Minireview

Genome interdependence in insect-bacterium symbioses

Evelyn Zientz*, Francisco J Silva† and Roy Gross*

Addresses: *Lehrstuhl für Mikrobiologie, Biozentrum, Universität Würzburg, Am Hubland, 97074 Würzburg, Germany. †Institut Cavanilles de Biodiversitat i Biologia Evolutiva, Universitat de València, Polígon La Coma, s/n. 46980 Paterna (Valencia), Spain.

Correspondence: Roy Gross. E-mail: roy@biozentrum.uni-wuerzburg.de

Abstract

Symbioses between unicellular and multicellular organisms have contributed significantly to the evolution of life on Earth. As exemplified by several studies of bacterium-insect symbioses, modern genomic techniques are providing exciting new information about the molecular basis and the biological roles of these complex relationships, revealing for instance that symbionts have lost many genes for functions that are provided by the host, but that they can provide amino acids that the host cannot synthesize.

Primary and secondary symbionts of insects

Many insects harbor intracellular bacteria in various tissues [1-3]. So-called primary symbionts reside in specialized 'host' cells called bacteriocytes; examples include the symbiosis of *Buchnera aphidicola* with aphids, of *Wigglesworthia* species with tsetse flies, of *Carsonella ruddii* with psyllids (small insects of the order Homoptera that suck sap from plant leaves), and of *Blochmannia* species with carpenter ants [4-7]. Bacteriocytes may form organ-like structures and are frequently associated with the midgut of the host animals. In general, primary insect-bacterium symbiosis appears to be obligate and mutually beneficial for the two partners [2,3]. In fact, it has not so far been possible to cultivate bacteriocyte-derived primary symbionts *in vitro*. Transmission of these symbiotic bacteria occurs vertically: the eggs or young embryos are infected by microorganisms derived from the mother.

In general, bacteriocyte symbioses appear to have a nutritional basis, as the host species in many cases feed on very specialized diets such as plant sap (aphids and psyllids) or blood (tsetse flies), both of which are poor in certain nutrients, such as amino acids or vitamins, that are essential for the host animals. In fact, 'curing' the insects of their symbionts by treatment with antibiotics may have severe consequences for their longevity, fecundity and/or development

of the animals. An extreme case of intracellular symbiosis was recently reported in mealybugs. Their bacteriocytes contain β-proteobacterial endosymbionts that themselves harbor intracellular bacteria belonging to the γ group of Proteobacteria [8].

In contrast to primary symbionts, the additional association of animals with so-called secondary symbionts occurs more sporadically and is not obligatory [2]. Bacteria in these cases can be found intra- and inter-cellularly not only in the midgut but also in many other tissues. Moreover, secondary symbionts can be cultivated *in vitro* and their transmission may occur horizontally, from one host to another [9,10].

Insights from phylogenetic analyses

As revealed by sequencing of 16S ribosomal DNA (rDNA), the primary symbionts of aphids, psyllids, tsetse flies and ants belong to the γ group of the Proteobacteria and appear to be members of the Enterobacteriaceae, which also includes many facultative intracellular pathogens of mammals and humans, such as *Salmonella typhimurium* [4,11-13]. Phylogenetic trees of the symbiotic bacteria reveal a strict co-speciation of the bacteria and their host animals - in other words each insect species is associated with its own species of bacterium - suggesting that the bacteriocyte

symbioses of each group of insects derive from a single parental infection.

From fossil records, the *Buchnera*-aphid symbiosis is estimated to have an age of 150 to 250 million years [4]. Although *Buchnera* isolates contain plasmids, genetic exchange with other bacteria is extremely limited or even nonexistent [14]. This long-lasting genetic isolation has led to intriguing consequences. For example, the 16S rRNA of these bacteria contains numerous destabilizing substitutions, which are found particularly in regions of the RNA molecule that, in free-living microorganisms, form highly conserved and very stable stem-loop structures [12,15]. In general, the mutation rate appears to be significantly higher in symbiotic bacteria than in free-living bacteria, resulting in a higher rate of non-synonymous substitutions [16,17]. These mutations affect the amino-acid composition of the respective proteins and may contribute to the generation of pseudogenes and to changes in protein function, for example by deletion of domains from multidomain proteins. The DNA polymerase I of *Escherichia coli* (encoded by the *polA* gene contains 928 amino acids and, in addition to the DNA polymerase activity, also has 3′-5′ and 5′-3′ exonuclease activities. In *Buchnera*, the *polA* gene encodes a truncated protein of only 286 amino acids, retaining only the 5′-3′ exonuclease activity.

