of glycosylation arose adventitiously and were subsequently selected for.

Conclusions and outlook

The sequential development of different glycoprotein functions we propose is necessarily somewhat speculative and not readily subject to direct experimental verification. Nevertheless, this scheme provides a reasonably satisfying explanation for a number of seemingly anomalous aspects of glycoprotein biosynthesis and a sometimes-bewildering multiplicity of proposed functions. As we learn more about the diversity of glycoprotein structure and the genetic complexity that underlies the biosynthetic machinery, we should be able to discern the evolving role of complex carbohydrates more clearly.

Acknowledgements

We thank Roger Dodd for help with preparation of figures and the Wellcome Trust and the Oxford Glycobiology Endowment for funding.

References

1 Drickamer, K. and Taylor, M. E. (1993) *Annu. Rev. Cell Biol.* 9, 237–264
2 Gahmberg, C. G. and Tolvanen, M. (1996) *Trends Biochem. Sci.* 21, 308–311
3 Hebert, D. N., Simons, J. F., Peterson, J. R. and Helenius, A. (1995) *Cold Spring Harbor Symp. Quant. Biol.* 60, 405–415
4 Fiedler, K. and Simons, K. (1995) *Cell* 81, 309–312
5 Hughes, R. C. (1983) *Glycoproteins*, Chapman and Hall
6 Kleene, R. and Berger, E. G. (1993) *Biochim. Biophys. Acta* 1154, 283–325
7 Natsuka, S. and Lowe, J. B. (1994) *Curr. Opin. Struct. Biol.* 4, 683–691
8 Kornfeld, R. and Kornfeld, S. (1985) *Annu. Rev. Biochem.* 54, 631–664
9 Kukuruzinska, M. A., Bergh, M. L. E. and Jackson, B. J. (1987) *Annu. Rev. Biochem.* 56, 915–944
10 Driouich, A., Faye, L. and Staehelin, L. A. (1993) *Trends Biochem. Sci.* 18, 210–214
11 Schachter, H. (1991) *Glycobiology* 1, 453–461
12 Moloney, D. J., Lin, A. I. and Haltiwanger, R. S. (1997) *J. Biol. Chem.* 272, 19046–19050
13 Marth, J. D. (1996) *Glycobiology* 6, 701–705
14 Yuen, C-T. *et al.* (1997) *J. Biol. Chem.* 272, 8924–8931
15 Kornfeld, S. (1992) *Annu. Rev. Biochem.* 61, 307–330
16 Mehta, D. P. *et al.* (1996) *J. Biol. Chem.* 271, 10897–10903
17 Drickamer, K. (1991) *Cell* 67, 1029–1032
18 Lasky, L. A. (1995) *Annu. Rev. Biochem.* 64, 113–139
19 Powell, L. D. and Varki, A. (1995) *J. Biol. Chem.* 270, 14243–14246
20 Parlati, F., Dominguez, M., Bergeron, J. J. M. and Thomas, D. Y. (1995) *J. Biol. Chem.* 270, 244–253
21 Itin, C., Roche, A. C., Monsigny, M. and Hauri, H. P. (1996) *Mol. Biol. Cell* 7, 483–493
22 Dwek, R. A. (1995) *Biochem. Soc. Trans.* 23, 1–25
23 Mer, G., Hietter, H. and Lefèvre, J-F. (1996) *Nat. Struct. Biol.* 3, 45–53
24 Thijssen-van Zuylen, C. W. E. M. *et al.* (1998) *Biochemistry* 37, 1933–1940
25 Blum, M. L. *et al.* (1993) *Nature* 362, 603–609

Conservation of gene order: a fingerprint of proteins that physically interact

Thomas Dandekar, Berend Snel, Martijn Huynen and Peer Bork

A systematic comparison of nine bacterial and archaeal genomes reveals a low level of gene-order (and operon architecture) conservation. Nevertheless, a number of gene pairs are conserved. The proteins encoded by conserved gene pairs appear to interact physically. This observation can therefore be used to predict functions of, and interactions between, prokaryotic gene products.

COMPLETELY SEQUENCED GENOMES provide us with an opportunity to study the evolution of genome organization at a comprehensive level. A variety of studies have focused on the conservation of gene order in evolution, and the authors have drawn different conclusions, depending on the phylogenetic distance between the species compared and on the genes that were analyzed[1–5]. For example, conservation of gene order between *Mycoplasma genitalium* and *Mycoplasma pneumoniae*[6] is likely to be a result of a lack of time for genome rearrangements after divergence of the two organisms from their last common ancestor. Hence, if one is interested in the selective constraints that preserve gene order, only relatively long evolutionary distances between the species compared should be considered. However, the distances should be small enough that a significant number of orthologous genes is still shared by the species.

