Evolution of mosaic operons by horizontal gene transfer and gene displacement in situ
Marina V Omelchenko†, Kira S Makarova†, Yuri I Wolf†, Igor B Rogozin† and Eugene V Koonin†

Addresses: †Department of Pathology, FE Hebert School of Medicine, Uniformed Services University of the Health Sciences, Bethesda, MD 20814-4799, USA. †National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD 20894, USA.

Correspondence: Eugene V Koonin. E-mail: koonin@ncbi.nlm.nih.gov

Published: 29 August 2003
Received: 22 April 2003
Revised: 26 June 2003
Accepted: 17 July 2003

Abstract

Background: Shuffling and disruption of operons and horizontal gene transfer are major contributions to the new, dynamic view of prokaryotic evolution. Under the ‘selfish operon’ hypothesis, operons are viewed as mobile genetic entities that are constantly disseminated via horizontal gene transfer, although their retention could be favored by the advantage of coregulation of functionally linked genes. Here we apply comparative genomics and phylogenetic analysis to examine horizontal transfer of entire operons versus displacement of individual genes within operons by horizontally acquired orthologs and independent assembly of the same or similar operons from genes with different phylogenetic affinities.

Results: Since a substantial number of operons have been identified experimentally in only a few model bacteria, evolutionarily conserved gene strings were analyzed as surrogates of operons. The phylogenetic affinities within these predicted operons were assessed first by sequence similarity analysis and then by phylogenetic analysis, including statistical tests of tree topology. Numerous cases of apparent horizontal transfer of entire operons were detected. However, it was shown that apparent horizontal transfer of individual genes or arrays of genes within operons is not uncommon either and results in xenologous gene displacement in situ, that is, displacement of an ancestral gene by a horizontally transferred ortholog from a taxonomically distant organism without change of the local gene organization. On rarer occasions, operons might have evolved via independent assembly, in part from horizontally acquired genes.

Conclusions: The discovery of in situ gene displacement shows that combination of rampant horizontal gene transfer with selection for preservation of operon structure provides for events in prokaryotic evolution that, a priori, seem improbable. These findings also emphasize that not all aspects of operon evolution are selfish, with operon integrity maintained by purifying selection at the organism level.
Background

Operons, clusters of co-transcribed genes that often encode functionally linked proteins, are the principal form of gene organization and regulation in prokaryotes [1,2]. Comparative analysis of bacterial and archaeal genomes has shown that only a few operons are conserved across large evolutionary distances. In general, gene order in prokaryotes is poorly conserved and prone to numerous rearrangements [3-6]. A detailed analysis of gene order conservation has shown that only 5-25% of the genes in bacterial and archaeal genomes belongs to gene strings (probable operons) shared by at least two distantly related species [7]. The presence of identical or similarly organized operons and suboperons in phylogenetically distant bacterial or archaeal lineages may result from three distinct evolutionary processes. Firstly, inheritance from the respective common ancestor - the core of the ribosomal protein superoperon is a case in point, but such conservation of operon organization is relatively rare; secondly, independent origin of identical operons or suboperons in different lineages; and thirdly, emergence of operons in a single lineage with subsequent dissemination by horizontal transfer. The potential central role of horizontal transfer in the evolution of operon organization of prokaryotic genomes is embodied in the ‘selfish operon model’ (SOM) [8-10]. This model posits that "the physical proximity of genes in an operon provides no selective benefit to the individual organism but does enhance the fitness of the gene cluster itself, as clusters can be efficiently inherited horizontally as well as vertically" [11]. Under SOM, operons are conceptually analogous to integrating viruses (phages), transposons and other mobile genetic elements, although coregulation of the genes in an operon could be an important selective factor that favors retention of operons during evolution.

Horizontal gene transfer (HGT) events have been classified into distinct categories of acquisition of new genes, acquisition of paralogs of existing genes and xenologous gene displacement whereby a gene is displaced by a horizontally transferred ortholog from another lineage (xenolog [12]). Each of these types of horizontal transfer is common among prokaryotes, but their relative contributions differ in different lineages. A potential central role of horizontal transfer in the evolution of operon organization of prokaryotic genomes is embodied in the ‘selfish operon model’ (SOM) [8-10]. This model posits that "the physical proximity of genes in an operon provides no selective benefit to the individual organism but does enhance the fitness of the gene cluster itself, as clusters can be efficiently inherited horizontally as well as vertically" [11]. Under SOM, operons are conceptually analogous to integrating viruses (phages), transposons and other mobile genetic elements, although coregulation of the genes in an operon could be an important selective factor that favors retention of operons during evolution.

Horizontal gene transfer (HGT) events have been classified into distinct categories of acquisition of new genes, acquisition of paralogs of existing genes and xenologous gene displacement whereby a gene is displaced by a horizontally transferred ortholog from another lineage (xenolog [12]). Each of these types of horizontal transfer is common among prokaryotes, but their relative contributions differ in different lineages. A potential central role of horizontal transfer in the evolution of operon organization of prokaryotic genomes is embodied in the ‘selfish operon model’ (SOM) [8-10]. This model posits that "the physical proximity of genes in an operon provides no selective benefit to the individual organism but does enhance the fitness of the gene cluster itself, as clusters can be efficiently inherited horizontally as well as vertically" [11]. Under SOM, operons are conceptually analogous to integrating viruses (phages), transposons and other mobile genetic elements, although coregulation of the genes in an operon could be an important selective factor that favors retention of operons during evolution.

Horizontal gene transfer (HGT) events have been classified into distinct categories of acquisition of new genes, acquisition of paralogs of existing genes and xenologous gene displacement whereby a gene is displaced by a horizontally transferred ortholog from another lineage (xenolog [12]). Each of these types of horizontal transfer is common among prokaryotes, but their relative contributions differ in different lineages. A potential central role of horizontal transfer in the evolution of operon organization of prokaryotic genomes is embodied in the ‘selfish operon model’ (SOM) [8-10]. This model posits that "the physical proximity of genes in an operon provides no selective benefit to the individual organism but does enhance the fitness of the gene cluster itself, as clusters can be efficiently inherited horizontally as well as vertically" [11]. Under SOM, operons are conceptually analogous to integrating viruses (phages), transposons and other mobile genetic elements, although coregulation of the genes in an operon could be an important selective factor that favors retention of operons during evolution.

