Opinion TRENDS in Microbiology Vol.12 No.5 May 2004

Have archaeal genes contributed to bacterial virulence?

Uri Gophna[1], Robert L. Charlebois[2] and W. Ford Doolittle[1]

[1]Genome Atlantic and Department of Biochemistry and Molecular Biology, Dalhousie University, 5850 College Street, Halifax, Nova Scotia, B3H 1X5, Canada
[2]NeuroGadgets Inc., Ottawa, Ontario, K1G 4B5, Canada

Although archaea have not been described as pathogens, it has recently been suggested that genes transferred from archaea might specifically contribute to bacterial virulence. Here, we survey 73 genomes of bacterial pathogens for the presence of genes originating from archaea. We describe 43 cases in which acquisition of archaeal genes by bacterial pathogens can be demonstrated. Although no *bona fide* virulence factors are among these acquired genes, several of them probably affect pathogen–host interactions.

Archaea are as diverse as bacteria and as widely distributed around the world. It is therefore mysterious that there are no diseases of archaeal origin. Eckburg and colleagues [1] have recently reasoned that such disorders might have been systematically overlooked, whereas Faguy [2] has suggested a different context in which archaea could be relevant to disease: as donors through lateral gene transfer (LGT) of virulence-promoting genes to pathogenic bacteria. When the genome of *Escherichia coli* O157:H7 was compared with that of K12, half a dozen O157:H7-specific genes could be recognized as having been derived from archaea [2]. In one case, an indirect argument for a role in infectivity or virulence could be made. Faguy argued on a more general basis that, for truly novel innovations in pathogenesis, bacteria might have often dipped into the archaeal gene pool. The idea has an unconventional appeal and would support claims for the medical relevance in studying archaea. There are nevertheless two reasons to be skeptical of such a notion. First, the divergence between bacteria and archaea is believed to be ancient and reflected in significant structural and biochemical differences between them, differences that should make successful LGT difficult. In particular, archaeal proteins involved in transcription and the DNA signals they recognize resemble their eukaryotic homologs much more than their bacterial counterparts (for reviews see Refs [3,4]). Nevertheless, recent sequencing of archaeal and bacterial genomes has shown that inter-domain transfer is not uncommon [5,6] and in some extreme cases even rampant. Consequently, a large fraction of *Thermotoga maritima* genes appears to have archaeal origins [7,8], and the archaeon *Methanosarcina mazei* bears a multitude of bacterial-derived genes, comprising up to a third of its genome [9].

Corresponding author: Uri Gophna (ugophna@dal.ca).

The second reason for skepticism – that archaea as non-pathogens have no need for virulence genes – can be countered by the observation that archaeal commensals can be found in animals (methanogens in ruminants and in our own intestines, and crenarchaeal symbionts in sponges [10–12]), and that the distinction between virulence genes and genes supporting harmless or even beneficial interactions with a host is vague and extremely species-dependent (Box 1).

In this review, we expand upon the search that was initiated by Faguy [2]. A total of 73 genomes from bacterial pathogens of animals and plants were analyzed for the presence of genes that originated from archaea using 'competitive matching' analysis; this software can be found at the NeuroGadgets Inc. website (http://www.neurogadgets.com/bioinformatics.php) and has been developed specifically to facilitate the detection of LGT in protein sequences (Box 2). Suspected transfer events were examined to see if they met specific criteria for archaea-to-bacteria gene transfer (Box 3). A total of 43 events of gene acquisition from archaea by bacterial pathogens were detected by this procedure (Table 1); only 12 of these appear to have been previously described in the literature.

Genes acquired from archaea by pathogenic bacteria

Few if any of the genes detected in our analysis can at present be directly or unequivocally linked to virulence. The vast majority of genes acquired by LGT from archaea are involved in primary metabolic functions: nucleotide and sugar metabolism, transport of metabolites and other biosynthetic functions. Most of these genes encode enzymes that could, in principle, function outside their genomic context without difficulty. We also detected genes involved in DNA/RNA modification or repair, which is more surprising because recognition signals for such activities are often taxon-specific. Another unexpected finding was the identification of laterally derived proteins with putative regulatory or signaling functions. However, inter-domain transfer of signaling proteins was previously observed in the opposite direction – bacteria to archaea and eukarya [13].

Acquired genes of particular interest

An adenylate cyclase homolog in the Lyme disease spirochete *Borrelia burgdorferi,* CyaB, appears to have been acquired from an ancestor of the genus *Pyrococcus*. CyaB belongs to the CYTH (CyaB–thiamine triphosphatases)

 0966-842X/$ - see front matter doi:10.1016/j.tim.2004.03.002

Box 1. What constitutes a virulence factor?