Gene order is already considerably disrupted when the protein-sequence identity shared by orthologs in two genomes is <50%[7]. We therefore analyzed genes from three sets of three completely sequenced genomes for which at least two of the intergenomic distances show less than 50% identity in shared orthologs (Fig. 1), which should be a sufficient test set for systematic studies. The genome sequences used (see Box 1) included those of proteobacteria (*Escherichia coli*[8], *Haemophilus influenzae*[9] and *Helicobacter pylori*[10]), Gram-positive bacteria (*M. genitalium*[11], *M. pneumoniae*[12] and *Bacillus subtilis*[13]) and archaea (*Methanococcus jannaschii*[14], *Methanobacterium thermoautotrophicum*[15] and *Archaeoglobus fulgidus*[16]).

To ensure that any conservation of gene order dates back to the earliest point at which the sequences compared diverged (rather than to more recent horizontal gene-transfer events) and hence reflects evolutionary constraints, we only considered genes that show the same order in a set of three genomes. For example, the urease operon is present in both *H. influenzae* and *H. pylori*, but is absent from *E. coli*. In *H. influenzae*, the G–C content of the urease-operon

T. Dandekar, **B. Snel**, **M. Huynen** and **P. Bork** are at the European Molecular Biology Laboratory, Postfach 102209, D-69012 Heidelberg, Germany; and **T. Dandekar**, **M. Huynen** and **P. Bork** are also at the Max-Delbrück-Centrum fuer Molekulare Medizin, Robert-Roessle Str. 10, 13122 Berlin-Buch, Germany.
Email: huynen@embl-heidelberg.de

 0968–0004/98/$19.00 PII: S0968-0004(98)01274-2

third coding position[9] differs significantly from that of other operons in the genome[7], which points to a recent horizontal operon transfer.

Conservation of gene clusters and pairs

In each of the three triple-genome sets compared, ~100 genes are conserved as pairs or clusters (Box 2). The direction of transcription in each conserved pair/cluster is the same for all genes. Furthermore, the order in which the genes are transcribed is always conserved, except in the case of the *B. subtilis* phosphotransfer protein (see Box 2). Physical interactions between the proteins that are encoded by a gene pair are apparent in most cases and can be predicted confidently for almost all of the remaining proteins (see below). Potential physical interactions can be divided into three categories, according to the evidence for their existence.

Experimentally confirmed interactions. For at least 75% of the conserved gene pairs, physical interactions between the encoded proteins have been demonstrated. These proteins include ribosomal proteins that are involved in the formation of the ribosomal particle[17] (and whose conserved gene order might reflect specific interactions during ribosome assembly), RNA polymerase subunits, ATP synthase subunits, transporter subunits, various enzymes/enzyme-subunits that interact (e.g. homoserine dehydrogenase) and cell-division proteins (see Box 2). The proteins all have basic cellular functions, and the existence of the genes that encode them in conserved gene pairs might reflect ancient interactions or genome architecture.

Predicted interactions. A further 20% of the conserved clusters/pairs encode proteins that are predicted to interact physically. In these instances, experimental evidence or the known biological context of the genes suggest that the proteins interact. For example, a gene pair that is conserved in proteobacteria encodes a surface-exclusion-protein homolog and a protein that shows significant similarity (enough to predict that the two proteins have very similar functions) to CDP ribitol pyrophosphorylase, an enzyme that is involved in cell-wall-surface synthesis[18]. Surface exclusion proteins in pheromone plasmids reduce conjugative transfer into plasmid-carrying strains and, like CDP ribitol pyrophosphorylase, are important and abundant cell-wall components[19]. Thus, a physical interaction between these proteins at the cell-wall surface, in an as-yet-unknown functional context, is likely.

