Microbiological Research 211 (2018) 31–46

Contents lists available at ScienceDirect

Microbiological Research

journal homepage: www.elsevier.com/locate/micres

Comparison of genomic islands in cyanobacteria: Evidence of bacteriophage-mediated horizontal gene transfer from eukaryotes

James S. Godde[a,b,*], Shakuntala Baichoo[c], Zahra Mungloo-Dilmohamud[c], Yasmina Jaufeerally-Fakim[b]

[a] Department of Biology, Monmouth College, 700 East Broadway, Monmouth, IL, 61462, USA
[b] Department of Agriculture and Food Science, Faculty of Agriculture, University of Mauritius, Reduit, Mauritius,
[c] Department of Digital Technologies, Faculty of ICDT, University of Mauritius, Reduit, Mauritius

ARTICLE INFO

Keywords:
Trans-domain lateral gene transfer
Horizontal gene transfer
Genomic islands
Microbial evolution

ABSTRACT

A number of examples of putative eukaryote-to-prokaryote horizontal gene transfer (HGT) have been proposed in the past using phylogenetic analysis in support of these claims but none have attempted to map these gene transfers to the presence of genomic islands (GIs) in the host. Two of these cases have been examined in detail, including an ATP sulfurylase (ATPS) gene and a class I fructose bisphosphate aldolase (FBA I) gene that were putatively transferred to cyanobacteria of the genus *Prochlorococcus* from either green or red algae, respectively. Unlike previous investigations of HGT, parametric methods were initially used to detect genomic islands, then more traditional phylogenomic and phylogenetic methods were used to confirm or deny the HGT status of these genes. The combination of these three methods of analysis- detection of GIs, the determination of genomic neighborhoods, as well as traditional phylogeny, lends strong support to the claim that trans-domain HGT has occurred in only one of these cases and further suggests a new insight into the method of transmission of FBA I, namely that cyanophage-mediated transfer may have been responsible for the HGT event in question. The described methods were then applied to a range of prochlorococcal genomes in order to characterize a candidate for eukaryote-to-prokaryote HGT that had not been previously studied by others. Application of the same methodology used to confirm or deny HGT for ATPS and FBA I identified a ⊗12 fatty acid desaturase (FAD) gene that was likely transferred to *Prochlorococcus* from either green or red algae.

1. Introduction

Transformation, conjugation, introgression, and phage-mediated transduction represent four ways that HGT is known to occur (Thomas and Nielsen, 2005; Arber, 2014). Although the concept was first described over seven decades ago, HGT remained merely an interesting genetic oddity until the turn of the 21 st century (Koonin et al., 2001; Arber, 2014). As a new genomics era was issued in with the advent of genome sequencing, the impact of HGT on prokaryotic evolution began to become much clearer. With the first bacterial genome sequenced in 1995 and the first archaeal genome published the following year, the field of comparative genomics was born (Jain et al., 2002). It was noticed that certain groups of genes in the archaea were more similar to their bacterial homologs, while others shared more similarity with eukaryotes (Jain et al., 2002). This also suggested that trans-domain HGT was likely to be more common than previously thought. HGT is now known to be ubiquitous and can explain the evolutionary history of prokaryotic genomes. Genomic islands, or GIs, represent direct evidence of the horizontal transfer of genes across species. GIs have been defined as segments of DNA ≥10 kb that have DNA parameters which differ from their surrounding genome (< 10kb segments are termed genomics islets) (Juhas et al., 2009). Other features of GIs are that they often 1) are inserted at tRNA genes, 2) are flanked by direct repeats, 3) harbor integrases, 4) carry insertion elements or transposons, and 5) carry genes which offer a selective advantage to the host (Juhas et al., 2009). The presence of specific genes on GIs and the selective advantage they confer allows further classification of GIs as pathogenicity, symbiosis, metabolic, fitness, or resistance islands, respectively (Hacker et al., 1997; Juhas et al., 2009).

