BMC Evolutionary Biology

Research

The role of laterally transferred genes in adaptive evolution
Pradeep Reddy Marri, Weilong Hao and G Brian Golding*

Address: Department of Biology, McMaster University, Hamilton, Ontario L8S 4K1, Canada

Email: Pradeep Reddy Marri - ReddyP@McMaster.CA; Weilong Hao - Weilong@Evol.McMaster.CA; G Brian Golding* - Golding@McMaster.CA

* Corresponding author

Abstract

Background: Bacterial genomes develop new mechanisms to tide them over the imposing conditions they encounter during the course of their evolution. Acquisition of new genes by lateral gene transfer may be one of the dominant ways of adaptation in bacterial genome evolution. Lateral gene transfer provides the bacterial genome with a new set of genes that help it to explore and adapt to new ecological niches.

Methods: A maximum likelihood analysis was done on the five sequenced corynebacterial genomes to model the rates of gene insertions/deletions at various depths of the phylogeny.

Results: The study shows that most of the laterally acquired genes are transient and the inferred rates of gene movement are higher on the external branches of the phylogeny and decrease as the phylogenetic depth increases. The newly acquired genes are under relaxed selection and evolve faster than their older counterparts. Analysis of some of the functionally characterised LGTs in each species has indicated that they may have a possible adaptive role.

Conclusion: The five Corynebacterial genomes sequenced to date have evolved by acquiring between 8 – 14% of their genomes by LGT and some of these genes may have a role in adaptation.

Background

Bacterial genomes are constantly under pressure from the selective challenges of their surroundings. To overcome these hardships, bacterial genomes evolve *via* mechanisms in the form of genome modifications by gene loss [1-3], gene genesis by duplication, modifying existing genes by mutations [4,5] or acquisition of new genes by lateral gene transfer (LGT) [6-13]. Recent studies indicate that LGT has a larger role in bacterial evolution than previously anticipated [14-19], accounting for anywhere between 1.6 – 32.6% of the genes in each individual genome [20].

Gene content varies dramatically even among strains belonging to a single bacterial species [21-23]; variations mostly resulting from gene loss [1-3] and/or acquisition of new genes by LGT [6-13]. LGT plays a significant role in the evolution of bacterial genomes and provides them with a ready-to-use novel gene pool that helps them to adapt faster to their ever changing surroundings and foray into new ecological niches. Documented evidence shows that laterally acquired genes can transform an otherwise avirulent bacteria into a virulent form [24,25], protect pathogenic bacteria against antibiotics [26], increase the metabolic diversity of the recipient bacteria [12,27-29] or

confer on it abilities to explore new challenging niches [30,31]. Keeping in mind the capability of LGTs to provide diverse adaptive features, we review some of the previous work done on lateral gene transfer in bacteria with an emphasis on the adaptive role of these laterally acquired genes. We also provide evidence about the transient nature of most of the laterally acquired genes based on a maximum likelihood modeling of the gene insertions/deletions at various stages during the evolution of the *Corynebacterium* species.

For the ease of discussion, we have classified the adaptive features into three major categories and will review each in turn: (1) Pathogenicity related features (2) Metabolic capabilities and (3) Survival under extreme environmental conditions.

Pathogenicity related features
There are many documented instances of the acquisition of virulence determinants in bacteria by the process of LGT [22,25,32-38], selected examples are discussed below. The acquisition of a 35 kb *eaeA* locus encoding proteins responsible for attaching and effacing lesions has transformed an avirulent *Escherischia coli* strain into an enteropathogenic strain [39], whereas the acquisition of pathogenicity islands (PAIs) ranging from 70–150 kb and encoding virulence realted proteins resulted in uropathogenic strains [40,41]. Lawrence and Ochman [7] have identified that about 18% of the genome of *E.coli* MG1655 was acquired by LGT and this laterally acquired DNA has "conferred properties permitting *E.coli* to explore otherwise unreachable ecological niches".

The genome of *Salmonella enterica* has two laterally acquired pathogenicity islands, SPI-1 and SPI-2 encoding proteins that help in apoptosis, entry into non-phagocytic cells and systemic infection [42], whereas *Bacillus cereus* genome has three laterally acquired genomic islands BCGI-1, BCGI-2 and BCGI-3 with genes encoding proteins responsible for antibiotic resistance, ferric anguibactin transport system and lantibiotic biosynthesis leading to a better survival of *B. cereus* inside the host [43]. The highly pathogenic strains of *Yersinia pestis* have a 102 kb High Pathogenicity Island (HPI) that contains the *hms* locus encoding the capacity to store hemin, yersinibactin-pesticin receptor and an iron-regulated high molecular weight protein enabling an increased level of pathogenicity and survival in their hosts [44]. The case with the *cag* pathogenicity island in *Helicobacter pylori* is similar. This laterally acquired region encodes many antigenic determinants and virulence factors indicating its role in pathogenesis [45].

A comparison of the virulent and benign strains of *Dichelobacter nodosus*, a principal causative agent of the ovine footrot, revealed that the acquisition of *vap* and *vrl* regions encoding several virulence related genes has transformed an otherwise benign strain into a virulent strain [46,47]. Similarly, the virulent strains of *Vibrio cholerae* acquired a 45 kb PAI that includes the *tcp-acf* gene cluster involved in colonization and the *toxT* gene involved in the regulation of cholera toxin. This region, absent in the corresponding avirulent strains, plays a role in virulence and host adaptation [48].

This data indicates the ability of the laterally acquired genes to transform an otherwise non-pathogenic bacteria into a pathogenic bacteria. Interestingly, all the examples involve the acquisition of an adaptive trait, virulence, by acquiring large islands of linked genes.

Metabolic adaptations
Primary metabolism
A recent study on the evolution of metabolic networks in *E. coli* by Pal et al [27] suggested that most of the laterally acquired genes have specific adaptive roles. Another study based on subtractive hybridisation by Espinosa and Kolter [49] has shown that the laterally acquired *gapC* gene has an adaptive role in the survival of *E. coli* in aquatic environments. Jahreis et al [50] demonstrated that the acquisition of the *CTscr94* locus (consisting of *scr* genes that encode a phosphoenol-dependent phosphotransferase system involved in the sucrose fermentation pathway) broadens the metabolic versatility of *Salmonella senftenberg* and confers on it the capability to utilise sucrose as the sole carbon source. Sullivan and Ronson [28] have demonstrated that the laterally acquired 500 kb symbiosis island from *Mesorhizobium loti* ICMP3153 confers symbiotic ability to non-symbiotic *Mezorhizobium* strains and enables the bacterium to expand its genome to exploit new environmental niches. Similarly, the acquisition of the cyanobacterial nitrogen fixation (*nif*) genes by *Wolinella succinogenes* might enable it to survive in environments outside the host [29].

Degradation of xenobiotics
The ability to degrade xenobiotics is a relatively new physiological development in bacteria and several instances of the adaptability of the bacterial communities to xenobiotics have been attributed to the acquisition of genes related to xenobiotic degradation by LGT [12,51,52].

Pseudomonas sp. strain B13 has a genomic island of about 105 kb containing the chlorocatechol degradative (*clc*) genes that have the ability to degrade 1,2-dichlorobenzene (CB). Ravatn et al [53] have demonstrated that the acquisition of the B13 specific *clc* genes by *Pseudomonas putida* increases its metabolic diversity and results in a better adaptation to higher concentrations of CB. Dejonghe et al [54] have demonstrated that the transfer of 2,4-

dichlorophenoxy acetic acid (2,4-D) degradation plasmids pEMT1 and pJP4 from *P. putida* UWC3 into the indigenous bacteria of two different horizons of 2,4-D contaminated soil enhanced the ability of the native bacteria to degrade 2,4-D, whereas de Lipthay et al [55] have demonstrated that lateral transfer of the gene (*tfdA*) encoding 2,4-dichlorophenoxyacetic acid dioxygenase into *Ralstonia eutropha* and other indigenous phenol degrading strains enhances the phenoxyacetic acid degradative capacity of the recipient strains in soil.

Poelarends et al [56] have indicated that the acquisition of the haloalkane dehalogenase gene (*dhaA*) was the key factor in the evolution of the haloalkane degradation ability of *Rhodococcus rhodochrous*, *Pseudomonas pavonaceae* and *Mycobacterium* sp. Similarly, the acquisition of the *pcpB* gene encoding pentalchlorophenol-4-monoxygenase by *Sphingomonas chlorophenolicum* confers on it the ability to degrade 2,3,4,6-tetrachlorophenol and survive under high concentrations of this chemical compound [57].

From the examples discussed above it is clear that some of the laterally acquired genes have the ability to increase the metabolic diversity of recipient bacteria enabling them to explore new ecological niches.

Survival under extreme environmental conditions
The genome sequence of the radiotolerant bacteria *Deinococcus radiodurans* has revealed that topoisomerase IB (that plays a role in recombination and for which a knockout mutant is more sensitive to UV radiation) and RNA-binding protein (that is involved in the regulation of UV related damage repair) were acquired by LGT from eukaryotes [58]. *D. radiodurans* has also acquired the genes belonging to the late embryogenesis abundant (*lea*) family from plants [58]. These genes give resistance against dessication in plants and might provide *D. radiodurans* with additional resistance to UV radiation, since there is a positive correlation between resistance to dessication and radioresistance [58].

A comparative genomic analysis of the peizophilic strains of *Photobacterium profundum* has identified that several genes that are common to all the strains and responsible for high pressure adaptation are probably laterally acquired. Some of these genes are upregulated under high pressure conditions possibly suggesting their involvement in high pressure adaptation of *P. profundum* [30]. The genome sequence of *Colwellia psychrerythraea* has revealed that the psychrophilic lifestyle of this organism is due to the acquisition of a few genes by LGT in addition to amino acid content variations. Though LGT was not a primary player in the adaptation of this organism to this lifestyle, some specific lateral acquisitions including some of the cold shock proteins (possibly acquired from *Vibrio* sp.

and *Shewanella onediensis*) and proteins responsible for the synthesis and degradation of complex high molecular weight organic compounds (acquired from *Ralstonia eutropha*) are certainly adaptive [31].

