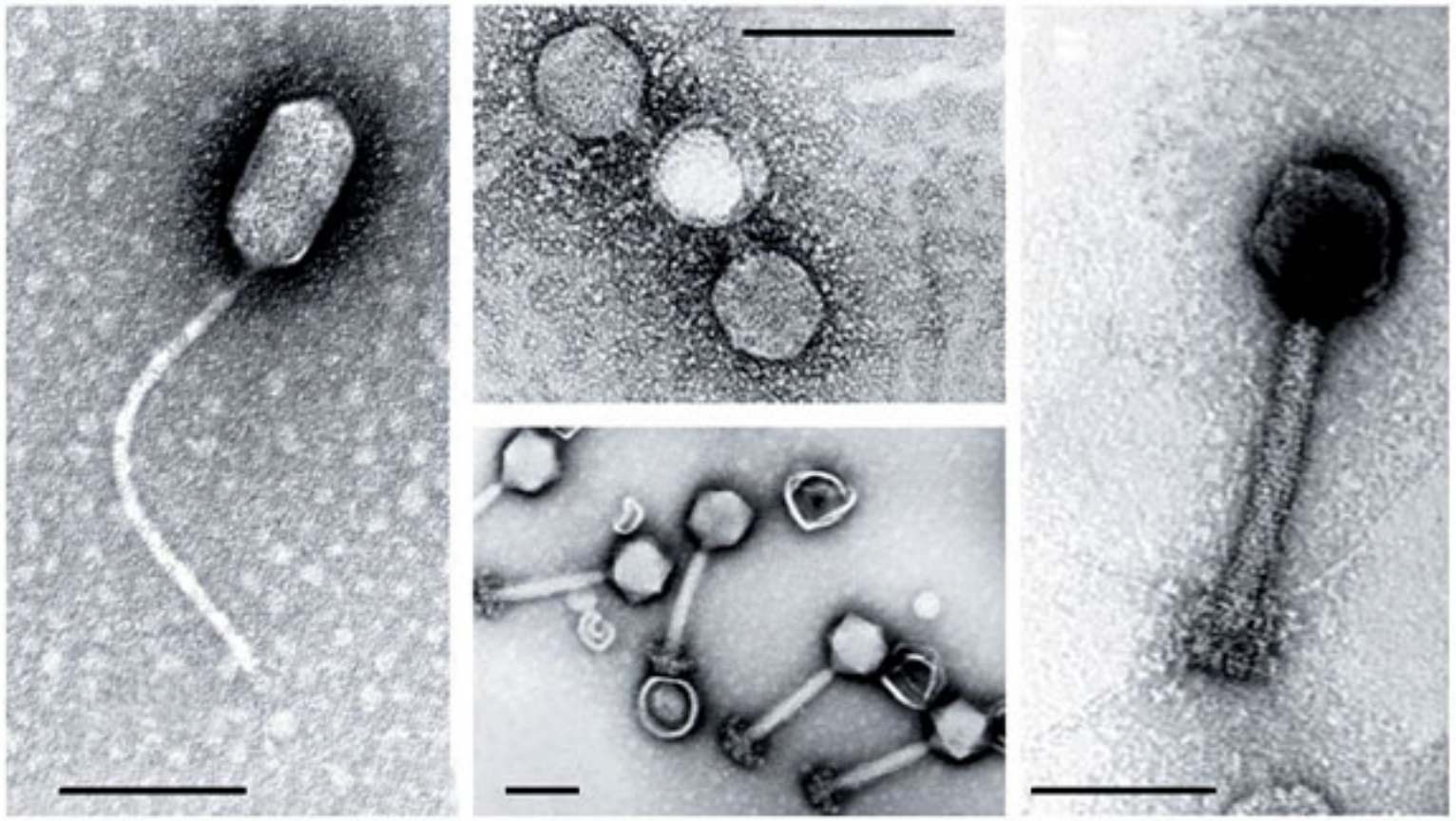
# Viruses in the Ocean



Genome sequences from the sea
Jed Fuhrman
Nature 424, 1001-1002(28 August 2003)
doi:10.1038/4241001a

Hurwitz *et al. Genome Biology* 2013, **14**:R123
http://genomebiology.com/2013/14/11/R123

Genome Biology

RESEARCH Open Access

# Metabolic reprogramming by viruses in the sunlit and dark ocean

Bonnie L Hurwitz[1,4], Steven J Hallam[2,3*] and Matthew B Sullivan[1*]

INSIGHT REVIEW NATURE|Vol 437|15 September 2005|doi:10.1038/nature04160

# Viruses in the sea

Curtis A. Suttle[1]

**Viruses exist wherever life is found. They are a major cause of mortality, a driver of global geochemical cycles and a reservoir of the greatest genetic diversity on Earth. In the oceans, viruses probably infect all living things, from bacteria to whales. They affect the form of available nutrients and the termination of algal blooms. Viruses can move between marine and terrestrial reservoirs, raising the spectre of emerging pathogens. Our understanding of the effect of viruses on global systems and processes continues to unfold, overthrowing the idea that viruses and virus-mediated processes are sidebars to global processes.**

# Executive Summary?

# Bonnie Hurwitz Appointed Program Director of Health Informatics at Arizona Health Sciences Center at UA

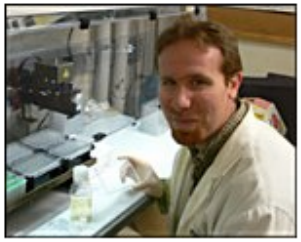
TUCSON, Ariz. – Bonnie L. Hurwitz, an expert in bioinformatics and computational biology, has been appointed program director of health informatics at the Arizona Health Sciences Center at the University of Arizona.