The accepted methods of HGT detection essentially fall into three categories: phylogenetic, parametric, and phylogenomic ones (Ragan, 2001; Azad and Lawrence, 2012). Phylogenetic approaches of creating multiple sequence alignments with either the DNA or protein sequences in question and creating phylogenetic trees from this data by various

Abbreviations: GI, genomic island; HGT, horizontal gene transfer

* Corresponding author at: Department of Biology, Monmouth College, 700 East Broadway, Monmouth, IL, 61462, USA.
E-mail address: jgodde@monmouthcollege.edu (J.S. Godde).

https://doi.org/10.1016/j.micres.2018.03.005
Received 20 November 2017; Received in revised form 11 February 2018; Accepted 17 March 2018
Available online 11 April 2018
0944-5013/

methods is the longest and most widely used method. Here, HGT is detected by anomalous placement of individuals within otherwise congruent trees (Godde and Bickerton, 2006). The main limitation of this approach is similar to disadvantages of using BLAST, the Basic Local Alignment Search Tool, to detect HGT; both are dependent on the breadth and depth of the sequence database being used (Azad and Lawrence, 2012). Indeed, some claims of putative HGT events have later been refuted once database coverage of the gene in question became more robust (Katz, 1996; Grauvogel et al., 2007). Parametric approaches examine the DNA sequence itself for atypical composition compared to the surrounding genome (Ragan, 2001; Azad and Lawrence, 2007, 2012). The parameters which are studied include nucleotide composition bias such as provided by% G + C, dinucleotide bias, codon usage bias, as well as tetranucleotide distributions, among others (Azad and Lawrence, 2012). Parametric analysis has the advantage over phylogenetics in that it does not rely on database coverage, only on the sequence of the individual genome in question. A main disadvantage of parametric analysis is that it is useful only in detecting fairly recent HGT events (within 10 million years or so) due to the process of amelioration, whereby foreign DNA slowly changes to resemble the host genome since they are now subject to the same mutational forces (Lawrence and Ochman, 1997). Phylogenomics provides evolutionary information by comparing entire genomes. The study of the distribution of genes among genomes (phyletic pattern analysis), the differences in genome content among close relatives, as well as the position of genes within genomic neighborhoods fall into this category (Ragan, 2001; Azad and Lawrence, 2007; Zhaxybayeva, 2009). This third class of HGT detection is typically used in conjunction with parametric analysis to help overcome some of its drawbacks, such as in the assignment of potential HGT events which fall close to the (admittedly, somewhat arbitrarily defined) threshold for given parameters (Azad and Lawrence, 2007, 2012). Similarity searches such as BLAST are still typically used to perform whole genome surveys and have been used to estimate the overall levels of HGT in prokaryotic genomes. One study found that levels of putative HGT were as high as 17% in cyanobactera (Ochman et al., 2000). Koonin et al. (2001) found that the same cyanobacterium, *Synechocystis*, exhibited 11% acquired genes, and ranked the archaea Halobacterium at the top of their list with 16% HGT. In addition, these researchers found that an average of one percent of prokaryotic genomes displayed similarity to eukaryotes, with both *Synechocystis* and *Halobacterium* exhibiting significantly higher levels of putative *trans*-domain HGT (Koonin et al., 2001). HGT is thus predicted to have played an integral role in the evolution of prokaryotic genomes, including the diversification which leads to speciation (Ochman et al., 2000).

