

Evidences of lateral gene transfer between archaea and pathogenic bacteria

Habib Bokhari^{1*}, Maryam Anwar¹, Hasan Bilal Mirza², Patrick Martin Gillevet³

¹Department of Biosciences, COMSATS Institute of Information Technology, Chakshahzad Campus, Park road Islamabad, Pakistan; ²Centre for Advanced Studies in Telecommunication (CAST), COMSATS Institute of Information Technology, Chakshahzad Campus, Park road Islamabad, Pakistan; ³Department of Environmental Sciences and Policy, George Mason University, Manassas, Virginia, USA; Habib Bokhari - Email: habib@comsats.edu.pk; *Corresponding author

Received June 09, 2011; Accepted June 27, 2011; Published July 06, 2011

Abstract:

Acquisition of new genetic material through horizontal gene transfer has been shown to be an important feature in the evolution of many pathogenic bacteria. Changes in the genetic repertoire, occurring through gene acquisition and deletion, are the major events underlying the emergence and evolution of bacterial pathogens. However, horizontal gene transfer across the domains i.e. archaea and bacteria is not so common. In this context, we explore events of horizontal gene transfer between archaea and bacteria. In order to determine whether the acquisition of archaeal genes by lateral gene transfer is an important feature in the evolutionary history of the pathogenic bacteria, we have developed a scheme of stepwise eliminations that identifies archaeal-like genes in various bacterial genomes. We report the presence of 9 genes of archaeal origin in the genomes of various bacteria, a subset of which is also unique to the pathogenic members and are not found in respective non-pathogenic counterparts. We believe that these genes, having been retained in the respective genomes through selective advantage, have key functions in the organism's biology and may play a role in pathogenesis.

Keywords: Phylogenetics, Lateral gene transfer, Pathogenesis

Background:

It is a common perception that only eukaryotes indulge in sex, whereas prokaryotes rely on vertical inheritance to meet new environmental challenges. In reality, prokaryotes are highly promiscuous, and the role of lateral gene transfer (LGT) as a driving force in prokaryotic evolution has been grossly underestimated [1]. The continuous ongoing process of evolution of bacteria is predominantly associated with the occurrence of events such as point mutations, genetic rearrangements, and horizontal gene transfer. A large number of human, animal and plant pathogens have evolved the capacity to produce unique set of virulence factors that are directly implicated in the process of establishing infection and disease. Additionally, many bacteria present in the environment express resistance traits against antibiotics. Both virulence factors and resistance determinants are subject to intra strain genetic and phenotypic variation. They are often encoded on unstable DNA regions. Thus, they can be readily transferred to bacteria of the same species or even to non-related prokaryotes. Archaea are widely distributed around the world but there are few documented diseases of archaeal origin. Difficulties in isolation, cultivation of archaea, and amenable genetic manipulation systems may contribute to a relative lack of knowledge. Different opinions exist to explain this possible anomaly and it has recently been agreed upon that pathogenic archaea might exist but have been systematically overlooked [2], whereas others [3] have suggested a different context in which archaea could be relevant to disease indirectly if not directly as donors through LGT of virulence-promoting genes to pathogenic bacteria. Attempt has been made in recent past partly by two different research groups [4, 5], both of which compiled literature data about archaea in possible association with human disease. Various toxin/antitoxin systems have been found in *Methanococcus jamastris*,

Archaeoglobus fulgidus, and *haloarchaea*. In addition, virulence genes for lipopolysaccharide biosynthesis and the *tadA* gene (e.g., required by *Actinobacillus actinomycetemcomitans* for nonspecific adherence) have been identified in archaea [3, 4]. In fact, methanogens have recently been linked to periodontal disease [5], a polymicrobial infection that affects the gums and supporting structures of the teeth and is characterized by periodontal pockets. Archaea, like bacteria, often have their genes arranged in the form of operons and hence are co-regulated and this arrangement would promote co-inheritance by LGT. The extent to which archaea are 'contaminated' by bacterial genes vary from species to species. LGT might account for some variation in archaeal genome sizes e.g. *Methanosarcina mazei* has an expanded genome of 4.10 Mb, 30% of which is bacterial in origin [6]. The comparison of complete genome sequences has already revealed that archaea are more than a sum of their (eukaryotic and bacterial) parts [7]. Archaeal genomes are believed to be a mosaic of molecular features, which are encoded by two different groups of genes: a lineage that codes for information processing which is eukaryotic in nature and a lineage that codes for operational (housekeeping) functions with a bacterial origin [8]. 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 [3]. For instance, it has been reported that a gene coding for a bifunctional catalase-peroxidase is likely a transfer from archaea to a variety of pathogenic bacteria, including *E. coli* O157:H7 [9]. Although not yet directly implicated in *E. coli* O157:H7 as a virulence factor, this enzyme has been implicated as a virulence factor in *Mycobacterium tuberculosis* [10] and in *Legionella pneumophila* [11]. Furthermore, this *E. coli* O157:H7 catalase-peroxidase has been associated with enterohaemorrhagic hemolysin in a variety of shiga-like toxin-producing