Horizontal gene transfer (HGT) events have been classified into distinct categories of acquisition of new genes, acquisition of paralogs of existing genes and xenologous gene displacement whereby a gene is displaced by a horizontally transferred ortholog from another lineage (xenolog [12]). Each of these types of horizontal transfer is common among prokaryotes, but their relative contributions differ in different lineages. A potential central role of horizontal transfer in the evolution of operon organization of prokaryotic genomes is embodied in the ‘selfish operon model’ (SOM) [8-10]. This model posits that "the physical proximity of genes in an operon provides no selective benefit to the individual organism but does enhance the fitness of the gene cluster itself, as clusters can be efficiently inherited horizontally as well as vertically" [11]. Under SOM, operons are conceptually analogous to integrating viruses (phages), transposons and other mobile genetic elements, although coregulation of the genes in an operon could be an important selective factor that favors retention of operons during evolution.

To assess the role of horizontal gene transfer in the evolution of operons systematically, we undertook phylogenetic analysis of members of highly conserved gene neighborhoods that are predicted to constitute operons [25]. We focused primarily on mosaic operons in which one or more of the genes apparently have been transferred from distantly related species such that the phylogeny of the transferred genes is obviously incongruent with the phylogeny of the remaining genes in the respective operons.

Results and discussion

Identification of horizontal gene transfer

Experimental data on operons in organisms other than E. coli and, to a lesser extent, B. subtilis are scarce. Therefore we used conserved gene pairs and connected gene neighborhoods associated with them as an approximation of operon organization of genes in other prokaryotic genomes. Several studies have suggested strongly that all gene pairs that are conserved in multiple genomes belong to the same operon [7,25,26]. Here we used an extremely conservative threshold (conservation of a gene pair in 10 genomes) to ensure that only genuine operons were analyzed. BLASTP searches for potential horizontal gene transfer identified 729 candidate genes (9% of all genes comprising conserved neighborhoods in 41 analyzed genomes), that is, genes whose encoded protein sequences were more similar to homologs from phylogenetically distant taxa than to those from the reference taxon (it might be worth noting that, throughout this analysis, we treated genes as atomic units and did not consider the relatively unlikely possibility of HGT for portions of genes). Phylogenetic analysis of these genes and their neighbors revealed different types of evolutionary events, some of which involve whole operons, whereas others seem to reflect operon mosaicity.

Probable horizontal transfer of whole operons or large portions of operons, when phylogenetic trees for all genes in a predicted operon had the same topology (which, however, was incompatible with the species tree) was identified in 35 neighborhoods - approximately one third of all analyzed neighborhoods. These events were classified into three categories: acquisition of a new (for the given lineage) operon, paralogous operon acquisition and xenologous operon
displacement [13]. Examples of all these classes of apparent operon transfer events are given in Table 1. These 35 neighborhood generally represented functional classes of genes known to be prone to HGT: transporters, general metabolism-related genes and signal transduction systems [13,15,17]. This seems to be a relatively low level of horizontal transfer in view of the purported selfish behavior of operons [9,10]. However, the strict threshold, described above, on the detection of conserved gene pairs undoubtedly led to many horizontally transferred operons being missed. Thus, the present analysis gives a conservative low bound of operon transfer.

| Operon acquisition | Recipient organism and correspondent genes | Probable source | Other probable recipients | Comment |
|--------------------|--------------------------------------------|-----------------|---------------------------|---------|
| Pyruvate-ferredoxin oxidoreductase | Thermotoga maritima TM0015-TM0018 | Archaea | Aae, Hpy, Bha/Sau | Apparently, the related operon for 2-oxoisovalerate oxidoreductase (TM1758-TM1759) was also transferred from archaea |
| Sulfate/molybdate transport | Bacillus halodurans BH3128-BH3130 | Gram-negative bacteria | - | No other such operons in Bacillus-Clostridium group members |
| Putative effector of murein hydrolase | Pyrococcus horikoshii PH1801-PH1802 | Bacteria | Pab, Mac | |
| Allophanate hydrolase subunits | Pyrococcus horikoshii PH0987-PH0988 | Bacteria | Pab | |

| Paralogous operon acquisition | Recipient organism and correspondent genes | Probable source | Other probable recipients | Comment |
|-------------------------------|--------------------------------------------|-----------------|---------------------------|---------|
| Dipeptide transporter | Vibrio cholerae VCO620-VCO616 | Thermotoga/Archaea | Tma | It has several another bacterial operons including VC1091-VC1095 |
| Ribonucleotide reductase alpha and beta subunit | Halobacterium sp. VNG2384G VNG2383G | Bacteria | - | Additional to "archeal:" Ribonucleotide reductase alpha subunit VNG1644G, beta subunit is apparently lost |
| Aromatic amino-acid biosynthesis | Halobacterium sp. VNG0384G VNG0386G | Bacteria | - | Paralogs of this pair are VNG1646G-VNG1647G |

| Xenologous operon displacement | Recipient organism and correspondent genes | Probable source | Other probable recipients | Comment |
|-------------------------------|--------------------------------------------|-----------------|---------------------------|---------|
| Histidine biosynthesis suboperon | Pseudomonas aeruginosa PA3151-PA3152 | Epsilon-Proteobacteria | - | |
| Panthothenate synthesis | Campylobacter jejuni CJ0297-CJ0298c | Gram-positive bacteria | - | |
| DNA repair SbcDC | Vibrio cholerae VCA0520-VCA0521 | Gram-positive bacteria | - | |
| DNA gyrase A and B | Halobacterium sp. VNG0887G-VNG0889G | Bacteria | Hbs, Tac, Tvo, Afu, | |
| Dipeptide transporter | Streptococcus pyogenes SP2000-SPy2004 | Gamma-Proteobacteria | - | |
| Glutamate synthase complex | Thermotoga maritima TM0394-TM0398 | Archaea | - | There is another homolog for gene TM0397 of possible archaeal origin |
| NADH:ubiquinone oxidoreductase | Halobacterium sp. VNG0635G-VNG0637G | Bacteria | - | |
| Phosphate transporter | Methanothermobacter thermoautotrophicum MTH1727-MTH1734 | Bacteria | - | |

Table 1
Examples of horizontally transferred operons
In addition, 19 predicted operons with different phylogenetic affinities of the constituent genes, that is, apparent mosaic operons, were identified (Table 2). Again, this is definitely a low bound - not only because of the high threshold set for the identification of conserved gene pairs, but also because this number includes only cases that were clearly resolved by phylogenetic tree analysis. In addition, we detected many uncertain cases where the different phylogenetic affinities of genes within an operon were not strongly supported (data not shown); at least some of these are probably also mosaic operons.