Interestingly, the genomes of primary symbionts have surprisingly low GC contents, below 30% [18]. This AT bias is especially obvious in spacer regions between coding sequences, but it is also present throughout protein-coding sequences, where it is noted particularly in third codon positions. Such unusual genomic features, which are not observed in closely related free-living bacteria, are likely to result from the accumulation of slightly deleterious mutations by genetic drift as a consequence of small effective population sizes and lack of recombination. It is not known, however, whether the effects of such unusually high AT contents and the high substitution rates are indeed detrimental to the fitness of endosymbiotic bacteria, although this would be expected for their free-living relatives. An extreme case of degenerative minimalism was recently reported for the psyllid endosymbiont *Carsonella*, which has an exceptionally low GC content (19.9%) and shows almost complete absence of intergenic spaces [6].

The secondary symbionts characterized in detail so far also belong to the Enterobacteriaceae [2,9,10]. In general, however, they do not show a strict co-speciation with their host animals, indicating frequent horizontal exchange or recent independent acquisition of these bacteria. Moreover, the evolutionary consequences of their association with insect hosts appear not to be as dramatic as for the primary symbionts. Also, although secondary symbionts show the same tendency to accumulate non-synonymous substitutions in protein-coding genes and have genomic DNA with a strong AT bias, these phenomena are much milder than in the primary symbionts, indicating that secondary symbionts have not reached such an advanced stage of accommodation with their host animals.

Rationalization of symbiotic genomes

Pulsed-field gel electrophoresis of the chromosomal DNA of primary symbionts reveals surprisingly small genome sizes, for example between 630 and 650 kilobases (kb) in the case of *Buchnera* isolates [19], between 705 and 770 kb in the genus *Wigglesworthia* [20] and around 800 kb in the genus *Blochmannia* (our unpublished results). These genome sizes are similar to the smallest genomes known so far, which were described for obligate parasitic *Mycoplasma* species [21]. Recently, genome-size reduction was recognized to be a general phenomenon observed in many bacteria living in obligate parasitic interactions with animals (including humans), and especially in obligate intracellular bacteria such as Chlamydiae and Rickettsiae [22-24]. The strict adaptation to a constant environment, such as is provided by the cytoplasm of a eukaryotic cell, may allow the bacteria to abolish many of the adaptive responses that are required by free-living bacteria, and may also allow them to reduce their anabolic capacity if they succeed in recruiting metabolic precursors from the host cell's metabolism (see below) [25]. A decrease in selection on the respective loci or, alternatively, increased levels of genetic drift, as observed for the vertically transmitted intracellular bacteria, may therefore favor the inactivation of such genes by accumulation of deleterious mutations followed by deletion [26,27].

The first results obtained with *Buchnera* isolates colonizing aphids from two phylogenetically distant subfamilies indicated relatively constant genome sizes (with a variability of less than 5%) [19], suggesting that genome-size reduction (rationalization) in *Buchnera* occurred very early and quickly after the stable establishment of its symbiosis with aphids. Recent results obtained with *Buchnera* isolates from aphids of three additional subfamilies have revealed, however, that the genome size of *Buchnera* is more variable and can even be less than that of the minimal genome reported for *Mycoplasma genitalium* ([21] and R. Gil, B. Sabater-Munoz, A. Latorre, F.J.S. and A. Moya, unpublished observations).

Clues from the *Buchnera* genome sequence

Compared with that of *E. coli*, the *Buchnera* genome has drastically fewer genes. In fact, the genome of *Buchnera* sp. strain APS comprises 600 genes on a circular chromosome and two small plasmids [28]. Virtually all *Buchnera* genes have highly related orthologs in *E. coli*, demonstrating their close phylogenetic relationship (Table 1). Most interestingly, *Buchnera* sp. strain APS has retained most of the biosynthetic machinery for amino acids that are essential for its host organism but has lost those for amino acids that are not essential for the host. Moreover, for some amino acids, such

Table 1

Gene repertoire of the *Buchnera* and *E. coli* genomes and their comparison with the results of the *W. pallidipes* gene scan based on an *E. coli* K12 gene array

