Evidence of Positive Darwinian Selection in Omp85, a Highly Conserved Bacterial Outer Membrane Protein Essential for Cell Viability

David A. Fitzpatrick, James O. McInerney

Department of Biology, National University of Ireland, Maynooth, Country Kildare, Ireland

Abstract. Omp85 is a highly conserved outer membrane protein found in all gram-negative bacteria. It is essential for bacterial cell viability and plays an integral function in the positioning and folding of other outer membrane proteins into the bacterial outer membrane. We have employed a maximum likelihood and a maximum parsimony approach to detect evidence of positive Darwinian selection in Omp85 homologues from 10 δ-proteobacteria and have identified 14 amino acid sites that show evidence of being under the influence of adaptive evolution. Interestingly all sites bar one are concentrated within surface loops of the protein that most likely interact with host immune response or the surrounding environment. Alternatively amino acids within membrane-spanning regions of the protein are found to be under purifying selection most likely as a result of structural constraints.

Key words: Omp85 — Positive selection

Introduction

The cell envelope of gram-negative bacteria consists of an outer membrane (OM) and an inner membrane separated by the peptidoglycan containing periplasm. Both membranes contain proteins (Voulhoux et al.

2003). Outer membrane proteins (OMPs) are synthesized in the cytoplasm by the Sec machinery (Manting and Driessen 2000) and have a β-barrel structure consisting of an even number of 8 to 22 membrane spanning β-strands with an antiparallel topology, which are connected by long and short loops forming β-hairpin structures (Voulhoux and Tommassen 2004). After translocation, the mature protein is released into the periplasm, where it folds and is subsequently inserted into the membrane. The molecular mechanism underlying the shuttle mechanism of transport of completed OMPs to OM has recently been explained. Initially it was thought that OMPs were translocated through membrane adhesion zones known as Bayer's junctions (Bayer 1968), but now there is very strong evidence to suggest that there is a specific proteinaceous machinery, specifically Omp85 and soluble chaperones that can dock transiently with the membrane surface (Kleinschmidt 2003; Tamm et al. 2001).

Omp85 homologues have been shown to be present in all gram-negative bacteria (Genevrois et al. 2003; Gentle et al. 2004; Stephens and Lammel 2001; Voulhoux and Tommassen 2004) and also eukaryote mitochondria, where they are involved in mitochondrial biogenesis. Omp85 can be divided into two domains, a periplasmic NH_2-terminal domain and a 12-stranded β-barrel domain at the COOH- terminus (Voulhoux et al. 2003). Knockout of Omp85 in *Neisseria* strains results in the accumulation of porin monomers; these OMPs are normally found as trimers (Voulhoux et al. 2003). Omp85 knockout also increases the accumulation of monomeric PilQ; this secretin, which is involved in type IV pili formation,

Correspondence to: James O. McInerney; *email:* james.o.mcinerney@may.ie

normally forms stable high molecular weight oligomers (Wolfgang et al. 2000). Immunofluoresence microscopy on knockout *Neisseria* strains has illustrated the importance of Omp85, as the binding of antibodies directed against a number of OMPs (PilQ, PorA, and PorB) was considerably weaker than wild types. Further, electron-dense material accumulates in the periplasm of *Neisseria* Omp85-depleted strains, and although the identity of this material could not be revealed (Genevrois et al. 2003), it is most likely misassembled OMPs (Voulhoux and Tommassen 2004). The fact that Omp85 is also located in the lipopolysaccharide biosynthetic operon (Genevrois et al. 2003; Voulhoux et al. 2003) reinforces the view that this gene has an important role in OMP biogenesis.