Below we describe in greater detail several case studies of putative mosaic operons; in each of these cases, in addition to the basic set of 41 species, we included in the analysis the apparent orthologs of the respective proteins from all prokaryotic species in which they were detected, in order to control for possible effects of taxon sampling. We found that, although the details of tree topology inevitably depended on the set of species analyzed, the conclusions regarding HGT were not affected by the inclusion of additional species.

Case studies of mosaic operons

Ribosomal protein L29 gene

In the previous study that prompted this work, we analyzed the phylogeny of several ribosomal proteins and found several cases of apparent horizontal transfer resulting in mosaic operon organization [20]. Horizontal transfer “in the heart of the ribosome” also has been independently described by others [21,22]. Here we report another case of a ribosomal protein operon with apparent in situ gene displacement (that is, displacement without change of the local gene arrangement) via HGT. Figure 1a shows the highly conserved gene arrangement around the gene for the large subunit protein L29. The phylogenetic trees for the flanking L16 and S7 genes showed largely congruent topologies without any indications of HGT (Figure 1b,d). In contrast in the L29 tree, unexpected clustering is seen for Aquifex aeolicus and both Rickettsia: the Aquifex branch is within the archaeal cluster, whereas the Rickettsia group is with Chlamydia, rather than with the rest of alpha-proteobacteria: the taxon where Rickettsia belong (Figure 1c). In situ displacement is the most likely mechanism behind this observation given that the structure of this operon is conserved in the majority of bacteria. The nature of the selective advantages conferred by this gene substitution is unclear, but the apparent sources of the transferred genes suggest that the displacements indeed might be adaptive. Aquifex apparently acquired the L29 gene from archaea, which could be related to the adaptation to the hyperthermal conditions, whereas Rickettsia probably captured the gene from other parasitic bacteria, such as Chlamydia. However, these observations also allow a non-adaptationist interpretation, under which the apparent source of acquired genes simply reflects the increased likelihood of gene exchange between the respective organisms due to co-habitation, with chance fixation of some of the transferred genes.

The ruvB gene of Mycoplasma

The genes for Holliday junction resolvase subunits RuvA and RuvB form an operon that is conserved in most of the sequenced bacterial genomes (Figure 2a). In the phylogenetic trees for RuvA and RuvB, the branch that includes Ureaplasma and Mycoplasma occupies drastically different positions. In contrast to RuvA, which belongs to the Gram-positive clade as expected (Figure 2b), mycoplasmal RuvB clusters with the epsilon-proteobacteria (Helicobacter and Campylobacter) and the mycoplasma-epsilon-proteobacteria clade further joins alpha-proteobacteria (Figure 2c). This clustering is strongly supported by bootstrap analysis and was shown to be robust using statistical tests of tree topology (Table 3). Thus, the ruvB gene seems to have undergone xenologous displacement in situ after the divergence of the mycoplasmal branch from the rest of Gram-positive bacteria. Notably, the gene exchange seems to have occurred between phylogenetically distant parasitic bacteria.

Undecaprenyl pyrophosphate synthase gene in the lipid biosynthesis operon of Rickettsia

In Rickettsia, the undecaprenyl pyrophosphate synthase gene (uppS), which belongs to a highly conserved doublet of lipid biosynthesis genes embedded in functionally diverse operons (Figure 3a), clusters with an unexpected assemblage of bacterial orthologs, including those from the spirochete Treponema pallidum and Fusobacterium nucleatum, but not with the ‘native’ taxon, alpha-proteobacteria (Figure 3b,c). Statistical testing of the tree topology showed that clustering of rickettsial uppS with those from other alpha-proteobacteria is highly unlikely (Table 3). The apparent in situ gene displacement of the uppS gene in Rickettsia was accompanied by a breakdown of the operon into three fragments (Figure 3a). The topology of the uppS tree suggests the possibility of multiple HGT events, although only the rickettsial genomes show evidence of gene displacement in situ. The emergence of gene displacement in bacterial parasites is noted here again.

NADH:ubiquinone oxidoreductase subunits in Halobacterium sp.

Gene organization in the NADH:ubiquinone oxidoreductase operon is highly conserved in all sequenced archaeal genomes and those of several groups of bacteria (Figure 4a). The nuoI gene of Halobacterium sp. shows an unexpected phylogenetic affinity with proteobacteria (Figure 4c), whereas the neighboring genes have the regular archaeal affinities (Figure 4b,d). The unusual phylogeny of halobacterial NuoI, which was strongly supported by statistical tests (Table 3), suggests in situ displacement by a proteobacterial gene. Notably, all three NADH:ubiquinone oxidoreductase subunits of the cyanobacteria unexpectedly grouped within the archaeal clusters of the respective trees (Figure 4b-d). These observations point to a complex history of HGT for the genes encoding all subunits of NADH:ubiquinone oxidoreductase.
### Table 2