The genome sequence of *Halobacterium* indicates that most of the components of the electron transport chain including NADH dehydrogenase (*nuo*), menaquinone (*men*) and cytochrome oxidase (*cox*) are similar to the corresponding genes in *E.coli* and *D.radiodurans*. This suggests the transfer of genes involved in aerobic respiration from eubacteria into *Halobacterium* by LGT [59]. Another thermophilic archaea, *Picrophilus torridus* has acquired the genes that have the ability to degrade the organic acids such as acetic and propionic acid, superoxide dismutase, peroxiredoxin and alkyl hydroperoxide reductase to cope with oxidative stress along with the genes that encode the main components of the electron transport chain [60].

In contrast to pathogenic adaptations, these adaptations have largely been acquired by the LGT of a single gene. The above examples spanning different bacterial genomes clearly emphasize the adaptive role of LGTs. However, the examples discussed above are only the success stories of laterally acquired genes. Not all of the laterally acquired genes are adaptive. Of the many genes that are acquired, only a few genes that are adaptive are retained and fixed in a population [61]. There have been some studies suggesting that most of the laterally transferred genes are not useful and are hence reduced to pseudogenes [62] and studies that indicated the inability of LGTs to translate into functional proteins [63]. Given the examples above emphasizing the adaptive role of LGTs and counter-examples that suggest that most of the LGTs are not adaptive, it is important to explore the adaptive nature of LGTs and how frequently they are adaptive, if at all.

Results
To answer the question on the adaptive role of LGTs, we have employed a maximum likelihood method to infer gene insertion/deletion rates at various stages during the evolution of five sequenced *Corynebacterium* genomes and have performed a systematic study on the species specific genes to further understand the role of the laterally acquired genes.

Insertion and deletion rates
Gene insertions and deletions in closely related genomes can be inferred by looking at the phyletic patterns of gene presence or absence on a phylogenetic tree. Maximum parsimony and maximum likelihood methods have been successfully employed to understand the gene insertions/deletions [18,64-70]. The maximum likelihood analysis [70] was performed using the phylogeny derived from a concatenated DNA sequence of the em fusA, gltS, infB,

lysS, rplB, rpoB, secY, serS and em ychF genes (Figure 1) and assuming that individual insertion and deletion events occur independently. The insertion/deletion rate of 1.18 (LnL = -10875.4) was obtained using a model which assumes a single constant insertion/deletion rate (μ) on all branches (Model I). Branch lengths are measured relative to the estimated number of base substitutions, suggesting that there is one gene gained/lost for every nucleotide substitution. This rate is higher than that obtained from the study on *Bacillus* (0.51) [70], and close to that of *Streptococcus* (1.17) [71]. A second model, Model II, that assumes different insertion/deletion rates on external and internal branches (Figure 1) resulted in a higher rate (1.26) on external branches compared to the internal branches (1.06; Table 1) indicating that gene movement is greater on the external branches. The improvement in the likelihood although small is significant ($\chi^2 = \Delta2$ LnL > 3.84 with d.f. = 1). A non-hierarchical model, assuming an irreversible gene loss, also shows a similar trend of a higher rate on external branches (1.33) over that on internal branches (0.60) (Table 1). The likelihood surface and the curvature of the likelihood surface are given as supplementary data (See Additional Files 9, 14 and 15).

A third model (Model III) considered a separate rate on each branch and also indicated that the external branches have an enhanced apparent insertion/deletion rate and that the rate of change varies greatly (Table 2). The large change in the likelihood from Model II to Model III with a small change in the number of parameters suggests that the fit of the model is significantly better. Interestingly, the rates on the external branches ($\mu3$, $\mu4$) leading to the two strains of C. glutamicum were higher compared to the other external branches. This is probably due to the larger genome size of C. glutamicum and smaller branch lengths. It is not due to the smaller number of genes differentiating these taxa since if these numbers of genes are halved, the rates remain comparable. Similarly, increased insertion/deletion rates on the branches leading to C. glutamicum strains ($\mu6$) and C. glutamicum and C. efficiens ($\mu7$) could be due to the fact that these represent the branches leading to species having larger genome sizes.

Species specific genes
Species specific genes were divided into two categories based on BLAST and phylogenetic clustering; (1) Lateral gene transfers and (2) possible multiple deletion or lateral transfers (MDLT) (Table 3). The unique genes that cluster with BLAST homologues (at an E-value cutoff of 1.0×10^{-10}) outside actinobacteria were considered as possible LGTs. Genes that were uniquely present in a *Corynebacterium* species but cluster with BLAST homologues from taxa within actinobacteria were considered as MDLT. The MDLT may possibly be instances of multiple deletions

that are specifically retained by a single species or cases of intra-group gene transfers. LGTs reported here do not include the ORFan genes (as defined by [72]; genes with no hits to the NCBI **nr** database) of these species. As a result, the number of LGTs presented here will be a large under estimate of the actual number of LGTs in each of these species. The list of LGTs and their putative function along with the entire list of species specific genes can be found at [73].

The present study identified 215 genes specific to the *C. diphtheriae* genome. Of the 215 genes, 48 were possibly acquired by lateral gene transfer whereas the remaining 167 are MDLT. The functions of most of these genes are unknown (81 out of 215 encode hypothetical proteins), however, a comparison of the results with earlier studies on *C. diphtheriae* [74] indicate that many of these genes might encode proteins that are responsible for virulence related traits.

The study identified 283 *C. efficiens* specific genes, 225 are designated as MDLT while the remaining 58 are possibly acquired by LGT. A majority (194/283) of the species specific genes encode hypothetical proteins.

A comparison with three other corynebacterial genomes revealed the presence of 659 genes that are uniquely present in *C. glutamicum* species. These are the genes that are present in both the strains of *C. glutamicum* and absent in other corynebacterial species. This does not include the set of genes that are unique to a single strain of *C. glutamicum*. Of the 659 genes, 74 were acquired by LGT, 206 were identified as MDLT due to their presence in other actinobacterial taxa. There were 377 genes present only in the two strains of *C. glutamicum* and did not have hits in NCBI. Many (346/659) of the species specific genes encode hypothetical proteins. Unlike the other corynebacterial species, *C. glutamicum* has 58 membrane proteins, 16 regulatory proteins and 15 ABC transport related proteins (as classified in the annotation) indicating a greater metabolic diversity compared to the other species.

The study revealed the presence of 323 genes that are specific to *C. jeikeium*. Of the 323 genes, 292 are MDLT whereas the remaining 31 genes were possibly acquired by LGT. Again, a majority (179 of 323) of the species specific genes encode hypothetical proteins, however, some of the genes whose function is identified revealed that these genes were involved in the uptake of iron and manganese and hence might have a role in virulence and pathogenesis.

Synonymous/non-synonymous changes and recent vs. ancient transfers
The synonymous substitution rate (K_s), non-synonymous substitution rate (K_a) and their ratio (K_a/K_s) was measured
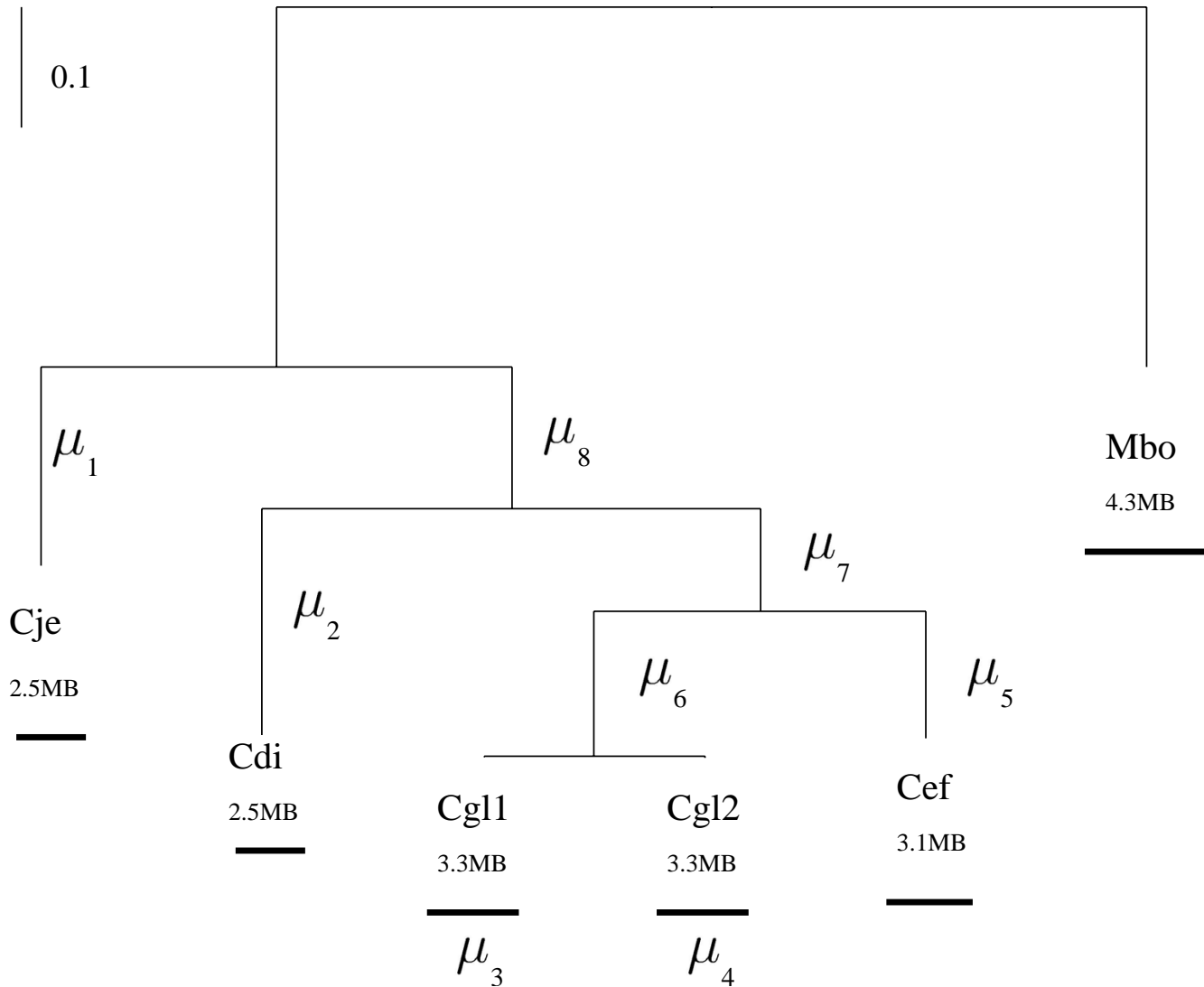
Figure 1
Bayesian tree obtained for the concatenated sequence of *fusA, gltS, infB, lysS, rplB, rpoB, secY, serS and* ychF genes. The abbreviations are: Cje (*Corynebacterium jeikeium*), Cdi (*C. diphtheriae*), Cgl$_1$ (*C. glutamicum* ATCC 13032, NCBI accession No. NC_006958), Cgl$_2$ (*C. glutamicum* ATCC 13032, NCBI accession No. NC_006958), Cef (*C. efficiens*), and Mbo (*Mycobacterium bovis*) as the outgroup. All branches are supported by posterior probability values of 100%. At least 80% of the trees for the commonly present single copy genes (534 genes in total) support branches 7 (with 82%) and 8 (with 86%), and trees for all genes support the remaining branches. **Model I**: Constant insertion/deletion rate on all branches. $\mu = \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5 = \mu_6 = \mu_7 = \mu_8$. **Model II**: Branches separated into external and internal branches. *External*: $\mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5 = \mu_6$; *Internal*: $\mu_7 = \mu_8$ **Model III**: Different insertion/deletion rate on each branch.

