Prevalence and evolution of core photosystem II genes in marine cyanobacterial viruses and their hosts

Matthew B Sullivan, Debbie Lindell, Jessica A Lee, Luke R Thompson, Joseph P Bielawski, Sallie W Chisholm (2006). PLoS Biology 4:e234

A tour

I. Overview

You are no doubt in the midst of finding articles pertinent to your projects. You'll usually rely on directed searches – searching for some set of key terms – but even then, you'll undoubtedly run across articles that you weren't looking for specifically but jump out at you as potentially interesting… or you ***should***. This is what currently distinguishes a human from a computer, the ability to notice something interesting that was not preprogrammed. If you tell a computer to search a sequence for the amino acid sequence CQEDCRALCTT, it will search for that sequence, and even if it encounters in the middle of the search ELVISISALIVEALIVE, it won't care. In contrast, you, a human, are capable of surprise.

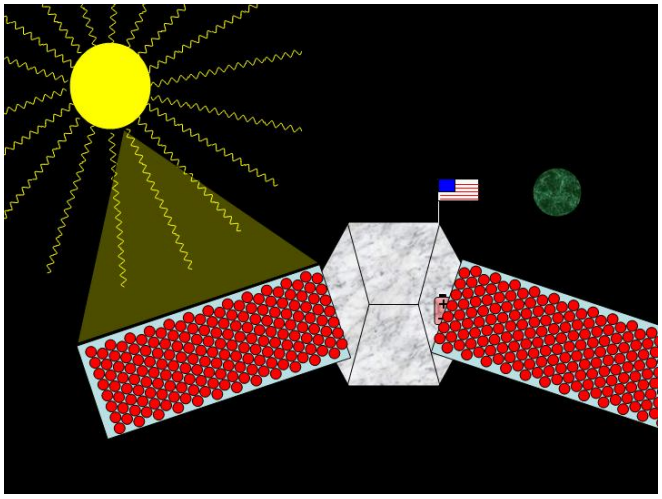
The article by Sullivan et al is not obviously connected to any of your projects, at least not from its title, but if you had some cause to glance through it, you might be surprised by what you find. I'm going to give you some cause to glance through it, by suggesting that the article may be worth your time.

Of course, I'm not suggesting that you actually ***read*** the whole thing. You're not going to be able to go through many articles if you insist on digesting every word of every section of every article. However, skim through it and perhaps linger on some section that strikes your fancy. That section might have something to do with localized variations in GC-content (if you're particularly interested in sequence bias) or the movement of genes from one organism to another (if you're particularly interested in mobile DNA). I wouldn't be surprised if everyone (almost) is interested in the phylogenies shown in the article, since everyone (almost) will find it desirable to generate a phylogeny for your own favorite proteins, perhaps comparing it to the phylogeny of mycobacteriophage as a whole.

II. Photosystem II genes

Regardless of what you consider most interesting about the article, it is clear that what Sullivan et al were most impressed by (as judged by their title and the content of the paper) was the presence of certain genes important in photosynthesis in the genomes of various phages. Why should phages carry with them photosynthesis genes?

We can talk about this if you like, but let me provide here a bit of photosynthesis lore that may help rationalize the phenomenon. Photosynthetic organisms face the problem that they live on sunlight, a very diffuse source of energy. If you had to run on sunlight, you wouldn't be able to gather in enough energy to get out of bed, even if you had green skin and slept outside during the day. Plants manage because they minimize their energy needs to nearly nothing and have elaborate systems for catching light. The same is true at the cellular level. In order for a cell to capture enough light to run metabolism, most of the proteins involved in the process must be devoted to absorbing light, just as the majority of the surface area of a solar powered space station might be energy panels that do nothing else (Fig. 1A).