**Examples of probable mosaic operons**

| Cluster | Species | Predicted operon | General operon function | Horizontally acquired genes | Probable source of horizontally acquired genes | Functions of horizontally acquired genes |
|---------|---------|------------------|-------------------------|-----------------------------|-----------------------------------------------|------------------------------------------|
| Cluster 1* | *Rickettsia prowazekii*  
*Rickettsia conorii*  
*Aquifex aeolicus* | RP633-661, RC0980-1008  
Aq001-021 | Ribosomal operon  
Ribosomal operon | RP651  
Aq018a | Chlamydia  
Archaea | L29  
L29 |
| Cluster 2 | *Rickettsia prowazekii*  
*Rickettsia conorii*  
*Ureaplasma urealyticum*  
*Mycoplasma leprae* | RP800-804, RC1234-1238  
UU128-138  
ML1139-1146 | F0F1-type ATPase  
F0F1-type ATPase  
F0F1-type ATPase | RP804  
RC1238  
UU128, UU132_1, UU133, UU134 | Gram-positive bacteria  
Gram-negative bacteria  
Gram-negative bacteria | Delta subunit  
Epsilon subunit, alpha subunit, delta subunit, delta subunit |
| Cluster 3 | *Rickettsia prowazekii*  
*Rickettsia conorii* | RP134-139, RC175-180 | Ribosomal proteins, transcription antiterminator, SecE | RP134  
RC175 | Gram-positive bacteria  
Preprotein translocase subunit | SecE |
| Cluster 5 | *Aquifex aeolicus* | Aq1968_1_2 two domains | Histidine biosynthesis | Gram-negative bacteria | Phosphoribosyl-AMP cyclohydrolase |
| Cluster 8 | *Methanococcus jannaschii*  
*Methanobacterium thermoautotrophicum*  
*Halobacterium sp.*  
*Bacillus subtilis*  
*Bacillus halodurans* | MJ1037-1038  
MTH1655-1661  
VNG0305-0309  
PabB-folK BH0090-0095 | Tryptophan biosynthesis  
Tryptophan biosynthesis  
Tryptophan biosynthesis  
Tryptophan biosynthesis | MJ1037  
MTH1660  
VNG0307G  
PabB, BH0090 | Bacteria  
Gram-negative bacteria  
Bacteria  
Gram-negative bacteria | Tryptophan synthase beta chain  
Tryptophan synthase alpha chain  
Tryptophan synthase beta chain |
| Cluster 9 | *Halobacterium sp.* | VNG0635G-0647G | NADH:ubiquinone oxidoreductase | VNG0640G | Gram-negative bacteria  
NADH dehydrogenase-like protein |
| Cluster 18 | *Rickettsia prowazekii*  
*Rickettsia conorii* | RP423-425, RC0588-0590 | Lipid metabolism | RP425, RC0590 | Spirochetes  
Undecaprenyl pyrophosphate synthase |
| Cluster 27 | *Halobacterium sp.* | VNG1306G-1310G | Succinate dehydrogenase/ fumarate reductase | VNG1310G | Actinobacteria  
Succinate dehydrogenase subunit C |
Table 2 (Continued)

Examples of probable mosaic operons

| Cluster 29 | Mycoplasma genitalium | MG461-466 MPN677-682 | Housekeeping | MG466 MPN682 | Gram-negative bacteria | Ribosomal protein L34 |
|------------|-----------------------|----------------------|--------------|-----------------|------------------------|-------------------------|

| Cluster 34 | Thermotoga maritima | TM0548-0556 | Leucine/isoleucine biosynthesis | TM0552 TM0554 | 2-Isopropylmalate synthase 3-Isopropylmalate dehydratase, small subunit 3-Isopropylmalate dehydratase, large subunit |
|------------|----------------------|-------------|---------------------------------|----------------|-------------------------|-------------------------|

| Cluster 41 | Thermotoga maritima | TM1243-1251 | Nucleotide metabolism | TM1243 | Archaea | Phosphoribosylaminomidazole-succinocarboxamide synthase |
|------------|---------------------|-------------|----------------------|--------|---------|-----------------------------------------------|

| Cluster 42 | Lactococcus lactis | L0104-0108 | Arginine biosynthesis | L0107 | Gram-negative bacteria | Acetylglutamate kinase |
|------------|------------------|-----------|----------------------|-------|------------------------|------------------------|

| Cluster 48 | Borrelia burgdorferi | BB0054-0061 | Carbohydrate metabolism (glycolysis, gluconeogenesis) | BB0057 | Gram-positive bacteria | Glyceraldehyde-3-phosphate dehydrogenase |
|------------|---------------------|------------|-----------------------------|--------|------------------------|-----------------------------------------------|

| Cluster 54 | Thermotoga maritima | TM1780-1785 | Arginine biosynthesis | TM1780 | Gram-negative bacteria | Argininosuccinate synthase |
|------------|---------------------|------------|----------------------|--------|------------------------|----------------------------|

| Cluster 63 | Mycoplasma pneumoniae, Mycoplasma genitalium | MPN573-574 MG391-392 | Molecular chaperones | MPN574 MG393 | Gram-negative bacteria | Heat shock protein (groES) |
|------------|------------------------------------------------|----------------------|-----------------|-------------------------|------------------------|------------------------|

| Cluster 82 | Mycoplasma pneumoniae, Mycoplasma genitalium | MPN535-536 MG358-359 | DNA replication, recombination and repair | MPN536 MG359 | Gram-negative bacteria | Holliday junction resolvase helicase subunit |
|------------|-----------------------------------------------|----------------------|-----------------|-------------------------|------------------------|-------------------------|
Lipopolysaccharide biosynthesis operon in Methanothermobacter thermoautotrophicus and Deinococcus radiodurans

The genes of the lipopolysaccharide biosynthesis (rfbABCD) operon appear to have been extensively and independently shuffled in many prokaryotic genomes and might have undergone multiple horizontal transfers. This conclusion is supported both by examination of the operon organization (Figure 5a) and by phylogenetic tree analysis (Figure 5b-e). The trees showed a clear affinity between the rfbA, rfbB, rfbC genes of Methanothermobacter thermoautotrophicum and Clostridium acetobutylicum (Figure 5b-d), with Fusobacterium nucleatum and Listeria monocytogenes joining the cluster in the case of rfbB (Figure 5b), whereas M. thermoautotrophicum RfbD clustered with its archaeal orthologs as expected (Figure 5e). The genes of the rfbABCD operon in Methanothermobacter are shuffled compared to the probable ancestral order, which is found in many bacteria and C. acetobutylicum also shows a rearrangement (Figure 5a). One likely scenario in this case is that M. thermoautotrophicum acquired the rfbABCD operon with the typical gene order from a bacterium of the clostridial lineage, which was followed by displacement of three resident genes and loss of one of the invading genes, accompanied by operon rearrangement. An alternative scenario is that the rearrangement occurred in the source bacterium of the clostridial group and Methanothermobacter acquired only the rfbABC portion, which might have inserted head-to-tail downstream of the original operon, followed by elimination of the resident rfbABC (Figure 5a).

