J Mol Evol (2002) 55:632–637
DOI: 10.1007/s00239-002-2358-5

JOURNAL OF MOLECULAR EVOLUTION

Simultaneous Horizontal Gene Transfer of a Gene Coding for Ribosomal Protein L27 and Operational Genes in *Arthrobacter* Sp.

Santiago Garcia-Vallvé, Francesc X. Simó, Miguel A. Montero, Lluís Arola, Antoni Romeu

Evolutionary Genomics Group, Department of Biochemistry and Biotechnology, Rovira i Virgili University, Pl. Imperial Tàrraco 1, E-43005 Tarragona, Catalonia, Spain

Received: 11 February 2002 / Accepted: 6 June 2002

Abstract. Phylogenetic analysis of bacterial L27 ribosomal proteins showed that, against taxonomy, the L27 protein from the Actinobacteria *Arthrobacter* sp. clusters with protein sequences from the *Bacillus* group. The L27 gene clusters in the *Arthrobacter* sp. genome with six genes responsible for creatinine and sarcosine degradation. Phylogenetic analyses of orthologue proteins encoded by three of these genes also showed a phylogenetic relationship with *Bacillus* species. Comparisons between the synonymous codon usage of the *Arthrobacter* sp. genes and those from complete genomes showed that *Arthrobacter* genes encoding the L27 ribosomal protein and the proteins responsible for the degradation of creatinine and sarcosine have a codon usage that is more similar to that of *Bacillus* species than that of *Arthrobacter*. We suggest that the *Arthrobacter* sp. genes encoding the L27 ribosomal protein and the proteins responsible for the degradation of creatinine and sarcosine were acquired simultaneously through horizontal gene transfer from an unknown *Bacillus* species.

Key words: Horizontal gene transfer — Codon bias — L27 ribosomal protein

Introduction

Horizontal gene transfer (HGT) is the physical exchange of genetic material between genomes. It has long been recognized as a promoter of prokaryotic genome evolution (Doolittle 1999; Koonin et al. 2001; Brochier et al. 2002). As most prokaryotic genes are regulated in operons, a given set of clustered genes on the chromosome would probably be mobilized in block if they were the object of an HGT event. The selective advantage would guide the fate of the transferred genes once they had been incorporated into the receiving genome (Ochman et al. 2000). The complexity hypothesis of HGT among genomes states that extensive horizontal transfer occurs for genes involved in housekeeping (operationals), whereas those involved in transcription, translation, and related processes (informationals) are seldom horizontally transferred (Jain et al. 1999).

One group of informational genes is those that encode the ribosomal proteins (r-proteins). Bacterial ribosomes contain about 35% protein, and the variety of r-proteins is enormous. In bacteria, an increased cellular demand for protein synthesis is met by increasing the number of ribosomes rather than by altering the activity of individual ribosomes (Green and Noller 1997). The proportion of cellular resources devoted to making ribosomes is so large, and the function of ribosomes so important, that cells must coordinate the synthesis of r-proteins and rRNA (Green and Noller 1997). Genes that encode the r-proteins occur in operons, which also contain the genes for the subunits of DNA primase, RNA polymerase, and protein elongation factors (Nomura 1984; Moore 1998). Individually, most of these r-proteins are small and highly conserved. Given the

Correspondence to: A. Romeu; *email:* romeu@quimica.urv.es

coordination and the conserved organization of genes coding r-proteins, HGT is considered to be rare for these genes (Makarova et al. 2001). In contrast, multiple HGT events seem to have affected the r-protein RpS14 (Brochier et al. 2000) and the evolution of other bacterial r-proteins deviates from straightforward vertical inherence and includes cases of differential gene loss and HGT (Makarova et al. 2001).

In this study we use molecular phylogeny and codon bias to assess the occurrence of HGT for a fragment of the *Arthrobacter* sp. genome. These regions include a gene coding for the r-protein L27 and a set of genes involved in catabolic pathways of creatinine and sarcosine. L27 belongs to a family of r-proteins which, on the basis of sequence similarities, groups eubacterial L27, plant chloroplast L27, algal chloroplast L27, and yeast mitochondrial L27 (Elhag and Bourque 1992) and is one of the proteins needed to assemble the large ribosomal subunit. The transfer of a group of genes may allow some informational ones to be acquired and maintained in the population. This seems to be the case for the *Arthrobacter* sp. L27 gene. Evolutionarily proximity between the *Arthrobacter* sp. sarcosine oxidase gene and that of the *Bacillus* sp. was suggested when the sarcosine oxidase gene from *Arthrobacter* sp. was cloned and sequenced (Nishida and Imanaka 1993).