Recently a number of membrane-associated proteins (Andrews and Gojobori 2004; Fares et al. 2001; Jiggins et al. 2002; Smith et al. 1995; Urwin et al. 2002) have been shown to be under the influence of positive Darwinian selection (adaptive evolution). Adaptive evolution is a process that encourages the retention of mutations that are beneficial to an individual or population (Creevey and McInerney 2002). For example, the replacement of a wild-type allele by a mutant allele with a higher fitness will occur through natural selection. In protein coding genes, positive selection is thought to be an ephemeral event frequently leading to the generation of novel protein function (Kinsella et al. 2003). At the DNA level, positive selection may be detected by comparing the rate of nonsynonymous (amino acid altering) nucleotide substitutions per nonsynonymous site (d_N) with that of synonymous substitution per synonymous site (d_S) (Hughes and Nei 1989). When $d_N/d_S > 1$, positive selection is said to be operating on the genes in question; alternatively, when $d_N/d_S < 1$, purifying selection is said to be operating.

As the membrane proteins of bacteria are exposed to the host immune system and other adverse environmental factors, it follows that they are under the greatest selective pressure for change. We postulate that Omp85 is an excellent candidate for displaying the effects of positive Darwinian selection, as it is surface-exposed, is an essential protein, and elicits an immune response in mouse models (Voulhoux et al. 2003). Because OMP85 is surface-accessible to anti-Omp85 antibodies, it has been suggested that this protein should be investigated for its vaccine potential (Genevrois et al. 2003; Stephens and Lammel 2001; Voulhoux and Tommassen 2004). Here we utilized a number of methods to determine if these genes are under the influence of adaptive evolution and show that sequence regions that encode surface-exposed loops exhibit an accelerated rate of nonsynonymous substitution that is compatible with positive Darwinian selection. Alternatively, structurally important transmembrane regions appear to be under the influence of purifying selection.

Materials and Methods

Sequences and Alignments

Omp85 homologues from 10 δ-proteobacteria were analyzed in this analysis. They are *Escherichia coli* gi_1786374, *Xanthomonas campestris* gi_21230822, *Xylella fastidiosa* gi_28198247, *Salmonella typhimurium* gi_16418729, *Pseudomonas aeruginosa* gi_9949808, *Pasteurella multocida* gi_12722432, *Vibrio cholerae* gi_9656813, *Haemophilus influenzae* gi_1573938, *Yersinia pestis* gi_15979119, and *Shigella flexneri* gi_24050380. Homologues from other bacteria were available but the sample was restricted to the δ-proteobacteria in an effort to limit the effects of saturation of synonymous sites that occurs between distantly related families; saturation of synonymous sites reduces the power of methods that detect for positive Darwinian selection. The δ-proteobacteria were selected, as they were the largest bacterial division in terms of available data.

Stop codons were removed from all sequences. The nucleotide sequences were translated into their amino acid equivalents and aligned using CLUSTALW 1.82 (Thompson et al. 1994). Gaps created in the amino acid alignment were transposed back to the nucleotide sequences to gain a codon-based alignment using the program Putgaps (http://bioinf.may.ie/software/putgaps). The codon alignment was corrected for obvious alignment ambiguity using the alignment editor Se-Al 2.0all (http://evolve.zoo.ox.ac.uk/software.html?id=seal). There is no evidence that the alignment suffers from a GC bias, as the final alignment passed a chi-square test of sequence homogeneity as implemented in Tree-Puzzle 5.1 (Schmidt et al. 2002). The Alignment is available on request from the authors.

Tree Reconstruction

The optimal model of sequence substitution was found by comparing the likelihood scores using Modeltest 3.04 (Posada and Crandall 1998). The optimal model of substitution was found to be TrNef+I+G. A maximum likelihood (ML) phylogeny that utilizes this model and the estimated parameters was reconstructed using the program PAUP 4.0b10 (Swofford 1998). A phylogeny was also reconstructed using a Bayesian framework as implemented in MR BAYES 3.0B4 (Huelsenbeck and Ronquist 2001), 500,000 generations of the mcmc chain were performed sampling every 100th generation, and the resultant trees were summarized using a majority rule consensus method implemented in PAUP 4.0b10 (Swofford 1998). Both methods reproduced identical trees with high branch supports (Fig. 1).