Another interesting case of mosaic structure of the same operon is seen in Deinococcus radiodurans (Figure 5a). Deinococcus RfbA shows clear affinity with proteobacteria (Figure 5d), whereas RfbD is of archaeal descent (Figure 5e), with RELL analysis revealing no competing topologies (Table 3). The remaining two genes of this operon in Deinococcus, rfbB (DRA0041) and rfbC (DRA0043), have uncertain phylogenetic affinities (Figure 5b, 5c). Thus, as in the case of M. thermoautotrophicus, this operon in Deinococcus was apparently formed through at least two events of xenologous gene displacement in situ and gene shuffling.

Leucine/isoleucine biosynthesis operon

Perhaps the most prominent case of mosaic operon organization is the leucine/isoleucine biosynthesis operon of several bacteria and archaea, particularly Thermotoga maritima.
Figure 1 (see legend on next page)
This is the only known branched chain amino acid biosynthesis operon, and it is partly conserved in a wide range of bacteria (Figure 6a). Following initial indications from the analysis of taxon-specific BLAST hits, we constructed phylogenetic trees for each of the genes of this operon. Unlike other bacteria, *Thermotoga* has two *leuA* paralogs, which are adjacent in the operon. The genes encoded by these paralogous genes show clearly distinct phylogenetic affinities: TM0552 belongs to a distinct clade within the archaeal domain, whereas TM0555 is part of a Gram-positive bacterial cluster (Figure 6b). This phylogenetic mosaic in *Thermotoga* extends further, with *LeuB* (TM0556) clustering with proteobacterial orthologs (Figure 6c), and *LeuC* (TM0554) and *LeuD* (TM0555) with archaeal orthologs (Figure 6d,e). All these affinities were strongly supported by two versions of bootstrap analysis (Table 3). The genes encoding *LeuA*, *LeuC*, and *LeuD* from *Thermotoga*, *Clostridium*, *Aquifex* and both *Pyrococcus abyssi* and *P. furiosus* belong to a well-defined clade, which also includes a medley of alpha-proteobacteria and cyanobacteria, within the archaeal domain in the respective trees (Figure 6b-e). Thus, this sub-operon apparently has been relatively recently horizontally spread among these organisms. *Pyrococcus abyssi* and *P. furiosus* probably acquired these genes after the divergence from the common ancestor with *P. horikoshii* because the latter has only the typical archaeal operon (Figure 6a).

Given the apparent propensity of *Thermotoga* (and other hyperthermophilic bacteria) for acquisition of archaeal genes via HGT, it seems most likely that the archaeal version of the *leuACD* suboperon originally entered the bacterial domain via *Thermotoga* or a related thermophilic bacterium. Formally, in *Thermotoga* these events could be classified as a combination of paralogous (sub)operon acquisition (TM0554-TM0555 in addition to another paralogous archaeal gene pair TM0291-TM0292) and xenologous gene displacements (genes TM0553, TM0556). In *Clostridium*, xenologous operon displacement seems to have occurred because the ancestral operon of the Gram-positive type apparently had been lost. The subsequent evolution of this operon in the four organisms proceeded along different paths. *Aquifex* has lost the operon structure even for the two subunits of 3-isopropylmalate dehydratase (*LeuB*, *LeuD*). Different genes in the operons of *P. abyssi* and *C. acetobutylicum* have been translocated and several genes probably have been independently accrued (Figure 6a). In both *P. abyssi* and *Thermotoga*, the original *LeuA* and *LeuB* genes within the *leuABCD* core seem to have been independently displaced by bacterial orthologs without a clear affinity.

Table 3

| Tree* | Diff lnL† | S.E.‡ | RELL-BP§ |
|-------|-----------|-------|----------|
| L19 original | 0.0 | ML | 0.8004 |
| IR2 | -12.6 | 7.7 | 0.0480 |
| 3R4 | -6.6 | 6.6 | 0.1516 |
| RuvB original | 0.00 | ML | 0.9631 |
| IR2 | -27.1 | 15.4 | 0.0369 |
| UppS original | 0.00 | ML | 0.9883 |
| IR2 | -29.3 | 12.8 | 0.0117 |
| NuoH original | 0.00 | ML | 0.8336 |
| IR2 | -7.4 | 7.9 | 0.1664 |
| RfbA original | 0.00 | ML | 1.0000 |
| IR2 | -151.1 | 25.0 | 0.0000 |
| RfdD original | 0.00 | ML | 0.9005 |
| IR2 | -17.0 | 13.3 | 0.0995 |
| LeuA original | 0.00 | ML | 1.0000 |
| IR2 | -150.2 | 25.8 | 0.0000 |
| 3R4 | -418.6 | 31.5 | 0.0000 |
| 5R6 | -245.0 | 27.8 | 0.0000 |
| LeuB original | 0.00 | ML | 0.9847 |
| IR2 | -52.9 | 18.1 | 0.0007 |
| 3R4 | -31.7 | 14.9 | 0.0146 |
| LeuC original | 0.00 | ML | 1.0000 |
| IR2 | -302.7 | 31.6 | 0.0000 |
| 3R4 | -493.1 | 32.1 | 0.0000 |
| LeuD original | 0.00 | ML | 1.0000 |
| IR2 | -66.6 | 17.2 | 0.0000 |
| 3R4 | -76.7 | 16.8 | 0.0000 |

*The numbers refer to local rearrangements of the tree as indicated on the corresponding figures. †Difference of the Log-likelihoods relative to the best tree. ‡Standard error of Diff lnL. §Bootstrap probability of the given tree calculated using the RELL method (Resampling of Estimated Log-likelihoods).
Figure 2