Functional annotation (COGs)*	Genes in *W. pallidipes* (gene array)	Genes shared by *W. pallidipes* and *Buchnera*	Genes in *Buchnera* sp. APS (genomic sequence)	Genes in *E. coli* K12 (genomic sequence)
Information storage and processing				
Translation, ribosomal structure and biogenesis	69	62	117	166
Transcription	32	10	16	245
DNA replication, recombination and repair	24	6	40	209
Cellular processes				
Cell division and chromosome partitioning	8	6	9	28
Post-translational modification, protein turnover, chaperones	19	13	32	117
Cell envelope biogenesis, outer membrane	28	8	24	200
Cell motility and secretion	21	10	42	136
Inorganic ion transport and metabolism	28	4	16	172
Signal transduction mechanisms	18	1	5	140
Metabolism				
Energy production and conversion	54	11	44	266
Carbohydrate transport and metabolism	53	11	31	333
Amino acid transport and metabolism	85	21	55	342
Nucleotide transport and metabolism	20	7	29	88
Coenzyme metabolism	27	13	32	117
Lipid metabolism	14	5	13	84
Secondary metabolite biosynthesis, transport and catabolism	16	2	4	89
Poorly characterized				
General function prediction only	42	5	29	302
Function unknown	15	0	20	255
Not in COGs	72	0	6	1000

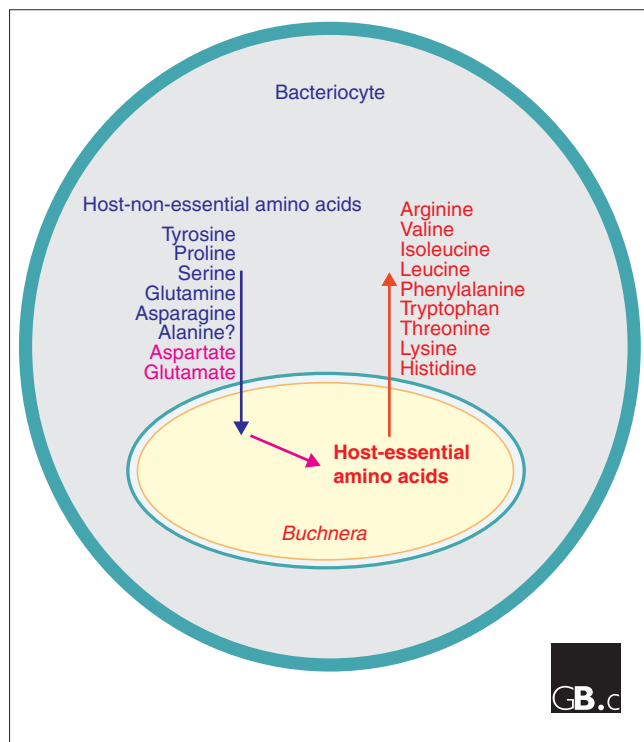*Functional annotations are from the Clusters of Orthologous Groups (COGs) database [34].

as glutamate and aspartate, which are non-essential for the host, the bacteria rely on their provision by the host, indicating that the biosynthetic pathways of both symbiotic partners are not only complementary but also mutually interdependent (Figure 1).

Buchnera has aerobic metabolism and contains all the genes encoding glycolytic enzymes, the pentose-phosphate cycle and aerobic respiration, but it lacks nearly all functions required for the tricarboxylic acid (TCA) cycle; *Buchnera* is apparently able to produce ATP, but it lacks all genes involved in fermentation and anaerobic respiration. There is a remarkable reduction in the repertoire of cell-membrane components and fatty acids and *Buchnera* is also not able to synthesize lipopolysaccharides or phospholipids and therefore has to import phospholipids or enzymes required for phospholipid biosynthesis from the host cell. Only very few transporter functions are present and, in accordance with the constant environmental conditions experienced within

the host cell, genes for regulatory systems or adaptive mechanisms are almost entirely missing [28].

Some of the evolutionary consequences mentioned above, such as the increased mutation rate, are related to the fact that *Buchnera* has very few proteins involved in DNA repair and recombination. Astonishingly, the *recA* gene, which is present in all other sequenced bacteria, is missing, although the *recBCD* genes are retained. Important repair and emergency systems, such as the *uvr* excision repair system and the SOS system, are incomplete, indicating that *Buchnera* is very susceptible to mutation. Eight pseudogenes have been identified in *Buchnera* sp. APS [28], but a larger number of pseudogenes was found in a *Buchnera* isolate from the aphid *Schizaphis graminum* [29].