Most host–bacteria interactions are benign and often beneficial for the host, but interactions with pathogenic bacteria can cause disease. The ability of pathogens to cause disease in the host is dependent upon the possession of specific mechanisms that are mediated by specific 'virulence factors', but this term is difficult to define accurately. Some virulence factors that are intuitively recognized include toxins, which are secreted proteins that alter the metabolism of host cells and damage tissues; adhesins that mediate binding to specific receptors on host cells and are required for colonization; and the secretion systems that are involved in their export. Other factors might not interact with specific host molecules, but contribute to virulence by allowing a pathogen to occupy an otherwise uninhabitable niche. These factors include systems that are used to scavenge iron from the host milieu and also the *Helicobacter pylori* urease, which neutralizes gastric acid by hydrolysis of urea to ammonia. For a gene to constitute a virulence factor it should mediate direct interaction with the host. It should also have a measurable impact on disease; specific enough so that bacteria with mutations in that gene will display an attenuated phenotype or that antibodies generated against the bacteria will be protective. Therefore, although in principle every surface molecule of a bacterium constitutes a potential virulence factor because it comes into contact with host proteins, only a fraction will affect virulence in a way that can be blocked specifically. If we try to restrict virulence factors to genes that are required for the ability to cause disease but not required for growth outside the host, we might exclude 'classical' virulence factors, such as adhesins, which besides mediating adherence of bacteria to host tissues often facilitate binding to solid surfaces and biofilm formation. This definition is also inadequate for those pathogens that cannot replicate outside their hosts. Whether something constitutes a virulence factor will depend on the species that contains it and other elements it uses during the infection process. Therefore, flagella might not be a virulence factor of non-pathogenic and some pathogenic *Escherichia coli* strains, instead they simply increase fitness as free-living bacteria, but they could contribute to the virulence of a subset of isolates. The presence of a single virulence factor rarely makes a bacterium virulent, and only a combination of virulence factors will determine its ability to cause disease.

domain-containing superfamily, which includes other prokaryotic adenylyl cyclases and mammalian thiamine triphosphatases [14]. Because (i) cyclic AMP signaling has not been identified in archaea, (ii) a cloned *Methanocaldococcus jannaschii* CyaB homolog shows no adenylyl cyclase activity *in vitro* [15], and (iii) *B. burgdorferi* lacks a cyclic AMP signaling system [16], we propose that this gene is unlikely to function as an adenylate cyclase. In addition, in many other cases the role of acquired genes remains mysterious because the receiving organism lacks other components of systems that they are usually part of. 3-Hydroxy-3-methylglutaryl coenzyme A reductase (HMG-CoA reductase or HMGR), which catalyzes the synthesis of mevalonate, is another example of this conundrum. This gene was previously identified in *Vibrio cholerae* as an acquisition from archaea [6]. *V. cholerae* does not have any other genes from the mevalonate isoprenoid biosynthesis pathway and the role of this enzyme remains elusive. Our survey indicated that this gene also has orthologs in both *Vibrio vulnificus* and *Vibrio parahaemolyticus* suggesting that an ancestral *Vibrio* first acquired it from an archaeon.

Box 2. Competitive matching: a tool for detecting lateral gene transfer (LGT)

Within the web-based genome analysis system NGIBWS (http://www.neurogadgets.com/bioinformatics.php) [31] is a method for 'competitive matching', where the open-reading frames (ORFs) of a genome are returned that match an ORF from a member of one group of genomes better than an ORF from any member of a second group of genomes. Parameters for the query include: (i) the BLASTP [32] expectation value threshold for a match, (ii) the minimum difference in normalized BLASTP scores [33] between hits to the two sets of target genomes, (iii) whether or not the ORF is required to be present in both sets of target genomes, and (iv) the choice itself of query genome and of the membership within the two sets of target genomes. In our case, we sought ORFs from bacterial pathogens that matched ORFs within the archaea better than they matched those within bacteria, with a BLASTP cutoff expectation value of e^{-5}, a minimum normalized BLASTP score difference of 0.1, and no requirement for ORFs to be matched in both sets.

Aminoacyl tRNA synthetases are an important part of the translational machinery and their evolutionary history often involves LGT. When BLAST hits within spirochetes were excluded from the analysis we could detect horizontal acquisition of phenylalanyl-tRNA synthetase by *B. burgdorferi* and *T. pallidum,* as identified in previous studies [16,17]. However, we failed to detect the previously reported acquisition of a class I lysyl-tRNA synthetase [18] by these organisms. Our competitive matching analysis stipulates that the normalized BLAST score of the closest archaeal homolog must be higher than that of the closest bacterial homolog by more than 0.1. In this case, the inclusion threshold was not met owing to a *Bacillus cereus* homolog (probably also a product of transfer from archaea), which had a score of 0.298 compared with the archaeal 0.365. This demonstrates that using conservative criteria in competitive matching inevitably results in missed transfer events (false negatives). Genes that are beneficial for recipients are more likely to be fixed in a population and transferred more than once to different bacterial phyla, making them more abundant in bacteria; this results in the original transfer being more difficult to detect (Figure 1).