for the orthologous genes between *C. glutamicum* and *C. efficiens*. The genes were divided into four groups based on the evolutionary time scale : genes present in *C. glutamicum* and *C. efficiens* only, genes present in *C. glutamicum, C. effciens* and *C. diphtheriae* only, genes present in all *Corynebacterium* species and absent in the outgroup and genes present in all the *Corynebacterium* species along with the outgroup. The genes present only in *C. glutamicum* and *C. efficiens* represent the putative most recently acquired

genes while the genes present in all the species including the outgroup represent the most ancient genes. The results indicate that the genes that are inferred to have arisen recently via LGT in the phylogeny have more non-synonymous substitution changes than those that were transferred somewhat more anciently (Figure 2A versus 2B is *P* < 0.01 in a Wilcoxon rank test). As genes are resident in the species for longer periods the non-synonymous changes continue to go down (Figure 2B versus 2C is *P* <

Table 1: Insertion/deletion rates inferred from the maximum likelihood analysis assuming different rates for external and internal branches.

Rate	Reversible		Deleted once[a]	
	MLE	Ln*L*	MLE	Ln*L*
constant μ	1.18	-10875.4	1.08	-10991.8
$\mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5 = \mu_6$ (external)	1.26	-10873.0	1.33	-10958.3
$\mu_7 = \mu_8$ (internal)	1.06		0.60	

[a]Genes can not be regained after deletion.

0.01, Figure 2C versus 2D is *P* < 0.01). A similar trend was observed for K_s and K_a/K_s where recently acquired genes had a higher rate of K_s (Figure 3) and K_a/K_s (Figure 4) compared to the genes that have been resident longer. The analysis was repeated removing all the genes that did not have any homologs in non-*Corynebacterium* genomes, assuming that the genes present only in *C. glutamicum* and *C. efficiens*, and have no hits in the NCBI **nr** database are potential annotation artefacts. The distribution of K_a, K_s, and K_a/K_s values remains remarkably similar even after removing the uniquely present ORFs (see Additional Files 10, 11, 12) confirming the robustness of the result that recently acquired genes have higher rates of K_a, K_s, as well as a higher K_a/K_s ratio.

Discussion

A maximum likelihood estimation of the gene insertion/deletion rate assuming a constant rate on all the branches was 1.18. This is higher than the inferred base substitution rates on the branches indicating that gene insertion/deletion plays a significant role in the evolution of these genomes. Genome rearrangement has been shown to have a minor role in the evolution of corynebacterial genomes [75] indicating that gene gain/loss might have a greater role in the evolution of *Corynebacterium* species. A model considering independent rates on each branch

confirmed the enhanced rate of insertions/deletions on the external branches clearly showing a decrease in the rate with an increase in phylogenetic depth. The enhanced rate on the external branches of the phylogeny might indicate the transient nature of many of the laterally acquired genes. However, the observed difference between the rates on external and internal branches is not as dramatic as reported for the *Bacillus* group [70]. This may be due to the different phylogenetic relationship between the studies. If the transient nature of the LGTs holds true, one should expect lower rates of insertions/deletions on long external branches compared to short external branches. In fact, the rate of ins/del on short external branches in the *Bacillus cereus* group [70] is higher than the rate on external branches in this study. Furthermore, the rate estimation of insertion/deletion is robust for different cutoffs used for determining gene homologues. Different cutoffs (expect value <10⁻²⁰ with match length > 85%, expect value <10⁻¹⁰ with match length > 70%, and expect value <10⁻⁵ with match length > 50%) show the similar trend that external branches tend to have higher rates of insertions/deletions than internal branches (see Additional Files 1, 2, 3, 4, 5, 6, 7).

In this study, the maximum likelihood estimation is based on a concatenated phylogeny of DNA sequences of

Table 2: Insertion/deletion rates inferred from the maximum likelihood analysis assuming different rates on each branch of the phylogeny.

Rate	Reversible		Deleted once[a]	
	MLE	Ln*L*	MLE	Ln*L*
constant μ	1.18	-10875.4	1.08	-10991.8
μ_1	1.37	-10578.0	1.29	-10618.8
μ_2	0.46		0.48	
μ_3	4.60		4.87	
μ_4	27.42		32.82	
μ_5	0.98		1.06	
μ_6	2.25		2.20	
μ_7	2.49		2.13	
μ_8	0.001		0.008	

[a]Genes can not be regained after deletion.

Table 3: Classification of unique genes in Corynebacterium species.

Genome	Total Genes	Species specific genes (Hypothetical Proteins)	LGTs	MDLT
C. diphtheriae	2272	215 (81)	48	167
C. efficiens	2950	283 (194)	58	225
C. glutamicum	3057	659 *(346)	74	206
C. jeikeium	2137	323 (179)	31	292

Numbers based on an E value cutoff of 1.0×10^{-10} with no length cutoff.
*377 genes present in *C. glutamicum* strains only

nine genes with Mycobacterium bovis as the outgroup. The phylogenetic topology is supported by ribosomal RNA sequences (data not shown) and by the commonly present genes (Figure 1). The topology is robust regardless of the outgroup chosen from Mycobacterium, which is the closest phylogenetic neighbor of Corynebacterium. Furthermore, possible alternative topologies of the commonly present genes were evaluated using nine concatenated genes. The best supported alternative topol-

