Approximate Search on Protein Structures for Identification of Horizontal Gene Transfer in Bacteria

Swetha Billa[1], Mark A. Griep[2], Peter Z. Revesz[1,*]
[1]Department of Computer Science and Engineering and [2] Department of Chemistry
University of Nebraska-Lincoln
*To whom correspondence should be address: revesz@cse.unl.edu

Abstract

Horizontal Gene Transfer (HGT) is defined as the movement of genetic material from one strain of species to another. Bacteria, being an asexual organism were always believed to transfer genes vertically. But recent studies provide evidence that shows bacteria can also transfer genes horizontally.

HGT plays a major role in evolution and medicine. It is the major contributor in bacterial evolution, enabling species to acquire genes to adapt to the new environments. Bacteria are also believed to develop drug resistance to antibiotics through the phenomenon of HGT. Therefore further study of HGT and its implications is necessary to understand the effects of HGT in biology and to study techniques to enable or disable the process based on its effects.

Methods to detect HGT events have been studied extensively but no method can accurately detect all the transfers between the organisms. This paper presents *an HGT identification method based on approximate searches on bacterial protein structures***.** This method makes use of Z-score similarities between the protein structures and also uses functions of BLAST and DaliLite to work with protein sequence and structural similarities. In addition, Jmol, a java viewer tool is used for visual structural comparisons and sequence alignment. We also present experimental results regarding HGTs between the *Firmicutes* bacterium *Bacillus subtilis* and various *Proteobacteria* bacteria.

1. Introduction

Horizontal gene transfer (HGT) or *lateral gene transfer* is the passing of genetic material from one organism to another, other than by descent in which genetic information travels through the generations as the cell divides. In nature, gene transfer occurs between two same species or closely related species via typical routes of reproduction, such as cross pollination of plants and interbreeding of animals. Such transfer is also called *vertical gene transfer*, since traits are passed on from parent to the offspring vertically.

Sometimes genes also move between different species, such as bacteria and plants, through a process unrelated to reproduction that is known as *horizontal gene transfer* (HGT). HGT can also occur between two closely related species.

HGT has first been described in a Japanese publication in 1959, which describes about the transfer of antibiotic resistance from one bacterium to another (Akiba et al. 1960). The phenomenon of HGT is quite significant in prokaryotes and some unicellular eukaryotes. Importance of HGT in the evolution of multicellular organisms has not been extensively studied.

1.1 How to determine HGT?

For a successful natural horizontal gene transfer, it would require stable integration of the gene into the genome, no disturbance of regulatory or genetic structures, expression and successive production of a functional protein (Susanna et al. 2006). There are two approaches to determine Horizontal Gene Transfer in a genome, I) Phylogenetic Comparison and II) Parametric Comparison. In Phylogenetic Comparison, different organisms are compared to find the similarity or dissimilarity. While in Parametric Comparison, genes that appear to be anomalous in their current genome context are thought to have been transferred or introduced from a foreign source (Lawrence and Ochman 2002).

1.2 Why is it important to study HGT?

Copyright © 2011, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

HGT plays a major role in bacterial evolution. *Antibiotic resistance (AR)* or antimicrobial resistance is a type of drug resistance where a microorganism is able to survive exposure to an antibiotic. The development of antibiotic resistance characteristics is often observed to develop much more rapidly than simple vertical inheritance of traits. Hence it is believed that development of antibiotic resistance among different bacteria is the result of HGT, as one bacterial cell acquires resistance and transfers those genes to other bacterial species (Frank-Kamenetskii 1993).

Antibiotic resistance (AR) poses a significant problem for the public health in the world. As more and more bacterium develop resistance to drugs, the need for alternative treatments increases. Controlling of *antibiotic resistance (AR)* in bacteria requires investigation of the antibiotic resistance mechanism (Song et al. 2008). Hence studies on HGT will help provide a greater incite on how this can be curbed.

1.3 Mechanisms of HGT

Exchange of genetic material can occur in 3 different ways in bacteria: *Transformation, Conjugation* and *Transduction*.

Transformation: A process of alteration of the gene by introducing foreign genetic material. This is more common in bacteria than in eukaryotes. This is the most common method of HGT used in laboratories to insert genes into bacteria for experimental purposes. Only short DNA can be exchanged through this process.