In situ displacement of the ruvB gene in Mycoplasma. (a) Organization of the Holliday junction resolvosome operon and surrounding genes in bacteria. COG0632 - Holliday junction resolvosome, DNA-binding subunit, COG2255 - Holliday junction resolvosome, DNA-binding subunit, COG0817 - Holliday junction resolvosome, endonuclease subunit, COG0392 - Predicted integral membrane protein, COG0282 - acetate kinase, COG0839 - NADH:ubiquinone oxidoreductase subunit 6 (chain J), COG0244 - ribosomal protein L10, COG0732 - restriction endonuclease S subunits, COG0809 - S-adenosylmethionine:tRNA-ribosyltransferase-isomerase, COG0772 - bacterial cell division membrane protein, COG0624 - acetylornithine deacetylase/succinyl-diaminopimelate desuccinylase and related deacylases, COG1487 - predicted nucleic acid-binding protein, COG1132 - ABC-type multidrug transport system, ATPase and permease components, COG0442 - prolyl-tRNA synthetase, COG0323 - DNA mismatch repair enzyme, COG1408 - predicted phosphohydrolases. The designations are as in Figure 1a. For species abbreviations, see Materials and methods. (b,c) Unrooted maximum-likelihood tree for RuvA (b) and RuvB (c); the designations are as in Figure 1b.
Figure 3

Genes with different phylogenetic affinities in the lipid biosynthesis operon of *Rickettsia*. (a) Organization of the lipid biosynthesis operon and surrounding genes in *Rickettsia prowazekii* and *Rickettsia conorii* (operons from three other alpha-proteobacteria are shown for comparison). COG0020 - undecaprenyl pyrophosphate synthase, UppS; COG0575 - CDP-diglyceride synthetase; COG0750 - predicted membrane-associated Zn-dependent proteases; COG233 - ribosome recycling factor; COG0528 - uridylic kinase; COG0745 - OmpR-like response regulator; COG0642 - signal transduction histidine kinase; COG0729 - outer membrane protein; COG2919 - septum formation initiator; COG0743 - 1-deoxy-D-xylulose 5-phosphate reductoisomerase.

The designations are as in Figure 1a. For species abbreviations, see Materials and methods. (b,c) Unrooted maximum-likelihood tree for UppS (b) and CdsA (c); the designations are as in Figure 1b.
Figure 4 (see legend on next page)
with any specific bacterial lineage (Figure 6a). The most likely scenario for evolution of this operon in *Thermotoga* is that it originated as a Gram-positive type operon and subsequently many genes (or sub-operons) have been displaced in situ through multiple horizontal transfers and a few additional genes have been inserted into the preexisting structure. The alternative but less likely hypothesis involves independent, de novo operon assembly from genes of different phylogenetic affinities. Several other apparent HGT events were detected during the analysis of the phylogenetic trees for leucine biosynthesis genes (DR1614 in *LeuD* tree, DR1610 in *LeuC* tree (Figure 6d,e)) but, in these cases, the acquired genes do not belong to conserved operons.

Conclusions

Intragenomic plasticity and inter-species horizontal mobility of operons are thought to be important facets of prokaryotic genome evolution. Indeed, the results presented here indicate that horizontal transfer of entire operons is the most likely explanation for most of the findings of co-localized ‘alien’ genes in a genome, which is generally consistent with SOM. However, a substantial fraction - approximately 35% - of operons with indications of horizontal transfer events appear to consist of genes with different phylogenetic affinities. Bearing artifacts of phylogenetic analysis, which can never be ruled out completely, but appear unlikely given the strong statistical support for the anomalous placement of the genes in question in phylogenetic trees, two evolutionary scenarios for the origin of such mosaic operons are conceivable. The first involves de novo assembly of operons, in part from genes acquired via HGT, whereas the second one postulates in situ xenologous displacement of genes within a resident operon. Analysis of mosaic operons suggested that both scenarios might apply, but in situ displacement is likely to be more frequent. In several cases, in situ displacement seems to have occurred between genomes of distantly related parasitic bacteria that might have shared a host. A sequence of events that is often considered as an alternative to HGT is an ancient duplication with subsequent differential loss of paralogs. However, in the cases analyzed here, this seems to be a particularly remote possibility because a tandem duplication followed by lengthy evolution of both paralogs within the operon would be required to mimic in situ displacement. Tandem pairs of paralogs are uncommon in operons and such a ‘smoking gun’ was not observed in any of the suspected cases of in situ displacement.

At first glance, in situ gene displacement seems highly unlikely: given the vast evolutionary distance separating the donor and recipient genomes, homologous recombination is out of the question. In cases when the displacing gene(s) is located on the periphery of an operon (for example, Figure 5a), a plausible mechanism could involve initial insertion of the invading gene in the vicinity of the resident operon, followed by deletion of intervening genes (provided these are non-essential). However, when the displacing gene is tucked between resident ones (for example, Figures 4a, 6a), displacement must have occurred with surgical precision. The only conceivable explanation seems to be that HGT is extremely common in the evolution of prokaryotes and so is intragenomic recombination, which provides for rare chance occurrences of in situ displacement. Conceivably, a horizontally acquired gene that displaces the resident ortholog without disruption of operon organization would have its chances of evolutionary fixation greatly increased, hence the apparent disproportional survival of the displacing genes. This explanation does not refute SOM as the conceptual framework explaining the origin of operons but emphasizes the ‘altruistic’ aspect of the evolution of operons whereby the operon integrity is maintained by strong purifying selection at the organism level.