In this new position, Hurwitz will provide coordination and leadership in the evaluation, design and implementation of statewide AHSC biomedical informatics health initiatives. In collaboration with major partners from the UA and organizations statewide, she will promote the use of state-of-the-art technologies and computational methods to advance biomedical informatics. Among other duties, she will develop

**Steven Hallam**
Assistant Professor

B.A., Sarah Lawrence College; Ph.D., University of California Santa Cruz; Postdoctoral, Massachusetts Institute of Technology

Email: shallam[at]interchange.ubc.ca



## Matthew B Sullivan

University of Arizona, Ecology and Evolutionary Biology Department
metagenomics - virus-host interactions - viral ecology - microbial genomics - phage biology
Verified email at email.arizona.edu
Homepage

**Curtis A. Suttle**
**Professor and Associate Dean, Research and Academic Services, Faculty of Science**
**Associate Dean Research, Faculty of Science**
**Biological Oceanography, Marine Microbiology, Environmental Virology**
*Office:* **Biodiversity 346** *Phone:* **604-822-8610** *Office2:* **Biodiversity** *Phone2:* **604-822-2273**
**1505-6270 University Boulevard, Biological Sciences Building**
*E-mail:* **Mail-Me**
*Personal Website:* **http://www.ocgy.ubc.ca/~suttle**

**How did the authors go about addressing the problem?**

OPEN ACCESS Freely available online PLOS BIOLOGY

# The *Sorcerer II* Global Ocean Sampling Expedition: Expanding the Universe of Protein Families

**Shibu Yooseph[1*], Granger Sutton[1], Douglas B. Rusch[1], Aaron L. Halpern[1], Shannon J. Williamson[1], Karin Remington[1], Jonathan A. Eisen[1,2], Karla B. Heidelberg[1], Gerard Manning[3], Weizhong Li[4], Lukasz Jaroszewski[4], Piotr Cieplak[4], Christopher S. Miller[5], Huiying Li[5], Susan T. Mashiyama[6], Marcin P. Joachimiak[6], Christopher van Belle[6], John-Marc Chandonia[6,7], David A. Soergel[6], Yufeng Zhai[3], Kannan Natarajan[8], Shaun Lee[8], Benjamin J. Raphael[9], Vineet Bafna[8], Robert Friedman[1], Steven E. Brenner[6], Adam Godzik[4], David Eisenberg[5], Jack E. Dixon[8], Susan S. Taylor[8], Robert L. Strausberg[1], Marvin Frazier[1], J. Craig Venter[1]**

**1** J. Craig Venter Institute, Rockville, Maryland, United States of America, **2** University of California, Davis, California, United States of America, **3** Razavi-Newman Center for Bioinformatics, Salk Institute for Biological Studies, La Jolla, California, United States of America, **4** Burnham Institute for Medical Research, La Jolla, California, United States of America, **5** University of California Los Angeles–Department of Energy Institute for Genomics and Proteomics, Los Angeles, California, United States of America, **6** University of California Berkeley, Berkeley, California, United States of America, **7** Physical Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley, California, United States of America, **8** University of California San Diego, San Diego, California, United States of America, **9** Brown University, Providence, Rhode Island, United States of America

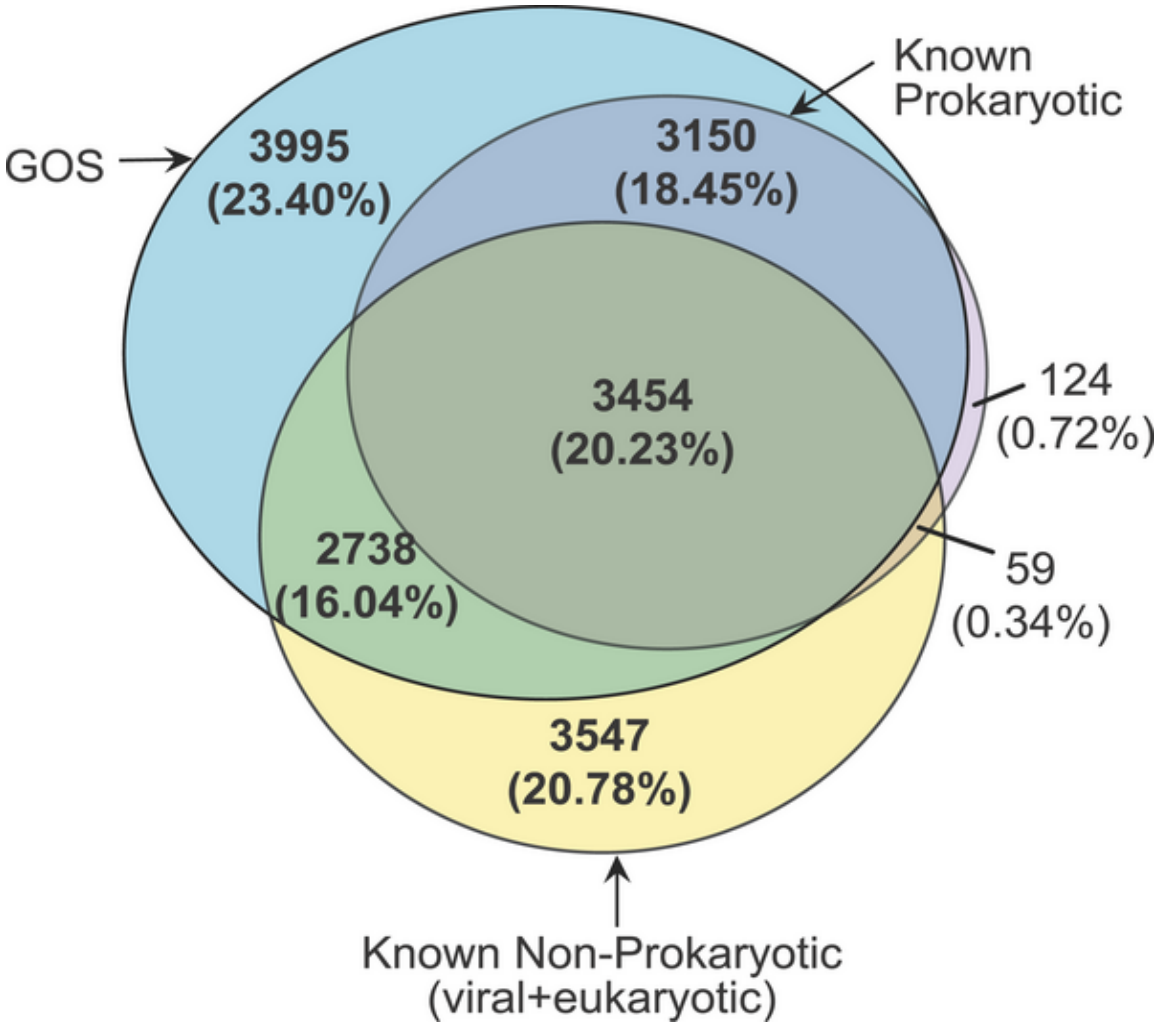
## Table 5.