Conjugation: A process in which a bacterial cell transfers genetic material to another cell through cell-cell contact. This can occur between distantly related bacteria or between a bacteria and eukaryotic cell. This process can transfer long fragments of DNA. The genes required for conjugation are usually found on a plasmid DNA.

Transduction: A process in which a DNA is moved from one bacterium to another by a bacterial virus. This bacterial virus is called a bacteriophage or simply phage. A phage inserts its DNS into a recipient and modifies its DNA. This method requires the donor and recipient to share the cell surface receptors. Hence it is usually seen in closely related bacteria. The length of the DNA transferred depends on the size of the phage head.

1.4 Overview of Existing Methods to Detect HGT

Compositional Methods

A gene which is horizontally transferred can contain recognizable signatures of its previous location since it comes from a different genomic background. Compositional methods use atypical nucleotide (Lawrence and Ochman 1997), atypical codon usage patterns (Lawrence and Ochman 1998) or their combination (Tsirigos and Rigoutsos 2005) to detect which genes in a genome have been horizontally gene transferred. Since over time the horizontally transferred genes adopt the signatures of the new genome, these methods can be used only on genes which have been transferred fairly recently. These methods are easily applicable to completely sequenced genomes. However, high rates of false positives and negatives have been observed in these methods.

Phylogeny-Based Method

Phylogeny-based detection of HGT is one of the most commonly used approaches for detecting HGT. It is based on the fact that HGT causes discrepancies in the gene tree as well as create conflict with the species phylogeny. So the methods that use this approach would compare the gene and species tress which would come up with a set of HGT events to explain the discrepancies among these trees.

When HGT occurs, the evolutionary history of the gene would not agree with the species phylogeny. The gene trees get reconstructed and their disagreements are used to estimate how many events of HGT could have occurred and the donors and recipients of the gene transfer.

Some of the issues when using this method for HGT detection are, determining if the discrepancy is actually a HGT and uniquely identifying the HGT scenario. The phylogenetic trees are only partially known and they are reconstructed using Phylogeny reconstruction techniques. The quality of this reconstruction which is usually done statistically has an impact on the HGT detection and sometimes could underestimate or overestimate the number HGT events.

Distance-Based Detection of HGT

The Distance-Based method incorporates distances typically used in the Phylogeny-based detection of HGT rather than the trees themselves. This method has many of the strengths of Phylogenetic approaches but avoids some of their drawbacks.

2. Methodology

We recently devised a HGT identification method based on approximate search on protein structures (Santosh et al. 2011). This method makes use of the fact that similar protein structures usually imply similar functionality. According to previous studies (Shortridge et al. 2011) during evolution the structure of the proteins remains

remarkably conserved in association with conserved functionality while DNA sequences are less conserved. Hence we can also expect a high level of conservation of the protein structures that are indirectly horizontally gene transferred from another organism. That is, while the transferred DNAs can change, they change in a way that leaves largely intact the protein structures that are built based on them.

The structure of the protein transferred may be different from proteins with similar functionality in the recipient organism. Hence to detect HGT, the goal would be identify anomalies in the structures of the proteins in an organism, with similar functionalities.

To identify these protein structure anomalies, we make use of the Cluster of Orthologous Group (COG) classification. According to this classification all proteins with similar functionality are categorized under the same COG number. Further, according to evolutionary theory they should have similar structures.

In this research, we extend the applicability of Santosh et al. (2011), which considered horizontal gene transfer between *E. coli* and other bacteria, to an efficient search for horizontal gene transfer between *Bacillus subtilis* and other bacteria. We consider two phyla of bacteria: (i) *Firmicutes,* and (ii) *Proteobacteria*. Most of the *Firmicutes* bacteria are gram positive. They are found in various environments and the group includes some notable pathogens. *Proteobacteria* is the largest and most diverse in the domain bacteria. This is an environmentally, geologically and evolutionarily important group. Most of the bacteria in *Proteobacteria* group are gram negative. *Firmicutes* and *Proteobacteria* diverged millions of years ago, and underwent random mutations during which they retained most of their native characteristics (Shortridge et al. 2011). Evidence of protein characteristics of bacteria belonging to one phyla being similar to the protein characteristics of bacteria in another phyla would indicate horizontal gene transfer.

2.1 The Method

We compared the *Firmicutes* bacterium *Bacillus subtilis* with *Proteobacteria* bacteria. We chose *Bacillus subtilis* because it has a large number of identified structures in the biological databases that were available for our research.

