

Structure-Based Phylogenies of the Serine β -Lactamases

Barry G. Hall, Miriam Barlow

Biology Department, Hutchison Hall, University of Rochester, Rochester, NY 14627-0211, USA

Received: 19 December 2002 / Accepted: 25 March 2003

Abstract. The serine β -lactamases present a special problem for phylogenetics because they have diverged so much that they fall into three classes that share no detectable sequence homology among themselves. Here we offer a solution to the problem in the form of two phylogenies that are based on a protein structure alignment. In the first, structural alignments were used as a guide for aligning amino acid sequences and in the second, the average root mean square distances between the alpha carbons of the proteins were used to create a pairwise distance matrix from which a neighbor-joining phylogeny was created. From those phylogenies, we show that the Class A and Class D β -lactamases are sister taxa and that the divergence of the Class C β -lactamases predated the divergence of the Class A and Class D β -lactamases.

Key words: Protein structure — Bayesian phylogeny — β -Lactamases

Introduction

Because microbes are ancient organisms, the phylogenies of microbial genes are often difficult to construct and there is often a great deal of uncertainty associated with those phylogenies. Estimated dates of the early divergences (Feng et al. 1997) put the division of eubacteria and archaeobacteria at 3.8 billion years ago, that of eukaryotes from archaeobacteria at

2.4 billion years ago and that of the gram-positive from gram-negative bacteria at 2.2 billion years ago. The divergence between plants and animals is estimated to have taken place 1.2 billion years ago and the divergence between protostome and deuterostome at 850 million years ago (Feng et al. 1997). One of the most recent events in microbial evolution, the divergence of *Salmonella* and *E. coli*, corresponds with the divergence of mammals, events which took place a little over 100 million years ago (Ochman and Wilson 1987). While it is fairly uncommon to see protostomes and deuterostomes displayed upon the same phylogeny and even less common to see plants and animals displayed together, it is very common to see gram-negative and gram-positive bacteria displayed upon the same tree. Because much of microbial history and evolution predates the occurrence of multicellular life, microbial phylogeneticists are more often confronted with the problem of determining whether or not real homology exists between two genes, than are those studying multicellular organisms. This problem is often reflected in published microbial gene trees. It is often the case that genes sharing no apparent sequence homology are aligned and presented in a phylogeny (Bush et al. 1995; Ogawara 1993; Rossolini et al. 2001; Shaw et al. 1993). Because it is possible to align any two nucleotide sequences, even when they are unrelated, it is also possible to infer relationships between genes whose relationship, if any, is so distant that no traces of that relationship remain in the sequences.

Microbial phylogenetics is further confounded by gene histories that do not reflect organismal histories. Horizontal transfer can make the question of where

to place the root of a tree difficult to answer because homologous genes from two distantly related organisms may be very closely related if the gene was moved into one of those two species through horizontal transfer (Koonin et al. 2001; Ragan 2001). This problem is particularly evident when constructing phylogenies of antibiotic resistance genes because many antibiotic resistance genes are located on broad host range plasmids (Carattoli 2001) where there is no reliable correlation between the origin of the gene and the organism in which that gene is found. Although the ancestry of microbes is known in many cases, the organismal pattern of descent may not tell us anything about the ancestry of a gene. In such cases it is sometimes possible to determine an outgroup for a phylogeny based on biochemical data or other information, but if there has been a great deal of horizontal transfer within a gene tree, there is little justification for placing a root anywhere on the tree, because there is so much uncertainty associated with the direction in which evolution occurred.

β -Lactamases are the primary means by which microorganisms acquire resistance to β -lactam antibiotics such as penicillins and cephalosporins. β -Lactamase genes are found on both chromosomes and plasmids and are widely distributed through both eubacterial and archeal kingdoms. As a result of their wide distribution and ancient history, β -lactamase genes present both the problem of homology and the problem of properly identifying an outgroup when constructing phylogenies. There are three classes of serine β -lactamases that share structural homology with each other (Ambler 1980; Jaurin and Grundstrom 1981; Ouellette et al. 1987) and with the DD peptidases (Medeiros 1997) or penicillin binding proteins. Class A includes the commonly encountered TEM β -lactamases, Class C includes the AmpC β -lactamases, and Class D includes the OXA β -lactamases. Because there is almost no detectable homology at the sequence level it is inappropriate to construct a phylogeny based on a sequence alignment that includes more than one class. Furthermore, there have been numerous horizontal transfer events that make it very difficult to infer the order of descent and the relationships that exist among the β -lactamases.