Materials and methods

Sequence data

Amino acid sequences from 41 completely sequenced prokaryotic genomes were extracted from the Genome division of the Entrez retrieval system [27] and used as the master species set for this analysis. Bacterial species abbreviations: *Aquifex aeolicus* (Aae), *Bacillus halodurans* (Bha), *Bacillus subtilis* (Bsu), *Streptococcus pyogenes* (Spy), *Staphylococcus aureus* (Sau), *Clostridium acetobutylicum* (Cab), *Borrelia burgdorferi* (Bbu), *Campylobacter jejuni* (Cje), *Chlamydia trachomatis* (Ctr), *Chlamydophila pneumoniae* (Cpn), *Deinococcus radiodurans* (Dra), *Escherichia coli* (Eco), *Haemophilus influenzae* (Hin), *Helicobacter pylori* (Hpy), *Lactococcus lactis* (Lla), *Mesorhizobium loti* (Mlo),...
Figure 5  (see legend on next page)
**Mycoplasma genitalium** (Mge), **Mycoplasma pneumoniae** (Mpn), **Mycobacterium leprae** (Mle), **Pasteurella multocida** (Pmu), **Neisseria meningitidis** (Nme), **Pseudomonas aeruginosa** (Pae), **Rickettsia prowazekii** (Rpr), **Rickettsia conorii** (Rco), **Synechocystis PCC6803** (Ssp), **Thermotoga maritima** (Tma), **Treponema pallidum** (Tpa), **Vibrio cholerae** (Vch), **Xylella fastidiosa** (Xfa), **Buchnera sp.** (Bsp), **Caulobacter crescentus** (Ccr), and **Ureaplasma urealyticum** (Uur). Archaeal species abbreviations: **Aeropyrum pernix** (Ape), **Archaeoglobus fulgidus** (Afu), **Halobacterium sp.** (Hsp), **Methanothermobacter thermoautotrophicum** (Mth), **Methanococcus jannaschii** (Mja), **Pyrococcus horikoshii** (Pho), **Pyrococcus abyssi** (Pab), **Thermoplasmia volcanium** (Tvo), **Thermoplasmia acidophilum** (Tac), **Sulfolobus solfataricus** (Sso). In addition, the following species were included in the case studies described in the text; bacteria: **Agrobacterium tumefaciens** (Atu), **Bifidobacterium longum** (Blo), **Brucella melitensis** (Rso), **Chlorobium tepidum** (Cte), **Enterococcus faecalis** (Efa), **Fusobacterium nucleatum** (Fnu), **Lactobacillus plantarum** (Lpl), **Leptospira interrogans** serovar (Lint), **Listeria innocua** (Lin), **Listeria monocytogenes** (Lmo), **Nitrosomonas europaea** (Neu), **Nos tor sp.** (Nsp), **Oceanobacillus iheyensis** (Oih), **Ralstonia solanacearum** (Rso), **Sino r hizobium meliloti** (Sme), **Streptomyces coelicolor** (Sco), **Thermoanaerobacter tengcongensis** (Tte), **Thermosynechococcus elongatus** (Tel), **Xanthomonas campestris** (Xea), **Shewanella oneidensis** (Son); archaea: **Methanopyrus kandleri** (Mka), **Methanosarcina acetivorans** (Mac), **Pyrococcus furiosus** (Pfu).

**Figure 5** (continued from previous page)

Genes with different phylogenetic affinities in the lipopolysaccharide biosynthesis operon of **Methanothermobacter thermoautotrophicum** and **Deinococcus radiodurans**. (a) Organization of the lipopolysaccharide biosynthesis operon in different prokaryotes. COG1091 - dTDP-4-dehydrorhamnose reductase; COG1209 dTDP-glucose pyrophosphorylase; COG1898 - dTDP-4-dehydrorhamnose 3,5-epimerase and related enzymes; COG1088 - dTDP-D-glucose 4,6-dehydratase. The designations are as in Figure 1a. For species abbreviations, see Materials and methods. (b-e) Unrooted maximum-likelihood tree for **RfbB** (b), **RfbC** (c), **RfbA** (d) and **RfbD** (e); the designations are as in Figure 1b.
Figure 6 (see legend on next page)
Reconstruction of gene neighborhoods

Gene neighborhoods for the 41 compared genomes were reconstructed as previously described [25]. Briefly, the collection of clusters of orthologous groups of proteins from complete genomes (COGs) [28] was used as the source of information on orthologous relationships for detecting conserved gene pairs. For the purpose of this analysis only 'highly conserved' gene pairs were considered, that is, those formed by genes from two COGs that were present in the same orientation and separated by less than three genes in at least 10 of the compared genomes. This conservative approach was adopted in order to ensure that all analyzed gene pairs belong to the same operon. At the next step, overlapping gene pairs were joined in triplets; each triplet was required to exist in at least one genome. Overlapping triplets were used to construct gene arrays by run search in an oriented graph; a gene array may or may not be found in its entirety in any available genome. Finally, gene arrays that shared at least three COGs were clustered into neighborhoods by using a single-linkage clustering algorithm [25]. Conserved gene pairs that did not belong to the reconstructed gene arrays were also analyzed.

Searching for candidate horizontally transferred genes

The protein sequences encoded by the genes of each neighborhood were searched against the non-redundant protein sequence database (NCBI, NIH, Bethesda) using the BLASTP program. The BLAST hits were analyzed to identify their potential phylogenetic affinity. For each protein, the best hits were identified to the taxon to which the given species belongs (hereinafter, reference taxon) and to other major taxa; hits to closely related species were disregarded (see Table S1 in the additional data file). Proteins that had more significant (lower E-value) hits to a non-reference taxon than to the reference taxon were considered candidates for horizontal transfer and the respective orthologous protein clusters were subject to further phylogenetic analysis as described in the next section. If phylogenetic analysis indicated that a particular gene was likely to be horizontally transferred,
Phylogenetic analysis

Multiple protein sequence alignments were constructed using the T-Coffee program [29] and positions containing >70% gaps were excluded. Distance trees were constructed by using the least-square method as implemented in the FITCH program of the PHYLIP package [30,31]. The least-square trees were subjected to maximum-likelihood local rearrangement using the ProtML program of the MOLPHY package, with the JTT-F model of amino acid substitutions [32,33]. The resulting trees are a surrogate for maximum-likelihood phylogenies; exhaustive maximum-likelihood tree construction is impractical for the number of species analyzed here. Bootstrap analysis was performed for each maximum-likelihood tree using the Resampling of Estimated Log-Likelihoods (RELL) method as implemented in MOLPHY [32-34]. Alternative placements of selected clades in maximum-likelihood trees were compared by using the rearrangement optimization (Kishino-Hasegawa) method as implemented in the ProtML program [34].

Additional data file

Additional data, including schematics of operon organization and phylogenetic trees for all gene clusters listed in Table 2, are available in an additional data file (Additional data file 1).