(verotoxin-producing) *E. coli* [9]. This correlation of the presence of catalase-peroxidase in many virulent but not in avirulent strains, suggests a direct role in the virulence of enterohaemorrhagic *E. coli*. In short there is no virulence factor known to date in archaea which directly implicates it as a pathogenic microorganism. Keeping in view the above mentioned fact, we have analyzed the available genome sequences of various pathogenic and nonpathogenic bacteria to determine if there are any potential archaeal genes that are present in pathogenic bacteria and absent in respective non-pathogenic bacteria. Thus, we hypothesize that pathogen genomes contain a number of archaeal-like genes of known or unknown functions that may contribute towards the pathogen's overall virulence.

Methodology:

In order to determine whether the acquisition of archaeal genes by lateral gene transfer is an important feature in the evolutionary history of pathogenic bacteria, we have developed a scheme of stepwise eliminations that identifies archaeal like genes in various bacterial genomes. Specifically, we compared BLASTP E-value/bit score for selected pathogenic bacterial hypothetical/putative/uncharacterized proteins and proteins conferring drug resistance against all subsets of GenBank as a preliminary screen for horizontal transfer candidates. Proteins from pathogenic bacteria that were absent or scored <95 bit score from corresponding non pathogen bacteria and scored >100 bit score when compared with archaea were selected as possible candidates for lateral gene transfer. In order to show that the traffic of LGT is from Archaea to Bacteria, those cases were selected where at least two archaeal genomes possess the suspected ORFs. In order to confirm horizontal transfer, representative protein sequences from the two kingdoms selected after steps 1-3 were aligned using CLUSTALX 1.83 and subjected to phylogenetic analyses. Two methods were used for the estimation of phylogeny: Neighbor joining package of PHYLIP and Bayesian estimation of phylogeny using the program MrBayes. All potential candidates for LGT from archaea which showed non-congruent phylogenies, i.e. grouping within the bacterial representatives with bootstrapping support of greater than 90% were classified as lateral transfers.

Data mining & Sequence alignment:

Initially the bacterial hypothetical or putative protein sequences (Table 1 see Supplementary material) were retrieved from the GOLD database and were used as queries to search for homologues against complete genome sequences of 461 bacterial and 39 archaeal species (20 May 2007 data) from the National Center for Biotechnology Information (NCBI) databases by BLASTP. In some instances archaeal hypothetical protein sequences were used to identify bacterial and archaeal homologues. All significant hits with >100 bit score were considered potential homologues, provided that archaeal and bacterial proteins were among the top best hits in Blast and there was no significant similarity to the respective non-pathogenic member and other members of eubacterial domain or cousins from same domain in case of archaea. Multiple sequence alignments of all the homologues were generated for sequence comparison and preliminary neighbor joining (NJ) as described below. We selected cases where pathogenic bacteria group more closely to archaea than their bacterial neighbors. To eliminate the possibility of such grouping by chance further screening was carried out where more than one archaeal types were present.

Phylogenetic analysis:

Sequence alignment was performed with the multiple sequence alignment software CLUSTAL X ver. 1.83. These alignments were subjected to two different phylogenetic approaches: neighbor joining and bayesian analysis. For neighbor joining, distance matrices were computed through ProtDist component of Phylip software suite. To confirm neighbor joining results the alignments were also subjected to Bayesian analysis using MrBayes. Trees from both analyses were drawn using TreeDyn.