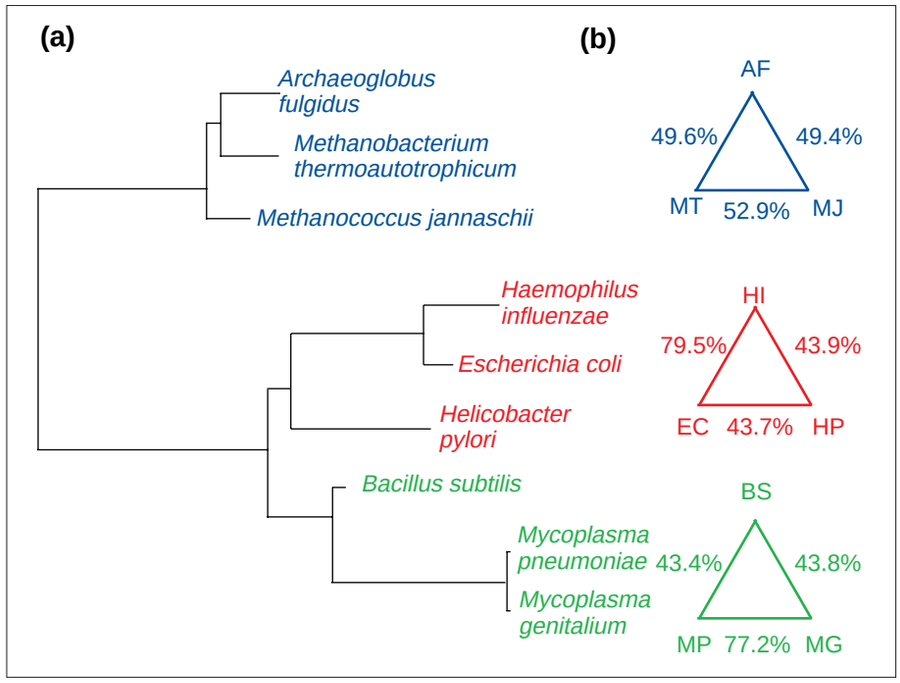

Figure 1

(a) A 16S-rRNA phylogenetic tree for the nine genomes analyzed. Each set of three genomes is color coded. Evolutionary distances are drawn to scale. **(b)** Triple comparisons[7] showing the average level of sequence identity for a set of 34 orthologous proteins that is shared by all the species analyzed. AF, *Archaeoglobus fulgidus*; BS, *Bacillus subtilis*; EC, *Escherichia coli*; HI, *Haemophilus influenzae*; HP, *Helicobacter pylori*; MG, *Mycoplasma genitalium*; MJ, *Methanococcus jannaschii*; MP, *Mycoplasma pneumoniae*; MT, *Methanobacterium thermoautotrophicum*.

In archaea, genes that encode the transcription factor TFIIS and a mutator protein T (MutT)-family homolog exist as a conserved gene pair. TFIIS increases fidelity in RNA-polymerase-II transcripts by enhancing 3′–5′-ribonuclease activity towards misincorporated nucleotides[20]. MutT-like proteins are 'housecleaning' enzymes – that is, ribo- or desoxy-nucleoside pyrophosphohydrolases that selectively recognize chemically aberrant nucleotides[21]. The crystal structures of both TFIIS and the MutT-like protein are known, and modeling studies suggest that the two proteins form a complex in which their β-sheets interact. A physical interaction between the two proteins might allow them to act in concert to enhance the fidelity of transcript synthesis in archaea.

Putative interactions. For the remaining conserved gene pairs (less than 5%), either no clear function has been assigned to the encoded proteins, or we are not aware of any evidence for an interaction (Box 2). For example, the MG221 and MG222 genes in *M. genitalium*, and their orthologs in *M. pneumoniae* and *B. subtilis*, form conserved gene pairs and are flanked by genes that encode proteins involved in cell division. Although the products of this gene pair might form a complex and participate in cell division, there is no evidence to support such a prediction. However, it is noteworthy that our triple comparisons reveal that negative controls (i.e. adjacent genes that do not interact) are not conserved as gene pairs.

History of the concept

The idea that physical interaction between encoded proteins is one of the reasons for evolutionary conservation of gene order has been around for a long time. Early studies on lambdoid bacteriophages showed that a conserved gene order could be correlated with

Box 1. GenBank accession numbers for the genomes compared

Archaeoglobus fulgidus AE000782
Bacillus subtilis AL009126
Escherichia coli U00096
Haemophilus influenzae L42023
Helicobacter pylori AE000511
Mycoplasma genitalium L43967
Methanococcus jannaschii L77117
Mycoplasma pneumoniae U00089
Methanobacterium thermoautotrophicum AE000666

physical interactions between the encoded proteins[22–24]. Subsequently, this idea was extended by analysis of completely sequenced genomes, such as those of *E. coli* and *H. influenzae*[25]. Our systematic comparison of nine complete genomes (Box 2) provides further strong support for this concept, although obviously not all physical interactions between proteins can be revealed by the conservation of gene order.