ogy (see Additional File 13) is significantly worse than the topology used in the study. However, the maximum likelihood estimation based on the alternative topology still supports that recently acquired genes have high rates of ins/del (see Additional File 8). In this model, insertion rate and deletion rate were assumed to be equal. This assumption was based on the fairly constant genome sizes of the closely related taxa and also to ensure that in the long term, genome sizes would not tend to zero or infin-
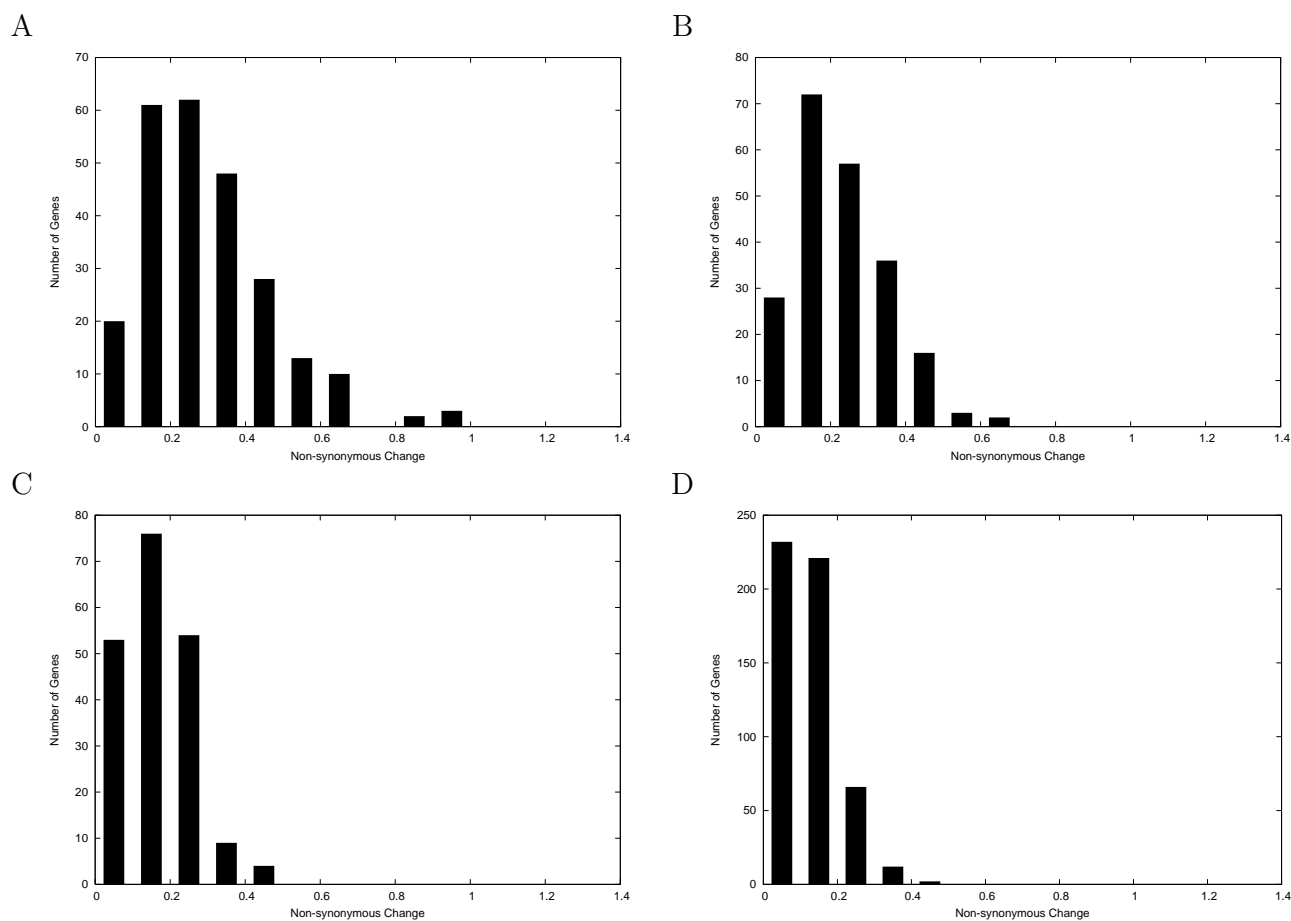

Figure 2
Non-synonymous change of different group-specific genes. Data is measured using the genes present in Cgl and Cef; genes present only in Cgl and Cef (A); genes present only in Cgl, Cef, and Cdi (B); genes present in only Cgl, Cef, Cdi, and Cje (C); genes present in all *Corynebacterium* taxa and Mbo (D).
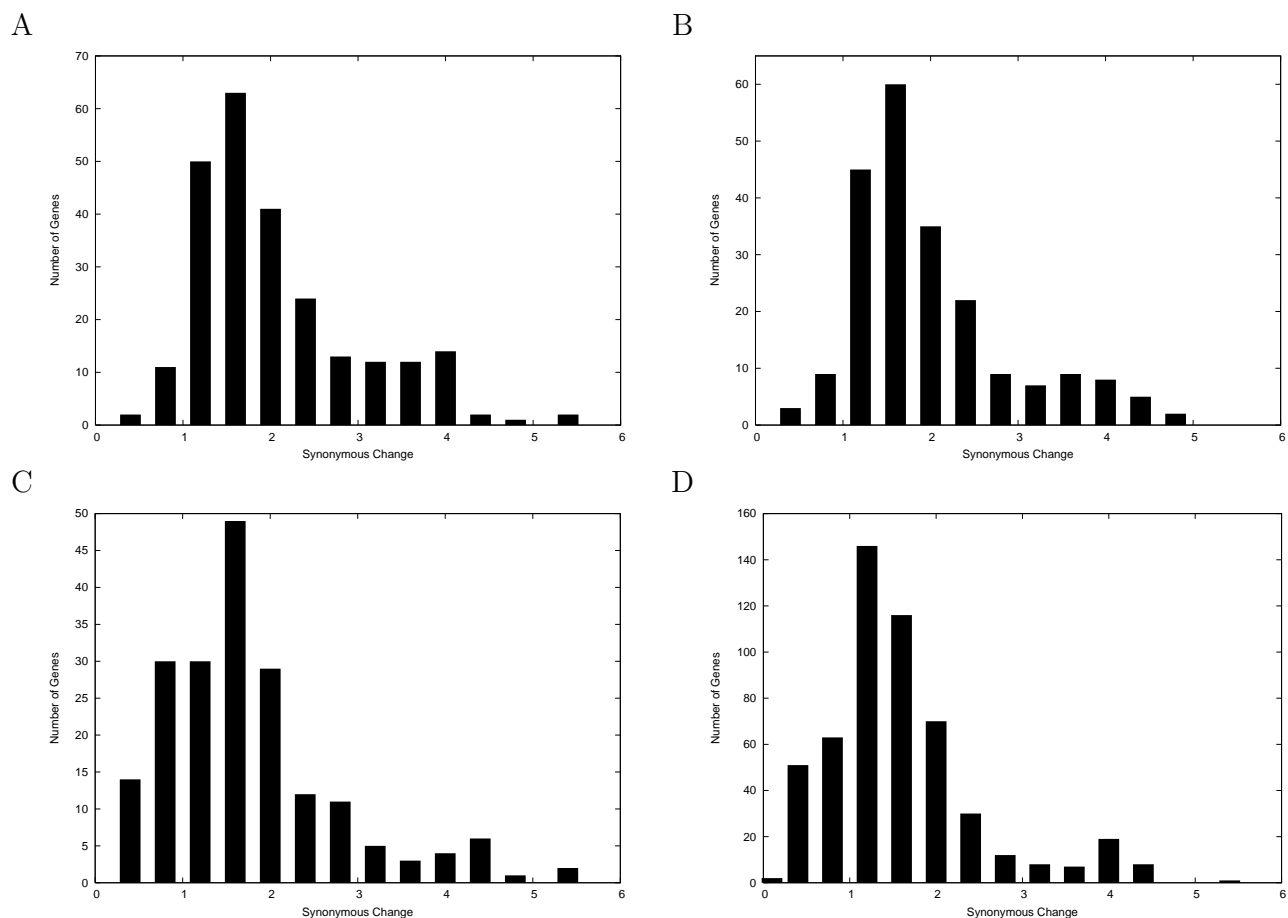
Figure 3
Synonymous change of different group-specific genes. Data is measured using the genes present in Cgl and Cef; genes present only in Cgl and Cef (A); genes present only in Cgl, Cef, and Cdi (B); genes present in only Cgl, Cef, Cdi, and Cje (C); genes present in all *Corynebacterium* taxa and Mbo (D).

ity. In the study, all insertions/deletions were assumed to be independent and the rate of insertion/deletion was estimated from the gene phyletic pattern. Because of the difficulty of inferring insertion/deletion events after many genome rearrangements, the number of gene insertions/ deletions rather than the actually insertion/deletion events was used in the maximum likelihood estimation. To overcome the simplistic assumptions in this study, more practical parameters, such as block ins/del rates, can be added to make the model more realistic in future studies.

The comparison of the synonymous (K_s) and non-synonymous (K_a) substitution rates of the genes that entered at various levels of the phylogeny indicate that the recently acquired genes evolve faster and have a higher proportion of synonymous as well as non-synonymous substitutions compared to their older counterparts (Figures 2,3). These results agree with earlier results indicating faster evolution

of the recently acquired genes [70-72]. One of the possible reasons for faster rates of evolution could be that the newly acquired genes are required in the new habitat and are evolving faster to adapt to their new roles/habitat. Alternatively, some of these genes could be under relaxed functional constraint as they are non-functional in the new environment and might be evolving faster until they are deleted [70]. A more recent study has demostrated that low GC content causes seletive silencing of foreign DNA [76]. It was found that recently acquired genes tend to have lower GC content compared with ancient ones (see Additional File 16). This also supports the elevated evolutionary rates of recently acquired genes.