| Novel Cluster ID | Inferred Function | | p-Value[a] | Neighboring Clusters with Contributing GO Annotation | | Other Neighbors of Interest[b] | Comments |
|---|---|---|---|---|---|---|---|
| | GO ID | Biological process | | Cluster ID | GO Annotation | | |
| 8837 | GO:0006260 | DNA replication | $4.70 \times 10^{-4}$ | 812 | ATPase involved in DNA replication | Phage Mu Mom DNA modification enzyme | Profile–profile match: DNA polymerase processivity factor |
| | | | | 2,655 | DNA polymerase family B | DNA methylase | |
| 12519 | GO:0006118 | Electron transport | $4.54 \times 10^{-3}$ | 1,362 | Cytochrome *c* oxidase subunit III | | Profile–profile match: PF03626— cytochrome *c* oxidase subunit IV; 3 predicted transmembrane helices |
| | | | | 1,771 | SCO1/SenC—biogenesis of photosynthetic systems | | |
| 11010151 | GO:0017004 | Cytochrome complex assembly | $\leq 1.00 \times 10^{-5}$ | 8,136 | Thioredoxin | | >20 diverse profile–profile matches, one of which is cytochrome *c* biogenesis factor ccmH_2 |
| | | | | 9,364 | Cytochrome *c* biogenesis protein | | |
| | | | | 1,317 | Cytochrome *c* assembly protein | | |
| 18456 | GO:0009252 | Peptidoglycan biosynthesis | $\leq 1.00 \times 10^{-5}$ | 1,252 | FAD binding domain | Extracytoplasmic function (ECF) sigma factor 24 | One predicted TM helix |
| | | | | 10,764 | UDP-N-acetylenolpyruvoylglucos-aminereductase | Viral RNA helicase | |
| 14219 | GO:0009628 | Response to abiotic stimulus | $3.10 \times 10^{-4}$ | 5,936 | Colicin V production protein | | Predicted soluble; exclusively neighbors to just two clusters. |
| | | | | | MatE multidrug efflux pump | | |
| 11480 | GO:0015031 | Protein transport | $3.00 \times 10^{-4}$ | 4,177 | MotA/TolQ/ExbB proton channel family | | Four predicted TM helices; Tol proteins facilitate transport of colicins, iron, and phage DNA |
| | | | | 9,569 | Biopolymer transport protein ExbD/TolR | | |
| 14360 | GO:0006777 | Mo-molybdopterin cofactor biosynthesis | $\leq 1.00 \times 10^{-5}$ | 9,745 | MoaC family | Sulfite oxidase | SAR11 blast match annotated as probable moaD; profile–profile matches to ThiS and molybdopterin converting factor; <.05% of sequences have PFAM match to ThiS family |
| | | | | 9,948 | MoaE protein | Predicted thioesterase | |
| | | | | 255 | Radical SAM superfamily | | |
| 8397 | GO:0017004 | Cytochrome complex assembly | $\leq 1.00 \times 10^{-5}$ | 8,136 | Thioredoxin | SMC superfamily (homologous to ABC family) | Blast match to "periplasmic or inner membrane–associated protein"; two predicted TM helices; 0.7% of sequences have PFAM match to cytochrome *c* biogenesis protein |
| | | | | 9,364 | Uncharacterized cytochrome *c* biogenesis protein | | |
| 13909 | GO:0015979 | Photosynthesis | $\leq 1.00 \times 10^{-5}$ | 13,990 | Photosystem II reaction centre N protein (psbN) | | Predicted soluble; single blast match to cyanophage P-SSM2 hypothetical protein; many phage proteins as minor neighbors |
| | | | | 5,184 | Photosynthetic reaction centre protein D1 (psbA) | | |
| | | | | 7,664 | Ferredoxin-dependent bilin reductase | | |

[a]p-Values were computed by simulating 100,000 neighbor cluster sets of equivalent size.
[b]Not all clusters could be mapped to a GO term.
doi:10.1371/journal.pbio.0050016.t005

Yooseph S, Sutton G, Rusch DB, Halpern AL, et al. (2007) The Sorcerer II Global Ocean Sampling Expedition: Expanding the Universe of Protein Families. PLoS Biol 5(3): e16. doi:10.1371/journal.pbio.0050016
http://www.plosbiology.org/article/info:doi/10.1371/journal.pbio.0050016

**Figure 3.**




Yooseph S, Sutton G, Rusch DB, Halpern AL, et al. (2007) The Sorcerer II Global Ocean Sampling Expedition: Expanding the Universe of Protein Families. PLoS Biol 5(3): e16. doi:10.1371/journal.pbio.0050016
http://www.plosbiology.org/article/info:doi/10.1371/journal.pbio.0050016