Detection of Recombination

Blocks of sequences that have incongruent phylogentic topologies due to recombination were searched for using the method of Crassly and Holmes (1997), which is implemented in PLATO 2.0.

Analysis of Selection

The ML approach of Yang et al. (2000) as implemented in the PAML package 3.13 (Yang 1997) was used to examine selection pressures acting on OMP85 homologues. This approach examines d_N and d_S codon-by-codon using various models of codon substitution that differ in how d_N/d_S ratio (ω) varied along sequences, as
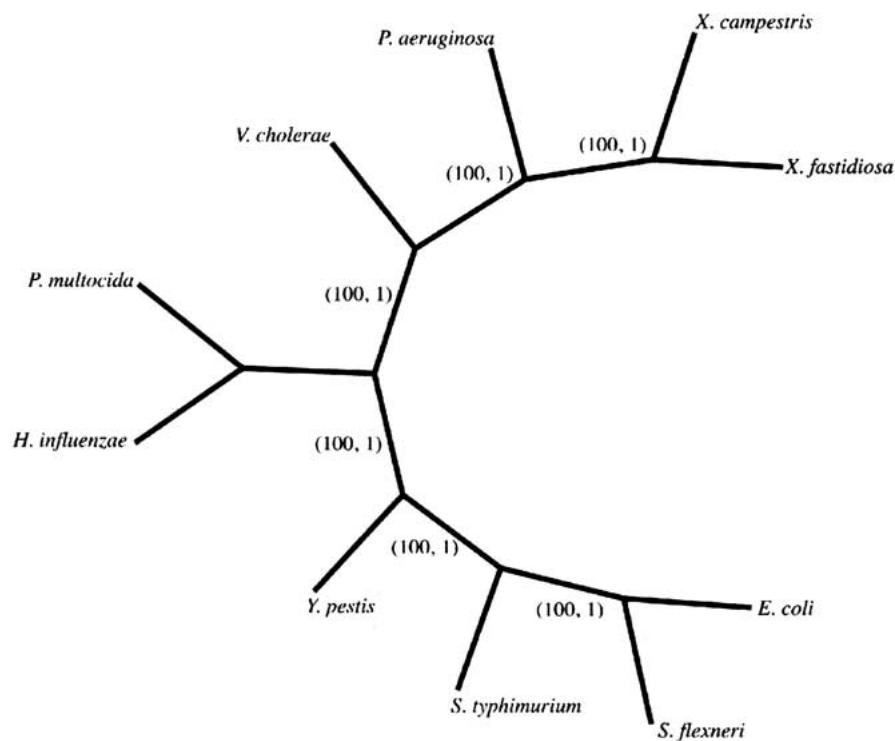
Fig. 1. Maximum likelihood phyiogenetic tree of the Omp85 gene from 10 δ-protobacteria. Numbers in parentheses correspond to 1000-pseudoreplicate bootstrap values and also branch supports from Bayesian analysis; both methods identified the same topology.

well as incorporating information about the phylogenetic relationships of the sequences in question so that comparisons are independent (Yang et al. 2000). The analysis consists of two major steps. The first step uses the likelihood ratio test (LRT) to test for positive selection (i.e., the presence of sites with ω > 1). This is achieved by comparing a null model that does not allow sites with ω > 1 and a more general model that does (Swanson et al. 2001). We used three LRTs. The first compares model M0 (one ratio), which assumes one ω for all sites, with model M3 (discrete $k = 3$), which assumes three site classes with independent co values estimated from the data. The second LRT compares M1 (neutral), which allows two site classes with values fixed at 0 and 1, with M2 (selection), which has an additional site class that allows ω > 1. The final LRT used compares M7 (beta), which allows for 10 site classes (each with ω < 1), with M8, which has an additional site class that allows ω > 1; the comparison of M7 *vs* M8 is the most stringent test of positive selection (Anisimova et al. 2001). In all LRTs good evidence for positive selection is found if the LRT indicates that models that allow for selection are significantly better than their null model; the level of significance is calculated as twice the difference of the likelihood scores ($2\Delta\ln L$) estimated by each model and the null distribution of these results can be approximated by χ^2 distribution with the number of degrees of freedom calculated as the difference in the number of estimated parameters between models (Yang et al. 2000).