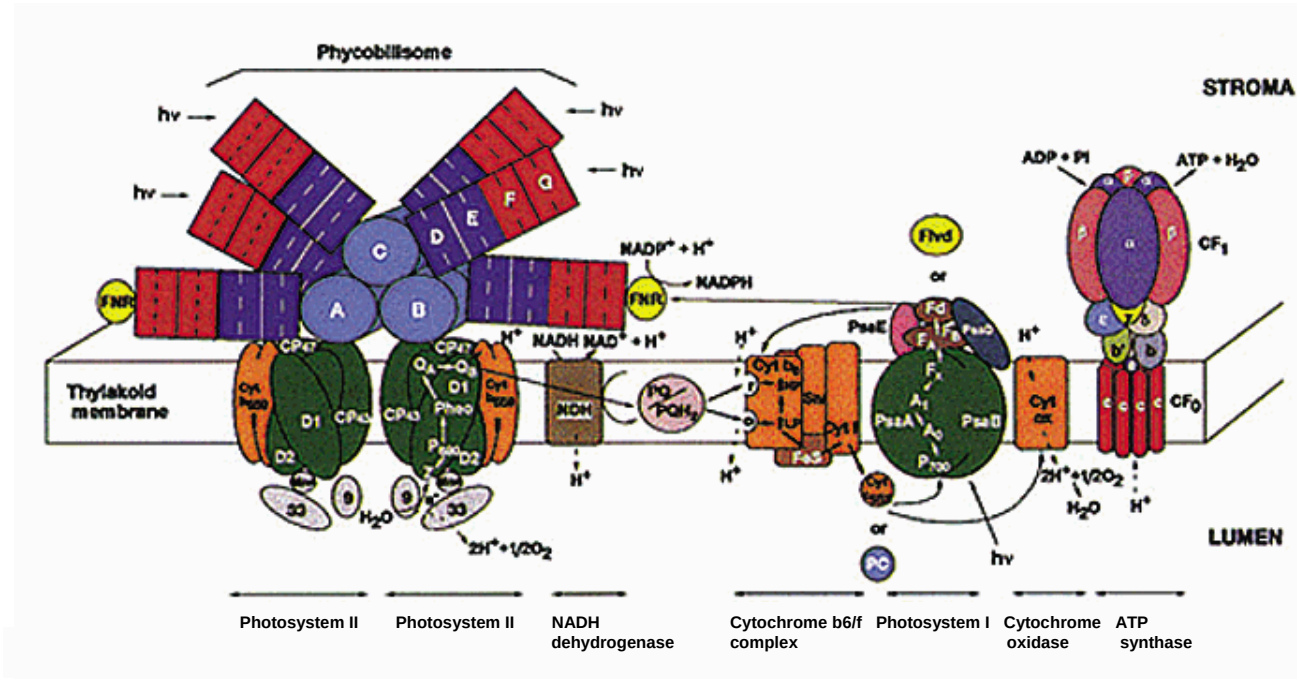


Fig. 1. Capture of light energy. (**A**) Expansive arrays of solar panels are necessary to capture enough light energy to drive the functions of a space station. (**B**) The photosynthesis machinery of most cyanobacteria. Light energy is captured by phycobiliproteins (comprising as much as 40% of the total cell protein), and the resulting high energy electrons are directed towards Photosystem II and the special chlorophyll contained within the D1 and D2 proteins. There's a good deal more of interest in this diagram, but for the present purposes, I'll leave it at that.

In photosynthetic cells, all of the energy absorbed by huge numbers of antenna proteins is funneled in the form of high energy electrons into a much smaller number of protein complexes, called photosystems: Photosystem I and Photosystem II (Fig. 1B). Two proteins, D1 (encoded by *psbA*) and D2 (encoded by *psbD*) are the ultimate targets, and they suffer for their service. The *psbA* gene is amongst the most highly expressed in photosynthetic cells, not because the cells need so much D1 protein but because D1 breaks down so rapidly.

Now enter the phage. One of the first things that many phage do is to disable host transcription, so that the transcriptional machinery can be devoted to the production of phage mRNA. But what about the many host proteins needed by the phage, e.g. the proteins in ribosomes or the enzymes of basic metabolism? No problem, the proteins live long enough for the phage to do what it needs to do, replicate, kill the cell, and leave. But not the short-lived D1 protein. If there is no *psbA* transcript to permit translation of new D1 protein to replace the old degraded D1 protein, then photosynthesis grinds to a halt. So the supply of metabolic energy dries up. So phage replication halts. What's a phage to do? Sullivan et al have something to say on this subject.

III. Phylogenetic analysis

Sullivan et al found that certain phages that infect marine cyanobacteria in the genus Prochlorococcus and the genus Synechococcus carry *psbA* and *psbD* genes. Where did the phage get the genes? Is there a phage version of the gene and a cyanobacterial version? Or did the phage get the gene from their cyanobacterial hosts? If so, which hosts?

These are questions that phylogenetic analysis might be able to answer. You may come up with similar questions and resort to similar tools in search of answers. One common question is whether proteins are related to each other in the same way as the organisms in which they reside are related. If so, then you might conclude that the genes encoding the proteins were passed down generation to generation and so accumulated mutations along with all the other genes in the cell. Alternatively, if the genes exhibit a very different family tree, then that is evidence that the gene may have been picked up by a cell from the outside, perhaps from a phage.