Operon conservation

Operons are a well-known aspect of genome organization in prokaryotes[26] and might also exist in animals such as *Caenorhabditis elegans*[27]. To date, an exhaustive operon architecture has not been established for any species, because comprehensive transcription maps are not available. The gene order in the many known operons in the genomes we have examined here is not conserved, apart from genes encoding proteins that interact physically. We cannot, however, exclude local gene rearrangements that interrupt gene order but preserve operon structure or even co-regulation. The Trp operon illustrates this point and the various other types of rearrangements, such as complete disruption of the operon and rearrangements of protein domain organization, that play a role in operon evolution (Fig. 2).

The only aspect of the Trp operon that is completely conserved among the nine genomes examined is the gene pair *trpB–trpA*, which encodes the two subunits of tryptophan synthase that interact and catalyze a single reaction. Co-regulation alone is unlikely to be a sufficient driving force for operon conservation. Regulatory sequences, in general, seem to evolve very rapidly in the species compared[7]. Note that a physical interaction between two proteins does not guarantee that the genes encoding these proteins exist as an evolutionarily conserved gene pair. In *Aquifex aeolicus*, which is close to the phylogenetic root of Bacteria, and *Synechocystis*, *trpA* and *trpB* are not adjacent.

Box 2. Confirmed, predicted and putative interactions involving proteins encoded by conserved gene pairs/clusters

Proteobacteria (94 proteins)

(1) Experimentally confirmed (74 proteins).

Ribosomal proteins[17]: Rps9 and Rpl13; initiation factor 3 (IF3), Rpl35 and Rpl20; Rpl21 and Rpl27; elongation factor G (EF-G), Rps7, and Rps12; Rpl7/12 and Rpl10; Rpl1 and Rpl11 are encoded by genes in a large cluster of ribosomal protein genes that includes the gene encoding SecY (L36).

ATP synthase[35]: AtpC, AtpD, AtpG, AtpA and AtpH.

Transporters: ABC transporter subunits[36]; dppB, dppC and dppD dipeptide transporter subunits.

Enzyme pairs/subunits: GroEL and GroES; FrdB and FrdA; NifS and NifU; biotin carboxylase and biotin carboxyl carrier protein; PheT and PheS; ModA and ModB; MraY and MurD; HslV and HslU; ThiD and ThiM; TrpA and TrpB; RpoB and RpoB′; TrpD and TrpE; MreC and MreB.

Regulation: FtsA and FtsZ. The exact ratio is important for division. FtsA acts as a link to the FtsZ ring[37].

(2) Predicted on the basis of experimental data or biological context (18 proteins).

A complex, involving rpl19, RNA methyltransferase and the 21k protein, that participates in ribosome maturation. The 21k protein is in fact a maturase and associates with ribosomal protein[38].

Clp protease (ClpAP) shares structural homology with the proteasome[39], and trigger factor is a prolyl isomerase that could be involved in protein degradation. The existence of the genes encoding these proteins as a conserved pair suggests that trigger factor interacts with ClpAP protease to eliminate misfolded proteins.

A membrane complex formed by glycosylating acyltransferase (LpxA), an acyl carrier protein and three protein-export membrane proteins is functionally plausible and suggested by a conserved gene cluster.

Other examples: CDP ribitol pyrophosphorylase and surface exclusion protein (see text); serine deaminase (SdaA) and the serine transporter (SdaC)[40]; YxjD, YxjE and a short-fatty-acid-chain transmembrane intake protein; the TolB membrane transporter and a peptidoglycan protein in the outer cell wall.

(3) Putative (2 proteins).

NusB and RibE (NusB might in fact facilitate translation of the highly structured *ribE* mRNA).

Gram-positive bacteria (109 proteins)

(1) Experimentally confirmed (83 proteins).

Ribosomal proteins[17]: L11 and L1; S12, S7 and EF-G; a large cluster contains genes that encode SecY and RNA-polymerase-α subunits; L35 and L20; L10 and L7/11; Rps9 and L13; L19 and tRNA methyltransferase; EF-Ts, *mukB* suppressor and ribosome-releasing factor.

ATP synthase: AtpC, AtpD, AtpG, AtpA, AtpH, AtpF, AtpE and AtpB.

Transporters: ATP transporter subunits; fructose permease IIBC component and phosphotransfer protein (in *Bacillus subtilis* the order of transcription is different); oligopeptide permease complex.