GlmU is one of two virulence-associated proteins in *Chlamydiae* that appear to have been acquired by LGT from archaea. GlmU, also known as UDP-*N*-acetylglucosamine pyrophosphorylase, is an important cell wall biogenesis enzyme involved in acetylation of glucosamine 1-phosphate and uridylation of *N*-acetylglucosamine 1-phosphate to produce UDP-*N*-acetylglucosamine [19]. GlmU homologs that have been found in the sequenced genomes of *Chlamydiae* share with several archaeal (but not other bacterial) homologs an insertion of up to 17 amino acids that occurs in a conserved region of the protein. Because of this, Griffiths and Gupta [20] suggested that these genes were acquired from an archaeon by the ancestor of *Chlamydiae*, an event with potential consequences for host–pathogen interactions. Despite having the entire gene cluster that is responsible for peptidoglycan biosynthesis, and despite producing a cell wall that is sensitive to penicillin and other inhibitors of

Box 3. Criteria for determining horizontal acquisition of genes from archaea

It is well known that the best BLAST-hit is not necessarily the closest phylogenetic relative. However, it is not always possible to get phylogenetic tree support, and not all topologies are easy to interpret. (Figure I; see also Figures 1 and 2 in the main text). We considered an open-reading frame (ORF) to be laterally derived if all of its full-length homologs are archaeal or if its two best BLAST hits are proteins from two different archaeal genera (disregarding same-species or in some cases same-genus homologs) and if one of the following conditions was met: (i) there are no bacterial homologs in the database; (ii) the closest bacterial homolog has a BLAST bit score that is less than half of the second best archaeal score; (iii) the three best hits are archaeal and the best archaeal score is at least 25% higher than best bacterial score; or (iv) the phylogenetic tree reconstruction shows that the protein clusters with an archaeal clade. It is important to note that a bacterial gene clustering with an archaeal gene from a single archaeal genus does not constitute sufficient support, as this could as easily represent an archaeal acquisition of a bacterial gene.

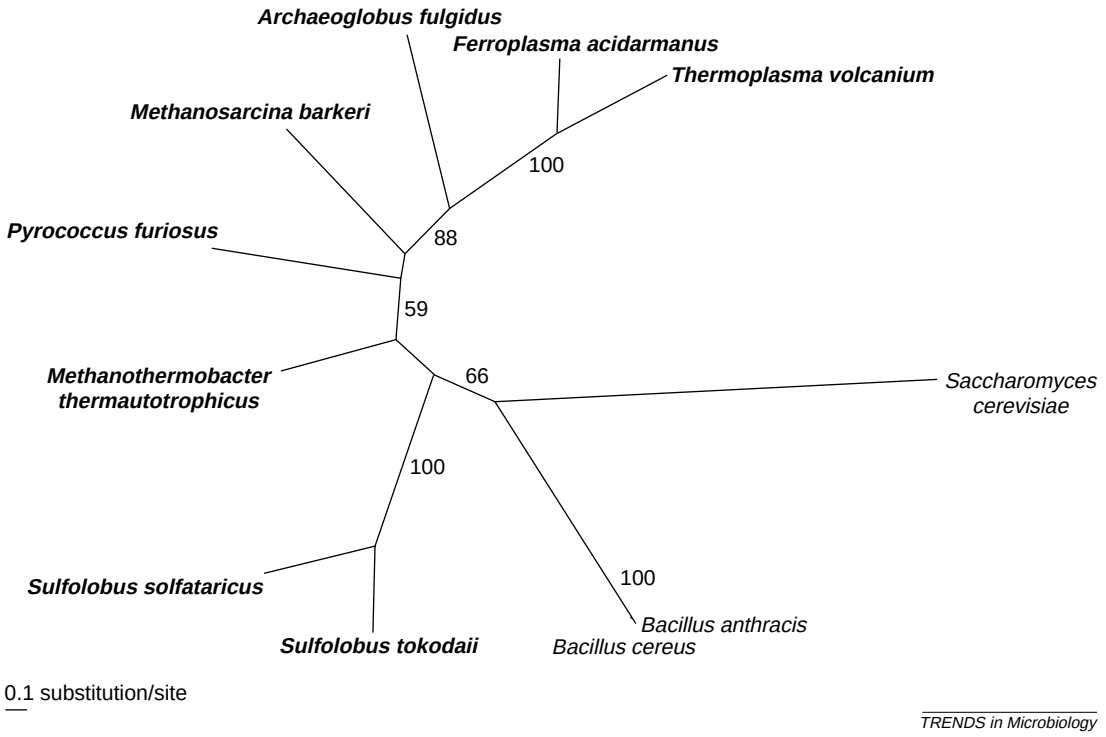

Figure I. The *Bacillus anthracis* and *Bacillus cereus* putative DNA/RNA methylase: in this topology there is no true bacterial cluster, and the bacterial protein does not cluster within any particular archaeal subgroup. The tree was reconstructed using the maximum likelihood method, using the JTT matrix with a gamma distributed among-sites rate variation model with eight categories. Archaeal sequences are in bold. Numbers indicate bootstrap support values (100 bootstrap trials) obtained with maximum likelihood distances.

peptidoglycan biosynthesis, chlamydia do not contain measurable amounts of this polymer [21]. The acquisition of a defective or incompatible GlmU from archaea that precludes the formation of a true peptidoglycan appears to be a plausible evolutionary event. Extensive cross-linking of outer-membrane proteins, which alters the bacterial surface, compensates for a lack of peptidoglycan. Therefore, the character of the chlamydial surface, with the obvious implications on host–pathogen interaction, might have been determined by a horizontal transfer event. Interestingly, we identified a second putative protein with two thymidylate synthase domains that appear to have also been acquired from archaea in the same cluster that encodes GlmU. This is in agreement with a study that shows the existence of thymidylate synthesis activity in *Chlamydiae* [22].