Stage 1

As the first stage of the method, we needed information about all the proteins that were studied in each of these bacteria. To get this data we made use of the PROFESS database (Triplet at al. 2009). Querying the PROFESS database we get the list of proteins studied in each of the bacteria and the COGs to which they belong to. The COG number uniquely identifies groups of proteins that have functional similarity.

Stage2

As the second stage of the method we perform a structural comparison of the proteins. This again is a two-step process, as in we first structurally compare proteins in each of the COGs with in each organism and then we structurally compare proteins in each of the COGs among the two organisms. DaliLite program was used for the structural comparisons. The DaliLite program takes the input of two PDB ids and applies structural comparison algorithms and provides a result in the form of a Z- score which is the index for measuring structural similarity in proteins.

There are 494 proteins for *Bacillus subtilis* and 3264 proteins for *Escherichia coli* that are documented in the PDB database. When we perform structural comparison for these two bacteria we are interested only in the common COGs between them. There are 88 common COGs among them. To perform pairwise structural comparison of proteins within each organism within the same COG, we would have $n * \frac{(n-1)}{2}$ pairs of PDB IDs, where n is the number of proteins in a given COG for a given organism.

And for comparison of proteins within a COG number in the two different organisms under consideration, we would have the cross product of the number of PDB IDs in that particular COG in each of the organisms. This has to be repeated for all the common COGs in the two organisms.

For all the pairs of PDB IDs obtained above, an alignment algorithm is applied to get a Z-score measure for each pair. The DaliLite tool is used to obtain this. When a pair-wise comparison is done using DaliLite it gives results based on multiple variations in the alignments of the two proteins. We choose the result set with the highest Z-score. In other words we use the score from the best alignment. The average Z-score is calculated within each COG. These average Z-scores are then normalized. By analyzing these normalized values we can identify anomalous COG numbers.

Since the average Z-scores are calculated within the same COGs, we expect the average Z-score for the same COG in two different organisms to be equal or have very little difference. If any large difference in the values of the average Z-score with in a same COG appears in the two organisms under consideration then it is unusual and further inspection of the proteins in that particular COG is required. For our research the threshold value for identifying this anomalous behavior is chosen to be 75%. If the average Z-score value of the first organism is less than or equal to 75% of the average Z-score value of the second organism then that particular COG is identified as an anomaly. After identifying all such COGs further analysis

of structures needs to be done to identify a possible candidate of HGT.

The table below shows sample data resulting from the comparison of *Bacillus subtilis* and *Escherichia coli*. In this example, COG 454 is considered anomalous because the average Z-score of *Bacillus subtilis* is only 39% of the average Z-score of *Escherichia coli*, which falls below our considered threshold value.

COG Number	*Bacillus subtilis*	*E.coli*	*Comparison Z-score*	*Bacillus subtilis Z-score Normalized*	*E.coli Z-score Normalized*	*Comparison Z-score Normalized*
454	12.09	35.7	9.71	0.34	1	0.27

Table 1: Example of Documented Data.

3. Analysis and Results

Analysis of proteins from *Bacillus subtilis*, which is gram positive, with other gram negative organisms needs to be done. The protein structures of *Bacillus subtilis* were compared with all the *Proteobacteria* (Gram negative) bacteria having more than 40 crystallized proteins in the PDB. There were 19 Gram negative organisms with number of crystallized proteins in them greater than 40. Of these 19 gram negative organisms only 5 organisms had matching COG numbers with the ones in *Bacillus subtilis*. The Gram negative organisms compared with *Bacillus subtilis* are:

1. *Escherichia coli*
2. *Pseudomonas aeruginosa*
3. *Pseudomonas putida*
4. *Haemophilus influenzae*
5. *Helicobacter pylori*

The protein structures of *Bacillus subtilis* are compared with the above 5 gram negative organisms. This comparison is performed only for the common COGs among the two different classes of bacteria i.e., 1 Gram positive organism and 5 Gram negative organisms. The following table gives the summary of the proteins structure comparisons performed in our preliminary analysis.