Materials and Methods

Protein and gene sequences were imported from the SwissProt and EMBL databases, respectively. Protein homologues were obtained by Blast search (Altschul et al. 1997). Sequences whose expected Blast e-values were less than e-10 were used for phylogeny. The sequence alignment of homologues was performed using the ClustalW software package (Thompson et al. 1994). The distance matrixes that served as input for neighbor-joining analyses were calculated using the point-accepted-mutation (PAM) Dayhoff substitution model. Assessed clade strength was calculated by bootstrap using 1000 replications. As a method of confirmation, maximum likehood trees were constructed using the Seqboot, Proml and Consense programs, as implemented in the phylogenetic package PHYLIP (Felsenstein 1988).

We compared codon usage by calculating the relative synonymous codon usage (RSCU) values for each gene. Codon usage analysis was made with (i) 59 genes from the Gram-positive high-G + C bacterium *Arthrobacter* sp. that correspond with all available genes except fragments and genes from plasmid pA3 (EMBL:AJ131246); (ii) a random sample of 60 genes from the Gram-positive high-G + C bacterium *Mycobacterium tuberculosis* H37Rv; (iii) a random sample of 60 genes from *Bacillus subtilis*; and (iv) a set of 25 highly expressed genes from *B. subtilis* defined by Garcia-Vallvé et al. (1999), plus the *rpmA* gene, which encodes the *B. subtilis* ribosomal protein L27. The genes of the random sample from *M. tuberculosis* and *B. subtilis* were selected from the genes of those species that were longer than 600 bp, were not predicted as horizontal transfer genes in our previous prediction (Garcia-Vallvé et al. 2000b), and did not deviate significantly in any amino acid or codon composition from the mean values (Garcia-Vallvé et al. 2000b). We compared the Pearson linear correlation coefficient (r) between the RSCU values of each pair of genes. As an arbitrary measurement of the pairwise distance (D) between the codon usage of genes, we used the following linear transformation of r: $D = (1 - |r|) \cdot 100$ (Garcia-Vallvé et al. 1999). Distances were located in a triangular matrix. We could then draw a dendrogram using the UPGMA algorithm. RSCU values were also compared for each *Arthrobacter* sp. gene with the mean values from bacterial complete genomes using the χ^2 value. The χ^2 value is a conventional measure of the difference between observed and expected values. In this analysis the minimum χ^2 value was considered as indicative of the putative origin of a gene.

Results

In the phylogenetic analyses of L27 r-protein, sarcosine oxidase, and creatinase protein sequences, the neighbor-joining and maximum likelihood tree topologies were highly consistent. The phylogenetic tree and the C-terminal region of the multiple alignment of homologues of *Arthrobacter* sp. L27 r-protein from bacteria are shown in Fig. 1. Although L27 ribosomal proteins are small (the complete multiple alignment used for tree construction contained 114 positions) and highly conserved, the monophyly of some of the bacterial phyla is retrieved. The relationships of some of these phyla, especially for the Proteobacteria species, which are not monophyletic, are not evolutionarily reliable as shown by the short branch length and low bootstrap values. According to the sequence availability, not all the bacterial taxa are equally represented in this study. *Aquificales*, *Thermotogales*, green sulfur bacteria, and the *Deinococcus* group are represented by only one L27 r-protein. Nevertheless, the phylogenetic analysis makes a distinction among the Cyanobacteria, Chlamydiales, Proteobacteria, and Gram-positive low- or high-G + C bacterial lineages. Several amino acid residues are strictly conserved along the multiple alignment (data not shown) and the only gaps found in the multiple alignment are located in the C-terminal region, after the last strictly conserved glycine residue. A distinctive feature of the L27 r-protein from Gram-positive low-G + C bacterial species is that the sequences are about 10 amino acid residues longer and have tails in the N-terminal end of the alignment. This is the reason for the different sequence number of this set of sequences in the fragment in Fig. 1. An interesting observation is the presence of the L27 r-protein from *Arthrobacter* sp. in the cluster, which is characteristic of the Gram-positive low-G + C bacterial sequences. This tree topology is consistent with structural features of the *Arthrobacter* sp. L27 r-protein, which match those of the Gram-positive low-G + C bacterial species analyzed. In particular, *Arthrobacter* sp. and *Bacillus subtilis* L27 r-protein sequences show 91% identity. Figure 1 shows the similarity between the two se-