The acquisition of an archaeal pyruvoyl-dependent arginine decarboxylase (PvlArgDC) by chlamydia is the second clear lateral transfer event that might have contributed to their virulence. *Chlamydiae* lack other genes that are required for arginine catabolism, but have genes that could allow the uptake of arginine and export of agmatine. Therefore, a possible role for PvlArgDC might be in raising the pH of host cells and suppressing nitric oxide production and apoptosis [23].

We detected an open-reading frame (ORF) in the genomes of both *Bacillus anthracis* and *Bacillus cereus* that contained two conserved but functionally uncharacterized domains: COG1683 and COG3272 (http://www.ncbi.nlm.nih.gov/COG). A cluster of bacterial homologs to this ORF, which probably originated in archaea, includes an ORF from a *Salmonella* pathogenicity island (PAI) and orthologs from bacteria that are considered to be non-pathogens, such as *Shewanella oneidensis* or *Wolinella succinogenes,* therefore its role in virulence remains questionable.

A CAAX (cysteine–aliphatic–aliphatic–any amino acid) protease, also present in *B. anthracis* and *B. cereus* and identified as derived from archaea, might have a role in interacting with eukaryotic host processes, because

Table 1. Open-reading frames acquired by bacterial pathogens from archaea

Category	Pathogen	Putative function	Accession numbers	Refs
Virulence-associated				
	Bacillus anthracis, Bacillus cereus	Unknown: has homolog in *Salmonella* PAI[a]	gi|30257169	
	Bacillus anthracis, Bacillus cereus	CAAX amino-terminal protease	gi|30257417	
	Chlamydiae (6 genomes)	Pyruvoyl-dependent arginine decarboxylase	gi|15605097	[23]
	Chlamydiae (6 genomes)	GlmU: cell wall synthesis	gi|15605360	[20]
	Mycobacterium tuberculosis	CLC type chloride channel	gi|13879634	
	Mycobacterium tuberculosis	Mannose-6-phosphate isomerase domain protein	gi|13883428	
Sugar metabolism and transport				
	Mycobacterium tuberculosis	F420-dependent glucose-6-phosphate dehydrogenase	gi|13879923	[34]
	Mycobacterium tuberculosis	Methylenetetrahydromethanopterin reductase-related	gi|13882808	[34]
	Clostridium perfringens	Hexosyltransferase	gi|18309481	
	Salmonella enterica serovar Typhimurium	Thermophilic glucose-6-phosphate isomerase	gi|16421304	
	Chlamydiae (6 genomes)	Alternative thymidylate synthase	gi|7190038	
DNA/RNA modification and repair				
	Bacillus anthracis, Bacillus cereus	DNA/RNA methylase	gi|30254647	
	Clostridium tetani	DNA repair (RAMP superfamily)	gi|28203585	
	Leptospira interrogans	CTAG modification methylase	gi|24214246	
Nucleotide metabolism				
	Borrelia burgdorferi	Adenine deaminase	gi|11496787	
	Borrelia burgdorferi	Cytidylate kinase (cmk-2)	gi|15595164	[16]
	Clostridium perfringens	Adenylate kinase-related protein	gi|18310344	
	Mycobacterium tuberculosis	NrdA ribonucleoside-diphosphate reductase, alpha subunit	gi|13880109	
	Mycobacterium tuberculosis	Nucleic acid-binding protein	gi|13881537	
	Treponema pallidum	Cytidylate kinase (cmk-2)	gi|15639333	[16]
	Yersinia pestis	A/G phosphoribosyltransferases	gi|15981923	
Signaling and regulation				
	Bacillus anthracis, Bacillus cereus	Helix– turn–helix regulator	gi|30257933	
	Coxiella burnetii	Phosphohistidine phosphatase SixA	gi|29540721	
	Leptospira interrogans	Phosphoesterase	gi|24216641	
	Mycobacterium tuberculosis	Nucleic acid-binding protein	gi|13881537	
Aminoacyl tRNA synthetases				
	Borrelia burgdorferi	Phenylalanyl-tRNA synthetase, alpha subunit (PheS)	gi|15594858	[17]
	Borrelia burgdorferi	Phenylalanyl-tRNA synthetase, subunit beta (PheT)	gi|15594859	[17]
	Treponema pallidum	Phenylalanyl-tRNA synthetase, alpha subunit (PheS)	gi|15639957	[17]
	Treponema pallidum	Phenylalanyl-tRNA synthetase, subunit beta (PheT)	gi|15639009	[17]
Metabolite transport				
	Coxiella burnetii	Inorganic phosphate transporter	gi|29540975	
	Coxiella burnetii	Major facilitator family transporter	gi|29541316	
	Shigella flexneri, Escherichia coli	Amino acid or amine transport protein	gi|30040230	
ATPase domain				
	Coxiella burnetii	AAA Atpase	gi|29541163	
	Clostridium tetani	V-type ATP synthase subunit	gi|28204367	
Other biosynthetic				
	Clostridium tetani	Threonine synthase	gi|28202742	
	Ralstonia solanacearum	Precorrin-6y methyltransferase (cobalamin synthesis)	gi|17431094	
	Vibrio (*3* genomes)	Hydroxy-3-methylglutaryl CoA reductase	gi|28809269	[6]
Other functions				
	Borrelia burgdorferi	Adenylate cyclase 2: unknown	gi|15595068	[14]
	Clostridium perfringens	Transmembrane protein	gi|18310348	
	Clostridium perfringens	Unknown: conserved hypothetical	gi|18310076	
	Clostridium tetani	LolA, Outer membrane lipoprotein-sorting protein	gi|28202771	
	Vibrio (*3* genomes)	Unknown: contains DUF358 domain, possible metal-binding	gi|9658504	
	Vibrio parahaemolyticus	Putative pyrophosphatase	gi|28806826	