In this study, the phyletic patterns were derived primarily from genome annotation. As described previously in [70], non-annotated genes were picked up via a TBLASTN search and genes that are uniquely present in only one studied genome were removed from the study. Further-
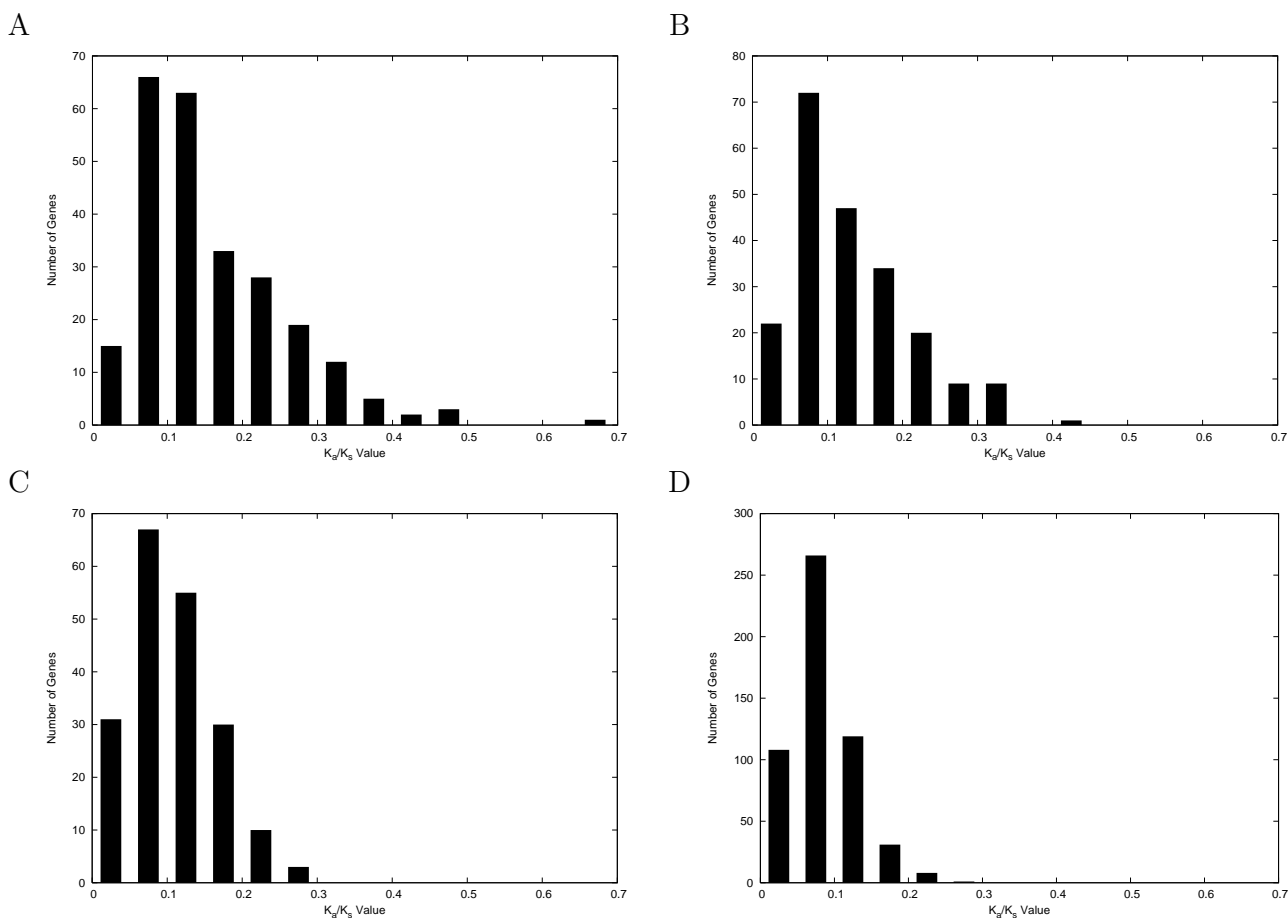
Figure 4
K_a/K_s ratio of different group-specific genes. Data is measured using the genes present in Cgl and Cef; genes present only in Cgl and Cef (A); genes present only in Cgl, Cef, and Cdi (B); genes present in only Cgl, Cef, Cdi, and Cje (C); genes present in all *Corynebacterium* taxa and Mbo (D).

more, a comparison was made by removing the ORFs only present in *C. glutamicum* and *C. efficiens* but not present in any other complete bacterial genomes (see Additional Files 10, 11, 12). The rates do not change remarkably after removing the Cgl-Cef group unique genes. This suggests that the fast rate of evolution of recently acquired genes is not an artifact of fast evolving non-gene ORFs in *C. glutamicum* and *C. efficiens*. A comparison of the corynebacterial genomes revealed that about 9 – 21% of the genes are specific to each species indicating that gene gain/ gene loss has a major contribution in the evolution of these genomes. The results indicate that more than 10 – 35% of the species specific genes have been acquired by lateral gene transfer while the remaining are identified as possible MDLT (as inferred by their presence in other actinobacterial genomes).

The LGTs broadly fall into two categories; functionally characterised genes and hypothetical proteins. The analysis of the functionally characterised genes reveals that the pathogenic species *C. diphtheriae* and *C. jeikeium* have recruited genes that help in survival, host attachment and virulence. *C. diphtheriae* has acquired the genes responsible for iron transport [77,78] and siderophore biosynthesis that are directly involved in virulence. The acquisition of the genes encoding fimbrial subunits might indicate an adaptive mechanism whereby it helps in the attachment to the host cell surface [79], whereas the acquisition of the lantibiotic biosynthesis genes might help it to defend itself from other bacteria [80]. *C. jeikeium* has acquired genes that help in iron uptake. In addition, it has also acquired the genes necessary for uptake of manganese, another important requirement in pathogenesis [81]. The acquisition of the gene *cbpA* encoding a collagen binding protein might help in the bacterium-host interaction [82],

whereas the presence of the genes *surA* and *surB*, encoding surface proteins, and the gene *acpA*, encoding an alkaline phosphatase, might help in virulence [83]. The acquisition of the neuraminidase encoding gene suggests that this might help *C. jeikeium* to prevent competition by other bacteria [84].

Unlike pathogenic species, the non-pathogenic species appear to have recruited many genes that are directly or indirectly involved in metabolic processes. *C. glutamicum* has acquired large numbers (53) of genes encoding membrane related proteins compared to the other species. Nishio et al [85] have indicated that many of the genes involved in amino acid biosynthesis are vertically inherited by *C. glutamicum* and only a few are acquired by lateral gene transfer. Our analysis identified only three genes (*brnE*, *mdh2*, *scrB*) that have been found to have a role in amino acid and vitamin biosynthesis [86]. The analysis of the LGTs in *C. efficiens* did not give many clues of their function as most of these genes encode hypothetical proteins.

We haven't considered in our study the ORFan genes [72] that account for about 10% of the genes in each of these species. Most of the recent studies on ORFans have confirmed that they are not a result of annotation errors but are in fact true genes, most of them being a part of genomic islands indicating their acquisition by LGT [23,72,87-89]. Given some of the examples reviewed here and based on the results obtained from this study on *Corynebacterium* species, it is compelling to suggest that at least some of these laterally acquired genes might have a role directly or indirectly in the adaptation of corynebacterial species.

Conclusion
We demonstrate that 13 – 20% of the protein coding genes are specific to each *Corynebacterium* species and that these species have evolved mostly by LGT and gene loss. Maximum likelihood analysis indicates that there are more lateral transfers inferred at the tips of the phylogeny and that most of the lateral gene transfers are transient. Recently acquired genes evolve faster compared to their native counterparts. The faster rate of evolution of these recently acquired genes might reflect adaptation to new niches or might reflect rapid gene decay.

Methods
Genome sequences used
Five *Corynebacterium* genome sequences were obtained from NCBI [90] to carry out the analysis. They are *C. jeikeium* (Cje; [91]), *C. diphtheriae* (Cdi; [74]), *C. glutamicum* (NCBI accession No. NC_003450, Cgl1; [92]), *C. glutamicum* (NCBI accession No. NC_006958, Cgl2; [86])

and *C. efficiens* (Cef; [93]). *Mycobacterium bovis* (Mbo; [94]) was used as the outgroup.

Maximum likelihood analysis
The evolutionary history of the *Corynebacterium* taxa was generated from the concatenated DNA sequences of *fusA*, *gltS*, *infB*, *lysS*, *rplB*, *rpoB*, *secY*, *serS and* ychF genes using Mr.Bayes ([95]; 200,000 generations sampled every 100 generations with a gamma distribution model and invariant class). The method to identify members of a gene family has been described in [66]. In brief, potential homologues were measured according to sequence similarities, and all paralogues in each genome were clustered as a single gene family and only one member was used for further analysis. The phyletic patterns (gene presence or absence in each genome) of all genes were used for the maximum likelihood analysis (see Additional File 1). The method to estimate likelihood has been described in [70]. Three different models, model I, II and III were used. Model I assumed a constant insertion/deletion rate on all the branches. In model II, the branches were separated into external and internal branches and rates were calculated separately for internal and external branches. In Model III the rates were calculated assuming each individual branch separately.

Identification of species specific genes
The protein sequences from all the genomes were compared to each other using BLASTP [96], with an E-value cutoff set at 1.0×10^{-20} and an additional criteria of match length set at 85% of the query sequence. A set of genes present uniquely in a species and absent in the outgroup were considered as specific to that species. The species specific genes were checked to ensure that they are not simply a result of annotation errors. In the case of *C. glutamicum* with two intraspecific strains, all the genes that are present in both the strains and absent in other genomes including the outgroup were considered species specific. The genes present uniquely in a single strain were not considered. The species specific genes identified are sensitive to the parameters used in any study and to the number of taxa sampled.

Lateral gene transfer
The unique proteins of each species were compared to the NCBI **nr** database using BLASTP and with the expect value cutoff set at 1.0×10^{-10} to identify homologs in other organisms. For each protein, the first 50 hits with an expect value less than 1.0×10^{-10} were chosen for further analysis regarding possible lateral gene transfers. The complete protein sequences of these 50 hits were extracted from the GenBank database and a multiple alignment was done using ClustalW [97]. The multiple alignment was used to generate a phylogenetic tree using the neighbor joining (NJ) method [98] as implemented in PHYLIP

[69]. The genes identified as LGTs by NJ trees were further confirmed by reconstructing phylogenetic trees using PROML (JTT + invariable class + γ distribution model) with rate paratmeter α calculated by Tree Puzzle [99].

Synonymous/non-synonymous substitution rate

Synonymous changes and non-synonymous changes were measured by the PAML package [100]. The tree length of synonymous changes was calculated as the sum of the branch lengths for the taxa only within Cgl and Cef using the maximum likelihood method from the PAML package. Genes were categorised into four groups based on their presence/absence in different taxa (and hence on the inferred time period when the genes were transferred). The four groups are characterised by genes present in Cgl and Cef; genes present in Cgl, Cef, and Cdi; genes present in Cgl, Cef, Cdi, and Cje; and genes present in Mbo and all *Corynebacterium* taxa. Single copy protein sequences and their corresponding DNA sequences were extracted from the annotated genomes. Protein sequences were aligned using ClustalW [97], and nucleotide sequence alignments were created from the protein alignments by replacing each amino acid with its corresponding codon.

Authors' contributions

PRM and WH contributed equally to this work. PRM, WH and GBG conceived the idea. WH performed the likelihood analysis and the measurement of K_a/K_s, PRM performed the LGT analysis. PRM, WH and GBG wrote the manuscript. All the authors have read and approved the manuscript.