Acknowledgements

We thank Jeffrey Lawrence for critical reading of the manuscript. Marina V. Omelchenko is supported by a grant from the US Department of Energy (Office of Biological and Environmental Research, Office of Science) grants DE-FG02 01ER62220 from the Genomes to Life Program.

References

1. Jacob F, Monod J: Genetic regulatory mechanisms in the synthesis of proteins. J Mol Biol 1961, 3:318-356.
2. Miller JH, Reznikoff WSE: The Operon. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory; 1978.
3. Mushagian AR, Koonin EV: Gene order is not conserved in bacterial evolution. Trends Genet 1996, 12:289-290.
4. Davdekar T, Snel B, Huynen M, Bork P: Conservation of gene order: a fingerprint of proteins that physically interact. Trends Biochem Sci 1998, 23:324-328.
5. Watanabe H, Mori H, Itoh T, Gojobori T: Genome plasticity as a paradigm of eubacterial evolution. J Mol Biol 1997, 4:457-564.
6. Itoh T, Takemoto K, Mori H, Gojobori T: Evolutionary instability of operon structures disclosed by sequence comparisons of complete microbial genomes. Mol Biol Evol 1999, 16:332-346.
7. Wolf YI, Rogozin IB, Kondrashov AS, Koonin EV: Genome alignment, evolution of prokaryotic genome organization, and prediction of gene function using genomic context. Genome Res 2001, 11:356-372.
8. Lawrence JG: Shared strategies in gene organization among prokaryotes and eukaryotes. Cell 2002, 110:407-413.
9. Lawrence JG, Roth JR: Selfish operons: horizontal transfer may drive the evolution of gene clusters. Genetics 1996, 143:1843-1860.
10. Lawrence J: Selfish operons: the evolutionary impact of gene clustering in prokaryotes and eukaryotes. Curr Opin Genet Dev 1999, 9:642-648.
11. Lawrence JG: Selfish operons and specification by gene transfer. Trends Microbiol 1997, 5:355-359.
12. Gray GS, Fitch WM: Evolution of antibiotic resistance genes: the DNA sequence of a kanamycin resistance gene from Staphylococcus aureus. Mol Biol Evol 1983, 1:57-66.
13. Kooin EV, Makarova KS, Aravind L: Horizontal gene transfer in prokaryotes - quantification and classification. Annu Rev Microbiol 2001, 55:709-742.
14. Aravind L, Tatusov RL, Wolf YI, Walker DR, Koonin EV: Evidence for massive gene exchange between archaeal and bacterial hyperthermophiles. Trends Genet 1998, 14:442-444. A published erratum appears in Trends Genet 1998, 15:41.
15. Jain R, Rivera MC, Lake JA: Horizontal gene transfer among genomes: the complexity hypothesis. Proc Natl Acad Sci USA 1999, 96:3801-3806.
16. Koonin KE, Clayton RA, Gill SR, Gwinn ML, Dodson RJ, Haft DH, Hickey EK, Peterson JD, Nelson WC, Ketchum KA, et al: Evidence for lateral gene transfer between archaea and bacteria from genome sequence of Thermotoga maritima. Nature 1999, 399:323-329.
17. Doolittle WF: Lateral genomics. Trends Cell Biol 1999, 9:MS-M8.
18. Garcia-Valle S, Romeu A, Palau J: Horizontal gene transfer in bacterial and archaeal complete genomes. Genome Res 2000, 10:1719-1725.
19. Gogarten JP, Doolittle WF, Lawrence JG: Prokaryotic evolution in light of gene transfer. Mol Biol Evol 2002, 19:2226-2238.
20. Makarova KS, Ponomarev VA, Koonin EV: Two or not two C: recurrent disruption of Zn-rubins, gene duplication, lineage-specific gene loss, and horizontal gene transfer in evolution of bacterial ribosomal proteins. Genome Biol 2001, 2:research0033.1-0033.14.
21. Broclier C, Philippe H, Moreira D: The evolutionary history of ribosomal protein RpS14: horizontal gene transfer at the heart of the ribosome. Trends Genet 2000, 16:529-533.
22. Broclier C, Batistea E, Moreira D, Philippe H: Eubacterial phylogeny based on translational apparatus proteins. Trends Genet 2002, 18:1-5.
23. Davies RL, Campbell S, Whitam TS: Mosaic structure and molecular evolution of the leukotoxin operon (lktCABD) in Mannheimia (Pasteurella) haemolytica, Mannheimia glucosidica, and Pasteurella trehalosi. J Bacterial 2002, 184:266-277.
24. Schnaidtman CA, Klena JD: Genetics of lipopolysaccharide biosynthesis in enteric bacteria. Microbiol Rev 1993, 57:655-682.
25. Rogozin IB, Makarova KS, Muravi J, Czabarka E, Wolf YI, Tatusov RL, Szekely LA, Koonin EV: Connected gene neighborhoods in prokaryotic genomes. Nucleic Acids Res 2002, 30:212-223.
26. Snel B, Lehmann G, Bork P, Huynen MA: STRING: a web-server to retrieve and display the repeatedly occurring neighbour-hood of a gene. Nucleic Acids Res 2000, 28:3442-3444.
27. ENTREZ Genome [http://www.ncbi.nlm.nih.gov/PMGifs/Genomes.org.html]
28. Tatusov RL, Natale DA, Garkavtsev IV, Tatusova TA, Shankavaram UT, Rao BS, Kryutin B, Galperin MY, Fedorova N, Koonin EV: The COG database: new developments in phylogenetic classification of proteins from complete genomes. Nucleic Acids Res 2001, 29:22-28.
29. Notredame C, Higgins DG, Heringa J: T-Coffee: a novel method for fast and accurate multiple sequence alignment. J Mol Biol 2000, 302:205-217.
30. Fitch WM, Margoliash E: Construction of phylogenetic trees. Science 1967, 155:279-284.
31. Felsenstein J: Inferring phylogenies from protein sequences by parsimony, distance, and likelihood methods. Methods Enzymol 1996, 266:418-427.
32. Hasegawa M, Kishino H, Saitou N: On the maximum likelihood method in molecular phylogenetics. J Mol Biol 1991, 215:368-389.