[a] *Salmonella* PAI, *Salmonella* pathogenicity island.

removal of a terminal CAAX motif is considered a eukaryotic mechanism that follows another post-translational eukaryotic modification (prenylation) and precedes methylation [24]. However, other bacterial proteins that share this motif have been identified in non-pathogens, such as cyanobacteria and *Caulobacter*.

A putative chloride channel that is common to *Mycobacterium tuberculosis* and *Mycobacterium bovis* clusters with homologs from acidophilic archaea (Figure 2). It has been demonstrated that chloride channels in bacteria might serve to increase resistance to extremely acidic conditions [25]. Some niches within a mammalian host are characterized by low pH, including phagosomes within macrophages, and consequently a chloride channel could be advantageous for a pathogen. However, because *M. tuberculosis*-containing phagosomes are not acidified [26],

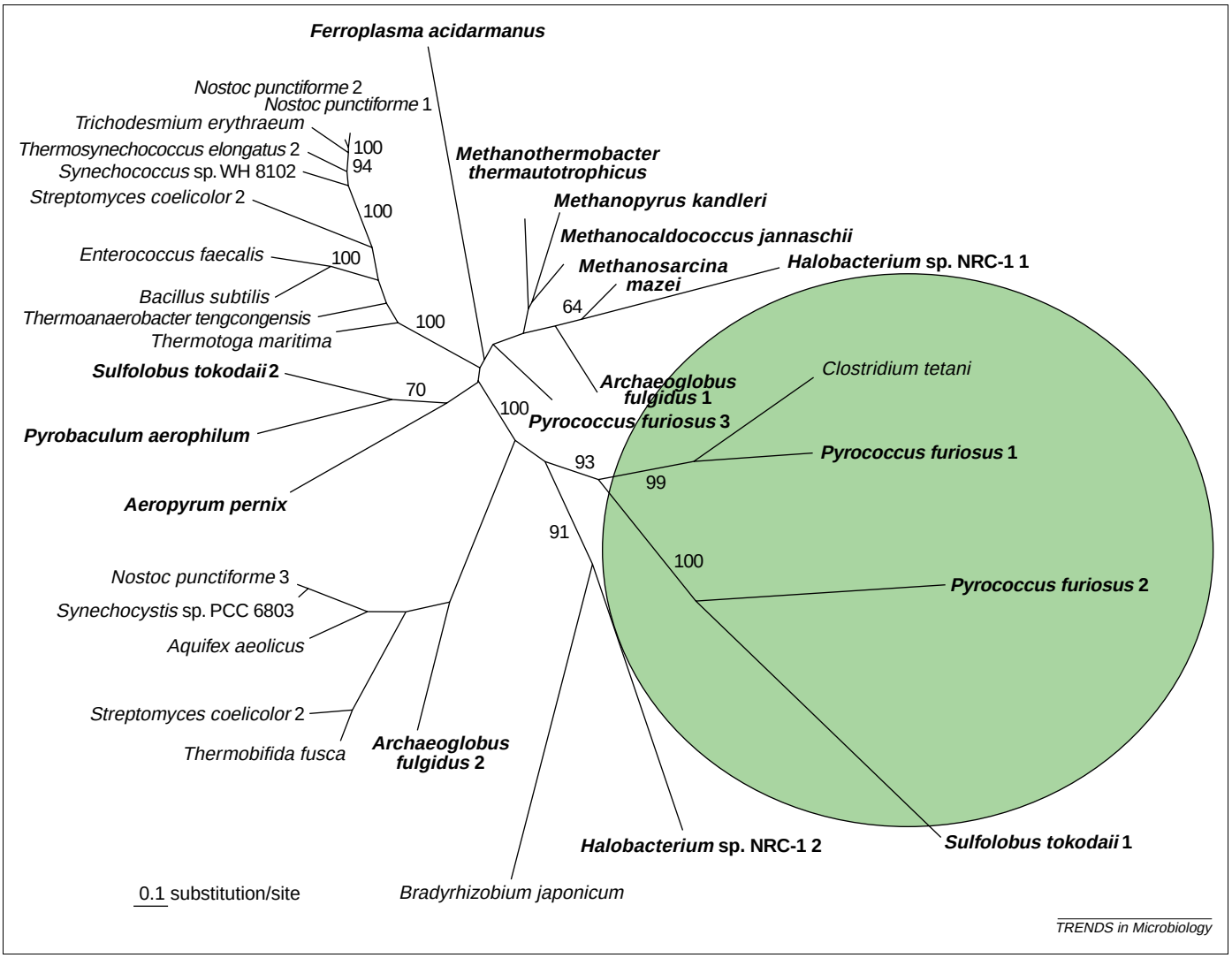

Figure 1. A phylogenetic tree of *Clostridium tetani* threonine synthase: a complex topology where not all archaeal proteins cluster in one group and more than one bacterial protein clusters with archaeal homologs. The tree was reconstructed using the maximum likelihood method, using the JTT matrix with a gamma distributed among-sites rate variation model with eight categories. Archaeal sequences are in bold. Green shading denotes clades of interest. A number following the taxon name is used when more than one homolog exists within the same genome. Numbers indicate bootstrap support values (100 bootstrap trials) obtained with maximum likelihood distances.

and phagosome–lysosome fusion is inhibited in *M. tuberculosis* infection [27], the role of this chloride channel in virulence remains questionable.