COG	**Bacterial Pairs**		**Findings**
	Number of Structures in Bacillus subtilis	*Number of Structures in E.coli*	
236	2	6	False hit because of protein complex
454	5	2	The Gram-positive protein structures are same with different ligands and the two Gram-negative proteins are same proteins crystalized twice
745	2	16	Substrate diversity
1057	2	2	The two Gram-positive protein structures are same and the two Gram-negative protein structures are same
1309	3	8	Substrate diversity
1925	7	8	False positive due to multiple protein conformations
	Number of Structures in Bacillus subtilis	*Number of Structures in Pseudomonas*	
1057	2	3	The two Gram-positive protein structures are of the same protein and the three Gram-negative proteins are same with different ligands
	Number of Structures in Bacillus subtilis	*Number of Structures in Pseudomonas*	
1309	3	5	Most likely a good example of HGT

4948	3	5	Most likely a good example of HGT
	Number of Structures in Bacillus subtilis	*Number of Structures in Haemophilus influenzae*	
2050	2	2	The two Gram-positive protein structures are of the same protein and one of the protein structures of the Gram-negative organism is a protein fragment.
	Number of Structures in Bacillus subtilis	*Number of structures in Helicobacter*	
745	2	4	The two Gram-positive protein structures are completely dissimilar. Two of the Gram negative structures are same with different conformations, one is a protein fragment.

Table 2: Summary of the HGT candidates among the compared protein structures.

3.1 Summary of Suspected HGTs

A further detailed analysis of the proteins in these candidate HGTs resulted in identification of the proteins 1VI0 in COG-1309 and 2GGE in COG-4948 as possible HGT to Bacillus subtilis.

PDB-ID	COG	ΔZ-score*	Receiving Bacteria	Donor Bacteria
1VI0	1309	3.49	*Bacillus subtilis*	*Pseudomonas putida*
2GGE	4948	8.49	*Bacillus subtilis*	*Unknown*

Table 3: Summary of Proteins suspected as HGT

* The ΔZ-score is the difference of the average comparison Z-scores of the HGT suspected protein with all the proteins in the opposite Gram organism and the average Z-scores of all the other proteins in the same COG as the suspected protein with all the proteins in the opposite Gram organism.

3.2 Detailed Analysis of suspected COGs

A detailed analysis of the suspected COGs suggested some HGT.

Detailed Analysis of COG-1309

COG-1309 from *Bacillus subtilis* includes 2 structures of putative transcriptional regulators (1RKT,1SGM) and one structure of transcriptional regulator (1VI0). Among these the 1VI0 had the most divergent structure according to the Z-score comparison.

COG-1309 from *Pseudomonas putida* includes five structures, all which are transcriptional regulators. All of the five proteins were closely related according to their Z-scores.

Bacillus subtilis protein 1VI0 was more similar to the five *Pseudomonas putida* proteins than it was to the other *Bacillus subtilis* proteins. Therefore, it is an excellent candidate to be a horizontally transferred gene product.

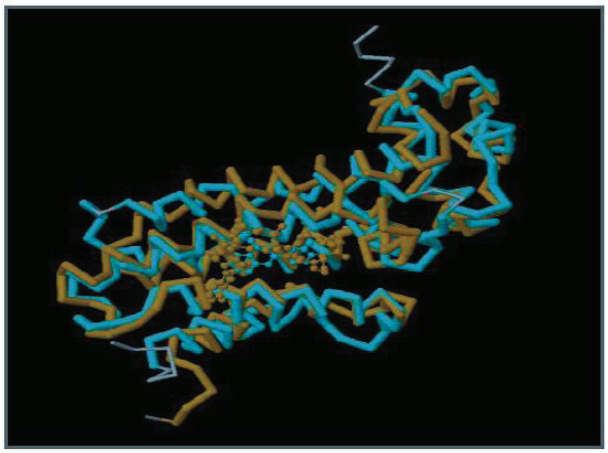

Figure 1: jFATCAT-rigid structure alignment results 1VI0 (*Bacillus subtilis*) vs. 2UXH (*Pseudomonas putida).*

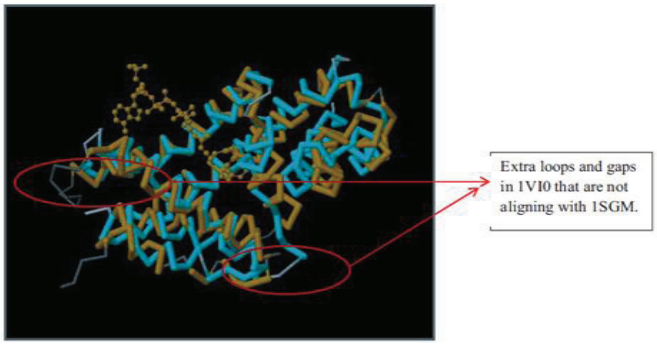

Figure 2: jFATCAT-rigid structure alignment results 1VI0 (*Bacillus subtilis*) vs. 1SGM (*Bacillus subtilis*).