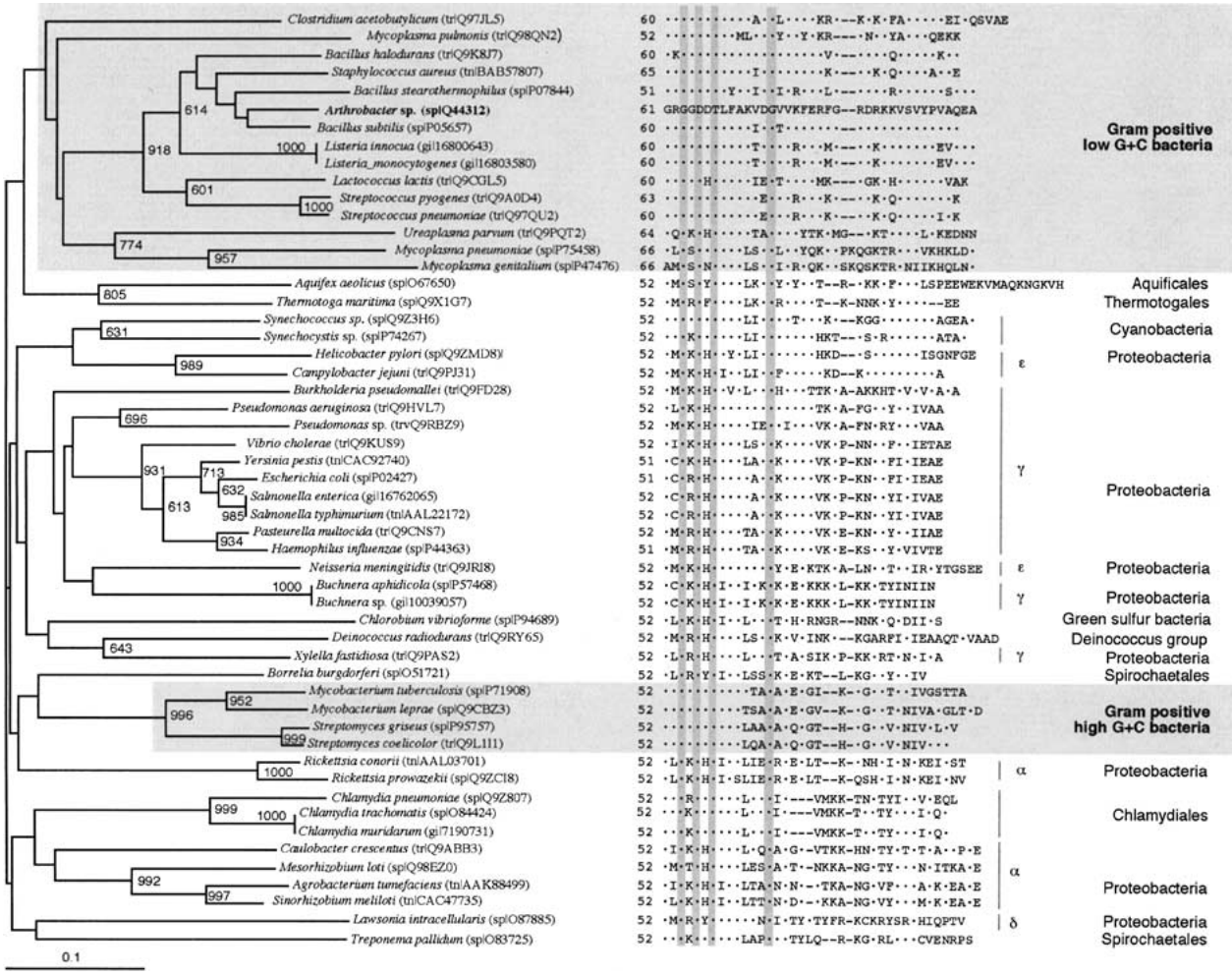

Fig. 1. Neighbor-joining phylogenetic tree and the C-terminal region of the multialignment of a selected group of L27 r-proteins. Branch lengths are drawn proportional to the amount of sequence change. The relationship between branch lengths and 0.1 matrix unit is shown by the *bar* at the bottom. Numbers at various nodes indicate bootstrap support values (only values above 500 are given). Protein sequences are named by their species and their ID SwissProt code or their TrEMBL or EMBL accession number. The anomalously clustered *Arthrobacter* sp. sequence is shown in *bold face*. Numbers on the *right* are the amino acid position from the N terminus. Dots correspond to residues identical to the *Arthrobacter* sp. fragment, used as the reference. Strictly conserved amino acids are *shaded* in the multialignment fragment. Taxonomy is indicated on the *right* and the monophyletic taxonomic clusters of Gram-positive low- or high-G+C bacteria are *shaded*. The C-terminal fragment of the *Arthrobacter* sp. L27 r-protein was taken from Nishida and Imanaka (1993).

quences at the C-terminal end, which is the most variable region of the multialignment, with a gap pattern characteristic of each taxonomic group. Since *Arthrobacter* sp. belongs to the Gram-positive high-G+C bacterium phylum, the present data strongly suggest that the *Arthrobacter* sp. L27 r-protein is a reasonable case of HGT from a Gram-positive low-G+C bacterium species to *Arthrobacter* sp.

The *Arthrobacter* sp. gene that codes the L27 r-protein forms a cluster within an 8127-bp DNA region of the *Arthrobacter* sp. genome with genes for sarcosine and creatine catabolism (EMBL: AB007122 or ASTE1826). In fact, the L27 gene is found upstream, separated by a palindromic sequence, of the *soxR* gene (coding for the negative regulator of *soxA*) and the *soxA* gene (coding for the sarcosine oxidase; EC 1.5.3.1). Downstream from *soxA* gene, there are (in this order) the *creB* gene, coding for a membrane