology scores to the same two *S. coelicolor* sequences.

FIGURE 1. Phylogenetic placement of inositol-1-phosphate synthases



The multiple sequence alignments used for generating the phylogenetic tree were constructed by excluding highly variable N-terminal and C-terminal amino acid stretches of the sequences. The tree was constructed after 100 cycles of bootstrapping using programs PROTDIST, SEQBOOT and CONSENSUS, available in the PHYLIP suite[5]. The tree was drawn using the TREEVIEW program[11]. The prokaryotic branches show an unusual mix of archaeal and eubacterial species. Abbreviations: Ppas, *Pichia pastoris*; Scer, *Saccharomyces cerevisiae*; Calb, *Candida albicans*; Dmel, *Drosophila melanogaster*; Lama, *Leishmania amazonesis*; Ehis, *Entamoeba histolytica*; Spol, *Spirodela polyrrhiza*; Zmay, *Zea mays*; Osat, *Oryza sativa*; Hvul, *Hordeum vulgare*; Bnap, *Brassica napus*; Pvul, *Phaseolus vulgaris*; A3tha, *Arabidopsis thaliana*; Atha, *Arabidopsis thaliana*; Taes, *Triticum aestivum*; A2tha, *Arabidopsis thaliana*; A4tha, *Arabidopsis thaliana*; Cpar, *Citrus paradisi*; Mcry, *Mesembryanthemum crystallinum*; Aper, *Aeropyrum pernix*; Tmar, *Thermotoga maritima*; Paby, *Pyrococcus abyssi*; Phor, *Pyrococcus horikoshii*; Aaeo, *Aquifex aeolicus*; Mthe, *Methanobacter thermoautotrophicum*; Scoe, *Streptomyces coelicolor*; Mlep, *Mycobacterium leprae*; Mtub, *Mycobacterium tuberculosis*; Aful, *Archaeglobulus fulgidus*; S2coe, *Streptomyces coelicolor*; S3coe, *Streptomyces coelicolor*. Scale bar represents 0.1% estimated sequence divergence.

The prokaryotic INO1p sequences with significant homology to the mycobacterial enzyme also included those from archaea *Methanobacter thermoautotrophicum*, *Archaeglobulus fulgidus*, *Pyrococcus horikoshii* and *Pyrococcus abyssi*. Interestingly, the *INO1* gene sequence is absent among the rest of the other available complete genome sequences of eubacteria and archaea, including *Methanococcus jannaschii*.

The available INO1p-related protein sequences were retrieved from non-redundant databases. A multiple alignment of all the retrieved INO1p and related sequences revealed the presence of a ~150 residue-long conserved domain towards the C-terminal of I1-P synthases. This domain possibly plays an important role in its structure–function relationship. Conversely, the homology at the N-terminal putative NAD-binding domain was far less[2]. Amino acid sequence analysis also exhibited a distinct difference in the size of eukaryotic and prokaryotic enzymes. The prokaryotic enzymes are 350–420 residues long, whereas the eukaryotic enzymes are >550 residues long.

Phylogenetic analysis of the INO1p family members after 100 cycles of bootstrapping showed distinct branches of eukaryotic and prokaryotic diversion. The eukaryotic branch showed an anticipated clustering of yeast and fungi I1-P synthase, followed by protozoan species and various plants. By contrast, the prokaryotic branch indicated a statistically significant but conflicting placement of various archaea and eubacterial INO1p (Fig. 1). Nearest to the node of the eukaryote–prokaryote diversion lie the two *S. coelicolor* enzyme sequences, along with *A. fulgidus*. Similarly, on the subsequent branching there is a mixed segregation of eubacterial and archaeal species. The branch containing *Mycobacteria* contains *M. thermoautotrophicum*, *Aquifex aeolicus* and a third protein of *S. coelicolor*. The remaining branch contains two *Pyrococci*, *Thermotoga maritima* and *Aeropyrum pernix* (Fig. 1). Placement of the species remained the same, irrespective of the algorithms or programs used for tree construction[5,6]. The mixed placement of archaea and eubacteria on all the three branches, with more than 95% bootstrapping confidence, might suggest a lateral transfer of the gene among diverse species.

One of the important factors in identifying the prokaryotic I1-P synthase was the location of the N-terminal NAD-binding site and a Cys residue present in its vicinity[2]. The Cys residue is reported to be crucial for electron transfer during the NAD-dependent conversion of G-6-P to I1-P (Ref. 7). Mysteriously, this Cys is absent in the *P. abyssi*, *P. horikoshii* and *T. maritima* sequences, which form the same branch of the prokaryotic phylogeny. The absence of the vital Cys residue indicates that the electron-transfer mechanism in these organisms might be different from other known prokaryotes. Interestingly, *Pyrococci* and *T. maritima* are both anaerobes, while *Mycobacteria* and *S. coelicolor* are aerobic in nature. Environmental adaptation and evolutionary pressure might be other immediate factors for the non-conserved Cys residue of I1-P synthases in anaerobes.

The N-terminal part of INO1 proteins shows a distinct pattern of conserved residues among different species. This domain is well conserved among the three branches of prokaryotic phylogeny, although it is not conserved across eukaryotes and prokaryotes (see http://bragg.imtech.ernet.in/shekhar/tig99/index.html). Thus, we

hypothesized that the N-terminal domain of the enzyme might have evolved independently in prokaryotes and eukaryotes. Moreover, high homology in the N-terminal domain among aerobes or among anaerobes, but only a weak similarity between them, suggests that this domain might have evolved separately in them.