Enzyme pairs/subunits: DNA gyrase subunits; RNA polymerase β and β′ subunits; hydroxymethyl-CoA-reductase and pro-lipoprotein diacylglyceryltransferase; thymidilate and folate reductase; pyruvate dehydrogenase; phosphoglycerate kinase and glycerinaldehyde-3-phosphate dehydrogenase; phosphoglycerate mutase and triosephosphate isomerase; pyruvate dehydrogenase; nitrogen-fixation enzymes; DNA helicase; Glu-tRNA amidotransferase (three subunits; the smallest was overlooked in genome sequencing but is also conserved in the cluster); GroEL homologs and GroES homologs.

Regulation: cell-division proteins (two different pairs).

(2) Predicted on the basis of experimental data or biological context (22 proteins).

Phe-tRNA synthetase might interact with IF3. Ribosome interactions similar to those involving Met-tRNA and the ribosome have been measured in initiation complexes[41].

6-Phosphofructokinase and pyruvate kinase. The product of the first enzyme activates the second. Physical coupling would therefore be advantageous.

Heat-shock-stress-response protein BS0069 and the salvage-pathway enzyme hpg transferase might be coupled.

Physical interaction at the folding stage?

Because the order in which genes within a conserved gene pair are transcribed is almost always conserved, it is reasonable to speculate that there is an interdependence of the folding of the proteins (co-translational folding[28]). Netzer and Hartl[29] have recently shown that cotranslational folding, at least for specific two-domain model polypeptides, works efficiently in a eukaryotic system. The examples given in Box 2 could be used in studies of cotranslational folding of protein chains that are translated in close proximity, and in parallel, from a polycistronic prokaryotic mRNA.

Co-adaptation at the molecular level

Co-adaptation[30–32] could select for clusters of genes in the genome and thus reduce the chance of genetic recombination perturbing co-adapted pairs of genes. In addition, genes whose products interact physically should also exhibit a lower rate of evolution, because of the selective constraints imposed by the interaction. Indeed, the degree of sequence conservation in conserved gene pairs is on average substantially higher than that in genes that do not

Box 2. Confirmed, predicted and putative interactions involving proteins encoded by conserved gene pairs/clusters (contd)

The ribosomal protein L34 and C5, a principal protein component of RNase P. Prokaryotic RNase P participates in ribosomal processing and maturation[42].

Other examples: His- and Asp-tRNA synthetase (tRNA synthetases are known to act in concert in eukaryotes[43]); protein phosphatase (or kinase) and transporter pump (brefeldin-resistance factor); a CinA homolog and phosphatidylglycerophosphate synthase (both participate in the SOS response); polytopic membrane protein P35 and two permeases; methylase and peptide-chain-release factor; Rpl19 and methyltransferase.

(3) Putative (4 proteins).

BS 1514 and BS 1515 (in *B. subtilis*, the flanking genes encode proteins that are both involved in cell division).

Cell-division protein BS2317 and the putative DNA-synthesis factor BS 2318.

Archaea (98 proteins)

(1) Experimentally confirmed (75 proteins).

Ribosomal proteins: Eight different ribosomal protein gene clusters – three contain different RNA polymerase subunit genes and one also contains the gene encoding the antiterminator NusA (Ref. 44) and S12, S7 and EF-G). The SecY translocase is encoded by a gene in another cluster.

ATP synthase: AtpB, AtpA, AtpF, AtpC, AtpE and AtpK.

Transporters: phosphate-transporter permease subunits; ABC transporter subunits; cobalt transporter subunits.

Enzymes: Trp-synthase subunits; methylviologen hydrogenase α and γ subunits; inosin dehydrogenase and cofactor ferredoxin; ferredoxin oxidoreductase α and β subunits; homoserin dehydrogenase subunit homologs.

Regulation: cell-division-protein-inhibitor subunits.

(2) Predicted on the basis of experimental data or biological context (12 proteins).

Chaperonin MJ0048 and the late-assembling[17] ribosomal protein L31. Association might increase fidelity of ribosome assembly in archaea.

Cell-division protein Ftsz, ribosomal L11 protein and antitermination factor and protein-translocation-factor homologs[37]. In Archaea, Ftsz associates with a complex that might allow efficient translation of this important protein.

Further examples: TFIIS and a MutT homolog (see text); a nucleic-acid-binding protein pair.

(3) Putative (11 Proteins).