Results:

Most investigations in the past, as reviewed above, have however, centered on transfer of known virulence genes and genes involved in establishing infections (antibiotic resistance, toxins, capsule, etc.) between close relatives who are both pathogens. However, many different functions – proteases, metabolic genes, oxygen protection genes, secretion systems, transporter genes, iron acquisition systems, antibiotic resistance can, under the right circumstances, contribute to enhance the virulence potential of pathogens. This is well summed up by Doolittle [1]: "lateral transfers have effectively changed the ecological and pathogenic character of bacterial species." A prerequisite to comprehending the pathogenicity mechanisms of an organism is the identification and examination of all its virulence genes. A total of 9 events

(Table 1) of gene acquisition from archaea by bacterial pathogens were detected by the above algorithm. All 9 candidates have homologues in pathogenic bacteria since they are either not present in corresponding non-pathogenic counterparts or their bit score is below the cutoff value.

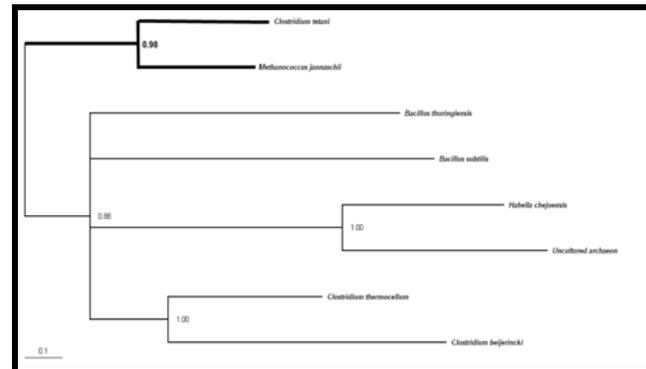


Figure 1: A phylogenetic tree of *Clostridium tetani*, spore coat polysaccharide biosynthesis protein spsG: a simple topology where the bacterial protein sequence groups strongly with archaeal sequence rather with other clostridium members or bacterial cousins. The tree was constructed using MrBayes.

Clostridium tetani spsG:

One unusual finding was the identification of a homologue of spore coat polysaccharide biosynthesis protein, spsG, from *Clostridium tetani* in *Methanocaldococcus jannaschii* (Figure 1). *Clostridium tetani* is a neuropathogenic Gram positive, spore forming bacterium predominantly found in the soil while *Methanocaldococcus jannaschii* is found in diverse environment such as flooded soils, human and animal gastrointestinal tracts, termites etc. In deep wound infections *Clostridium tetani* occasionally causes spastic paralysis in humans and animals commonly known as tetanus disease and is caused by the secretion of a potent neurotoxin called tetanus. Spores are produced by many species of bacilli and clostridia in response to severe external stress. These highly resilient dormant spores are able to withstand extremes of temperature, radiation, chemical assault, and even the vacuum of outer space. Upon the return of favorable environmental conditions, spores can readily convert to actively growing vegetative cells through a process known as germination. These abilities enable spores not only to survive in extreme conditions but, in some species, to cause significant disease. In contrast *C. acetobutylicum*, a nonpathogenic solvent producer does not harbor the gene spsG. Strong grouping of spsG with *Methanocaldococcus jannaschii* suggests that it has been acquired through LGT and plays a significant role in the establishment and maintenance of a pathogenic lifestyle of *C. tetani*. Furthermore there is a strong similarity of the spsG neighboring genes N-acetylneuraminase synthase and spore coat polysaccharide biosynthesis protein (spsF) of *C. tetani* and *M. jannaschii*. Recently, spsG-like sequences from an uncultured archaeon and CMP-N-acetylneuraminic acid synthetase of *Hahella chejuensis* have been shown to be very similar to each other suggesting the possibility of an LGT event. CMP-N-acetylneuraminic acid synthetase has also been reported to be present in milk microbial communities [12].

Listeria innocua protease:

The human gut is pre-occupied by a complex community of trillions of microorganisms representing all three known domains of life: Bacteria, Archaea, and Eukarya. *Listeria innocua* is widespread in the environment and in food. This species has to date never been described in association with human disease and is generally considered to be nonpathogenic and noninvasive sharing 84% of its genes with *L. monocytogenes*, an enteroinvasive human pathogen that can cross the intestinal, blood-brain and placental barriers [13]. Only recently, *Listeria innocua* has been found to be associated with severe bacteremia and the virulence factor in this particular case is protease. Thus, proteases may play important roles in parasite life cycles and host-parasite interactions. The presence of very similar proteases between *Methanocorpusculum labreanum* and *Listeria innocua* makes protease another possible candidate of LGT (supplementary figures 2a and 2b available with authors).