The second major step identifies codon sites under positive selection when the LRT suggests their presence. This is achieved using the Bayes theorem to calculate the posterior probabilities that sites are from different ω classes (Nielsen and Yang 1998). Positions with a high probability of coming from the class with ω > 1 are considered likely to be under positive selection (Swanson et al. 2001).

ML models that allow for heterogeneity in the d_N/d_S ratio among lineages were also tested. The simplest model is the one-ratio model, M0. The most general model is the free-ratio model, which assumes as many ω parameters as the number of braches in the tree; this model is a parameter-rich model (Yang 1998). Models that fit between these two extremes include the two-ratio and three-

ratio models; these allow predefined lineages to have a different ω value from the rest of the tree (Yang 1998). All of these methods were utilized in an effort to determine whether positive selection has acted along specific lineages within the δ-proteobacteria Omp85 homologues. This was achieved by separate analyses which labeled the 17 internal branches of the phylogenetic tree (Fig. 1) individually to see if any branch has a co that deviates significantly from the others.

Recent studies have shown that under certain conditions ML methods can be sensitive to violations of assumptions made in models that test for positive selection. These sensitivities under certain conditions can result in false positives (Suzuki and Nei 2001a, 2004). To account for this we applied a maximum parsimony (MP) method to test for adaptive evolution. We applied the model of Li (1993) and used a sliding window procedure (Fares et al. 2002) as implemented in SWAPSC (Fares 2004). This method infers a statistically optimum codon-window size and slides it along the alignment. Each window step is then tested for the significance of nonsynonymous substitutions to synonymous substitutions and the nonsynonymous-to-synonymous rate ratio ω; in this manner positively, neutrally, or negatively evolving sites can be located. This method also has the ability to test for saturation of synonymous substitutions (Fares 2004); if any sites are highlighted, they can be removed.

Results

Recombination and Saturation Analysis

The effects of recombination on methods that detect positive selection have been documented (Anisimova et al. 2003). In an effort to ensure that these pitfalls do not affect our results, we utilized the method of Grassly and Holmes (1997) to ensure that no recombination has occurred within our dataset. This

Table 1. Results of the ML analysis of Omp85 using a variety of models

Model	Estimates of parameters	ln$L$	2Δln$L$	Positively selected codons
M0 (one ratio)	$\omega_1 = 0.0761$	19,967.76		None
M1 (neutral)	$p_1 = 0.15$, $p_2 = 0.85$, $\omega_1 = 0$, $\omega_2 = 1$	$-20,780.4$		Not allowed
M2 (selection)	$p_1, = 0.15$, $p_2 = 0.81$, $p_3 = 0.03$, $\omega_1 = 0$, $\omega_2 = 1$, $\omega_3 = 6.1$	$-19,649.4$	(M1 *vs* M2) $p < 0.0005$	286, 564, 673, 674, 675, 676, 735, 736, 737, 738, 740, 743, 745
M3 (discrete $k = 3$)	$p_1 = 0.38$, $p_2 = 0.57$, $p_3 = 0.04$, $\omega_1 = 0.02$, $\omega_2 = 0.1$, $\omega_3 = 3.65$	$-19,603.4$	(M0 *vs* M3) $p < 0.0005$	286, 564, 673, 674, 675, 676, 735, 736, 737, 738, 740, 742, 743, 745
M7 (beta)	$p = 0.90064$, $q = 8.27904$	$-19,632.1$		Not allowed
M8 (beta & ω)	$p = 1.08$, $q = 13.88$, $p_2 = 0.038$, $\omega_2 = 4.98$	$-19,586.9$	(M7 *vs* M8) $p < 0.0005$	286, 564, 673, 674, 675, 676, 735, 736, 737, 738, 740, 742, 743, 745

Note. p_1, p_2, and p_3 refer to the proportion of sites in categories 1, 2, and 3, respectively. ω refers to the d_N/d_S ratios in these categories of sites. p and q are beta estimates. Models M1 and M7 do not allow positively selected sites.

method analyzes an alignment for sequence blocks within the alignment that deviate significantly from a predefined topology (the optimal tree found earlier). No blocks were found to deviate significantly from the imposed phylogeny.