Cleavage and polyadenylation-specific factor MJ1237A and proteasome subunit MJ1236 might participate in a proteasome complex; MJ0591/MT0686 and MJ0592/MT0685) might also participate in a proteasome complex.

GTP cyclohydrolase II and cofactor.

Nodulin 35 (uricase II), hsp70 and L21 (the latter two could help in conformational changes – L21 does this during ribosome assembly).

Conserved protein (unknown function) and rps19.

17 ribosomal proteins are conserved as pairs or small clusters between all species. In addition, the archaea and proteobacteria share *trpB–trpA*; the archaea and Gram-positive bacteria share RNA polymerase β and β′ subunits. The six eubacteria share bigger ribosomal protein clusters, F1-ATPase, and *pheT–pheS* (a putative partner in Gram-positive bacteria) and the *groEL–groES* pair (45 proteins altogether).

Analysis

We compared proteobacterial (*Escherichia coli*, *Haemophilus influenzae* and *Helicobacter pylori*), Gram-positive bacterial (*M. genitalium*, *M. pneumoniae* and *Bacillus subtilis*) and archaeal (*Methanococcus jannaschii*, *Methanobacterium thermoautotrophicum* and *Archaeoglobus fulgidus*) genomes using the Smith-Waterman algorithm run on a parallel biocellerator machine (see http://shag.embl-heidelberg.de:8000/Bic/docs/bicINFO.html). Orthologous genes were identified on the basis of relative sequence identity. 425, 266 and 585 orthologous triplets were identified in the proteobacteria, the Gram-positive bacteria and the archaea respectively. Within sets of orthologs, genes that show a conserved gene order were analyzed using various sequence-analysis techniques[45]. *B. subtilis* genes are numbered sequentially from dnaA (BS0001).

exist as conserved gene pairs. For example, the average degree of sequence identity shared by orthologs that exist as conserved gene pairs in *E. coli* and *H. pylori* is 46%; the equivalent figure for those that do not exist as conserved gene pairs is 38%. Similar differences (both mean and median values) in sequence identity are found in the other sets of genomes compared and are highly significant ($p \ll 0.01$, using the Wilcoxon rank sum test).

Compilation of data for pairs of conserved genes thus provides a database for the study of co-adaptation at the molecular level (e.g. for analysis of compensatory mutations). Ancient duplications as a source for the observed gene clusters (the so-called Natal model), alone are not a sufficient explanation for the number of observed gene pairs, because the majority of paralogous proteins do not cluster.

Note that physical interaction is only one of many constraints on protein structure and evolution; an extensive discussion of the factors that contribute to protein structure and evolution can be found elsewhere[33].

Applications of conserved gene pairs to functional searches

Conservation of gene order could be routinely used as a tool for predicting both physical interactions between proteins and protein function. It can be exploited in several ways.

(1) If the products of both genes have only been tentatively assigned functions, a conserved gene order can be used to predict both physical interaction and function. For example, the protein products of the *H. influenzae snzA–snzB* gene pair (HI1647 and HI1648, respectively) tentatively have been assigned functions as a phosphate-binding protein in amino acid synthesis and as a glutamine amidotransferase in nucleotide synthesis, respectively[34]. A direct interaction between the partners is probable, given that this gene pair is conserved in an archaeon [*Archeoglobus fulgidus* (genes AF508 and AF509)] and a Gram-positive bacterium [*Bacillus subtilis* (genes *yaaD* and *yaaE*)].

(2) If the function of the product of one gene in a conserved gene pair is known, it can hint at the function of the product of the neighboring gene. For example, in Gram-positive bacteria, the *nifU* gene and a neighboring gene exist as a conserved gene pair. The product of the neighboring gene had not been assigned a function; however, the hypothesis that it was involved in nitrogen fixation was confirmed by subsequent sequence analysis: the protein, annotated as 'hypothetical protein', is a NifS homolog. A cluster of three genes, which encode ribosomal protein L19, an RNA methyltransferase and a hypothetical protein, in Gram-positive bacteria is another example. The hypothetical protein should be a protein that associates with the other two. A literature search reveals that this protein has in fact recently been shown to be ribosome associated[35]. In several other cases, the fact that a protein is encoded by a gene that exists in a conserved gene pair has allowed, or at least speeded up, functional assignment.