Clostridium tetani multi drug efflux pump:

Previous study has shown that resistance-nodulation division, (RND) and small multidrug resistance (SMR) pumps are not present in *M. jannaschii*, and these

systems have so far been identified only in eubacteria. To our knowledge this report suggests for the first time that such efflux pumps exist in archaea *Methanosarcina burtonii*, which group strongly with *C. tetani* (supplementary figures 3a and 3b available with authors). This means that efflux pumps have been encoded within bacterial genomes for hundreds of millions or even billion of years ago through an LGT event from archaea to bacteria. This also puts to rest the previous belief about efflux pumps arising in response to drug use. It is possible that this multidrug transporter in *C. tetani* and other virulent strains is used for the transport of antibiotics or toxic substances.

Flavobacterium virulence protein:

The strong similarity of Flavobacterium hypothetical protein, which possesses the features of virulence protein, with the one from *Methanosarcina mazei* suggests an LGT event (supplementary figures 4a and 4b available with authors). The Flavobacterium spp. has been studied in the past for multiple factors. Flavobacterium spp. are gram-negative non-fermenting aerobic bacilli that are mainly distributed in water and soil. They are not normally a component of human microflora and are rarely isolated from clinical specimens. It is a facultative human pathogen, capable of biodegrading nylon oligomers [13], and is found in contaminated soils [14]. As it happens with several microorganisms, Flavobacterium spp. can undergo adaptation. Some skin infections as well as serious, even lethal cases of bacteraemia have been reported to have caused by Flavobacterium often in association with catheter sepsis [15]. Furthermore due to the production of metallo-beta-lactamase, Flavobacterium spp. is resistant to carbapenem antibiotics. In addition capsule components have been shown to form part of the virulence mechanism of certain pathogenic Flavobacterium spp. and also play an important role in the adhesion and biofilm formation of certain bacterial species [16].

Shewanella putrefaciens restriction modification system:

Bacteria like *S. enterica* and *E. coli*, within the Enterobacteriaceae were initially described to possess type I R-M systems. However, this list gets bigger with the inclusion of functional type I systems in a wide variety of bacterial taxa, such as *Helicobacter pylori*, *Neisseria gonorrhoeae*, *Lactococcus lactis* and *Mycoplasma pulmonis* [17]. During this study we have seen the clustering of type I R-M system among bacteria such as *Shewanella putrefaciens* and archaea such as *Methanospirillum hungatei* (supplementary figures 5a and 5b available with authors). This clustering of *Shewanella* and other bacteria with archaea becomes more relevant to the scope of current study as recent study suggested a *Shewanella* species role as human pathogen [18]. The occurrence of RM systems that are shared between the two different domains suggests that R-M systems are readily acquired through LGT. Thus, it contradicts the fact that R-M systems would pose restriction barriers in gene flow. R-M systems are a remarkable characteristic of bacterial species and are probably involved in the adaptation of these bacteria to different environmental conditions as many R-M genes show repeats within their coding sequences, indicating that their expression is under the control of phase variation mechanisms.

Discussion:

Lateral gene transfer is the major mechanism for acquisition of new virulence genes in pathogens and archaea in this regard, by virtue of their diverse evolutionary history and environments which may provide a pool of potential virulence genes to bacterial pathogens. The new discipline of genomics stimulated interest in these exotic microorganisms, as biologists started finding their genes of interest in a new context. The genes described here, having been retained in the genome through selective advantage, most likely play a key role in the biology of pathogenic bacteria. The strong selection for some of the functions that these 9 transferred genes encode may provide clues regarding

virulence of the bacterium. However, study demonstrates the need for a systematic, comprehensive approach to the study of LGT based on first principles, i.e. rigorous inference and statistically based comparison of molecular phylogenetic trees. In addition with the availability of more archaeal genomes as well as non pathogenic bacteria, a tree-based approach will become both more challenging and more rewarding. The confirmation of archaea having clear association with disease is rather a difficult proposition as suggested by Gophna *et al.* [19] mainly due to the lack of understanding of their interaction with host or animal model systems in which to evaluate virulence. However, a simple approach like ours can indirectly implicate them as the predominant engine of variation in prokaryotes and the catalyst for the emergence of new bacterial pathogens. Further confirmation to our findings based on our hypothesis will come from experimental data where the role of such potential virulence factors can be validated and hence significance of LGT established.