A mannose-isomerase domain protein of *M. tuberculosis* has also been acquired from archaea. Because mannose-containing glycolipids are an important component of mycobacterial cell walls [28] and are involved in host–pathogen recognition of *M. tuberculosis* [29], any possible effect exerted on the biogenesis of these glycolipids by this putative protein could affect virulence.

Virulence versus metabolic versatility

We surveyed complete genomes from bacterial pathogens of both animals and plants for genes acquired by LGT from the archaea. Applying conservative criteria, we detected 43 putative transfer events, most of which have not been previously identified. At least a broad putative function can be applied to the majority of ORFs detected on the basis of similarity results and conserved domain content. Most have putative functions in the primary metabolism of bacterial pathogens. Conversely, very few genes could be linked to bacterial virulence and 'classical' virulence factors (e.g. toxins or adhesins) could not be determined to be of archaeal origin. Our inability to detect archetypical virulence factors can be regarded as trivial as to date an archaeal pathogen has not been identified [1], and of the sequenced archaeal genomes only the genus *Methanosarcina* is associated with a host organism. Furthermore, failure to identify transferred virulence genes could be attributed to several limitations inherent to our dataset:

- Our search for lateral transfer events stipulates that the closest match (highest normalized BLAST hit score) will be an archaeon rather than a bacterial species. This is essential for automation purposes but doubtlessly excludes many laterally acquired genes, such as the class I lysyl-tRNA synthetase of *B. burgdorferi*. Genes acquired early by the ancestors of several pathogens, or more than once by separate bacterial lineages, will often not be detected.
- Complete bacterial genomes used in our study outnumber archaeal genomes 119 to 16, therefore, there is a