Saturation of synonymous sites is also an important issue when trying to detect adaptive evolution events by the criterion that d_N/d_S is >1 as saturation can lead to the underestimation of d_S and an inflation of the d_N/d_S ratio (Lynn et al. 2004). The moving window program SWAPSC (Fares 2004) was used in an effort to test for saturation of synonymous sites. No saturated sites were located and nearly 97% of the sites show strong purifying selection.

Analysis of Selection

ML analysis of the selection pressures acting on the Omp85 gene provided strong evidence for positive selection (Table 1). The one-ratio model (M0) predicted a ω value of 0.0761, which would indicate very strong purifying selection; however, LRTs indicate that models that allow for positively selected sites increased the likelihood scores when compared with models that do not permit positive selection (Table 1). M2 (selection) fits the data significantly better than M1 (neutral) and locates ~3% of all sites in a class that is under the influence of positive selection ($\omega = 6.1$). M3 (discrete) and M8 (beta) also fit the data better than M0 (one ratio) and M7 (beta), respectively. Both of these models indicate that ~3% of sites have a ω value of 3.65 and 4.98, respectively, a value indicative of strong positive selection. Bayesian posterior probabilities identified 14 positively selected codons for M2, M3, and M8, these codons were almost identical in all case, except M2 did not infer site 740. Codon-based likelihood models that allow for different d_N/d_S ratios among evolutionary lineages (Yang 1998) were used so as to determine whether particular lineages in the datasets are undergoing positive selection. In all cases none of the LRTs performed allowed for positive selection on any one lineage exclusively (results not shown).

All of the sites inferred to be under the influence of positive Darwinian selection using the ML method were also detected using the sliding window approach (results not shown). The sliding window approach also inferred additional sites, which may be under the influence of positive Darwinian selection. The majority of these sites (15 in total) lie between codon 564 and codon 590. The ML method had also put these sites in a class with $\omega > 1$, but they were not considered significant, as their Bayesian posterior probabilities were less than 0.95. From the positive-selected amino acid sites, an average ω value of 2.87 was found. MP methods are very conservative (Suzuki and Nei 2002) and may be subject to the problem of possible convergences (Lynn et al. 2004). Despite this problem, results from ML and MP methods are in good agreement with one another and therefore do not affect any conclusions derived in this study, as we will only consider those sites inferred by both methods to be under the influence of adaptive evolution.