As a partial solution to the alignment problems associated with very distantly related genes, some phylogenies have been reconstructed with alignments of protein crystal structures (Breitling et al. 2001; Caetano-Anolles 2002; Johnson et al. 1990). Two of the methods that exist for creating structure-based phylogenies are (1) to create a pairwise distance matrix of the average distance between the alpha carbons of two aligned structures and then use that distance matrix to construct a phylogeny using the neighbor-joining method (Saitou and Nei 1987) or (2) to use the structural alignment as a guide for aligning

Table 1. Accession and PDB numbers for crystallized β -lactamases

Name	Class	PDB No.	DNA accession No.
TEM-1	A	1M40	AF309824
Bla-B	A	1HZO	D37831
BEPEN	A	1I2S	V00093
PSE-4	A	1G68	J05162
Cfre-AmpC	C	1FR6	AF492447
PER-1	A	1E25	Z21957
blaF	A	1MFO	L25634
Nmc-A	A	1BUE	Z21956
Toho-1	A	1BZA	AB038771
SABLA	A	1BSG	M28303
AmpC-Eco	C	2BLS	AF124202
DD-Pep	DD-pep	3PTE	M26842
AmpC-Ent	C	1BLS	AF411148
blaZ	A	3BLM	AP003139
Oxa-10	D	1K4F	XXU37105
Oxa-13	D	1H8Z	AF315786
SHV-1	A	1SHV	M59181
Estb	Esterase	1CI9	U33634

protein sequences with the assumption that amino acids located at structurally homologous sites are truly homologous. In this paper we have used both of these methods to reconstruct phylogenies of the serine β -lactamases and infer their order of descent.

Methods

The Structures and Sequences

Table 1 lists the PDB and GenBank accession numbers of the β -lactamases and the DD peptidase and the esterase that were used for this phylogeny. Those sequences were chosen to include the three major classes of the serine β -lactamases, the DD peptidases which are known structural homologs of the serine β -lactamases, and an esterase as a known outgroup. The VAST Structure Neighbors utility at the NCBI web site (<http://www.ncbi.nlm.nih.gov/Structure/RESEARCH/iucrabs.html>) was used to determine what structures were homologous and therefore appropriate to include in our analysis. We included only structures that had a *p*-value, or probability of getting a particular structural alignment by chance alone, that was $\leq 10^{-5.5}$.

Distance Matrix and Neighbor-Joining Tree Reconstruction

A distance matrix was created by determining all pairwise VAST calculations of the root mean square (RMS) distances between the alpha carbons of the crystal structures. Because RMS distances are sensitive to the length of alignable sequence, the RMS distances were normalized by dividing the RMS distances by the number of aligned residues. That matrix of normalized RMS distances was used as the input for a neighbor-joining tree that was estimated by PAUP* (Swofford 2000).

Alignment

To create a protein alignment based on the crystal structure alignment, in VAST we selected the option of viewing the crystal

structure alignment in hypertext which displays an amino acid alignment of the homologous regions of the protein. We copied that protein alignment from the Internet browser window into a BBEedit 7.0 document, and proceeded to manually delete any regions of the protein alignment where the aligned amino acids were not structurally homologous (denoted by lowercase letters). For this editing process either the commercial program BBEedit 7.0 (Bare Bones Software, Inc.; <http://www.barebones.com/>) or the freely available BioEdit 5.0.9 (<http://www.mbio.ncsu.edu/BioEdit/bioedit.html>) is the preferred tool because it is possible to select and delete columns from text (rectangular selection). We then converted the alignment to PHYLIP format by deleting all numbering of the alignment. At the beginning of each line we ensured that there were exactly 10 characters, consisting of the sequence name and spaces, before each sequence. At the top of the alignment we first entered the number of taxa contained within the alignment and then, on the same line, we entered the number of characters contained within an aligned sequence and saved the document as a text file. We imported the file into PAUP* and were then able to export the file in Nexus format and proceed with a Bayesian reconstruction (Huelsbeck and Ronquist 2001) of the phylogeny. (This method is given in detail to facilitate the process for those who wish to align amino acid sequences based on protein crystal structure.)