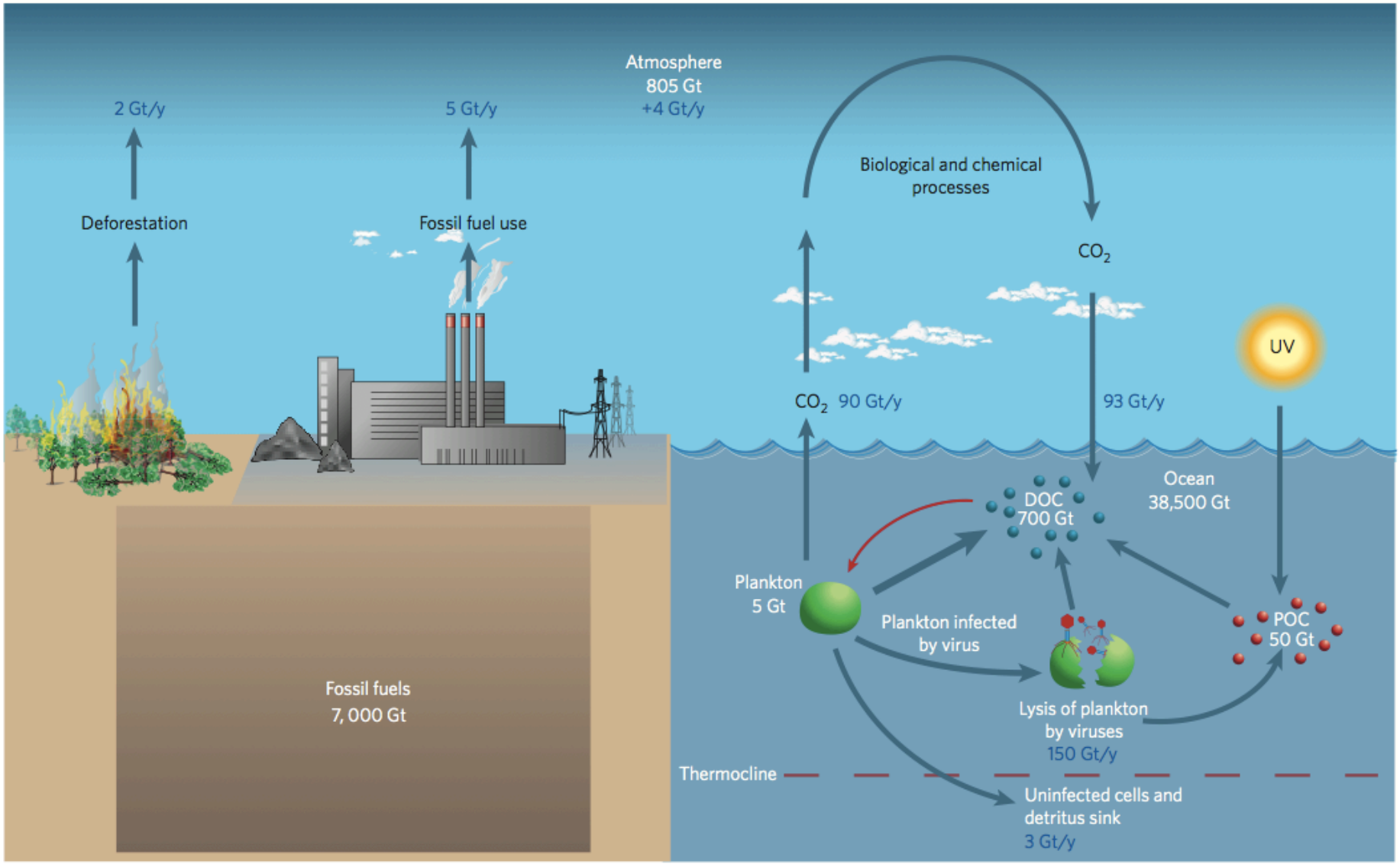



**Figure 4 | Viruses can affect the efficiency of the biological pump.** Viruses cause the lysis of cells, converting them into particulate organic carbon (POC) and dissolved organic carbon (DOC). This reduces the rate atwhich C sinks from the surface layer into the deep ocean where the carbon is trapped for millennia (biological pump). Instead the carbon is retained in the surface waters where it is photo-oxidized and respired, in chemical equilibrium with the atmosphere. The net effect is a faster rate of $CO_2$ build-up in the atmosphere than would occur if the POC were 'exported' to the deep ocean.

**What is the specific hypothesis being addressed?**

## Sample Collection

- Seawater collected from 4 regions around the Pacific Ocean at varying depths (5 to 4500 meters), nutrients and season

## Sample Preparation

- Glass Filtration
- .22um to pre-filter the seawater
- Concentrated by FeCl precipitation
- DNase and CsCl purification
- DNA extraction using Wizards PCR DNA Purification Resin and Minicolumns
- Modified Linker Amplification (LA) protocol

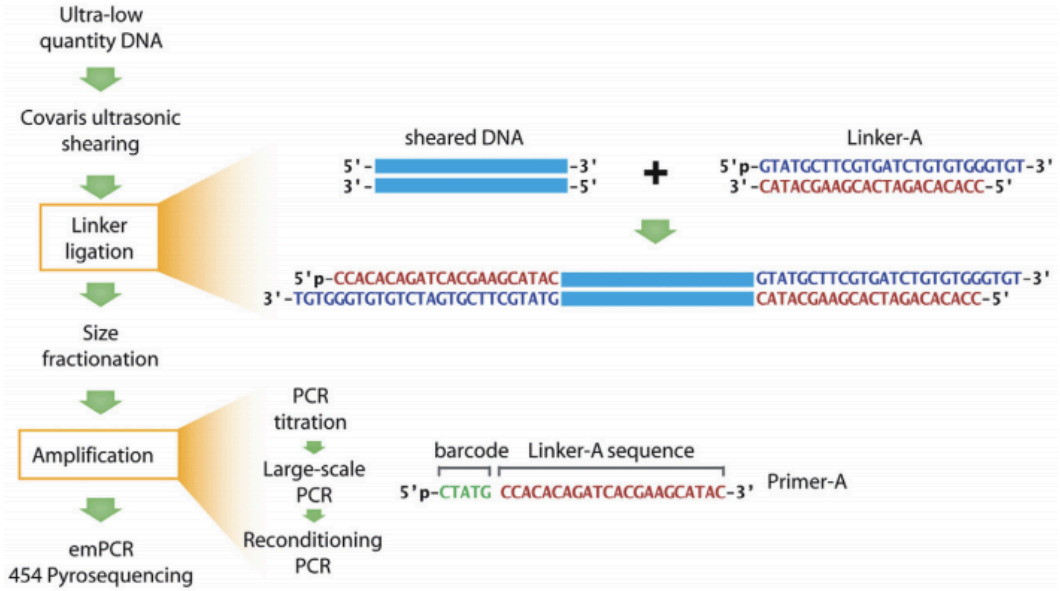



**Fig. 1.** Linker amplification (LA) method schema. This study assesses an optimized LA method, with particular focus on providing new bar-codes in the linker ligation step to facilitate pooling of samples, as well as quantitative evaluation of the impact of amplification on resulting isolate and community DNA genomic sequencing.