Additional material

Additional File 1
The phyletic patterns of the Corynebacterium *taxa. (cut-off: expect value less than 10^{-20} and match length over 85%)*
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2148-7-S1-S8-S1.pdf]

Additional File 2
The phyletic patterns of the Corynebacterium *taxa (cut-off: expect value less than 10^{-10} and match length over 70%)*
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2148-7-S1-S8-S2.pdf]

Additional File 3
The phyletic patterns of the Corynebacterium *taxa (cut-off: expect value less than 10^{-05} and match length over 50%)*
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2148-7-S1-S8-S3.pdf]

Additional File 4
Insertion/deletion rates inferred from the maximum likelihood analysis assuming different rates for external and internal branches (cut-off: expect value less than 10^{-10} and > 70% match length)
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2148-7-S1-S8-S4.pdf]

Additional File 5
Insertion/deletion rates inferred from the maximum likelihood analysis assuming different rates on each branch of the phylogeny (cut-off: expect value less than 10^{-10} and > 70% match length)
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2148-7-S1-S8-S5.pdf]

Additional File 6
Insertion/deletion rates inferred from the maximum likelihood analysis assuming different rates for external and internal branches (cut-off: expect value less than 10^{-05} and > 50% match length)
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2148-7-S1-S8-S6.pdf]

Additional File 7
Insertion/deletion rates inferred from the maximum likelihood analysis assuming different rates on each branch of the phylogeny (cut-off: expect value less than 10^{-05} and > 50% match length)
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2148-7-S1-S8-S7.pdf]

Additional File 8
Insertion/deletion rates inferred from the maximum likelihood analysis assuming different rates for external and internal branches. Likelihood was estimated using the alternative topology and a cutoff with expect value less than 10^{-20} and > 85% match length.
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2148-7-S1-S8-S8.pdf]

Additional File 9
Curvature of likelihood change in different models.
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2148-7-S1-S8-S9.pdf]

Additional File 10
Non-synonymous change of the genes present in Cgl and Cef. A, genes present in Cgl and Cef; B, genes present in Cgl and Cef and having matches in the database. Genes show similar rates of evolution by only looking at the genes having matches elsewhere the genome database.
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2148-7-S1-S8-S10.pdf]

Additional File 11

Synonymous change of the genes present in Cgl and Cef. A, genes present in Cgl and Cef; B, genes present in Cgl and Cef and having matches in the database. Genes show similar rates of evolution by only looking at the genes having matches elsewhere the genome database.
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2148-7-S1-S8-S11.pdf]

Additional File 12

K_a/K_s *ratio of the genes present in Cgl and Cef. A, genes present in Cgl and Cef; B, genes present in Cgl and Cef and having matches in the database. Genes show similar rates of evolution by only looking at the genes having matches elsewhere the genome database.*
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2148-7-S1-S8-S12.pdf]

Additional File 13

An alternative topology of the concatenated genes. The insertion/deletion rates on external branches and internal branches were studied.
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2148-7-S1-S8-S13.pdf]

Additional File 14

The likelihood value shows a continuous pattern of change as the insertion/deletion rate changes in the reversible gain/loss model.
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2148-7-S1-S8-S14.pdf]

Additional File 15

The likelihood surface with different rates of insertion/deletion on external/internal branches in the reversible gain/loss model.
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2148-7-S1-S8-S15.pdf]

Additional File 16

Decreased GC content of recently acquired genes.
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2148-7-S1-S8-S16.pdf]

Acknowledgements

References

1. Cole ST, Eiglmeier K, Parkhill J, James KD, Thomson NR, Wheeler PR, Honore N, Garnier T, Churcher C, Harris D, Mungall K, Basham D, Brown D, Chillingworth T, Connor R, Davies RM, Devlin K, Duthoy S, Feltwell T, Fraser A, Hamlin N, Holroyd S, Hornsby T, Jagels K, Lacroix C, Maclean J, Moule S, Murphy L, Oliver K, Quail MA, Rajandream MA, Rutherford KM, Rutter S, Seeger K, Simon S, Simmonds M, Skelton J, Squares R, Squares S, Stevens K, Taylor K, Whitehead S, Woodward JR, Barrell BG: **Massive gene decay in leprosy bacillus.** *Nature* 2001, **409:**1007-1011.
2. Ogata H, Audic S, Renesto-Audiffren P, Fournier PE, Barbe V, Samson D, Roux V, Cossart P, Weissenbach J, Claverie JM, Raoult D: **Mechanisms of evolution in *Rickettsia conorii* and *R. prowazekii*.** *Science* 2001, **293:**2093-2098.
3. Foster J, Ganatra M, Kamal I, Ware J, Makarova K, Ivanova N, Bhattacharyya A, Kapatral V, Kumar S, Posfai J, Vincze T, Ingram J, Moran L, Lapidus A, Omelchenko M, Kyrpides N, Ghedin E, Wang S, Goltsman E, Joukov V, Ostrovskaya O, Tsukerman K, Mazur M, Comb D, Koonin E, Slatko B: **The Wolbachia genome of *Brugia malayi* : endosymbiont evolution within a human pathogenic nematode.** *PLoS Biol* 2005, **3:**.
4. Sokurenko EV, Chesnokova V, Dykhuizen DE, Ofek I, Wu XR, Krogfelt KA, Struve C, Schembri MA, Hasty DL: **Pathogenic adaptation of *Escherichia coli* by natural variation of the FimH adhesin.** *Proc Natl Acad Sci USA* 1998, **95:**8922-8926.
5. Feldgarden M, Byrd N, Cohan FM: **Gradual evolution in bacteria: evidence from *Bacillus* systematics.** *Microbiology* 2003, **149:**3565-3573.
6. Lawrence JG: **Selfish operons and speciation by gene transfer.** *Trends Microbiol* 1997, **5:**355-359.
7. Lawrence JG, Ochman H: **Molecular archaeology of the *Escherichia coli* genome.** *Proc Natl Acad Sci USA* 1998, **95:**9413-9417.
8. Lawrence JG: **Gene transfer, speciation, and the evolution of bacterial genomes.** *Curr Opin Microbiol* 1999, **2:**519-523.
9. de Koning AP, Brinkman FS, Jones SJ, Keeling PJ: **Lateral gene transfer and metabolic adaptation in the human parasite *Trichomonas vaginalis*.** *Mol Biol Evol* 2000, **17:**1769-1773.
10. Ochman H, Lawrence JG, Groisman EA: **Lateral gene transfer and the nature of bacterial innovation.** *Nature* 2000, **405:**299-304.
11. Lawrence JG, Ochman H: **Reconciling the many faces of lateral gene transfer.** *Trends Microbiol* 2002, **10:**1-4.
12. Springael D, Top EM: **Horizontal gene transfer and microbial adaptation to xenobiotics: new types of mobile genetic elements and lessons from ecological studies.** *Trends Microbiol* 2004, **12:**53-58.
13. Hughes AL, Friedman R: **Poxvirus genome evolution by gene gain and loss.** *Mol Phylogenet Evol* 2005, **35:**186-195.
14. Lan R, Reeves PR: **Gene transfer is a major factor in bacterial evolution.** *Mol Biol Evol* 1996, **13:**47-55.
15. Gogarten JP, Doolittle WF, Lawrence JG: **Prokaryotic evolution in light of gene transfer.** *Mol Biol Evol* 2002, **19:**2226-2238.
16. Jain R, Rivera MC, Moore JE, Lake JA: **Horizontal gene transfer accelerates genome innovation and evolution.** *Mol Biol Evol* 2003, **20:**1598-1602.
17. Kunin V, Ouzounis CA: **The balance of driving forces during genome evolution in prokaryotes.** *Genome Res* 2003, **13:**1589-1594.
18. Mirkin BG, Fenner TI, Galperin MY, Koonin EV: **Algorithms for computing parsimonious evolutionary scenarios for genome evolution, the last universal common ancestor and dominance of horizontal gene transfer in the evolution of prokaryotes.** *BMC Evol Biol* 2003, **3:**2.
19. Snel B, Huynen MA, Dutilh BE: **Genome trees and the nature of genome evolution.** *Annu Rev Microbiol* 2005, **59:**191-209.
20. Koonin EV, Makarova KS, Aravind L: **Horizontal gene transfer in prokaryotes: Quantification and classification.** *Annu Rev Microbiol* 2001, **55:**709-742.
21. Lawrence JG, Hendrickson H: **Genome evolution in bacteria: order beneath chaos.** *Curr Opin Microbiol* 2005, **8:**572-578.
22. Gressmann H, Linz B, Ghai R, Pleissner KP, Schlapbach R, Yamaoka Y, Kraft C, Suerbaum S, Meyer TF, Achtman M: **Gain and Loss of Multiple Genes During the Evolution of *Helicobacter pylori*.** *PLoS Genet* 2005, **1:**.
23. Tettelin H, Masignani V, Cieslewicz MJ, Donati C, Medini D, Ward NL, Angiuoli SV, Crabtree J, Jones AL, Durkin AS, Deboy RT, Davidsen TM, Mora M, Scarselli M, Margarit y Ros I, Peterson JD, Hauser CR, Sundaram JP, Nelson WC, Madupu R, Brinkac LM, Dodson RJ, Rosovitz MJ, Sullivan SA, Daugherty SC, Haft DH, Selengut J, Gwinn ML, Zhou L, Zafar N, Khouri H, Radune D, Dimitrov G, Watkins K, O'Connor KJ, Smith S, Utterback TR, White O, Rubens CE, Grandi G, Madoff LC, Kasper DL, Telford JL, Wessels MR, Rappuoli R, Fraser

CM: **Genome analysis of multiple pathogenic isolates of** *Streptococcus agalactiae* **: implications for the microbial "pan-genome".** *Proc Natl Acad Sci USA* 2005, **102:**13950-13955.