Bacillus subtilis versus each other			
	1RKT	1SGM	1VI0
1RKT		12.3	15.5
1SGM			15
1VI0			

Pseudomonas putida proteins versus each other					
	2UXH	2UXI	2UXO	2UXP	2UXU
2UXH		32	32.4	32.5	32.3
2UXI			32.3	32.4	32.3
2UXO				32.8	32.5
2UXP					32.5
2UXU					

Bacillus subtilis versus Pseudomonas putida proteins					
	2UXH	2UXI	2UXO	2UXP	2UXU
1RKT	15.8	15.8	15.9	15.8	15.9
1SGM	11.9	11.9	11.9	11.9	11.9
1VI0	17.3	17.2	17.3	17.5	17.5

Table 4: COG-1309 in Comparison between *Bacillus subtilis* and *Pseudomonas putida*.

To further confirm that this is a genuine case of HGT, we compare the 3-D structure of the protein 1VI0. Sequence alignments with all the proteins in *Pseudomonas putida* with all other proteins in *Bacillus subtilis* in the COG-1309 are done. This is done using the Jmol tool.

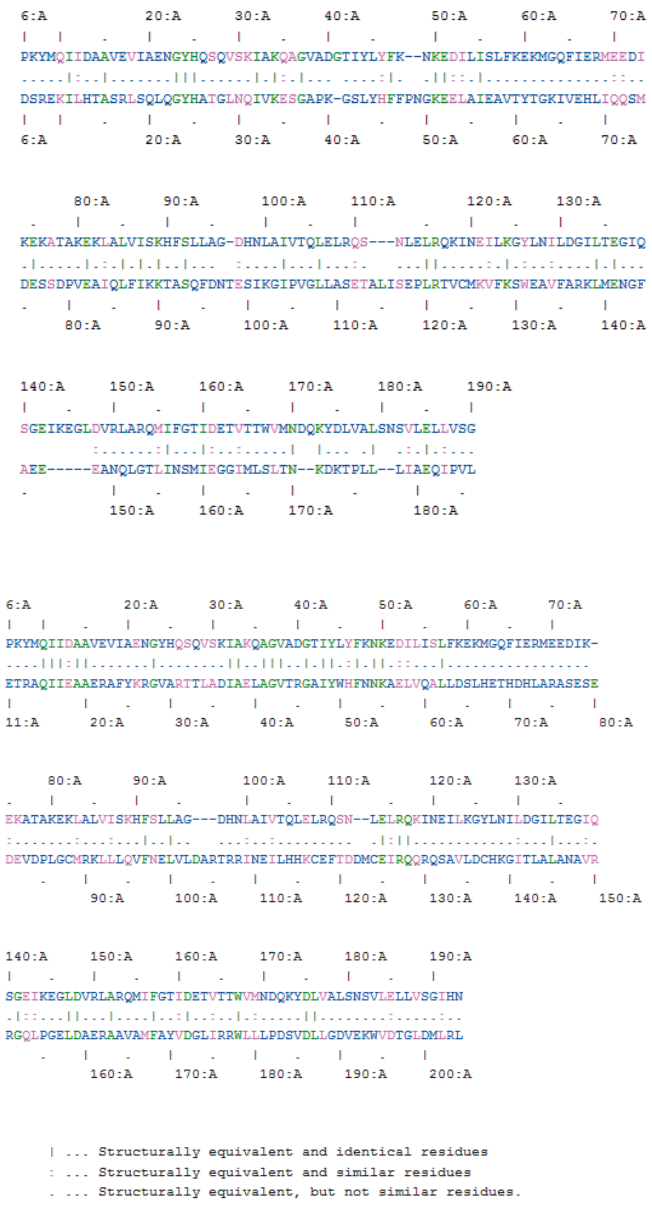

Figure 3: (top) Sequence alignment results 1VI0 (*Bacillus subtilis*) vs. 1SGM (*Bacillus subtilis*). (bottom) Sequence alignment results 1VI0 (*Bacillus subtilis*) vs. 2UXH (*Pseudomonas putida*).