transporter; the *creA* gene, coding for creatinase (EC 3.5.3.3); the *crnA* gene, coding for creatininase (EC 3.5.2.10); and the *crnB* gene, also coding for a membrane transporter. Because of the close physical relationship between these genes, we were interested in determining whether the *Arthrobacter* sp. gene cluster for creatinine and sarcosine degradation also shows evidence of HGT events. However, not all genes involved in the sarcosine and creatinine operons are equally distributed among the bacterial species. For example, the complete *Bacillus subtilis* and *Bacillus halodurans* genomes do not have the genes for the catabolism of creatinine and sarcosine. Also, in many cases the Blast hits for these *Arthrobacter* sp. proteins were sequences with different biological functions. Nevertheless, we properly aligned

each set of homologous proteins and analyzed the branching arrangement of the resulting phylogenies. As far as a possible HGT event is concerned, the phylogenetic analyses of the creatinase CRNA and the membrane transporters CREB and CRNB were considered ambiguous, partially because of the patchy distribution of some proteins (CRNA) and difficulties in determining orthologues for others (transporters CRNB and CREB). The phylogenetic trees obtained from CRNA, CRNB, and CREB (results not shown) suggest that the evolutionary history of these proteins is complicated and may include multiple duplications, losses, and transfer events. It is also difficult to define orthologues of the transcription regulator SOXR because any other regulator of a sarcosine oxidase gene is not known. The SOXR protein sequence has the highest similarity to the LysR family of transcription regulators. Specifically, the phylogenetic tree obtained from a group of bacterial regulators of transcription (results not shown) clearly clusters (bootstrap value, 995) the SOXR protein from *Arthrobacter* sp. with *B. subtilis* and *B. halodurans* LysR transcription regulators, which suggests the *Bacillus* origin of this protein.