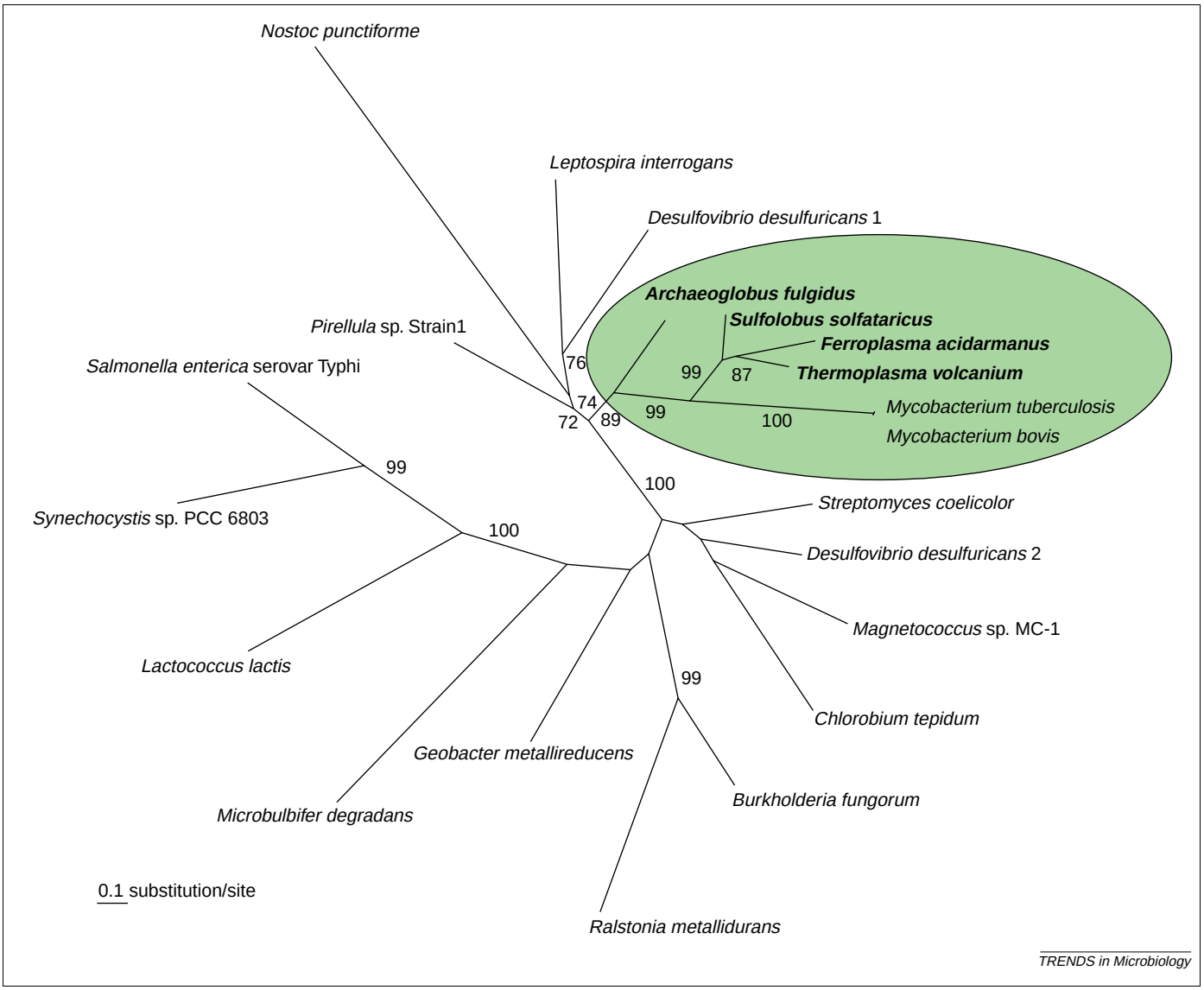

Figure 2. A phylogenetic tree of the *Mycobacterium tuberculosis* chloride channel: an easy to interpret topology where the bacterial protein sequence in question clusters among archaeal proteins, with strong statistical support. The tree was reconstructed using the maximum likelihood method, using the JTT matrix with a gamma distributed among-sites rate variation model with eight categories. Archaeal sequences are in bold. Green shading denotes clades of interest. A number following the taxon name is used when more than one homolog exists within the same genome. Numbers indicate bootstrap support values (100 bootstrap trials) obtained with maximum likelihood distances.

strong bias toward bacterial ORFs, increasing the chance of 'false negatives' occurring.

• Of the 16 archaeal genomes included in our study, 13 are thermophiles, whereas virtually all bacterial pathogens in this study are mesophiles. Proximity between viable mesophiles and thermophiles might be rare and the function of thermophilic proteins, especially enzymes, in mesophilic organisms is also less likely. It is also possible that further LGT events will be detected once more genomes of mesophilic archaea become available for analysis.

Nevertheless, the relative scarcity of virulence-associated genes acquired from archaea and the weak support provided for a role in virulence of such genes (detected by analyses that assessed their contribution to virulence) prevents us from supporting Faguy's hypothesis. Furthermore, even the virulence-associated genes of *E. coli* O157:H7 that had been designated previously as being acquired from archaea [2] no longer meet our criteria for archaea–bacteria LGT. They now have several bacterial non-pathogen relatives in the updated non-redundant protein databases.

By contrast, our results appear to highlight the role of archaea as a reservoir of metabolic innovation for bacteria, including those that are pathogens. Newly acquired metabolic capabilities might contribute not only to fitness outside hosts but also to within-host survival. The contribution could be direct, by creating a metabolite that somehow interferes with host processes in a manner beneficial for the pathogen or indirect, by allowing growth in an otherwise uninhabitable host niche, as suggested previously for the chloride channel in *M. tuberculosis* and the pyruvoyl-dependent arginine decarboxylase of *Chlamydiae*. As for acquisition of *bona fide* (or obvious) virulence factors through horizontal gene transfer, introduction of innovation to the pathogen gene pool could be more easily facilitated by genes procured from

the victims themselves (eukaryotes) as has been suggested before [30].

Acknowledgements

UG is supported by a Killam Postdoctoral Fellowship, WFD by the Canada Research Chair Program, and this research by Genome Atlantic and the Canadian Institutes for Health Research. We thank Yan Boucher for his expert help in phylogenetic analysis.