 *Environmental Microbiology*, **14**, 2526–2537

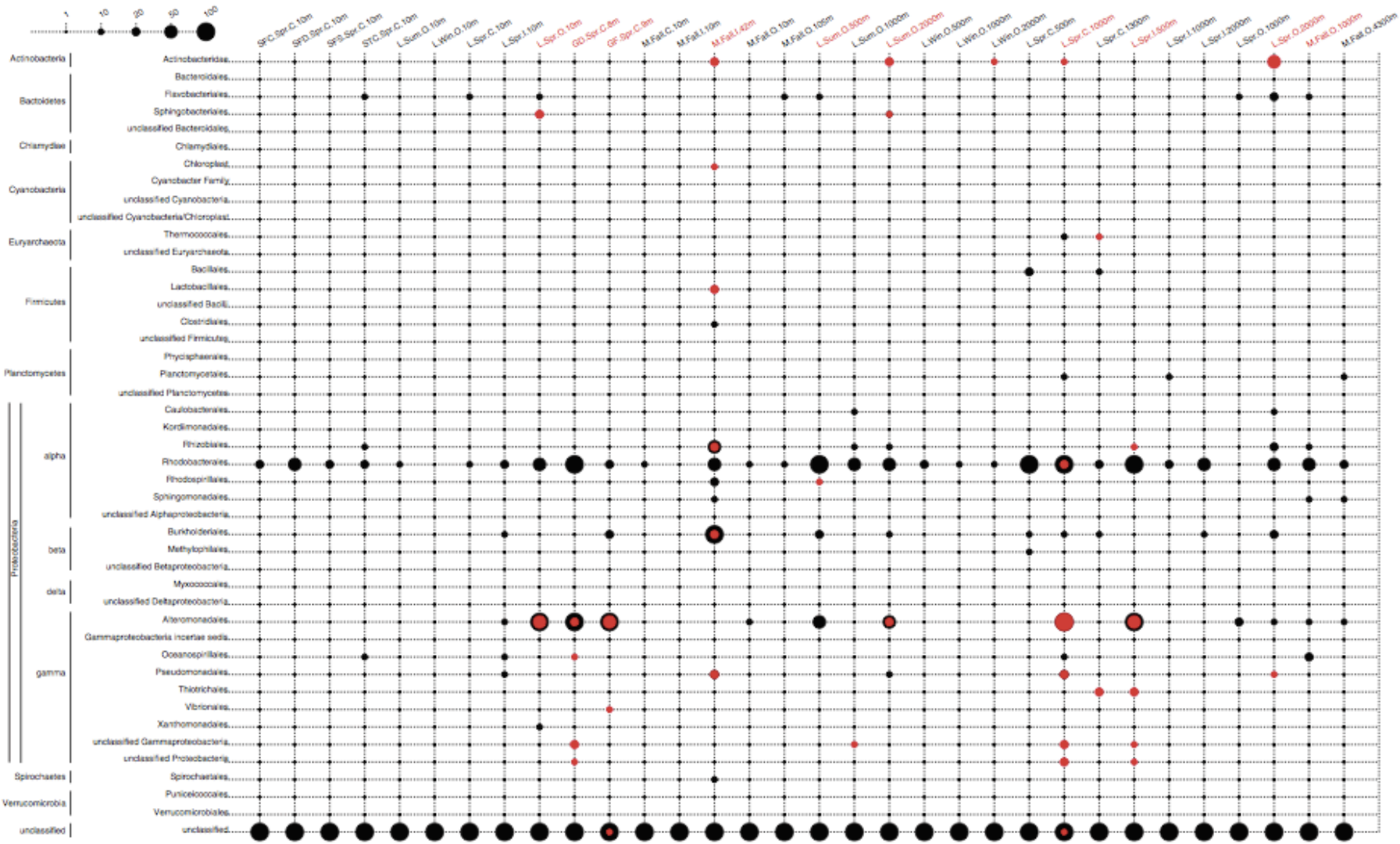



**Figure 1 Taxonomic distribution of viral metagenomic read hits to small subunit 16S ribosomal DNA and carbon metabolism genes by bacterial order.** 16S hits are noted in red and carbon metabolism gene hits are noted in black. Samples and metadata are further described by Hurwitz and Sullivan [21].