A similar detailed analysis was done on the COG 4948 and observed that protein 2GGE must have been probably been transferred horizontally but not from *Pseudomonas putida.*

3.3 False Positives

A situation where erroneously a positive result is observed is termed as false positive. During our analysis we noticed many situations that might cause false positives. They are listed as follows:

1. **Protein Fragments:** Many of the PDB-ids in the Protein Data Bank correspond to protein domains and protein fragments. The structural comparison of these domains and protein fragments with the whole protein sometimes leads to falsely suspecting a protein for HGT. Good examples of this case are COG-2050 and COG-745.

2. **Substrate Diversity:** The COG's enzyme specificity is fixed within the COG but the substrate specificity is diverse. Good examples for this case are COG-745 and COG-1309.

3. **Conformation changes:** There are two or more conformations of the same protein. Example: COG-1925 and COG-745.

4. **HGT from other sources:** There are some cases in which a protein is identified as possible HGT but not exactly from the organism with which we are comparing. Example: Protein 2GGE in COG-4948.

5. **Different Subunits:** Different subunits of a multi subunit enzyme have very dissimilar structures and with the structure-based method these could look like a possible candidate of HGT, but they are not.

4. Conclusion and Future Work

4.1 Conclusion

We extended Santosh et al. (2011) and presented an improved protein structure based method to detect horizontal gene transfer. We tried to identify possible HGT in *Firmicutes* from *Proteobacteria*. Various cases of false positives have been identified and documented. This method cannot be evaluated for efficiency over other methods for two reasons. First, because it uses a completely different approach to identify HGT, as in it uses protein structures rather than complete genomes used in other techniques. Secondly, each of the techniques used to identify HGT do not yield the same result set. Automation of the procedure to identify HGT was possible only to a certain extent after which the data had to be analyzed manually, which took substantial amount of time. Automation of the entire procedure would be complex to implement as careful analysis and structural visualization of each candidate for HGT was required to zero in on a participant of HGT.

4.2 Future Work

The false positives discussed earlier can cause erroneous results. This method can be improved by eliminating the cases for false positives.

Accuracy of our method also depends on the accuracy of sources from which data is collected for various organisms. Unfortunately we cannot guarantee this. The main source of data for this research was the PDB database. Many underlying problems exist with this database, some of which are as follows:

1. Like any other biological database, PDB is incomplete, as in it does not contain complete protein structure information for all the organisms. It's a constant growing collection of sets of protein structure data. So there is limited flexibility when choosing organisms.

2. Since it relies on entries from various biologists and biochemists, same proteins may be crystallized multiple times, resulting in duplicated entries (multiple PDB IDs for the same protein).

3. Some proteins have been crystallized with and without ligands and substrates, each appear with a unique PDB-id.

4. Protein domains and protein fragments appear with unique PDB-id.

5. Some proteins have been mutated at only one or a few residues, but each structure has a unique PDB-id.

As the quality of the biological databases used increases, so can the efficiency of our method be improved.
This research was based on COG classification, which is a generalized classification. But researchers are moving away from this classification to more specific types of classification of proteins such as GO and eggNOG . Some of the databases have already gotten rid of this classification. Our method can also be applied and tested with these classifications to prove its efficiency. Following similar procedures to identify HGT with these new classifications might provide interesting results.

The DaliLite tool used in this research for structural comparison of proteins can be replaced with CPASS program which compares ligand defined active sites to determine sequence and structural similarity.

This research can be scaled to other organisms belonging to other classifications of phyla. As more genomic data of

organisms becomes available in the biological databases, this research can be used to identify more cases of HGT.

Scalability of this research might help to answer other intriguing questions, such as:

1. Which proteins have more probability of being horizontally gene transferred?
2. What is the functionality of such proteins?
3. Which organism has the highest percentage of HGT proteins?
4. What are the conditions that would enable a horizontal gene transfer?
5. What is rate of occurrence of the HGT?

Identifying the reasons and causes behind the occurrence of HGT can be an interesting way to extend this research. Each method to detect HGT follows a different approach. Comparison and statistical analysis to see the accuracy of each of the methods could also provide interesting results.