The genetic make-up of *Buchnera* demonstrates that this organism is auxotrophic for several key compounds of the primary metabolism and is therefore entirely dependent on

Figure 1
A model for the mutual dependence of amino-acid biosynthetic pathways between *Buchnera* and its host cell. *Buchnera* is located in a vacuole of the bacteriocyte. According to its gene repertoire, it is able to synthesize amino acids that are essential for the host organism (red). On the other hand, *Buchnera* appears to require external supply of several amino acids that are not essential to the host (blue). Among these amino acids are aspartate and glutamate (magenta), which, together with other amino-acid precursors, seem to be imported from the host cytoplasm by the endosymbionts in order to enable biosynthesis of the respective host-essential amino acids.

its host. The genomic information on *Buchnera* also demonstrates, however, that this relationship is in fact mutually beneficial to both partners, because the symbiosis enables the host to occupy an ecological niche that is extremely poor in certain nutrients that can not be synthesized by the animal itself. The genetic isolation of these bacteria precludes any return to an independent lifestyle and may indicate that they face increasing degeneration problems (a phenomenon known as Muller's ratchet), which may cause a progressive loss of their identity as independent organisms and might finally convert them into new types of cellular organelles [30].

Parasitic and symbiotic microorganisms
As already pointed out, genome rationalization is observed in several symbiotic and parasitic microorganisms, especially in bacteria living in obligate association with a host.

The genome sequences of several pathogenic microorganisms are available, including those of various *Mycoplasma*, *Rickettsia* and *Chlamydia* species. Comparative analysis of these genomes has revealed interesting differences particularly with regard to metabolism [31]. For example, whereas the *Buchnera* genome has retained a significant number of biosynthetic functions beneficial for the host, in particular those for host-essential amino acids and cofactors, the parasitic microorganisms have greatly reduced their repertoire of such genes, largely relying on supply by the host. To compensate for these metabolic insufficiencies, parasite genomes encode relatively high numbers of transport systems, such as transporters for specific amino acids. Some parasites, such as *Rickettsia*, have even gained access to the host cell's ATP pool by using specific transport systems that are very likely to be derived from the eukaryotic cell; these organisms may therefore be considered energy parasites, even though they are able to generate ATP themselves.

These data indicate that a parasitic lifestyle is generally associated with not only a significant loss of anabolic genes but also the retention of catabolic and transport functions. In contrast, symbiont genomes may have been selected for keeping specific anabolic functions that are of benefit for the host [28,29,31]. As exemplified by *Mycobacterium leprae*, however, there are exceptions to this 'rule'. The genome of the obligate intracellular human parasite *M. leprae* has a size of about 3.3 megabases (Mb), and contains an extraordinary number of pseudogenes (over 1,100). In contrast to the genomes of other obligate parasites, the degenerate *M. leprae* genome retains almost complete sets of anabolic pathways, including those involved in nucleoside and amino-acid biosynthesis, but it has lost many factors involved in catabolism and energy metabolism [32].

Pathogenic microorganisms are, in general, well equipped with a variety of different surface structures, which are frequently involved in the protection of the bacteria against host-defense mechanisms. In marked contrast, the *Buchnera* genome is astonishingly poor in genes involved in the biosynthesis of membrane components and surface structures. This indicates that *Buchnera* cells are quite fragile, and suggests that such symbiotic microorganisms may have reached an intracellular niche devoid of significant host-defense mechanisms [28].

Gene arrays - an alternative to sequencing?
The genome sequence of *Buchnera* sp. APS has demonstrated the enormous power of modern genomics to characterize the symbiotic associations of uncultured bacteria with their host organisms. The genomic sequences of other *Buchnera* isolates, of *Wigglesworthia* and of *Blochmannia* are currently being determined and will provide interesting insights into the evolution of the respective symbioses and into the different adaptation strategies of the bacteria to