As all Omp85 homologues have been shown to be highly conserved (Gentle et al. 2004; Voulhoux et al. 2003; Voulhoux and Tommassen 2004), and in the absence of tertiary structures we instead used the Omp85 topology of Voulhoux et al. (2003) as a working model on which to plot the positively selected sites that were inferred by both methods. All but one positively selected site (site 281) are found in the regions that are membrane-exposed. Two loops in particular account for the majority (12 of 14) of the positively selected sites (Fig. 2). The remaining positively selected site is located in the periplasmic NH$_2$-terminal domain. As an aside, the additional sites (between codon 564 and codon 590) inferred by the moving window analysis but not by the ML analysis are also found in membrane-exposed regions.
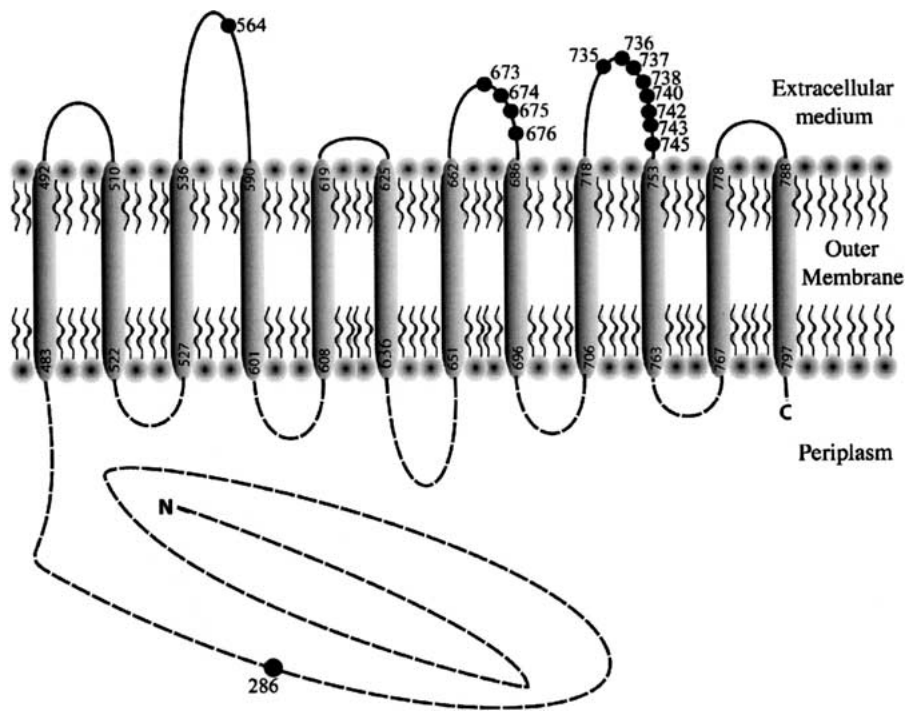
Fig. 2. Topology of Omp85 as predicted by Voulhoux et al. (2003). Periplasmic domains are represented by dashed lines. Exposed domains are represented by solid lines. The first and last amino acid of each β-strand are indicated. Amino acid sites that are under the influence are shown as large black dots. The amino (N) and carboxy (C) termini of the protein are indicated. This model is not to scale.

Discussion

The methods used to detect positive selection in this analysis included ML and MP. Each method has it limitations but in general the ML method has been considered the most robust criterion for detecting adaptive evolution in protein coding genes (Yang 2002). Recent criticisms (Suzuki and Gojobori 1999; Suzuki and Nei 2001b, 2004) suggest that the ML method produces many false positives. To account for this problem we used robust topologies and also a MP approach which has been shown to be conservative. High levels of recombination can also affect ML analysis (Anisimova et al. 2003), yet we did not find any signs of recombination events. This is in agreement with the initial characterization of this gene family, which found no hypervariable or recombining regions (Voulhoux et al. 2003).

ML and MP analysis of selection pressures found that approximately 3% of all codon sites have been under the influence of positive Darwinian selection. ML methods that can test for positive selection along particular lineages were also implemented in this analysis, yet we did not find any evidence for lineage-specific variation of selective pressures. Transmembrane regions of proteins are likely to be highly conserved (Jiggins et al. 2002). This analysis failed to locate any positively evolving sites in transmembrane regions. This finding is probably due to important structural constraints in anchoring Omp85 in the OM and its vital role in cell viability. One positively evolving site was found in the periplasmic NH_2-terminal domain; it is impossible to draw any definitive conclusions from this result in the absence of a tertiary structure, but possible advantages may include an increased ability to fold membrane proteins. When we mapped the remaining positively evolving sites onto a two-dimensional model it became clear that the majority of sites were found in surface-exposed loops, thus revealing the evolutionary processes acting on Omp85. Surface-exposed loops are predicted to be recognized by host responses (Urwin et al. 2002) and are therefore under the greatest pressure for change in an attempt to stay one step ahead of their host/environment in what can be described as an "arms race" (Jiggins et al. 2002). OMP85 has been suggested as a possible vaccine target, as it is highly conserved among the gram-negative bacteria and elicits an immune response. An ideal vaccine candidate should be one that is not amenable to structural change, and broadly speaking this is the case with OMP85. However, we have provided evidence for possible structural change by means of positive Darwinian selection in two membrane-exposed loops of this protein. Information regarding the possible binding sites of anti Omp85-antibodies is not available but it is reasonable to postulate that the two membrane-exposed loops are targets for host antibodies. It is feasible that amino acid altering substitutions at antigen binding sites may occur, thereby rendering vaccines developed from such proteins obsolete. We would suggest that this kind of event is much more likely in those proteins that have historically shown an ability to evolve under positive selection. We therefore propose that an analysis of putative antigen binding sites and a search for his-