References

1 Eckburg, P.B. *et al.* (2003) Archaea and their potential role in human disease. *Infect. Immun.* 71, 591–596
2 Faguy, D.M. (2003) Lateral gene transfer (LGT) between Archaea and *Escherichia coli* is a contributor to the emergence of novel infectious disease. *BMC Infect. Dis.* 3, 13
3 Grabowski, B. and Kelman, Z. (2003) ARCHAEAL DNA REPLICATION: Eukaryal proteins in a bacterial context. *Annu. Rev. Microbiol.* 57, 487–516
4 Bell, S.D. *et al.* (2001) Basal and regulated transcription in Archaea. *Biochem. Soc. Trans.* 29, 392–395
5 Boucher, Y. *et al.* (2001) Bacterial origin for the isoprenoid biosynthesis enzyme HMG-CoA reductase of the archaeal orders Thermoplasmatales and Archaeoglobales. *Mol. Biol. Evol.* 18, 1378–1388
6 Boucher, Y. and Doolittle, W.F. (2000) The role of lateral gene transfer in the evolution of isoprenoid biosynthesis pathways. *Mol. Microbiol.* 37, 703–716
7 Nelson, K.E. *et al.* (1999) Evidence for lateral gene transfer between Archaea and bacteria from genome sequence of *Thermotoga maritima*. *Nature* 399, 323–329
8 Nesbø, C.L. *et al.* (2001) Phylogenetic analyses of two "archaeal" genes in *Thermotoga maritima* reveal multiple transfers between archaea and bacteria. *Mol. Biol. Evol.* 18, 362–375
9 Deppenmeier, U. *et al.* (2002) The genome of *Methanosarcina mazei*: evidence for lateral gene transfer between bacteria and archaea. *J. Mol. Microbiol. Biotechnol.* 4, 453–461
10 Tajima, K. *et al.* (2001) Phylogenetic analysis of archaeal 16S rRNA libraries from the rumen suggests the existence of a novel group of archaea not associated with known methanogens. *FEMS Microbiol. Lett.* 200, 67–72
11 Florin, T. (1993) Methanogenesis in the human large intestine. *Gut* 34, 573–574
12 Schleper, C. *et al.* (1998) Genomic analysis reveals chromosomal variation in natural populations of the uncultured psychrophilic archaeon *Cenarchaeum symbiosum*. *J. Bacteriol.* 180, 5003–5009
13 Koretke, K.K. *et al.* (2000) Evolution of two-component signal transduction. *Mol. Biol. Evol.* 17, 1956–1970
14 Iyer, L.M. and Aravind, L. (2002) The catalytic domains of thiamine triphosphatase and CyaB-like adenylyl cyclase define a novel superfamily of domains that bind organic phosphates. *BMC Genomics* 3, 33
15 Sismeiro, O. *et al.* (1998) *Aeromonas hydrophila* adenylyl cyclase 2: a new class of adenylyl cyclases with thermophilic properties and sequence similarities to proteins from hyperthermophilic archaebacteria. *J. Bacteriol.* 180, 3339–3344
16 Subramanian, G. *et al.* (2000) Comparative genome analysis of the pathogenic spirochetes *Borrelia burgdorferi* and *Treponema pallidum*. *Infect. Immun.* 68, 1633–1648
17 Wolf, Y.I. *et al.* (1999) Evolution of aminoacyl-tRNA synthetases – analysis of unique domain architectures and phylogenetic trees reveals a complex history of horizontal gene transfer events. *Genome Res.* 9, 689–710
18 Ibba, M. *et al.* (1997) Archaeal-type lysyl-tRNA synthetase in the Lyme disease spirochete *Borrelia burgdorferi*. *Proc. Natl. Acad. Sci. U. S. A.* 94, 14383–14388
19 Pompeo, F. *et al.* (2001) Dissection of the bifunctional *Escherichia coli* N-acetylglucosamine-1-phosphate uridyltransferase enzyme into autonomously functional domains and evidence that trimerization is absolutely required for glucosamine-1-phosphate acetyltransferase activity and cell growth. *J. Biol. Chem.* 276, 3833–3839
20 Griffiths, E. and Gupta, R.S. (2002) Protein signatures distinctive of chlamydial species: horizontal transfers of cell wall biosynthesis genes glmU from archaea to chlamydiae and murA between Chlamydiae and *Streptomyces*. *Microbiology* 148, 2541–2549
21 Chopra, I. *et al.* (1998) Antibiotics, peptidoglycan synthesis and genomics: the chlamydial anomaly revisited. *Microbiology* 144, 2673–2678
22 Fan, H.Z. *et al.* (1991) Biochemical evidence for the existence of thymidylate synthase in the obligate intracellular parasite *Chlamydia trachomatis*. *J. Bacteriol.* 173, 6670–6677
23 Graham, D.E. *et al.* (2002) *Methanococcus jannaschii* uses a pyruvoyl-dependent arginine decarboxylase in polyamine biosynthesis. *J. Biol. Chem.* 277, 23500–23507
24 Bracha, K. *et al.* (2002) The *Arabidopsis* AtSTE24 is a CAAX protease with broad substrate specificity. *J. Biol. Chem.* 277, 29856–29864
25 Iyer, R. *et al.* (2002) A biological role for prokaryotic ClC chloride channels. *Nature* 419, 715–718
26 Crowle, A.J. *et al.* (1991) Evidence that vesicles containing living, virulent *Mycobacterium tuberculosis* or *Mycobacterium avium* in cultured human macrophages are not acidic. *Infect. Immun.* 59, 1823–1831
27 Smith, I. (2003) *Mycobacterium tuberculosis* pathogenesis and molecular determinants of virulence. *Clin. Microbiol. Rev.* 16, 463–496
28 Patterson, J.H. *et al.* (2003) Mannose metabolism is required for mycobacterial growth. *Biochem. J.* 372, 77–86
29 Maeda, N. *et al.* (2003) The cell surface receptor DC-SIGN discriminates between *Mycobacterium* species through selective recognition of the mannose caps on lipoarabinomannan. *J. Biol. Chem.* 278, 5513–5516
30 Stebbins, C.E. and Galan, J.E. (2001) Structural mimicry in bacterial virulence. *Nature* 412, 701–705
31 Charlebois, R.L. *et al.* (2003) Characterization of species-specific genes using a flexible, web-based querying system. *FEMS Microbiol. Lett.* 225, 213–220
32 Altschul, S.F. *et al.* (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25, 3389–3402
33 Clarke, G.D.P. *et al.* (2002) Inferring genome trees by using a filter to eliminate phylogenetically discordant sequences and a distance matrix based on mean normalized BLASTP scores. *J. Bacteriol.* 184, 2072–2080
34 Heiss, G. *et al.* (2003) Homologous npdGI genes in 2,4-dinitrophenol- and 4-nitrophenol-degrading *Rhodococcus* spp. *Appl. Environ. Microbiol.* 69, 2748–2754