(3) If the functions of the products of both genes are known, the fact that the

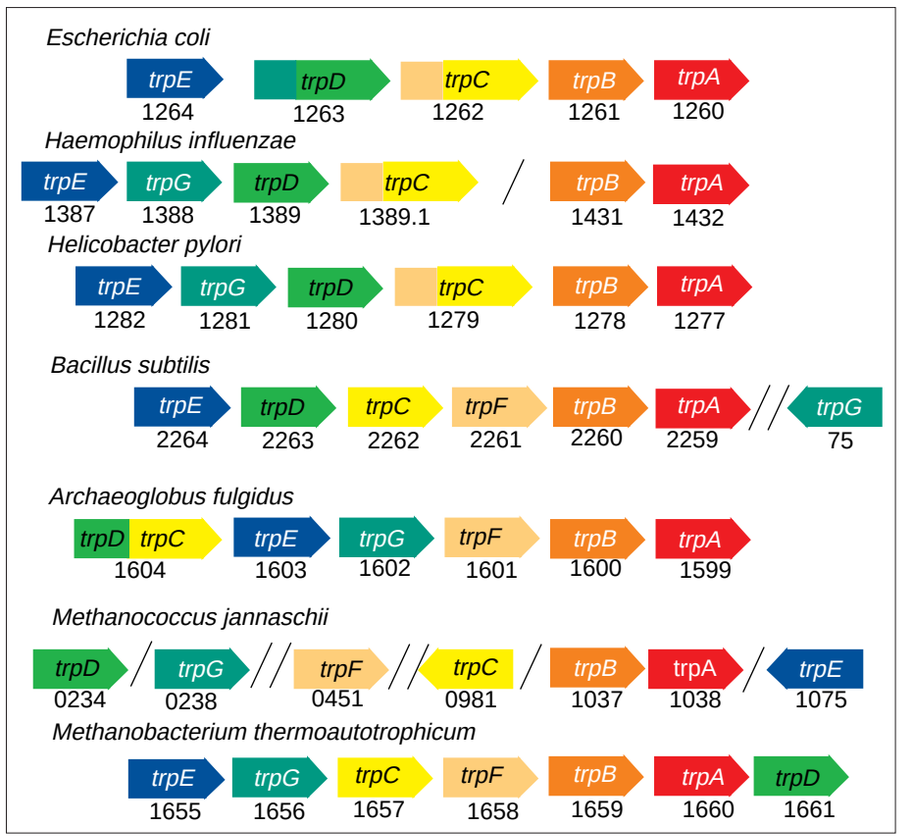

Figure 2

Structure of the tryptophan operon in different organisms. Arrows indicate the direction of transcription. Black lines indicate disruption of the operon by intervening genome sequences; double lines indicate a separation of more than 50 genes. The proteins encoded by the genes shown follow: *trpA*, tryptophan synthase α chain; *trpB*, tryptophan synthase β chain; *trpC*, indol-3-glycerol phosphate synthetase; *trpD*, anthranilate phosphoribosyltransferase; *trpE*, anthranilate synthase component I; *trpF*, anthranilate phosphoribosylisomerase; *trpG*, anthranilate synthase component II. Gene numbers are indicated and are consecutive along the genome. In the proteobacteria, the *trpC* and *trpF* genes are fused. The *trpG* and *trpD* genes in *Escherichia coli*, and the *trpC* and *trpD* genes in *Archaeoglobus fulgidus*, are also fused. The only feature of the Trp operon that is conserved across all seven genomes is the *trpA–trpB* gene pair.

genes exist as a conserved gene pair might reveal novel functional aspects.

In summary, completely sequenced genomes not only provide us with information about the evolution of genome organization and about constraints at higher order levels, but also, by revealing gene context, provide additional information about the function of individual gene products.

Further information, including details for different genomes (e.g. amino acid sequences and gene names) and a detailed description of our differential genome-analysis method can be found at http://www.bork.embl-heidelberg.de/Genome/conserved_pairs