Bayesian Phylogenetic Reconstruction

Phylogenies were constructed from protein alignments by the Bayesian methods (Mau and Newton 1997; Mau et al. 1999; Rannala and Yang 1996) as implemented by the program MrBayes v3.0b3 (Huelsbeck and Ronquist 2001). An explanation of the Bayesian method, and a comparison with the more classical distance and parsimony methods, is presented in *Phylogenetic Trees Made Easy* (Hall 2001). MrBayes is available at no charge from <http://brahms.biology.rochester.edu/software.html>. The rate matrix for the amino acid data was set to "mixed" so that the program estimates the optimum rate matrix from the data by using the command "prset aamodel = mixed." Four chains, with a "temperature" of 0.2 for the heated chains, were run for 510,000 generations, sampling trees every 100 generations. The ln likelihood of the trees had converged on a constant value by generation 10,000, i.e., after saving 100 trees. The consensus tree, with branch lengths, was calculated from the final 5,000 trees visited, well after convergence had occurred.

One of the advantages of Bayesian inference of phylogeny is that the results are easy to interpret. For example, the sum of the posterior probabilities of all trees is 1. Moreover, the posterior probability of any single clade is simply the sum of the posterior probabilities of all trees that contain that clade. The consensus trees calculated by MrBayes do not include the posterior probabilities of the clades, thus each entire set of trees was imported into PAUP* (Swofford 2000) and the same trees used by MrBayes to calculate a consensus were used to calculate a 50% majority rule consensus in PAUP* (Swofford 2000). The resulting tree shows the posterior probabilities of the clades, i.e., the percent of time that those taxa are included in the clade.

The consensus trees calculated by MrBayes were imported into PAUP* for the purposes of displaying and printing the tree.

Results and Discussion

The goal in creating an alignment is to align either homologous regions or amino acids or nucleotides to each other. Because there is mutational saturation among the amino acid sequences of the three classes

of serine β -lactamases and the DD peptidases, it is impossible to detect homologous amino acids by direct sequence comparison. Despite the abundance of amino acid substitutions that have occurred since the divergence of the DD peptidases and the serine β -lactamases, the structures of the β -lactamases and DD peptidases still have a large amount of detectable homology. By aligning the crystal structures of those proteins, it is possible to identify homologous amino acids because those amino acids occur at structurally homologous positions in the protein.

To our knowledge a multiple alignment algorithm does not yet exist for protein structures. Because we were unable to use this preferred method for alignment, we used the VAST master/slave alignment algorithm to generate our structural alignment. With this method one structure is used as the "master" and all remaining structures are aligned to it. For our "master" in this alignment, we used the Est B esterase because it was the predetermined outgroup for the phylogeny. EstB was chosen as the outgroup sequence because it is the most functionally distant relative of the serine β -lactamases and DD-peptidases that is a structural neighbor of all of the other sequences in the phylogeny. Ideally an outgroup sequence is more distantly related to each of the other sequences than any of those sequences are to each other. That ideal situation occurs only when a molecular clock is operating so that all of the ingroup sequences are evolving at about the same rate. The serine β -lactamases deviate from a molecular clock as the result of several episodes of intense positive selection during the early evolution of the Class D β -lactamases (Barlow and Hall 2002b) and of the Class A β -lactamases (Barlow and Hall, unpublished results). As a result, the outgroup EstB sequence is more closely related to the DD-peptidase and the Class C β -lactamases than some of the Class A β -lactamases are to the Class D β -lactamases. Using the master-slave alignment with EstB as the master, any region of homology shared between any one β -lactamase or DD peptidase and EstB should be shared among all of the β -lactamases and DD peptidases.

The Bayesian phylogeny is shown in Fig. 1 and the neighbor-joining phylogeny is shown in Fig. 2. Despite the different methods used to create those phylogenies, their topologies are highly similar. In both phylogenies, the Class A and Class D β -lactamases are sister taxa with the Class C β -lactamases diverging first. In the neighbor-joining phylogeny, the Class C β -lactamases and the DD peptidases are sister taxa, and in the Bayesian phylogeny, the Class C taxon is shown as an isolated group. In the Bayesian phylogeny the node where the DD peptidases diverge from the β -lactamases has a posterior probability of 0.76 (Fig. 1), meaning that the topology at that node was recovered 76% of the time.