Koonin et al. (2001) estimated that about one percent of prokaryotic genomes could be the result of HGT from eukaryotic sources. While the percentage may seem low, the implications for potential HGT events are enormous. NCBI currently lists 3.8 million prokaryotic sequences in its Gene database. If this estimate is correct, that means there are perhaps 38,000 known genes which are the result of HGT from eukaryotes! Koonin et al. (2001), for their part, did list almost 100 putative eukaryote-to-prokaryote HGT events which were identified by such methods as detecting the unexpected ranking of sequence similarity among homologs, unexpected phylogenetic tree topologies, unusual phyletic patterns, as well as the conservation of gene order between distant taxa. Koonin's laboratory then grouped these putative transfers into three main categories: aminoacyl-tRNA sythetases (26), proteins and domains involved in signal transduction (12), and “functionally diverse genes” (21). One would expect that, within the intervening years, a number of the miscellaneous genes that make up the final category would have been investigated using one or more of the HGT detection methods discussed above. A search of the literature, however, reveals a single case: that of a class I fructose-bisphosphate aldolase, or *FBA I*, the transfer of which has been supported by phylogenetic and phylogenomic methods (Rogers et al., 2007). That is not to say that a dozen or more other eukaryote-to-prokaryote transfers have been proposed since then, using mostly phylogenetic arguments (Jenkins et al., 2015; Andersson et al., 2003; Cazalet et al., 2004; Da Lage et al., 2004; Ruiz-González and Marín, 2004, Bond et al., 2005; Richards et al., 2006; Guljamow et al., 2007; Almeida et al., 2008; Patron et al., 2008; Takishita and Inagaki, 2008; Lurie-Weinberger et al., 2010; Wu and Zhang 2011; Duplouy et al., 2013; Gomez-Valero and Buchrieser, 2013). One surprising thing about these studies is that none of them employed parametric methods to investigate the HGT in question. Parametric methods were thus used to locate GIs within specific bacterial genomes to determine whether purported HGT events involving eukaryote-to-prokaryote transfers mapped to these islands. Investigations began with a gene of interest that has been purportedly transferred between eukaryotes and cyanobacteria: *ATPS*, which encodes ATP sulfurylase, an enzyme involved in purine, selenoamino acid, and sulfur metabolism (Patron et al., 2008). Work then turned to the one example from Koonin et al. (2001) that has been further described in the literature: the transfer of *FBA I* into cyanobacteria (Rogers et al., 2007). FBA, sometimes just called aldolase, is an enzyme involved in the fourth step of glycolysis, where it cleaves fructose 1,6-bisphosphate into glyceraldehyde 3-phosphate, or GADP, and dihydroxyacetone phosphate, or DHAP.

Using cyanobacteria as a model system has a number of advantages. They have apparently been subject to not only some of the highest levels of HGT found in bacteria, but they also contain some of the highest levels of putative trans-domain HGT as well. For HGT to occur, the organisms in question have to live in close proximity to one another and it is noteworthy that the aquatic environment houses cyanobacteria as well as various amoeba and other protists whose contributions to HGT have been documented by others (see Raoult and Koonin, 2012 and accompanying articles). Rogers et al. (2007) found that the class I FBA usually associated with eukaryotes was found in two marine cyanobacteria: *Prochlorococcus* and *Synechococcus* where it typically accompanied, but sometimes replaced, the class II FBA typically found in bacteria. *Prochlorococcus* and *Synechococcus* are two closely related genera of marine cyanobacteria, the former of which is believed to be the smallest photosynthetic organism and is known to have the smallest genome of any free-living phototroph (Partensky et al., 1999; Biller et al., 2015). Both genera can be referred to as photosynthetic picoplankton based on their small size and are particularly prevalent in oligotrophic regions of the oceans where nutrients are scarce (Partensky et al., 1999). Despite their small size, the massive numbers of marine unicellular cyanobacteria enable them to account for 20–40% of carbon fixation in the oceans; *Prochlorococcus* is believed to be the most abundant photosynthetic organism on earth (Palenik et al., 2003, Biller et al., 2015). It is likely due to their importance in the global ecology that *Prochlorococcus* genomes are so well represented in current databases. NCBI's Assembly database houses 201 genomes from different strains of *Prochlorococcus* and 82 genomes from different strains of *Synechococcus*. One group in particular has deposited the vast majority of *Prochlorococcus* genomes from single cell genomics methods applied to environmental samples taken from ocean water (Malmstrom et al., 2012; Kashtan et al., 2014).

In all, the clear advantages of using marine photosynthetic picoplankton as test case to determine whether putative eukaryote-to-prokaryote HGT events could be confirmed using parametric analysis were evident and further study was merited. One goal of this study was to test a set of methods which could then be applied to a wider set of prokaryotic genomes in order to identify more representatives of the multitude of *trans*-domain transfers that have been predicted to have occurred. Another aim was to gain novel insights into the specific mechanisms of transmission which have led to gene transfer in these previously reported cases of HGT.