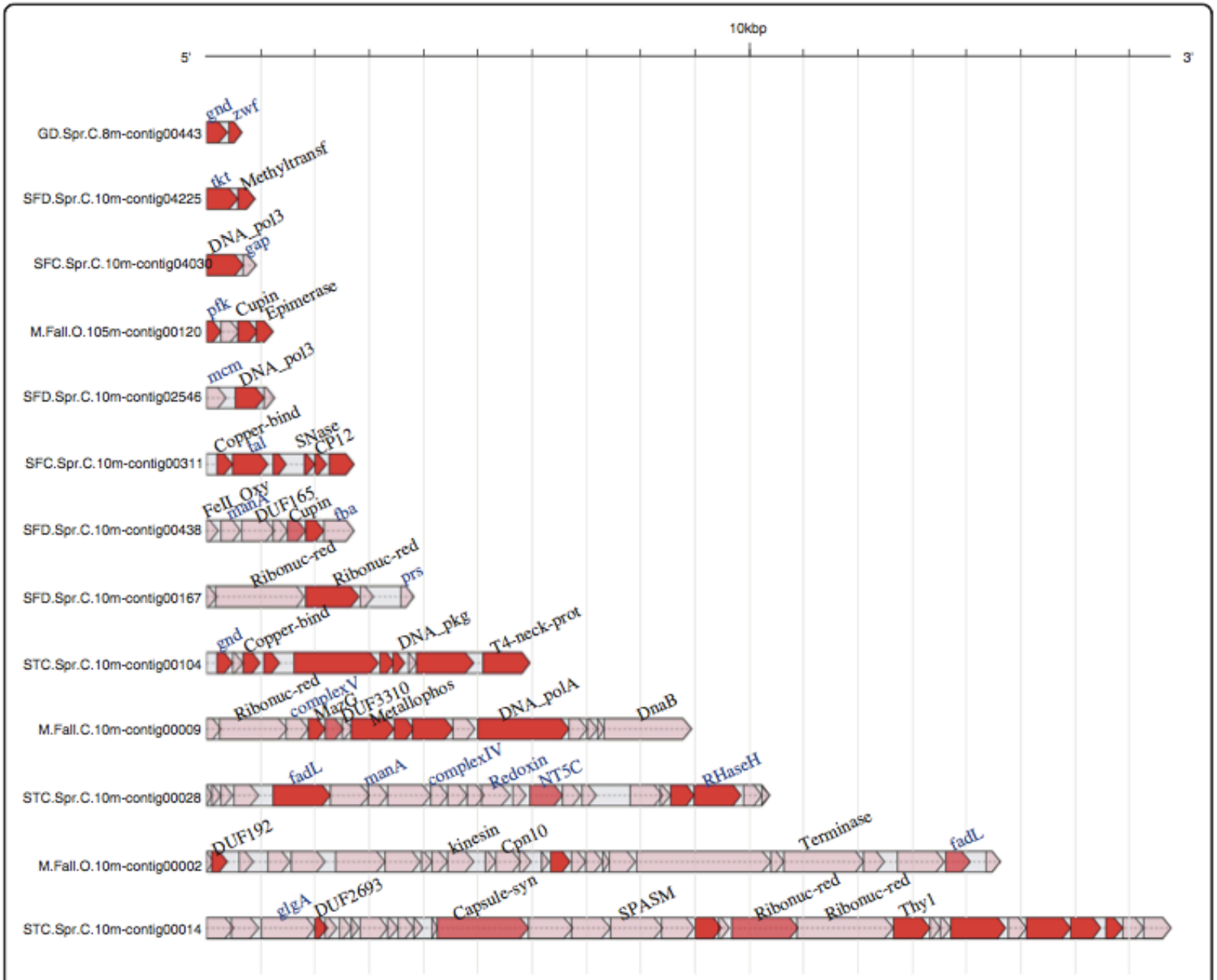



**Figure 2 Representative contigs containing carbon metabolism genes.** Example contigs containing carbon metabolism shown in blue, in context with other genes shown in black. Genes are colored based on superkingdom annotation: red, viral; light red, bacterial; pink, no superkingdom.