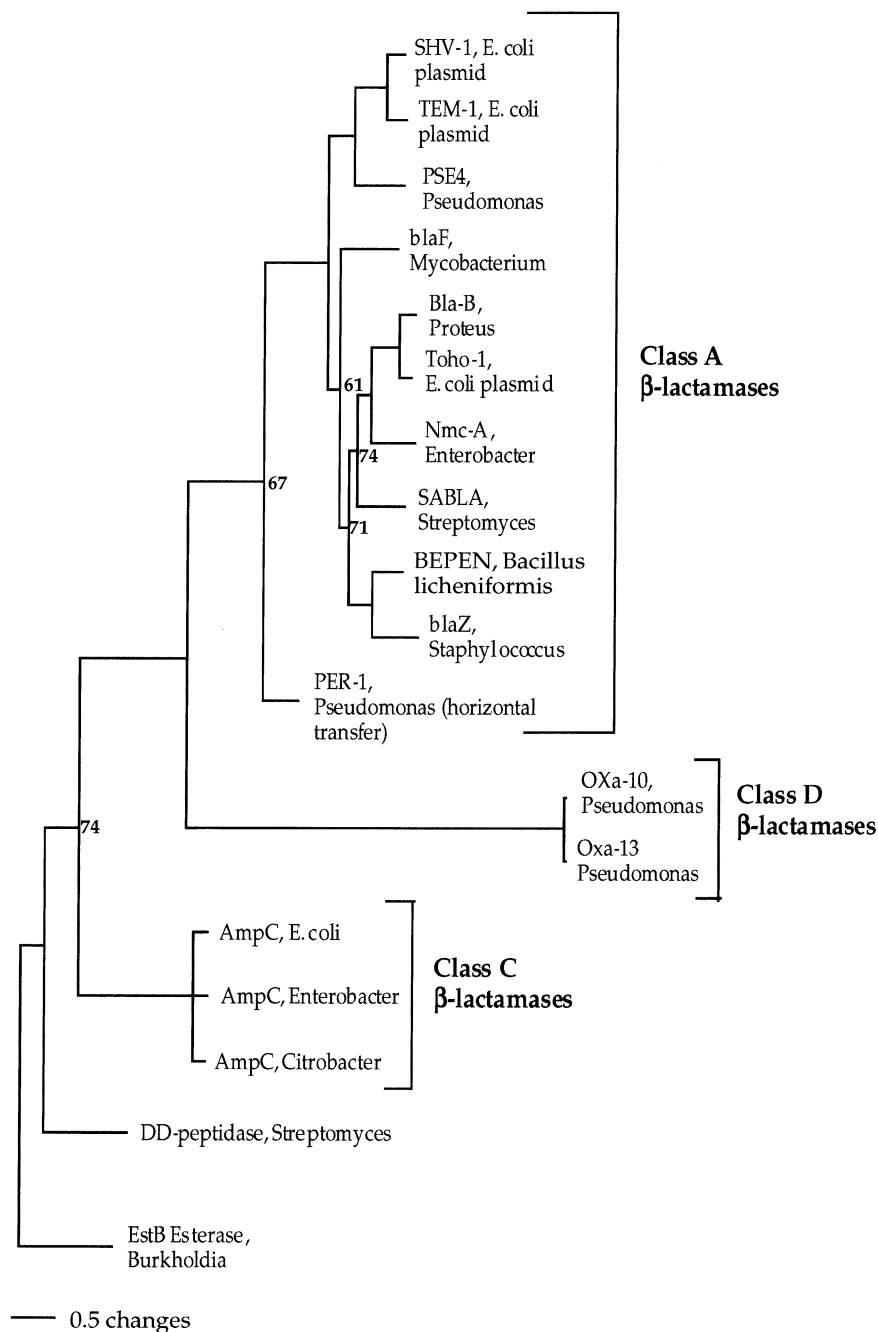


Fig. 1. Bayesian phylogeny of the serine β -lactamases determined from a structurally based alignment of the amino acids. Esterase B was used as the outgroup to root the tree. Except as indicated by *boldface numbers* adjacent to the nodes, the confidence on all clades was >0.85 .

In the other 24% of the trees sampled, the branching order at that node was the same as the neighboring tree.