On the other hand, the sarcosine oxidase (*soxA*) and creatinase (*creA*) genes are clustered in the *Bacillus* sp. genome (EMBL: BSSO). Figure 2 shows the phylogenetic trees of homologues of *Arthrobacter* sp. sarcosine oxidase SOXA and homologues of *Arthrobacter* sp. creatinase CREA. Homologues of *Arthrobacter* sp. sarcosine oxidase included other sarcosine oxidase proteins from bacteria and archaea. The complete multiple alignment used for tree construction contained 404 positions. The sarcosine oxidase phylogenetic tree (Fig. 2A) shows three major clusters with high bootstrap values. One branch groups two sequences from archaeal *Pyrococcus* species, two sequences from the Gram-positive high-G+C bacterium *Streptomyces*, and one from the α-subdivision Proteobacteria *Rhizobium loti*. The second branch is defined by a set of Enterobacteriaceae, γ-subdivision Proteobacteria sequences. Interestingly, in the third branch the bootstrap value of 1000 firmly supports the cluster of *Arthrobacter* sp. sarcosine oxidase with that of the *Bacillus* sp., a Gram-positive low-G+C bacterium. If we now turn to the group of genes for creatinine degradation, the top six hits of the Blast search with *Arthrobacter* sp. creatinase CREA protein provide bacterial creatinase sequences (two γ- and one β-subdivision Proteobacteria, one Flavobacterium, and one Gram-positive low-G+C bacterium). Further down the list of significant Blast hits were other biological functions such as peptidase and aminopeptidase. Analogously, the phylogenetic tree (Fig. 2B) shows that a bootstrap value of 1000 supports one tight cluster of these six creatinase sequences, in which the branches of both the *Arthrobacter* sp. and the *Bacillus* sp. sequences are derived from the same node with a bootstrap of 1000. The complete multiple alignment used for tree construction contained 454 positions. As there are few representative sequences of each group of interest, the sarcosine oxidase (SOXA) and creatinase (CREA) sequences are not particularly suitable for phylogenetic analysis. However, their phylogenetic analysis supports a transfer from a *Bacillus* species to *Arthrobacter* sp. Since these genes cluster together with the L27 gene in the *Arthrobacter* sp. genome, these genes may have been acquired simultaneously in a single transfer.

Anomalous nucleotide composition is also widely used to detect genes acquired by HGT (Garcia-Vallvé et al. 1999, 2000a, b). So to study the putative origin of the *Arthrobacter* sp. L27, *soxA*, *soxR*, and *creA* genes and the genes (*creB*, *crnA*, and *crnB*) for which the phylogenetic analyses were ambiguous, we compared the codon usage of genes from *B. subtilis* and the Gram-positive high-G+C bacteria *Arthrobacter* sp. and *M. tuberculosis* H37Rv Values. In our analyses of these data (dendrogram not shown), genes from the three species can be clearly distinguished using the relative synonymous codon usage (RSCU) values. However, the set of *Arthrobacter* genes that form the sarcosine–creatinine operon do not have the standard *Arthrobacter* codon usage. The *Arthrobacter* sp. gene that encodes the L27 r-protein shows a codon usage similar to that of the highly expressed genes from *B. subtilis*. The *Arthrobacter* sp. *soxR*, *soxA*, *creB*, *creA*, *crnA*, and *crnB* genes show a codon usage similar to that of class 1 genes (Moszer 1998) from *B. subtilis*. This hypothesis was also confirmed using the χ^2 distance between the synonymous codon usage of all available *Arthrobacter* sp. genes and the mean values of each available complete genome from Gram-positive species, including *Arthrobacter* sp. (results not shown). Thus, the minimum distance of the *Arthrobacter* sp. genes corresponds to the *Arthrobacter* sp., *M. tuberculosis*, *Mycobacterium leprae* mean values. On the other hand, the minimum distance of the *Arthrobacter* genes that form the sarcosine–creatinine operon, except for the *soxA* gene, corresponds to the *B. subtilis* or *Bacillus halodurans* mean values. *B. subtilis*, highly expressed genes are the minimum distance for the *Arthrobacter L27* gene. In this analysis when all available complete genomes were added, the minimum distance of the *Arthrobacter* genes that form the sarcosine–creatinine operon, except for the *crnB* gene, did not change.