the various host organisms. Meanwhile, attempts have been made to obtain information about the gene content of symbionts in the absence of complete genome sequence data, using commercially available *E. coli* K12 gene arrays. As mentioned above, many primary and secondary symbionts of insects are closely related to *E. coli* and may be considered to be variants of a common ancestor with *E. coli*, with extensive deletions. It therefore seems feasible to use such gene arrays for hybridization experiments with symbiont DNA preparations. So far, data are available for *Wigglesworthia pallidipes* and *Sodalis glossinidius*, the primary and secondary symbionts of tsetse flies, respectively [18,33]. The use of array data for the interpretation of the biology of these organisms must be taken with some caution, however, especially if the base composition of the tested bacteria is very different from *E. coli*, as it is in the case of the AT-rich *W. pallidipes* genome. Only about 85% of the gene content of the *W. pallidipes* genome could be detected with the *E. coli* array, and some essential genes such as several aminoacyl tRNA synthetases and ribosomal proteins that were expected to be present in the genome were not found [20]. One additional problem of array technology is that recently evolved pseudogenes would hybridize almost as strongly as functional genes, leading to incorrect interpretations of the capabilities of the organism. This problem is of particular importance in the case of microorganisms with genomes on the way towards rationalization, as these may carry an unusually high number of pseudogenes.

A rough comparison of the gene content of *Buchnera* sp. APS and *W. pallidipes* using their genomic sequence and the gene-array data, respectively, reveals a quite striking observation (Table 1). The number of genes shared by the two genomes is extremely low (around 200), even considering that this number may be somewhat underestimated because of several *Wigglesworthia* genes that, so far, could not be detected in the array experiments. This implies that, from the thousands of genes of the common ancestor, each endosymbiont has retained a surprisingly small minimum set of primary essential genes plus an additional set of 300-400 specific genes. These latter genes are candidates for genes that are required for the specific adaptation to different symbiotic lifestyles.

In conclusion, much can be learnt from genome sequences about the evolution and lifestyles of symbiotic and parasitic bacteria, and a surprising amount may be discovered even without a full genome sequence. What is emerging so far is that symbiosis is associated with loss of different sets of genes than in parasitism.

Acknowledgements

References

1. Buchner P: *Endosymbiosis of animals with plant microorganisms.* New York: Interscience Publishers Inc.; 1965.
2. Moran NA: **Bacterial menageries inside insects.** *Proc Natl Acad Sci USA* 2001, **98:**1338-1340.
3. Moran NA, Baumann P: **Bacterial endosymbionts in animals.** *Curr Opin Microbiol* 2000, **3:**270-275.
4. Baumann P, Baumann L, Lai CY, Rouhbakhsh D, Moran NA, Clark MA: **Genetics, physiology, and evolutionary relationships of the genus *Buchnera*: intracellular symbionts of aphids.** *Annu Rev Microbiol* 1995, **49:**55-94.
5. Aksoy S: **Tsetse - a haven for microorganisms.** *Parasitol Today* 2000, **16:**114-118.
6. Clark MA, Baumann L, Thao ML, Moran NA, Baumann P: **Degenerative minimalism in the genome of a psyllid endosymbiont.** *J Bacteriol* 2001, **183:**1853-1861.
7. Schröder D, Deppisch H, Obermayer M, Krohne G, Stackebrandt E, Hölldobler B, Goebel W, Gross R: **Intracellular endosymbiotic bacteria of *Camponotus* species (carpenter ants): Systematics, evolution and ultrastructural analysis.** *Mol Microbiol* 1996, **21:**479-489.
8. von Dohlen CD, Kohler S, Alsop ST, McManus WR: **Mealybug beta-proteobacterial endosymbionts contain gamma-proteobacterial symbionts.** *Nature* 2001, **412:**433-436.
9. Aksoy S, Chen X, Hypsa V: **Phylogeny and potential transmission routes of midgut-associated endosymbionts of tsetse (Diptera:Glossinidae).** *Insect Mol Biol* 1997, **6:**183-190.
10. Sandström JP, Russell JA, White JP, Moran NA: **Independent origins and horizontal transfer of bacterial symbionts of aphids.** *Mol Ecol* 2001, **10:**217-228.
11. Chen X, Li S, Aksoy S: **Concordant evolution of a symbiont with its host insect species: molecular phylogeny of genus *Glossina* and its bacteriome-associated endosymbiont, *Wigglesworthia glossinidia*.** *J Mol Evol* 1999, **48:**49-58.
12. Spaulding AW, von Dohlen CD: **Psyllid endosymbionts exhibit patterns of co-speciation with hosts and destabilizing mutations in ribosomal RNA.** *Insect Mol Biol* 2000, **10:**57-67.
13. Sauer C, Stackebrandt E, Gadau J, Hölldobler B, Gross R: **Systematic relationships and cospeciation of bacterial endosymbionts and their carpenter ant host species: proposal of the new taxon *Candidatus* Blochmannia gen. nov.** *Int J Syst Evol Microbiol* 2000, **50:**1877-1886.
14. van Ham RC, Gonzalez-Candelas F, Silva FJ, Sabater B, Moya A, Latorre A: **Postsymbiotic plasmid acquisition and evolution of the *repA*1-replicon in *Buchnera aphidicola*.** *Proc Natl Acad Sci USA* 2000, **97:**10855-10860.
15. Lambert JD, Moran NA: **Deleterious mutations destabilize ribosomal RNA in endosymbiotic bacteria.** *Proc Natl Acad Sci USA* 1998, **95:**4458-4462.
16. Ochman H, Elwyn S, Moran NA: **Calibrating bacterial evolution.** *Proc Natl Acad Sci USA* 1999, **96:**12638-12643.
17. Buades C, Michelena JM, Latorre A, Moya A: **Accelerated evolution in bacterial endosymbionts of aphids.** *Int Microbiol* 1999, **2:**11-14.
18. Clark MA, Moran NA, Baumann P: **Sequence evolution in bacterial endosymbionts having extreme base compositions.** *Mol Biol Evol* 1999, **16:**1586-1598.
19. Wernegreen JJ, Ochman H, Jones IB, Moran NA: **Decoupling of genome size and sequence divergence in a symbiotic bacterium.** *J Bacteriol* 2000, **182:**3867-3869.
20. Akman L, Aksoy S: **A novel application of gene arrays: *Escherichia coli* array provides insights into the biology of the obligate endosymbiont of tsetse flies.** *Proc Natl Acad Sci USA* 2001, **98:**7546-7551.
21. Herrmann R, Reiner B: ***Mycoplasma pneumoniae* and *Mycoplasma genitalium*: a comparison of two closely related bacterial species.** *Curr Opin Microbiol* 1998, **1:**572-579.
22. Andersson SG, Kurland CG: **Reductive evolution of resident genomes.** *Trends Microbiol* 1998, **6:**263-268.
23. Andersson JO, Andersson SG: **Insights into the evolutionary process of genome degradation.** *Curr Opin Genet Dev* 1999, **9:**664-671.
24. Moran NA, Mira A: **The process of genome shrinkage in the obligate symbiont, *Buchnera aphidicola*.** *Genome Biol* 2001, **2:**research0054.1-0054.12.
25. Goebel W, Gross R: **Intracellular survival strategies of mutualistic and parasitic prokaryotes.** *Trends Microbiol* 2001, **9:**267-273.