24. Wren BW: **Microbial genome analysis: insights into virulence, host adaptation and evolution.** *Nat Rev Genet* 2000, **1:**30-39.

25. Hacker J, Kaper JB: **Pathogenicity islands and the evolution of microbes.** *Annu Rev Microbiol* 2000, **54:**641-679.

26. Alcaine SD, Sukhnanand SS, Warnick LD, Su WL, McGann P, McDonough P, Wiedmann M: **Ceftiofur-resistant** *Salmonella* **strains isolated from dairy farms represent multiple widely distributed subtypes that evolved by independent horizontal gene transfer.** *Antimicrob Agents Chemother* 2005, **49:**4061-4067.

27. Pal C, Papp B, Lercher MJ: **Adaptive evolution of bacterial metabolic networks by horizontal gene transfer.** *Nat Genet* 2005, **37:**1372-1375.

28. Sullivan JT, Ronson CW: **Evolution of rhizobia by acquisition of a 500-kb symbiosis island that integrates into a phe-tRNA gene.** *Proc Natl Acad Sci USA* 1998, **95:**5145-5149.

29. Baar C, Eppinger M, Raddatz G, Simon J, Lanz C, Klimmek O, Nandakumar R, Gross R, Rosinus A, Keller H, Jagtap P, Linke B, Meyer F, Lederer H, Schuster SC: **Complete genome sequence and analysis of** *Wolinella succinogenes.* *Proc Natl Acad Sci USA* 2003, **100:**11690-11695.

30. Campanaro S, Vezzi A, Vitulo N, Lauro FM, D'Angelo M, Simonato F, Cestaro A, Malacrida G, Bertoloni G, Valle G, Bartlett DH: **Laterally transferred elements and high pressure adaptation in** *Photobacterium profundum* **strains.** *BMC Genomics* 2005, **6:**122.

31. Methe BA, Nelson KE, Deming JW, Momen B, Melamud E, Zhang X, Moult J, Madupu R, Nelson WC, Dodson RJ, Brinkac LM, Daugherty SC, Durkin AS, DeBoy RT, Kolonay JF, Sullivan SA, Zhou L, Davidsen TM, Wu M, Huston AL, Lewis M, Weaver B, Weidman JF, Khouri H, Utterback TR, Feldblyum TV, Fraser CM: **The psychrophilic lifestyle as revealed by the genome sequence of** *Colwellia psychrerythraea* **34H through genomic and proteomic analyses.** *Proc Natl Acad Sci USA* 2005, **102:**10913-10918.

32. Dobrindt U, Hacker J: **Whole genome plasticity in pathogenic bacteria.** *Curr Opin Microbiol* 2001, **4:**550-557.

33. Hentschel U, Hacker J: **Pathogenicity islands: the tip of the iceberg.** *Microbes Infect* 2001, **3:**545-548.

34. Hacker J, Carniel E: **Ecological fitness, genomic islands and bacterial pathogenicity. A Darwinian view of the evolution of microbes.** *EMBO Rep* 2001, **2:**376-381.

35. Hacker J, Blum-Oehler G, Muhldorfer I, Tschape H: **Pathogenicity islands of virulent bacteria: structure, function and impact on microbial evolution.** *Mol Microbiol* 1997, **23:**1089-1097.

36. Poly F, Threadgill D, Stintzi A: **Identification of** *Campylobacter jejuni* **ATCC 43431-specific genes by whole microbial genome comparisons.** *J Bacteriol* 2004, **186:**4781-4795.

37. Lima WC, Van Sluys MA, Menck CF: **Non-gamma-proteobacteria gene islands contribute to the** *Xanthomonas* **genome.** *OMICS* 2005, **9:**160-172.

38. Gill SR, Fouts DE, Archer GL, Mongodin EF, Deboy RT, Ravel J, Paulsen IT, Kolonay JF, Brinkac L, Beanan M, Dodson RJ, Daugherty SC, Madupu R, Angiuoli SV, Durkin AS, Haft DH, Vamathevan J, Khouri H, Utterback T, Lee C, Dimitrov G, Jiang L, Qin H, Weidman J, Tran K, Kang K, Hance IR, Nelson KE, Fraser CM: **Insights on evolution of virulence and resistance from the complete genome analysis of an early methicillin-resistant** *Staphylococcus aureus* **strain and a biofilm-producing methicillin-resistant** *Staphylococcus epidermidis* **strain.** *J Bacteriol* 2005, **187:**2426-2438.

39. Swenson DL, Bukanov NO, Berg DE, Welch RA: **Two pathogenicity islands in uropathogenic** *Escherichia coli* **J96: cosmid cloning and sample sequencing.** *Infect Immun* 1996, **64:**3736-3743.

40. McDaniel TK, Jarvis KG, Donnenberg MS, Kaper JB: **A genetic locus of enterocyte effacement conserved among diverse enterobacterial pathogens.** *Proc Natl Acad Sci USA* 1995, **92:**1664-1668.

41. Blum G, Ott M, Lischewski A, Ritter A, Imrich H, Tschape H, Hacker J: **Excision of large DNA regions termed pathogenicity islands from tRNA-specific loci in the chromosome of an** *Escherichia coli* **wild-type pathogen.** *Infect Immun* 1994, **62:**606-614.

42. Groisman EA, Ochman H: **How** *Salmonella* **became a pathogen.** *Trends Microbiol* 1997, **5:**343-349.

43. Zhang R, Zhang CT: **Identification of genomic islands in the genome of Bacillus cereus by comparative analysis with** *Bacillus anthracis.* *Physiol Genomics* 2003, **16:**19-23.

44. Fetherston JD, Perry RD: **The pigmentation locus of** *Yersinia pestis* **KIM6+ is flanked by an insertion sequence and includes the structural genes for pesticin sensitivity and HMWP2.** *Mol Microbiol* 1994, **13:**697-708.

45. Censini S, Lange C, Xiang Z, Crabtree JE, Ghiara P, Borodovsky M, Rappuoli R, Covacci A: **cag, a pathogenicity island of** *Helicobacter pylori***, encodes type I-specific and disease-associated virulence factors.** *Proc Natl Acad Sci USA* 1996, **93:**14648-14653.

46. Cheetham BF, Tattersall DB, Bloomfield GA, Rood JI, Katz ME: **Identification of a gene encoding a bacteriophage-related integrase in a vap region of the** *Dichelobacter nodosus* **genome.** *Gene* 1995, **162:**53-58.

47. Billington SJ, Johnston JL, Rood JI: **Virulence regions and virulence factors of the ovine footrot pathogen,** *Dichelobacter nodosus.* *FEMS Microbiol Lett* 1996, **145:**147-156.

48. Kovach ME, Shaffer MD, Peterson KM: **A putative integrase gene defines the distal end of a large cluster of ToxR-regulated colonization genes in** *Vibrio cholerae.* *Microbiology* 1996, **142:**2165-2174.

49. Espinosa-Urgel M, Kolter R: *Escherichia coli* **genes expressed preferentially in an aquatic environment.** *Mol Microbiol* 1998, **28:**325-332.

50. Jahreis K, Bentler L, Bockmann J, Hans S, Meyer A, Siepelmeyer J, Lengeler JW: **Adaptation of sucrose metabolism in the** *Escherichia coli* **wild-type strain EC3132.** *J Bacteriol* 2002, **184:**5307-5316.

51. Diaz E: **Bacterial degradation of aromatic pollutants: a paradigm of metabolic versatility.** *Int Microbiol* 2004, **7:**173-180.

52. Top EM, Springael D: **The role of mobile genetic elements in bacterial adaptation to xenobiotic organic compounds.** *Curr Opin Biotechnol* 2003, **14:**262-269.

53. Ravatn R, Zehnder AJ, van der Meer JR: **Low-frequency horizontal transfer of an element containing the chlorocatechol degradation genes from Pseudomonas sp. strain B13 to** *Pseudomonas putida* **F1 and to indigenous bacteria in laboratory-scale activated-sludge microcosms.** *Appl Environ Microbiol* 1998, **64:**2126-2132.

54. Dejonghe W, Goris J, El Fantroussi S, Hofte M, De Vos P, Verstraete W, Top EM: **Effect of dissemination of 2,4-dichlorophenoxyacetic acid (2,4-D) degradation plasmids on 2,4-D degradation and on bacterial community structure in two different soil horizons.** *Appl Environ Microbiol* 2000, **66:**3297-3304.

55. de Lipthay JR, Barkay T, Sorensen SJ: **Enhanced degradation of phenoxyacetic acid in soil by horizontal transfer of the tfdA gene encoding a 2,4-dichlorophenoxyacetic acid dioxygenase.** *FEMS Microbiol Ecol* 2001, **35:**75-84.

56. Poelarends GJ, Kulakov LA, Larkin MJ, van Hylckama Vlieg JE, Janssen DB: **Roles of horizontal gene transfer and gene integration in evolution of 1,3-dichloropropene- and 1,2-dibromoethane-degradative pathways.** *J Bacteriol* 2000, **182:**2191-2199.

57. Tiirola MA, Wang H, Paulin L, Kulomaa MS: **Evidence for natural horizontal transfer of the pcpB gene in the evolution of polychlorophenol-degrading sphingomonads.** *Appl Environ Microbiol* 2002, **68:**4495-4501.

58. Mattimore V, Battista JR: **Radioresistance of** *Deinococcus radiodurans* **: functions necessary to survive ionizing radiation are also necessary to survive prolonged desiccation.** *J Bacteriol* 1996, **178:**633-637.