Discussion

The prediction of horizontally transferred genes using an atypical nucleotide composition is based on the

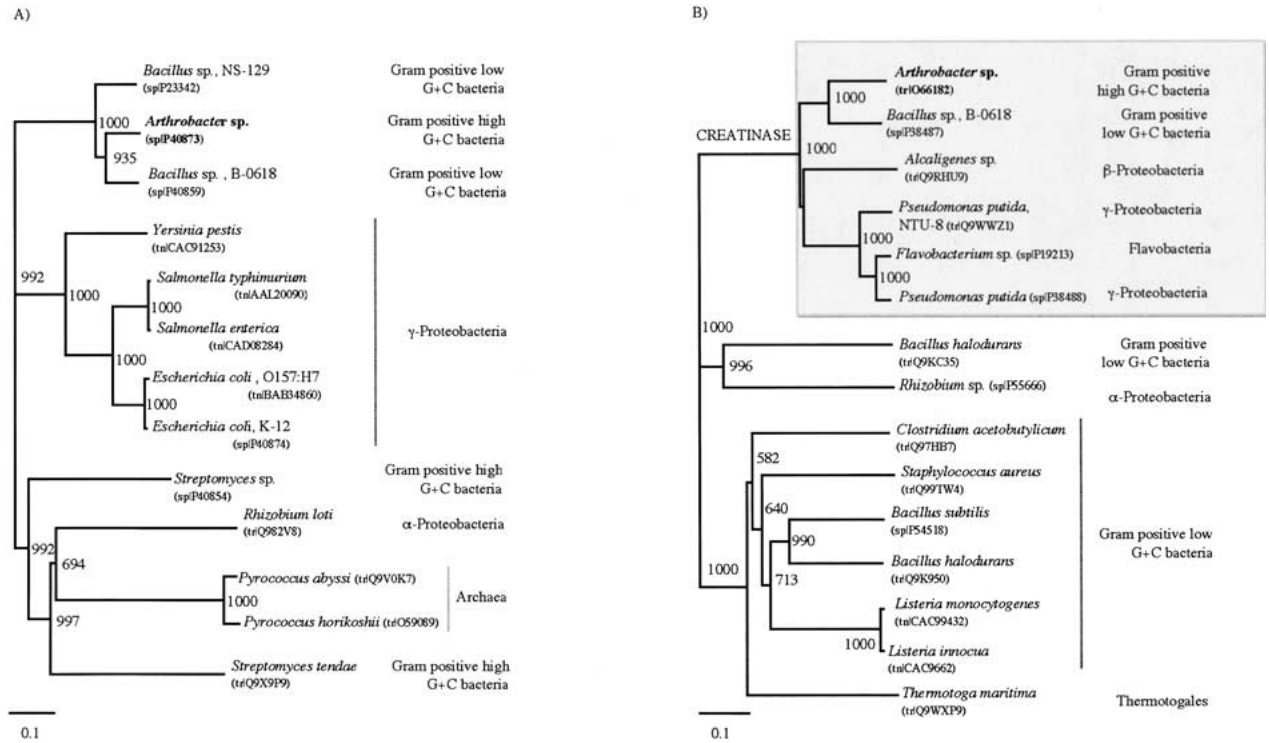

Fig. 2. Phylogenetic trees of homologues of *Arthrobacter* sp. (**A**) sarcosine oxidase SOXA and (**B**) creatinase CREA protein sequences. Tree specifications, taxonomic annotations. In tree B, branches corresponding to the creatinase sequences are *boxed*.

hypothesis that genes from distantly related species have a different nucleotide composition. This approach, however, cannot predict all acquired genes in a genome unambiguously (Garcia-Vallvé et al. 2000b). Transfers cannot be detected between genomes with similar compositions or genes that have adjusted to the base composition and codon usage of the resident genome (the amelioration process) (Lawrence and Ochman 1997). In addition, an unusual composition can also be caused by factors other than HGT—such as selection, mutation bias, intragenomic base content variation (Guindon and Perriere 2001), and the direction of transcription relative to the replication origin (Eisen 2000). To avoid these limitations, multiple lines of evidence, such as combining measures based on base composition with a phylogenetic approach (Koski et al. 2001), should be used (Eisen 2000). An typical G+C content and pattern of codon usage should be considered helpful (Wang 2001), especially for detecting the origin of the transferred genes or donor organism (Garcia-Vallvé et al. 2000a). The present data, which do not rely only on the information from the anomalous nucleotide composition but integrate analyses of phylogenetic tree topology, strongly suggest that a group of genes that includes those that code for the r-protein L27 and sarcosine- and creatinine-related catabolism proteins is a reasonably firm case for blocking HGT from *Bacillus* to *Arthrobacter* species. This identification shows that HGT of informational genes, especially those that code r-proteins, may be more common than previously suspected.