26. Silva FJ, Latorre A, Moya A: **Genome size reduction through multiple events of gene disintegration in *Buchnera* APS.** *Trends Genet* 2001, **17:**615-618.

27. Mira A, Ochman H, Moran NA: **Deletional bias and the evolution of bacterial genomes.** *Trends Genet* 2001, **17:**589-596.

28. Shigenobu S, Watanabe H, Hattori M, Sakaki Y, Ishikawa H: **Genome sequence of the endocellular bacterial symbiont of aphids *Buchnera* sp. APS.** *Nature* 2000, **407:**81-86.

29. Tamas I, Klasson LM, Sandström JP, Andersson SG: **Mutualists and parasites: how to paint yourself into a (metabolic) corner.** *FEBS Lett* 2001, **498:**135-139.

30. Andersson JO: **Is *Buchnera* a bacterium or an organelle?** *Curr Biol* 2000, **10:**R866-R868.

31. Moran NA, Wernegreen JJ: **Life style evolution in symbiotic bacteria: insights from genomics.** *Trends Ecol Evol* 2000, **15:**321-326.

32. Cole ST, Eiglmeier K, Parkhill J, James KD, Thomson NR, Wheeler PR, Honore N, Garnier T, Churcher C, Harris D, *et al.*: **Massive gene decay in the leprosy bacillus.** *Nature* 2001, **409:**1007-1011.

33. Akman L, Rio RVM, Beard CB, Aksoy S: **Genome size determination and coding capacity of *Sodalis glossinidius*, an enteric symbiont of tsetse flies, as revealed by hybridization to *Escherichia coli* gene arrays.** *J Bacteriol* 2001, **183:**4517-4525.

34. **COGs: phylogenetic classification of proteins encoded in complete genomes** [http://www.ncbi.nlm.nih.gov/COG]