An earlier study (Barlow and Hall 2002a) found that the Class C β -lactamases are present only in the Proteobacteria division of the gram-negative eubacteria. Because the Class A β -lactamases are found in both the gram-positive and the gram-negative eubacteria, the finding that Class C β -lactamases diverged prior to the divergence of Class A and Class D implies that Class C arose before the gram-positive and gram-negative divergence. The Class D β -lactamases, which arose after the Class C β -lactamases, are found in gram-negatives outside of the Proteo-

bacteria group (Barlow and Hall 2002b). The Proteobacteria are the most recent group of gram-negative bacteria (Brown et al. 2001). The absence of Class C β -lactamases from both gram-positive and gram-negative bacteria outside of the Proteobacteria group means that either Class C β -lactamases were lost from the gram-positive lineage and from each of the gram-negative lineages that precede the Proteobacteria or that Class C β -lactamases simply have not yet been detected in those lineages. Since the loss model would require many loss events we expect that Class C β -lactamases or their homologs will be found in the genomes of gram-positive and nonproteobacteria gram-negative bacteria.

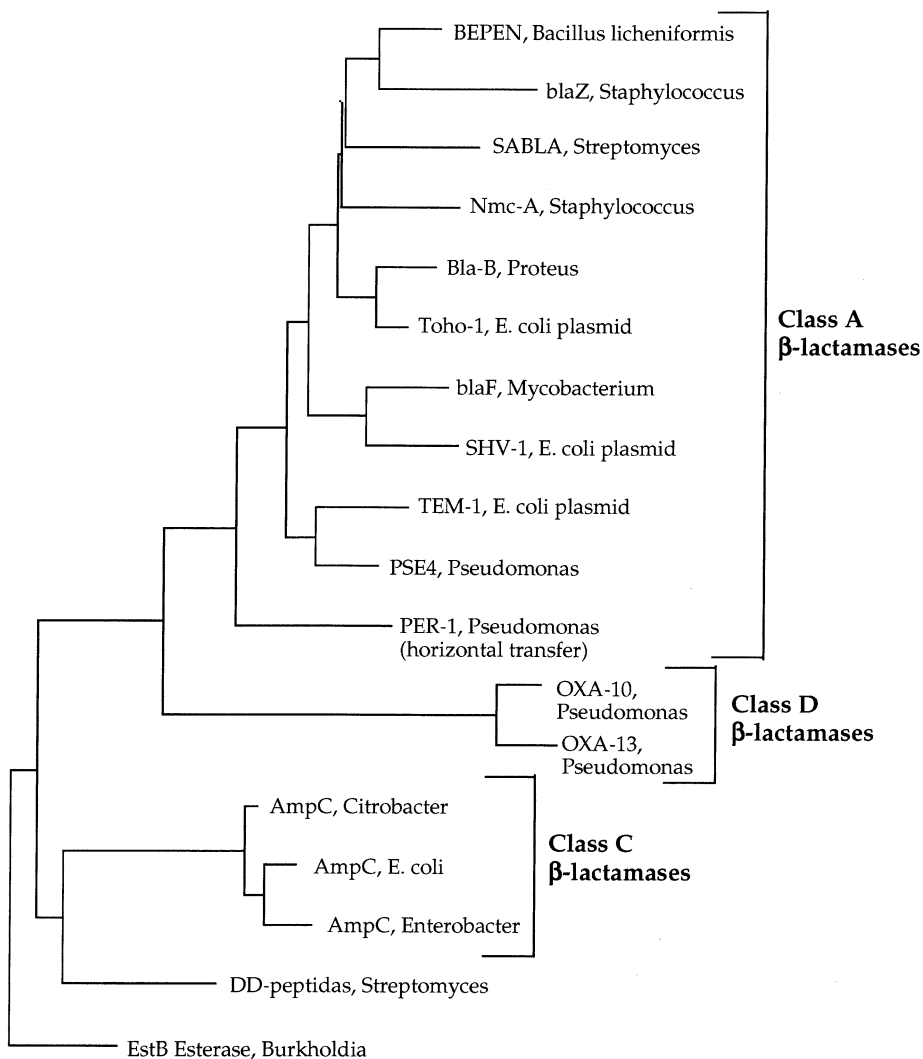


Fig. 2. Neighbor-joining phylogeny of the serine β -lactamases based on the root-mean-square distances between the alpha carbons of the amino acids in the structurally aligned proteins.

We do not take the branch lengths in the Bayesian phylogeny seriously because by removing the regions that did not align, we probably deleted the regions of the proteins that evolve most rapidly. The branch lengths of the RMS distances tree are even less meaningful because they are based on RMS distances which are affected to some extent by the resolution of the crystal structures and because they have never been correlated with a molecular clock.