In bacterial genomes, many genes involved in catabolic pathways are clustered on the chromosome and regulated as operons. These operon structures in bacteria are said to be the creations of HGT (Lawrence and Roth 1996). The clustered operational genes of the *Arthrobacter* sp. described here are involved in the metabolism of creatine and sarcosine, and it has been suggested that the *creA* and *creB* genes and the *crnA* and *crnB* genes form an operon (Nishiya et al. 1998). The advantage for the acquisition of this set of genes is clearly that the *Arthrobacter* sp. was able to acquire a new catabolic pathway in which creatinine is hydrolyzed to glycine and formaldehyde. The role (if any) of the transporter proteins encoded by the *creB* and *crnB* genes in the catabolism of creatinine and sarcosine is not known, but the amino acid transporter encoded by the *crnB* gene may act as a glycine transporter. The region upstream from the cluster, however, depicts the gene coding for the r-protein L27.

Our genomic and phylogenetic analyses support the previously described evolutionary relationship between the *Arthrobacter* sp. sarcosine oxidase and that of the *Bacillus* sp. enzyme (Nishida and Imanaka 1993) and suggest that this is due to an HGT event. Homologue genes of sarcosine and creatinine catabolism are absent in the genomes of *B. subtilis* (Kunst et al. 1997) and *B. halodurans* (Takami et al. 2000), and sarcosine oxidase and creatinase genes are clus-

tered in a *Bacillus* species. The rationale behind all these schemes is that in this HGT the donor of the gene cluster may have been either an ancestral species of the *Bacillus* genus or an unknown *Bacillus* species in which gene rearrangements formed the appropriate cluster structure of creatinine and sarcosine metabolism genes. In the first case, further rearrangements, gene divergence, and gene loss created the current arrangement of these genes in the *Bacillus* genomes.

It is well known that the gene that codes the L27 protein forms a transcriptional unit with the gene encoding another ribosomal protein termed L21. The two genes are clustered in all eubacteria for which the complete genome is available, despite their phylogenetic distances. Interestingly, in all the complete genomes belonging to the *Bacillus*/*Staphylococcus* group (which includes the *Bacillus* species), *Spirochaetales*, and *Thermotogales*, there is an ORF between the L21 and the L27 genes. This ORF encodes a protein that belongs to the COG2868 family (Tatusov et al. 2001). In the Gram-positive high-G+C bacteria (which include the *Arthrobacter* species) and other species not included in the above groups, the L21 gene is always found upstream from the L27 gene. If it were possible to obtain the complete genome sequence for *Arthrobacter* sp., we could check whether a native L27 is still present or whether the acquired L27 gene is tightly linked to an L21 gene and to a gene of the COG2868 family. The presence of a gene of the COG2868 family between the acquired L21 and L27 genes in the *Arthrobacter* sp. genome would be further proof of a transfer from a *Bacillus*/*Staphylococcus* species to *Arthrobacter*.

Acknowledgments. We thank Kevin Costello (of the Language Service at our university) for his help with writing the manuscript. This work is part of the CYOP5A1O project of the DGI of the Spanish government.