torical adaptive evolution within these sites seems to be a sensible precautionary measure prior to the expensive process of developing a vaccine.

References

Andrews TD, Gojobori T (2004) Strong positive selection and recombination drive the antigenic variation of the PilE protein of the human pathogen Neisseria meningitidis. Genetics 166:25–32

Anisimova M, Bielawski JP, Yang Z (2001) Accuracy and power of the likelihood ratio test in detecting adaptive molecular evolution. Mol Biol and Evol 18:1585–1592

Anisimova M, Nielsen R, Yang Z (2003) Effect of recombination on the accuracy of the likelihood method for detecting positive selection at amino acid sites. Genetics 164:1229–1236

Bayer ME (1968) Areas of adhesion between wall and membrane of Escherichia coli. J Gen Microbiol 53:395–404

Creevey CJ, McInerney JO (2002) An algorithm for detecting directional and non-directional positive selection, neutrality and negative selection in protein coding DNA sequences. Gene 300:43–51

Fares MA (2004) SWAPSC: Sliding window analysis procedure to detect selective constraints. Bioinforrnatics:bth303

Fares MA, Moya A, Escarmis C, Baranowski E, Domingo E, Barrio E (2001) Evidence for positive selection in the capsid protein-coding region of the foot-and-mouth disease virus (FMDV) subjected to experimental passage regimens. Mol Biol Evol 18:10–21

Fares MA, Elena SF, Ortiz J, Moya A, Barrio E (2002) A sliding window-based method to detect selective constraints in protein-coding genes and its application to RNA viruses. J Mol Evol 55:509–521

Genevrois S, Steeghs L, Roholl P, Letesson JJ, van der Ley P (2003) The Omp85 protein of Neisseria meningitidis is required for lipid export to the outer membrane. EMBO J 22:1780–1789

Gentle I, Gabriel K, Beech P, Waller R, LithgowT (2004) The Omp85 family of proteins is essential for outer membrane biogenesis in mitochondria and bacteria. J Cell Biol 164:19–24

Grassly NC, Holmes EC (1997) A likelihood method for the detection of selection and recombination using nucleotide sequences. Mol Biol Evol 14:239–247

Huelsenbeck JP, Ronquist F (2001) MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics 17:754–755

Hughes AL, Nei M (1989) Nucleotide substitution at major histocompatibility complex class II loci: Evidence for overdominant selection. Proc Natl Acad Sci USA 86:958–962

Jiggins FM, Hurst GD, Yang Z (2002) Host-symbiont conflicts: Positive selection on an outer membrane protein of parasitic but not mutualistic Rickettsiaceae. Mol Biol Evol 19:1341–1349

Kinsella RJ, Fitzpatrick DA, Creevey CJ, McInerney JO (2003) Fatty acid biosynthesis in Mycobacterium tuberculosis: Lateral gene transfer, adaptive evolution, and gene duplication. Proc Nat Acad Sci USA 100:10320–10325

Kleinschmidt JH (2003) Membrane protein folding on the example of outer membrane protein A of Escherichia coli. Cell Mol Life Sci Cmls 60:1547–1558