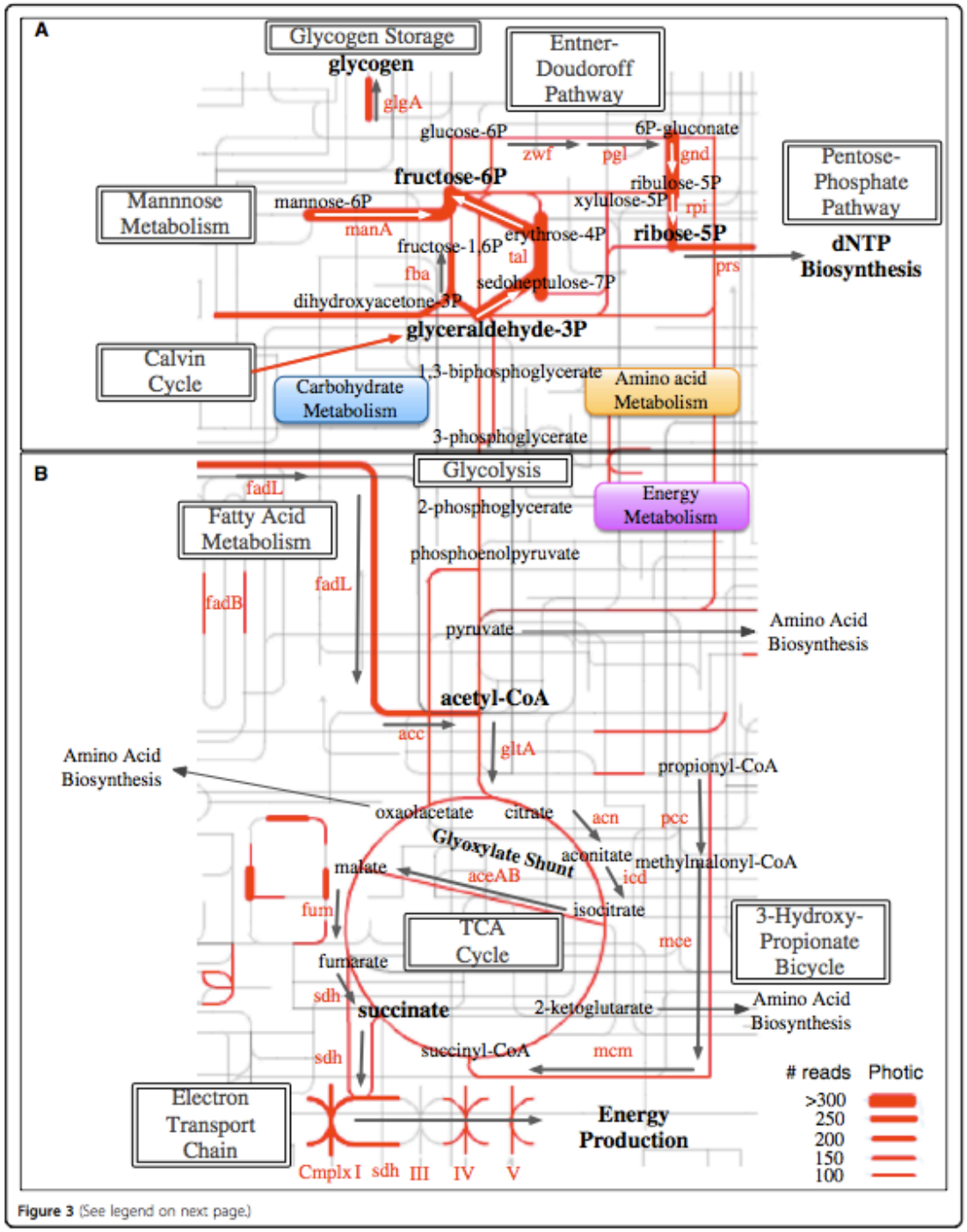



**Figure 3** (See legend on next page.)

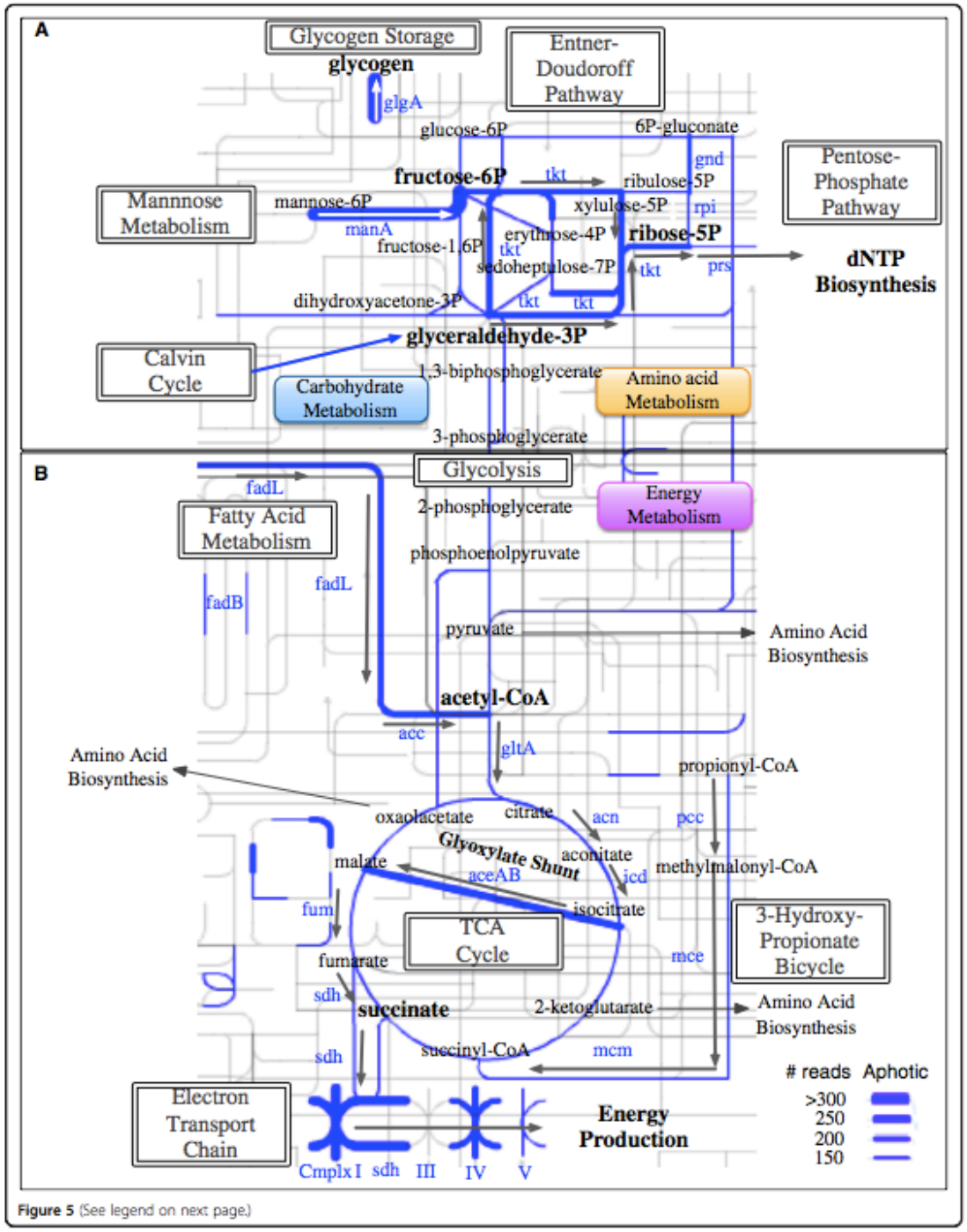
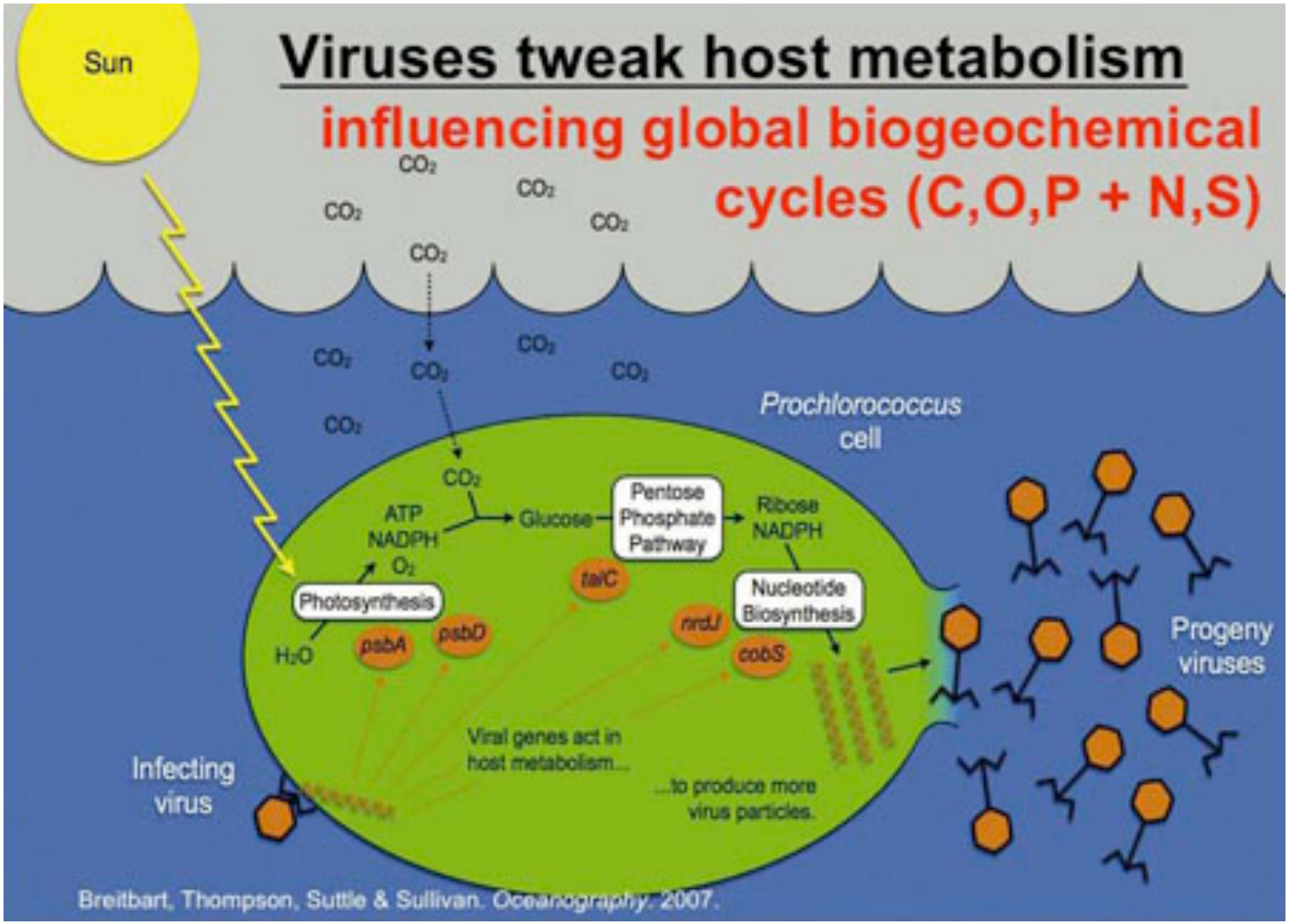



**Figure 5** (See legend on next page.)

**What are particular challenges associated with these method(s)?**

**What were the take-home results, and do folks believe them?**


Sun
Viruses tweak host metabolism
influencing global biogeochemical cycles (C,O,P + N,S)
CO2
Prochlorococcus cell
ATP
NADPH
O2
Glucose
Pentose Phosphate Pathway
Ribose
NADPH
Photosynthesis
H2O
psbA
psbD
talC
nrdJ
Nucleotide Biosynthesis
cobS
Progeny viruses
Infecting virus
Viral genes act in host metabolism...
...to produce more virus particles.
Breitbart, Thompson, Suttle & Sullivan. Oceanography. 2007.