In making these judgments, you will need to have access to a family tree of mycobacteriophage. To my knowledge, the best such tree available is shown in Fig. 2.

IV. Activities

- Pick a favorite part of the article and delve into it.
- Look for opportunities to check the claims of the article yourself, by identifying specific genes within PhAnToMe/BioBIKE.
- Compare a tree of your favorite protein (or tRNA) with the tree of mycobacteriophages. Are they concordant? Is there any evidence of acquisition of exchange of genes amongst phage or between phages and their hosts?

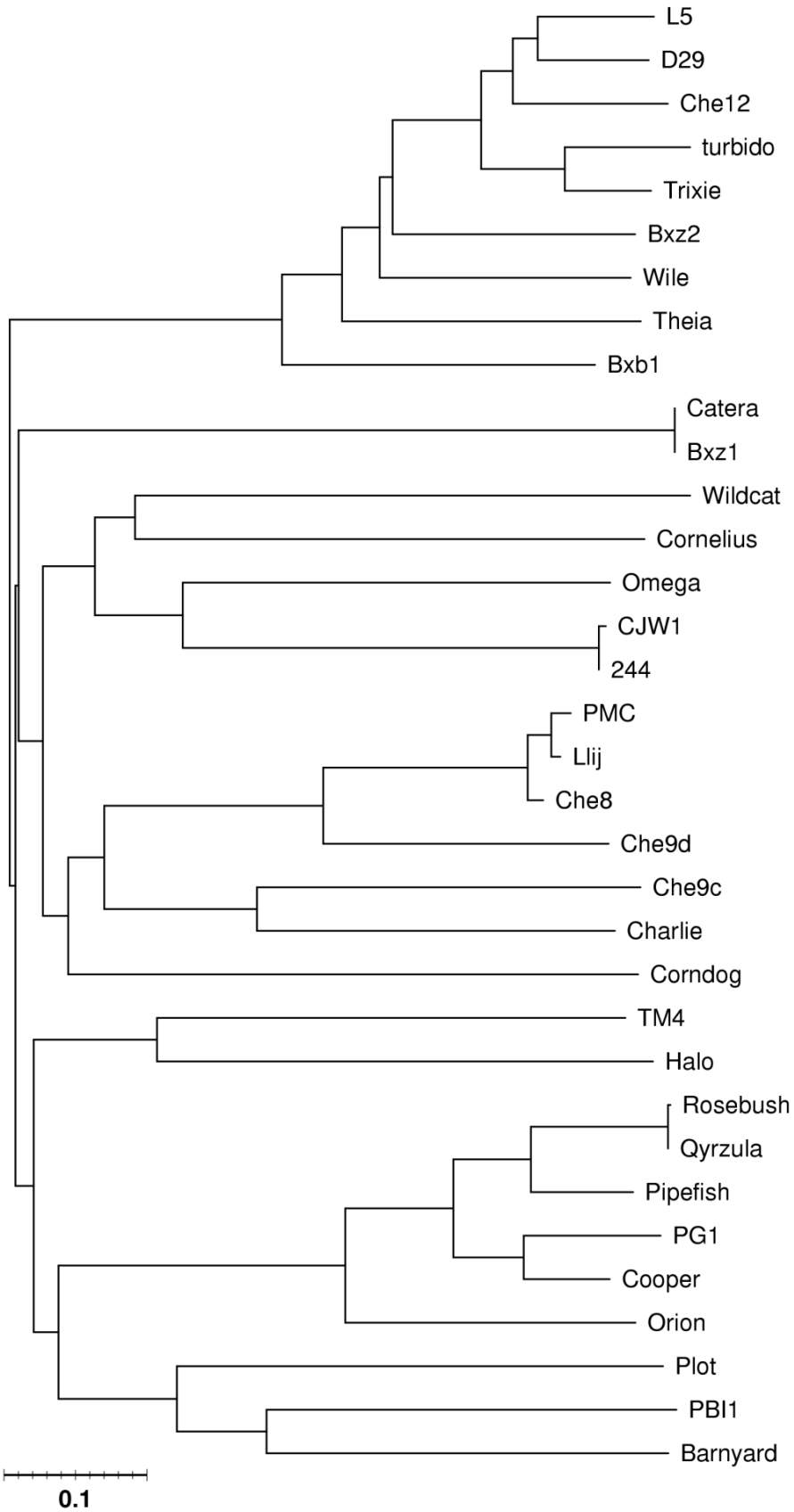


Fig 2: Phage Proteomic Tree based on number of shared proteins between pairs of the mycobacteriophages. From Abi-Falah R (2010), unpublished work.