References

Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. Nucleic Acids Res 25:3389–3402

Brochier C, Philippe H, Moreira D (2000) The evolutionary history of r-protein RpS14. Trends Genet 16:529–533

Brochier C, Bapteste E, Moreira D, Philippe H (2002) Eubacterial phylogeny based on translational apparatus proteins. Trends Genet 18:1–5

Doolittle WF (1999) Phylogenetic classification and the universal tree. Science 284:2124–2129

Eisen JA (2000) Horizontal gene transfer among microbial genomes: New insights from complete genome analysis. Curr Opin Genet Dev 10:606–611

Elhag GA, Bourque DP (1992) Nuclear-encoded chloroplast ribosomal protein L27 of *Nicotiana tabacum*: cDNA sequence and analysis of mRNA and genes. Biochemistry 31:6856–6864

Felsenstein J (1988) Phylogenies from molecular sequences: Inference and reliability. Annu Rev Genet 22:521–565

Garcia-Vallve S, Palau J, Romeu A (1999) Horizontal gene transfer in glycosyl hydrolases inferred from codon usage in *Escherichia coli* and *Bacillus subtilis*. Mol Biol Evol 9:1125–1134

Garcia-Vallve S, Romeu A, Palau J (2000a) Horizontal gene transfer of glycosyl hydrolases of the rumen fungi. Mol Biol Evol 17:352–361

Garcia-Vallve S, Romeu A, Palau J (2000b) Horizontal gene transfer in bacterial and archaeal complete genomes. Genet Res 10:1719–1725

Green R, Noller JF (1997) Ribosomes and translation. Annu Rev Biochem 66:679–716

Guindon S, Perriere G (2001) Intragenomic base content is a potential source of biases when searching for horizontally transferred genes. Mol Biol Evol 18:1838–1840

Jain R, Rivera MC, Lake JA (1999) Horizontal gene transfer among genomes: The complexity hypothesis. Proc Natl Acad Sci USA 96:3801–3806

Koonin EV, Makarova KS, Aravind L (2001) Horizontal gene transfer in prokaryotes: Quantification and classification. Annu Rev Microbiol 55:709–742

Koski LB, Morton RA, Golding GB (2001) Codon bias and base composition are poor indicators of horizontally transferred genes. Mol Biol Evol 18:404–412

Kunst F, Ogasawara N, Moszer I, et al. (1997) The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*. Nature 390:249–256

Lawrence JG, Ochman H (1997) Amelioration of bacterial genomes: Rates of change and exchange. J Mol Evol 44:383–397

Lawrence JG, Roth JR (1996) Selfish operons: Horizontal transfer may drive the evolution of gene clusters. Genetics 143:1843–1860

Makarova KS, Ponomarev VA, Koonin EV (2001) Two C or not two C: Recurrent disruption of Zn-ribbons, gene duplication, lineage-specific gene loss, and horizontal gene transfer in evolution of bacterial ribosomal proteins. Genome Biol 2(9): research 0033.1 Genome Biology is an electronic publication (www.genomebiology.com)

Moore PB (1998) The three-dimensional structure of the ribosome and its components. Annu Rev Biophys Biomol Struct 27:35–58

Moszer I (1998) The complete genome of *Bacillus subtilis*: From sequence annotation to data management and analysis. FEBS Lett 430:28–36

Nishida Y, Imanaka T (1993) Cloning and sequencing of the sarcosine oxidase gene from *Arthrobacter* sp. TE1826. J Ferment Bioeng 75:239–244

Nishiya Y, Toda A, Imanaka T (1998) Gene cluster for creatinine degradation in *Arthrobacter* sp. TE1826. Mol Gen Genet 257:581–586

Nomura M (1984) The control of ribosome synthesis. Sci Am 250:102–114

Ochman H, Lawrence JG, Groisman EA (2000) Lateral gene transfer and the nature of bacterial innovation. Nature 405:299–304

Takami H, Nakasone K, Takaki Y, et al. (2000) Complete genome sequence of the alkaliphilic bacterium *Bacillus halodurans* and genomic sequence comparison with *Bacillus subtilis*. Nucleic Acids Res 28:4317–4331

Tatusov RL, Natale DA, Garkavtsev IV, Tatusova TA, Shankavaram UT, Rao BS, Kiryutin B, Galperin MY, Fedorova ND, Koonin EV (2001) The COG database: New developments in phylogenetic classification of proteins from complete genomes. Nucleic Acids Res 29:22–28

Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTALW: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. Nucleic Acids Res 22:4673–4680

Wang B (2001) Limitations of compositional approach to identifying horizontally transferred genes. J Mol Evol 53:244–250