To explore the possibility that the highly conserved C-terminal domain might have descended vertically among the different organisms, and fused to the independently evolved N-terminal domain in a later event, we constructed phylogenies of these two domains separately. The procedure to construct the phylogenetic tree was identical to that used earlier. The placement of species on these trees was once again identical to the tree shown in Fig. 1. Therefore, the two domains were integrated into the genome of the common ancestor, and were not exchanged individually among species before fusing with one another.

The mixed placement of various species in the INO1p phylogeny might suggest its lateral transfer among different archaea and eubacteria. The evolution of I1-P synthases between *S. coelicolor* and *A. fulgidus*, between *T. maritima* and *P. horikoshii* or *P. abyssi*, and between *M. thermoautotrophicum* and *M. tuberculosis* or *Mycobacterium leprae*, appears to be complex. The complete genome sequence of *T. maritima* recently suggested similar complex phenomena between *T. maritima* and archaea[8]. Horizontal transfer of genes among eubacteria and archaea has been a topic of considerable debate[9,10]. It has been suggested that a variety of similar genes among archaea and hyperthermophilic eubacteria might have descended vertically[9]. The phylogeny of I1-P synthases strongly discounts a vertical ancestry as all three branches among prokaryotes show strong bootstrapping confidence values, and conservation of key functional domains within the species in each branch. Moreover, all the three branches show a mixed placement of archaea and eubacteria (Fig. 1). In conclusion, there appears to be a complex pattern of evolution of the I1-P synthase gene among archaea and eubacteria.

Acknowledgements

We thank T. Chakrabarti for useful discussions. Financial assistance was provided by IMTECH and CSIR.

References

1 Kozloff, L.M. *et al.* (1991) Phosphatidylinositol, a phospholipid of ice nucleating bacteria. *J. Bacteriol.* 173, 2053–2060
2 Bachhawat, N. and Mande, S.C. (1999) Identification of the *INO1* gene of *Mycobacterium tuberculosis H37Rv* reveals a novel class of inositol-1-phosphate synthase enzyme. *J. Mol. Biol.* 291, 531–536
3 Culbertson, M.R. and Henry, S.A. (1975) Inositol-requiring mutants of *Saccharomyces cerevisiae. Genetics* 80, 23–40
4 Altschul, S.F. *et al.* (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25, 3389–3402
5 Felsenstein, J. (1993) PHYLIP – Phylogeny Inference Package (version 3.5c). Department of Genetics, University of Washington, Seattle
6 Strimmer, K. and von Haeseler, A. (1996) Quartet puzzling: a quartet maximum likelihood method for reconstructing tree topologies. *Mol. Biol. Evol.* 13, 964–969
7 Gumber, S.C. *et al.* (1984) *myo*-Inositol-1-phosphate synthase from pine pollen: sulfhydryl involvement at the active site. *Arch. Biochem. Biophys.* 231, 372–377
8 Nelson, K.E. *et al.* (1999) Evidence for lateral gene transfer between archaea and bacteria from genome sequence of *Thermotoga maritima. Nature* 399, 323–329
9 Kyrpides, N.C. and Olson, G.J. (1999) Archaeal and bacterial hyperthermophiles. Horizontal gene exchange or common ancestry? *Trends Genet.* 15, 298–299
10 Arvind, L. *et al.* (1998) Evidence for massive gene exchange between archaeal and bacterial hyperthermophiles. *Trends Genet.* 14, 442–444
11 Page, R.D.M. (1996) TREEVIEW: An application to display phylogenetic trees on personal computers. *Comp. Appl. Biosci.* 12, 357–358

Great chieftain o' the fungal-race!

First International Fission Yeast Meeting, Edinburgh, UK, 25–30 September 1999

Appropriately, the first international celebration of fission yeast was held in Edinburgh, an historic home of *Schizosaccharomyces pombe* research. The early days of experimentation with fission yeast were recalled by Murdoch Mitchison (University of Edinburgh, UK), and Urs Leupold [Bern, Switzerland, in a letter read by Paul Nurse, (Imperial Cancer Research Fund, London, UK)], both founding fathers of *S. pombe* research. Current interest in *S. pombe* was reflected by the assembly of more than 400 enthusiasts in a friendly, dynamic meeting organized by Stuart MacNeill (Institute of Cell and Molecular Biology, Edinburgh, UK), with help from Paul Nurse and Mitsuhiro Yanagida (Kyoto University, Kyoto, Japan). The meeting was timely, given that sequencing is almost complete (http://www.sanger.ac.uk/projects/S_pombe), and the post-genome future was discussed. All aspects of the biology of fission yeast were covered, from cell cycle to structure and signalling. Here, we review some of the many highlights.

DNA replication was a major focus of the meeting. Fission yeast origins are frequently localized at promoters (Francisco Antequera, University of Salamanca, Spain). This coincidence apparently does not reflect a transcriptional requirement for origin activity but, more likely, the shared need for a particular chromatin configuration[1]. Initiation of replication must be tightly regulated to occur only once per cycle. Origin recognition proteins (Orp – counterparts of the origin recognition complex) form complexes that bind origins constitutively[2], perhaps through the A/T

Janet Partridge
janet.partridge@hgu.mrc.ac.uk

Robin Allshire
robin.allshire@hgu.mrc.ac.uk

MRC Human Genetics Unit, Western General Hospital, Edinburgh, UK EH4 2XU.