Li WH (1993) Unbiased estimation of the rates of synonymous and nonsynonymous substitution. J Mol Evol 36:96–99

Lynn DJ, Lloyd AT, Fares MA, O'Farrelly C (2004) Evidence of positively selected sites in mammalian alpha-defensins. Mol Biol Evol 21:819–827

Manting EH, Driessen AJ (2000) Escherichia coli translocase: the unravelling of a molecular machine. Molecular Microbiology 37:226–38

Nielsen R, Yang Z (1998) Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. Genetics 148:929–936

Posada D, Crandall KA (1998) MODELTEST: Testing the model of DNA substitution. Bioinformatics 14:817–818

Schmidt HA, Strimmer K, Vingron M, von Haeseler A (2002) TREE-PUZZLE: Maximum likelihood phylogenetic analysis using quartets and parallel computing. Bioinformatics 18:502–504

Smith NH, Maynard-Smith J, Spratt BG (1995) Sequence evolution of the porB gene of Neisseria gonorrhoeae and Neisseria meningitidis: Evidence of positive Darwinian selection. Mol Biol Evol 12:363–370

Stephens RS, Lammel CJ (2001) Chlamydia outer membrane protein discovery using genomics. Curr Opin Microbiol 4:16–20

Suzuki Y, Gojobori T (1999) A method for detecting positive selection at single amino acid sites. Mol Biol Evol 16:1315–1328

Suzuki Y, Nei M (2001a) Reliabilities of Parsimony-based and likelihood-based methods for detecting positive selection at single amino acid sites. Mol Biol Evol 18:2179–2185

Suzuki Y, Nei M (2001b) Reliabilities of parsimony-based and likelihood-based methods for detecting positive selection at single amino acid sites. Mol Biol Evol 18:2179–2185

Suzuki Y, Nei M (2002) Simulation study of the reliability and robustness of the statistical methods for detecting positive selection at single amino acid sites. Mol Biol Evol 19:1865–1869

Suzuki Y, Nei M (2004) False positive selection identified by ML-based methods: Examples from the Sig1 gene of the diatom Thalassiosira weissflogii and the tax gene of a human T-cell lmphotropic virus. Mol Biol Evol msh098

Swanson WJ, Yang Z, Wolfner MF, Aquadro CF (2001) Positive Darwinian selection drives the evolution of several female reproductive proteins in mammals. Proc Nat Acad Sci USA 98:2509–2514

Swofford DL (1998) PAUP*: Phylogenetic analysis using parsimony (*and other methods). Sinquer Sunderland, MA

Tamm LK, Arora A, Kleinschmidt JH (2001) Structure and assembly of beta-barrel membrane proteins. J Biol Chem 276:32399–32402

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

Urwin R, Holmes EC, Fox AJ, Derrick JP, Maiden MC (2002) Phylogenetic evidence for frequent positive selection and recombination in the meningococcal surface antigen PorB. Mol Biol Evol 19:1686–1694

Voulhoux R, Tommassen J (2004) Omp85, an evolutionarily conserved bacterial protein involved in outer-membrane-protein assembly. Res Microbiol 155:129–135

Voulhoux R, Bos MP, Geurtsen J, Mols M, Tommassen J (2003) Role of a highly conserved bacterial protein in outer membrane protein assembly. Science 299:262–265

Wolfgang M, Van Putten JP, Hayes SF, Dorward D, Koomey M (2000) Components and dynamics of fiber formation define a ubiquitous biogenesis pathway for bacterial pili. EMBO 19:6408–6418

Yang Z (1997) PAML: A program package for phylogenetic analysis by maximum likelihood. Comp Appl Biosci Cabios 13:555–556

Yang Z (1998) Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution. Mol Biol Evol 15:568–573

Yang Z (2002) Inference of selection from multiple species alignments. Curr Opin Genet Dev 12:688–694

Yang Z, Nielsen R, Goldman N, Pedersen AM (2000) Codon-substitution models for heterogeneous selection pressure at amino acid sites. Genetics 155:431–449