References

1 Kolsto, A. B. (1997) *Mol. Microbiol.* 24, 241–248
2 Koonin, E. V. and Galperin, M. Y. (1997) *Curr. Opin. Genet. Dev.* 7, 757–763
3 Siefert, J. L. *et al.* (1997) *J. Mol. Evol.* 45, 467–472
4 Tamames, J., Casari, G., Ouzounis, C. and Valencia, A. (1997) *J. Mol. Evol.* 44, 66–73
5 Watanabe, H., Mori, H., Itoh, T. and Gojobori, T. (1997) *J. Mol. Evol.* 44, S57–S64
6 Himmelreich, R. *et al.* (1997) *Nucleic Acids Res.* 25, 701–712
7 Huynen, M. A. and Bork, P. *Proc. Natl. Acad. Sci. U. S. A.* (in press)
8 Blattner, F. R. *et al.* (1997) *Science* 277, 1453–1462
9 Fleischmann, R. D. *et al.* (1995) *Science* 269, 496–512
10 Tomb, J. F. *et al.* (1997) *Nature* 388, 539–547
11 Fraser, C. *et al.* (1995) *Science* 270, 349–548
12 Himmelreich, R. *et al.* (1996) *Nucleic Acids Res.* 24, 4420–4449
13 Kunst, F. *et al.* (1997) *Nature* 390, 249–256
14 Bult, C. *et al.* (1996) *Science* 273, 1058–1073
15 Smith, D. R. *et al.* (1997) *J. Bacteriol.* 179, 7135–7155
16 Klenk, H. P. *et al.* (1997) *Nature* 390, 364–370
17 Noller, H. F. and Nomura, M. (1996) in *E. coli and Salmonella* (Neidhardt, F. C., ed.), pp. 167–186, FC ASM Press
18 Cheah, S. C., Hussey, H. and Baddiley, J. (1981) *Eur. J. Biochem* 118, 497–500
19 Olmsted, S. B., Erlandsen, S. L., Dunny, G. M. and Wells, C. L. (1993) *J. Bacteriol.* 175, 6229–6237
20 Jeon, C. and Agarwal, K. (1996) *Proc. Natl. Acad. Sci. U. S. A.* 93, 13677–13682
21 Bessman, M. J., Frick, D. N. and O'Handley, S. F. (1996) *J. Biol. Chemistry* 271, 25059–25062
22 Casjens, S. R. and Hendrix, R. (1974) *J. Mol. Biol.* 90, 20–25
23 Botstein, D. (1980) *Ann. New York Acad. Sci.* 354, 484–490
24 Campbell, A. (1994) *Annu. Rev. Microbiol.* 48, 193–222
25 Mushegian, A. R. and Koonin, E. V. (1996) *Trends Genet.* 12, 289–290
26 Jacob, F. (1997) *C. R. Acad. Sci. (Ser. III)* 320, 199–206
27 Zorio, D. A., Cheng, N. N., Blumenthal, T. and Spieth, J. (1994) *Nature* 372, 270–272
28 Thanaraj, T. A. and Argos, P. (1996) *Protein Sci.* 5, 1594–1612
29 Netzer, W. J. and Hartl, F. U. (1997) *Nature* 388, 343–349
30 Fisher, R. A. (1930) *The Genetical Theory of Natural Selection*, Clarendon Press
31 Wallace, B. (1991) *J. Hered.* 82, 89–95
32 Lawrence, J. G. and Roth, J. R. (1996) *Genetics* 143, 1843–1860
33 Kimura, M. (1983) *The Neutral Allele Theory of Molecular Evolution*, Cambridge University Press
34 Galperin, M. Y. and Koonin, E. Y. (1997) *Molecular Microbiology* 24, 443–445.
35 Harold, F. M. and Maloney, P. C. (1996) in *E. coli and Salmonella* (Neidhardt, F. C., ed.), pp. 283–306, FC ASM Press
36 Eym, Y., Park, Y. and Park, C. (1996) *Mol. Microbiol.* 21, 695–702
37 Lutkenhaus, J. and Mukherjee, A. (1996) in *E. coli and Salmonella* (Neidhardt, F. C., ed.), pp. 1615–1626, FC ASM Press
38 Bylund, G. O., Persson, B. C., Lundberg, L. A. and Wikstrom, P. M. (1997) *J. Bacteriol.* 179, 4567–4574
39 Kessel, M. (1995) *J. Mol. Biol.* 250, 587–594
40 McFall, E. and Newmann, E. B. (1996) in *E. coli and Salmonella* (Neidhardt, F. C., ed.), pp. 358–379, FC ASM Press
41 Abdurashidova, G. G. *et al.* (1985) *Mol. Biol.* 19, 553–557
42 Morrissey, J. P. and Tollervey, D. (1995) *Trends Biochem. Sci.* 20, 78–82
43 Agou, F. and Mirande, M. (1997) *Eur. J. Biochem.* 243, 259–267
44 Keener, J. and Nomuras, M. (1996) in *E. coli and Salmonella* (Neidhardt, F. C., ed.), pp. 1417–1431, FC ASM Press
45 Bork, P. and Gibson, T. J. (1996) *Methods Enzymol.* 266, 162–184