The main use of crystal structure based phylogenies is to infer branching orders of distantly related taxa where information about homology at the sequence level has been lost. Fine scale phylogenetic analysis of closely related genes should be based on direct sequence alignments whenever possible. However, great care must be taken to make sure that there is sufficient homology at the sequence level to produce an accurate and meaningful alignment. It is often not obvious whether there is sufficient similarity among sequences to imply homology. One way to assess sequence-based homology is to use BLAST (<http://www.ncbi.nlm.nih.gov/blast/index.html>) (Altschul et al. 1990, 1997;

Tatusova and Madden 1999) to align a pair of sequences, rejecting as non-homologous those that produce no significant alignment, or, more conservatively, those whose "E-score" is $<10^{-4}$ in an alignment of at least 60% of the length of the shorter sequence.

Because antibiotic resistance genes are often of very ancient origin (Barlow and Hall 2002b), and because the antibiotic resistance field tends to group genes as much by functional as by sequence similarity, those constructing phylogenies of resistance genes will frequently be confronted by the problem of deciding whether or not a set of sequences have sufficient sequence homology to appear on a single tree. Phylogenies of antibiotic resistance genes are often presented in which completely inappropriate sequence alignments have been used to construct a tree. For instance, the AAC(6') group of aminoglycoside acetyltransferases are typically discussed as though they are homologs and are placed onto a single tree (Hannecart-Pokorni et al. 1997; Shaw et al. 1992, 1993; Wu et al. 1997), whereas they actually belong to three distinct families that exhibit no detectable se-

quence homology among each other (Salipante and Hall 2003). Similarly, the metallo- β -lactamases, which have previously been categorized as a single group, Class B (Ambler 1980; Galleni et al. 2001; Rasmussen and Bush 1997), actually belong to two separate phylogenetic groups (Hall et al. 2003). We present this phylogeny not only to bring together the three classes of serine β -lactamases, but also to present the tools we used to create it in an effort to discourage misleading phylogenies and to emphasize that there are appropriate tools that can be used to infer the relationships of taxa that diverged long ago.

Acknowledgments. This study was supported by Grant GM60761 from the National Institutes of Health. We are grateful to Reviewer 1 for insightful comments on an earlier version of this paper.