59. Kennedy SP, Ng WV, Salzberg SL, Hood L, DasSarma S: **Understanding the adaptation of** *Halobacterium* **species NRC-1 to its extreme environment through computational analysis of its genome sequence.** *Genome Res* 2001, **11:**1641-1650.

60. Futterer O, Angelov A, Liesegang H, Gottschalk G, Schleper C, Schepers B, Dock C, Antranikian G, Liebl W: **Genome sequence of** *Picrophilus torridus* **and its implications for life around pH 0.** *Proc Natl Acad Sci USA* 2004, **101:**9091-9096.

61. Gogarten JP, Townsend JP: **Horizontal gene transfer, genome innovation and evolution.** *Nat Rev Microbiol* 2005, **3:**679-687.

62. Liu Y, Harrison PM, Kunin V, Gerstein M: **Comprehensive analysis of pseudogenes in prokaryotes: widespread gene decay and failure of putative horizontally transferred genes.** *Genome Biol* 2004, **5:**R64.

63. Taoka M, Yamauchi Y, Shinkawa T, Kaji H, Motohashi W, Nakayama H, Takahashi N, Isobe T: **Only a small subset of the horizontally transferred chromosomal genes in *Escherichia coli* are translated into proteins.** *Mol Cell Proteomics* 2004, **3**:780-787.

64. Daubin V, Lerat E, Perriere G: **The source of laterally transferred genes in bacterial genomes.** *Genome Biol* 2003, **4**:R57.

65. Daubin V, Moran NA, Ochman H: **Phylogenetics and the cohesion of bacterial genomes.** *Science* 2003, **301**:829-832.

66. Hao W, Golding GB: **Patterns of bacterial gene movement.** *Mol Biol Evol* 2004, **21**:1294-1307.

67. Neyman J: **Molecular studies of evolution: A source of novel statistical problems.** In *Statistical Decision Theory and Related Topics* Edited by: *Gupta SS, Yackel J. New York, USA: Academic Press*; 1971:1-27.

68. Gu X: **Maximum-likelihood approach for gene family evolution under functional divergence.** *Mol Biol Evol* 2001, **18**:453-464.

69. Felsenstein J: *Inferring phylogenies Sunderland, Mass.: Sinauer Associates, Inc*; 2004.

70. Hao W, Golding GB: **The fate of laterally transferred genes: Life in the fast lane to adaptation or death.** *Genome Res* 2006, **16**:636-643.

71. Marri PR, Hao W, Golding GB: **Gene gain and gene loss in *Streptococcus* : Is it driven by habitat?** *Mol Biol Evol* 2006, **23**:2379-2391.

72. Daubin V, Ochman H: **Bacterial genomes as new gene homes: the genealogy of ORFans in E. coli.** *Genome Res* 2004, **14**:1036-1042.

73. **Supplementary Information** [http://evol.mcmaster.ca/Coryne bacterium.html]

74. Cerdeno-Tarraga AM, Efstratiou A, Dover LG, Holden MT, Pallen M, Bentley SD, Besra GS, Churcher C, James KD, De Zoysa A, Chillingworth T, Cronin A, Dowd L, Feltwell T, Hamlin N, Holroyd S, Jagels K, Moule S, Quail MA, Rabbinowitsch E, Rutherford KM, Thomson NR, Unwin L, Whitehead S, Barrell BG, Parkhill J: **The complete genome sequence and analysis of *Corynebacterium diphtheriae* NCTC13129.** *Nucleic Acids Res* 2003, **31**:6516-6523.

75. Nakagawa I, Kurokawa K, Yamashita A, Nakata M, Tomiyasu Y, Okahashi N, Kawabata S, Yamazaki K, Shiba T, Yasunaga T, Hayashi H, Hattori M, Hamada S: **Genome sequence of an M3 strain of *Streptococcus pyogenes* reveals a large-scale genomic rearrangement in invasive strains and new insights into phage evolution.** *Genome Res* 2003, **13**:1042-1055.

76. Navarre WW, Porwollik S, Wang Y, McClelland M, Rosen H, Libby SJ, Fang FC: **Selective Silencing of Foreign DNA with Low GC Content by the H-NS Protein in *Salmonella*.** *Science* 2006, **313**:236-238.

77. Qian Y, Lee JH, Holmes RK: **Identification of a DtxR-regulated operon that is essential for siderophore-dependent iron uptake in *Corynebacterium diphtheriae*.** *J Bacteriol* 2002, **184**:4846-4856.

78. Andrews SC, Robinson AK, Rodriguez-Quinones F: **Bacterial iron homeostasis.** *FEMS Microbiol Rev* 2003, **27**:215-237.

79. Pallen MJ, Lam AC, Antonio M, Dunbar K: **An embarrassment of sortases – a richness of substrates?** *Trends Microbiol* 2001, **9**:97-102.

80. Cotter PD, Hill C, Ross RP: **Bacterial lantibiotics: strategies to improve therapeutic potential.** *Curr Protein Pept Sci* 2005, **6**:61-75.

81. Storz G, Imlay JA: **Oxidative stress.** *Curr Opin Microbiol* 1999, **2**:188-194.

82. Esmay PA, Billington SJ, Link MA, Songer JG, Jost BH: **The *Arcanobacterium pyogenes* collagen-binding protein, CbpA, promotes adhesion to host cells.** *Infect Immun* 2003, **71**:4368-4374.

83. Reilly TJ, Baron GS, Nano FE, Kuhlenschmidt MS: **Characterization and sequencing of a respiratory burst-inhibiting acid phosphatase from *Francisella tularensis*.** *J Biol Chem* 1996, **271**:10973-10983.

84. Camara M, Boulnois GJ, Andrew PW, Mitchell TJ: **A neuraminidase from *Streptococcus pneumoniae* has the features of a surface protein.** *Infect Immun* 1994, **62**:3688-3695.

85. Nishio Y, Nakamura Y, Usuda Y, Sugimoto S, Matsui K, Kawarabayasi Y, Kikuchi H, Gojobori T, Ikeo K: **Evolutionary process of amino acid biosynthesis in *Corynebacterium* at the whole genome level.** *Mol Biol Evol* 2004, **21**:1683-1691.

86. Kalinowski J, Bathe B, Bartels D, Bischoff N, Bott M, Burkovski A, Dusch N, Eggeling L, Eikmanns BJ, Gaigalat L, Goesmann A, Hartmann M, Huthmacher K, Kramer R, Linke B, McHardy AC, Meyer F, Mockel B, Pfefferle W, Puhler A, Rey DA, Ruckert C, Rupp O, Sahm H, Wendisch VF, Wiegrabe I, Tauch A: **The complete *Corynebacterium glutamicum* ATCC 13032 genome sequence and its impact on the production of L-aspartate-derived amino acids and vitamins.** *J Biotechnol* 2003, **104**:5-25.

87. Siew N, Fischer D: **Analysis of singleton ORFans in fully sequenced microbial genomes.** *Proteins* 2003, **53**:241-251.

88. Siew N, Fischer D: **Structural biology sheds light on the puzzle of genomic ORFans.** *J Mol Biol* 2004, **342**:369-373.

89. Hsiao WW, Ung K, Aeschliman D, Bryan J, Finlay BB, Brinkman FS: **Evidence of a Large Novel Gene Pool Associated with Prokaryotic Genomic Islands.** *PLoS Genet* 2005, **1**:.

90. **National Center for Biotechnology Information** [http://www.ncbi.nlm.nih.gov/]

91. Tauch A, Kaiser O, Hain T, Goesmann A, Weisshaar B, Albersmeier A, Bekel T, Bischoff N, Brune I, Chakraborty T, Kalinowski J, Meyer F, Rupp O, Schneiker S, Viehoever P, Puhler A: **Complete genome sequence and analysis of the multiresistant nosocomial pathogen *Corynebacterium jeikeium* K411, a lipid-requiring bacterium of the human skin flora.** *J Bacteriol* 2005, **187**:4671-4682.

92. Ikeda M, Nakagawa S: **The *Corynebacterium glutamicum* genome: features and impacts on biotechnological processes.** *Appl Microbiol Biotechnol* 2003, **62**:99-109.

93. Nishio Y, Nakamura Y, Kawarabayasi Y, Usuda Y, Kimura E, Sugimoto S, Matsui K, Yamagishi A, Kikuchi H, Ikeo K, Gojobori T: **Comparative complete genome sequence analysis of the amino acid replacements responsible for the thermostability of *Corynebacterium efficiens*.** *Genome Res* 2003, **13**:1572-1579.

94. Garnier T, Eiglmeier K, Camus JC, Medina N, Mansoor H, Pryor M, Duthoy S, Grondin S, Lacroix C, Monsempe C, Simon S, Harris B, Atkin R, Doggett J, Mayes R, Keating L, Wheeler PR, Parkhill J, Barrell BG, Cole ST, Gordon SV, Hewinson RG: **The complete genome sequence of *Mycobacterium bovis*.** *Proc Natl Acad Sci USA* 2003, **100**:7877-7882.

95. Huelsenbeck JP, Ronquist F: **MRBAYES: Bayesian inference of phylogenetic trees.** *Bioinformatics* 2001, **17**:754-755.

96. Altschul SF, Madden TL, Schffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ: **Gapped BLAST and PSI-BLAST: a new generation of protein database search programs.** *Nucleic Acids Res* 1997, **25**:3389-3402.

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

98. Saitou N, Nei M: **The neighbor-joining method: a new method for reconstructing phylogenetic trees.** *Mol Biol Evol* 1987, **4**:406-422.

99. Strimmer K, von Haeseler A: **Quartet puzzling: A quartet maximum-likelihood method for reconstructing tree topologies.** *Mol Biol Evol* 1996, **13**:964-969.

100. Yang Z: **PAML: a program package for phylogenetic analysis by maximum likelihood.** *Comput Appl Biosci* 1997, **13**:555-556.