References

Akiba T.; Koyama K.; Ishiki Y.; Kimura S.; and Fukushima T. 1960. On the mechanism of the development of multiple-drug resistant clones of Shigella. *Japanese Journal of Microbiology.* 4:219–27.

Berman, H.M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T.N.; Weissig, H.; Shindyalov, I.N.; and Bourne, P.E. 2000. The Protein Data Bank. *Nucleic Acids Research*, 28: 235-42, 2000.

Bourne, P.E.; Addess, K.J.; Bluhm, W.F.; Chen, L.; Deshpande, N.; Feng, Z.; Fleri, W.; Green, R.; Merino-Ott, J.C.; Townsend-Merino, W.; Weissig, H.; Westbrook, J.; Berman, H.M. 2004. The distribution and query systems of the RCSB Protein Data Bank. *Nucleic Acids Research* 32: 223-225.

Frank-Kamenetskii, M.D.; and Liapin, L. 1993. *Unraveling DNA: The Most Important Molecule of Life*. Reading, Mass.: Perseus Books.

Koonin, E.V.; Senkevich, T.G.; and Dolja, V.V. 2006. The ancient virus world and evolution of cells. *Biology Direct* 1:29.

Lawrence J.G.; and Ochman H. 1997. Amelioration of bacterial genomes: rates of change and exchange," *Journal of Molecular Evolution*, 383-97.

Lawrence, J.G.; and Ochman, H. 1998. Molecular archaeology of the Escherichia coli genome," *Proceedings of the National Academy of Sciences USA*, 9413-9417.

Lawrence, J.G.; and Ochman, H. 2002. Reconciling the many faces of lateral gene transfer". *Trends in Microbiology*, 10:1-4.

Powers, R.; Copeland, J.C.; Germer, K.; Mercier, K.A.; Ramanathan, V.; and Revesz, P. 2006. Comparison of Protein Active Site Structures for Functional Annotation of Proteins and Drug Design," *PROTEINS: Structure, Function, and Bioinformatics*, 65(1):124-135.

Revesz, P.Z., 2010. *Introduction to Databases: From Biological to Spatio-Temporal.* New York: Springer.

Santosh, V.; Griep, M.; Revesz, P.Z. 2011. Protein Structure-Based Method for Identifying Horizontal Gene Transfer. In *Proceedings of the 4th International C* Conference on Computer Science and Software Engineering,* 9-16. New York, NY: ACM Press.

Shortridge, M. D.; Triplet, T.; Revesz, P.Z.; Griep, M.A.; and Powers, P. 2011. Bacterial protein structures reveal phylum dependent divergence. *Computational Biology and Chemistry*, 35(1): 24-33.

Song L.; Ning Y.; Zhang Q.; Yang C.; Gao G.; and Han J. 2008. Studies on antimicrobial resistance transfer in vitro and existent selectivity of avian antimicrobial-resistant enterobacteriaccae in vivo. *Agricultural Sciences in China*, 7:636-640.

Susanna, K.A.; den Hengst, C.D., Hamoen, L.W.; and Kuipers, O.P. 2006. Expression of transcription activator ComK of Bacillus subtilis in the heterologous host Lactococcus lactis leads to a genome-wide repression pattern: A case study of horizontal gene transfer. *Applied Environmental Microbiology* 72(1): 404–411.

Tatusov, R.L.; Galperin, M.Y.; Natale, D.A.; and Koonin, E.V. 2000. The COG database: A tool for genome-scale analysis of protein functions and evolution. *Nucleic Acids Research*, 28: 33-36.

Tatusov, R.L.; Natale, D.A.; Garkavtsev, I.V.; Tatusova, T.A.; Shankavaram, U.T.; Rao, B.S.; Kiryutin, B.; Galperin, M.Y.; Fedorova, N.D.; and Koonin E.V. 2001. The COG database: New developments in phylogenetic classification of proteins from complete genomes. *Nucleic Acids Research*, 29: 22-28.

Triplet, T.; Shortridge, M.; Griep, M.; Stark, J.; Powers, R., and Revesz, P. 2010. PROFESS: a PROtein Function, Evolution, Structure and Sequence database. *Database: The Journal of Biological Databases and Curation*. doi no. 10.1093/baq011.

Tsirigos, A.; and Rigoutsos, I. 2005. A new computational method for the detection of horizontal gene transfer events. *Nucleic Acids Research*, 33: 922–933.