References

- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol* 215:403–410
- 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
- Ambler RP (1980) The structure of beta-lactamases. *Philos Trans R Soc Lond B Biol Sci* 289:321–331
- Barlow M, Hall BG (2002a) Origin and evolution of the AmpC β -lactamases of *Citrobacter freundii*. *Antimicrob Agents Chemother* 46:1190–1198
- Barlow M, Hall BG (2002b) Phylogenetic analysis shows that the OXA β -lactamase genes have been on plasmids for millions of years. *J Mol Evol* 55:314–321
- Breitling R, Laubner D, Adamski J (2001) Structure-based phylogenetic analysis of short-chain alcohol dehydrogenases and reclassification of the 17beta-hydroxysteroid dehydrogenase family. *Mol Biol Evol* 18:2154–2161
- Brown JR, Douady CJ, Italia MJ, Marshall WE, Stanhope MJ (2001) Universal trees based on large combined protein sequence data sets. *Nat Genet* 28:281–285
- Bush K, Jacoby GA, Medeiros AA (1995) A functional classification scheme for beta-lactamases and its correlation with molecular structure. *Antimicrob Agents Chemother* 39:1211–1233
- Caetano-Anolles G (2002) Tracing the evolution of RNA structure in ribosomes. *Nucleic Acids Res* 30:2575–2587
- Carattoli A (2001) Importance of integrons in the diffusion of resistance. *Vet Res* 32:243–259
- Feng D-F, Cho G, Doolittle RF (1997) Determining the divergence times with a protein clock: Update and reevaluation. *Proc Natl Acad Sci USA* 94:13028–13033
- Galleni M, Lamotte-Brasseur J, Rossolini GM, Spencer J, Dideberg O, Frere JM (2001) Standard numbering scheme for class B beta-lactamases. *Antimicrob Agents Chemother* 45:660–663
- Hall BG (2001) Phylogenetic trees made easy: A how-to manual for molecular biologists. Sinauer Associates, Sunderland, MA
- Hall BG, Salipante S, Barlow M (2003) The metallo- β -lactamases fall into two distinct phylogenetic groups. *J Mol Evol* 57:249–254
- Hannecart-Pokorni E, Depuydt F, de wit L, van Bossuyt E, Content J, Vanhoof R (1997) Characterization of the 6'-N-aminoglycoside acetyltransferase gene aac(6)-Im [corrected] associated with a sulI-type integron. *Antimicrob Agents Chemother* 41:314–318
- Huelsenbeck JP, Ronquist F (2001) MrBayes: Bayesian inference of phylogeny. *Bioinformatics* 17:754–755
- Jaurin B, Grundstrom T (1981) AmpC cephalosporinase of *Escherichia coli* K-12 has a different evolutionary origin from that of beta-lactamases of the penicillinase type. *Proc Natl Acad Sci USA* 78:4897–4901
- Johnson MS, Sutcliffe MJ, Blundell TL (1990) Molecular anatomy: Phyletic relationships derived from three-dimensional structures of proteins. *J Mol Evol* 30:43–59
- Koonin EV, Makarova KS, Aravind L (2001) Horizontal gene transfer in prokaryotes: Quantification and classification. *Annu Rev Microbiol* 55:709–742
- Mau B, Newton M (1997) Phylogenetic inference for binary data on dendrograms using Markov chain Monte Carlo. *J Comput Graph Stat* 6:122–131
- Mau B, Newton M, Larget B (1999) Bayesian phylogenetic inference via Markov chain Monte Carlo methods. *Biometrics* 55:1–12
- Medeiros AA (1997) Evolution and dissemination of beta-lactamases accelerated by generations of beta-lactam antibiotics. *Clin Infect Dis* 24:S19–S45
- Ochman H, Wilson AC (1987) Evolution in bacteria: Evidence for a universal substitution rate in cellular genomes. *J Mol Evol* 26:74–86
- Ogawara H (1993) Phylogenetic tree and sequence similarity of beta-lactamases. *Mol Phylogenet Evol* 2:97–111
- Ouellette M, Bissonnette L, Roy PH (1987) Precise insertion of antibiotic resistance determinants into Tn21-like transposons: Nucleotide sequence of the OXA-1 beta-lactamase gene. *Proc Natl Acad Sci USA* 84:7378–7382
- Ragan MA (2001) Detection of lateral gene transfer among microbial genomes. *Curr Opin Genet Dev* 11:620–626
- Rannala B, Yang ZH (1996) Probability distribution of molecular evolutionary trees: A new method of phylogenetic inference. *J Mol Evol* 43:304–311
- Rasmussen BA, Bush K (1997) Carbapenem-hydrolyzing beta-lactamases. *Antimicrob Agents Chemother* 41:223–232
- Rossolini GM, Condemi MA, Pantanella F, Docquier JD, Amicosante G, Thaller MC (2001) Metallo-beta-lactamase producers in environmental microbiota: New molecular class B enzyme in *Janthinobacterium lividum*. *Antimicrob Agents Chemother* 45:837–844
- Saitou N, Nei M (1987) The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Mol Biol Evol* 4:406–425
- Salipante SJ, Hall BG (2003) Determining the limits of the evolutionary potential of an antibiotic resistance gene. *Mol Biol Evol* 20:653–659
- Shaw KJ, Rather PN, Sabatelli FJ, et al. (1992) Characterization of the chromosomal aac(6)-Ic gene from *Serratia marcescens*. *Antimicrob Agents Chemother* 36:1447–1455
- Shaw KJ, Rather PN, Hare RS, Miller GH (1993) Molecular genetics of aminoglycoside resistance genes and familial relationships of the aminoglycoside-modifying enzymes. *Microbiol Rev* 57:138–163
- Swofford DL (2000) PAUP*. Phylogenetic analysis using parsimony (*and other methods). Sinauer Associates, Sunderland, MA
- Tatusova TA, Madden TL (1999) BLAST 2 sequences, a new tool for comparing protein and nucleotide sequences. *FEMS Microbiol Lett* 174:247–250
- Wu HY, Miller GH, Blanco MG, Hare RS, Shaw KJ (1997) Cloning and characterization of an aminoglycoside 6'-N-acetyltransferase gene from *Citrobacter freundii* which confers an altered resistance profile. *Antimicrob Agents Chemother* 41:2439–2447