Conclusion:

A few of the genes detected during this study have yet to be directly or indirectly linked to virulence. However it appears that there is diverse range of genes acquired by LGT from archaea e.g. metabolic functions, enzymes like proteases, antibiotic resistance genes, restriction modification system and hypothetical proteins whose role has yet to be elucidated.

Competing Interests:

The authors declare no competing interests.

Authors' Contributions:

All authors have contributed equally to this research.

Acknowledgements:

We thank the Higher Education Commission, Life sciences core group of Pakistan, COMSATS Institute of Information Technology and George Mason University for making collaborative interaction possible between the authors.

References:

- [1] Doolittle WF. *Science* 1999 **284**: 2124 [PMID: 10381871]
- [2] Vianna ME *et al.* *J Clin Microbiol.* 2006 **44**: 1274 [PMID: 16597851]
- [3] Faguy DM. *BMC Infect Dis.* 2003 **3**: 13 [PMID: 12816550]
- [4] Eckburg PB *et al.* *Infect Immun.* 2003 **71**: 591 [PMID: 12540534]
- [5] Cavicchioli R *et al.* *Bioessays.* 2003 **25**: 1119 [PMID: 14579252]
- [6] Koonin EV *et al.* *Annu Rev Microbiol.* 2001 **55**: 709 [PMID: 11544372]
- [7] Doolittle WF *et al.* *Philos Trans R Soc Lond B Biol Sci.* 2003 **358**: 39 [PMID: 12594917]
- [8] Makarova KS *et al.* *Genome Res.* 1999 **9**: 608 [PMID: 10413400]
- [9] Stephan R *et al.* *J Appl Microbiol.* 2000 **88**: 335 [PMID: 10736003]
- [10] Yu S *et al.* *Protein Sci.* 2002 **11**: 58 [PMID: 11742122]
- [11] Ciancotto NP. *Int J Med Microbiol.* 2001 **291**: 331 [PMID: 11727817]
- [12] Henriques AO & Moran CP Jr. *Methods.* 2000 **20**: 95 [PMID: 10610808]
- [13] Negoro S. *Appl Microbiol Biotechnol.* 2000 **54**: 461 [PMID: 11092619]
- [14] Ortiz-Hernandez L *et al.* *Int J Environ Poll.* 2001 **17**: 147
- [15] Hsueh PR *et al.* *Clin Infect Dis.* 1995 **21**: 1337 [PMID: 8589173]
- [16] Boliver R & Abramovits W. *J Infect Dis.* 1989 **159**: 150 [PMID: 2909636]
- [17] Binnewies TT *et al.* *Funct Integr Genomics.* 2006 **6**: 165 [PMID: 16773396]
- [18] Pagani L *et al.* *J Clin Microbiol.* 2003 **41**: 2240 [PMID: 12734291]
- [19] Gophna U *et al.* *Trends Microbiol.* 2004 **12**: 213 [PMID: 15120140]

Edited by P Kanguane

Citation: Bokhari *et al.* *Bioinformation* 6(8): 293-296 (2011)

License statement: This is an open-access article, which permits unrestricted use, distribution, and reproduction in any medium, for non-commercial purposes, provided the original author and source are credited.

Supplementary material:

Table 1: Shows the ORFs and their possible functions

Sr #	Category Pathogen	Putative Function	Accession numbers	Matches in Archaea	Bit score
a)	Virulence associated				
	Clostridium tetani	Spore coat polysaccharide biosynthesis protein spsG	gi 28203804	Methanocaldococcus jannaschii	180
	Hahella chejuensis	//	gi 83647929	Uncultured archaea	119
	Streptomyces avermitilis	beta-propeller repeat protein	gi 29832287	Cluster of Archaea	>200
	Pelodictyon phaeoclathratiforme	Protein of unknown function DUF1016	gi 68548684	Methanococcoides burtonii	505
	Listeria innocua	putative intracellular protease ThiJ/PfpI	gi 16799541	Methanocorpusculum labreanum	334
b)	Antibiotic resistance				
	Clostridium tetani	Na ⁺ driven multidrug efflux pump	gi 28210172	Methanococcoides burtonii	326
c)	Other Functions				
	Cluster of bacteria	hypothetical protein	gi 94995108 gi 42526112 gi 57240880	Methanoculleus marisnigri	424
	Clostridium beijerinckii	Protein of unknown function	gi 82748509	Methanococcus maripaludis	>200
	Flavobacterium spp.	hypothetical protein	gi 126661814	Methanosarcina mazei	384