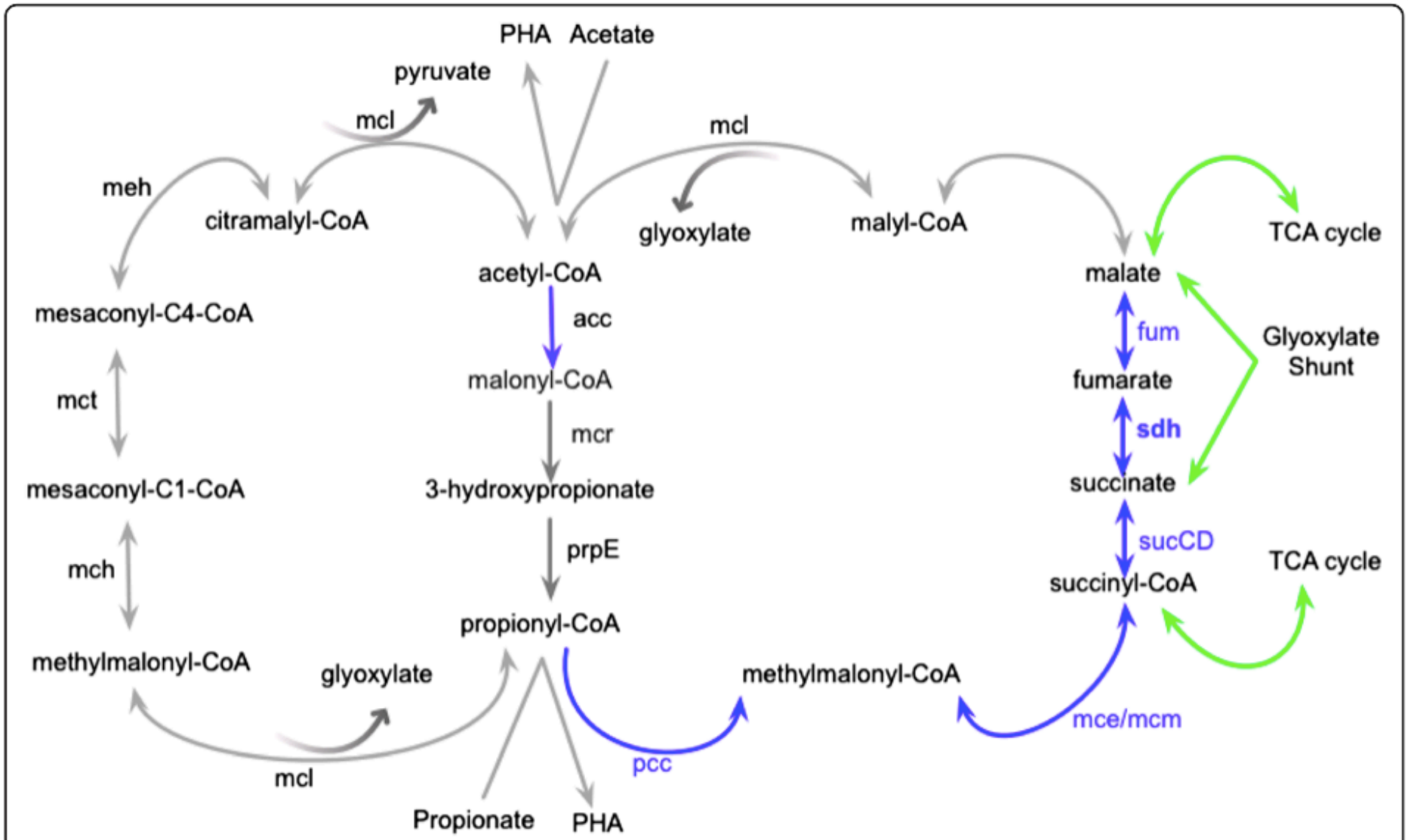



**Figure 4 Overview of Pacific Ocean Virome (POV)-encoded 3-hydroxypropionate Bicycle enzymes.** Enzyme names are listed as in Additional file 5: Table S3. The figure complements and highlights the pathways shown in Figures 3 and 5. The enzyme *acc* can also play a role in fatty acid metabolism.

**Table 1 Pacific Ocean Virome viral samples used in searching for carbon metabolism genes[a]**

| Sample | 16S + carbon metabolism genes in the same bacterial order | 16S hits mostly toa single bacterial species | 16S hits to many bacterial species | Random read recruitment to top bacterial genomes | Cellular contamination? | Reads, n |
|---|---|---|---|---|---|---|
| Aphotic | | | | | | |
| L.Win.O.1000m[b] | No | No | No | No | None | 147,537 |
| L.Win.O.2000m[b] | No | No | No | No | None | 125,896 |
| L.Spr.C.500m[b] | No | No | No | No | None | 136,876 |
| L.Spr.C.1300m[b] | No | No | No | No | None | 98,478 |
| L.Spr.I.1000m[b] | No | No | No | No | None | 122,565 |
| L.Spr.I.2000m[b] | No | No | No | No | None | 49,914 |
| L.Sum.O.1000m[b] | No | No | No | No | None | 70,596 |
| L.Spr.O.1000m[b] | No | No | No | No | None | 101,179 |
| L.Win.O.500m[b] | No | No | No | No | None | 167,616 |
| M.Fall.O.4300m[b] | No | No | No | No | None | 144,588 |
| L.Spr.C.1000m | Yes | Yes | No | Yes | High GTA[c] | 97,126 |
| L.Spr.I.500m | Yes | Yes | No | Yes | Low GTA[d] | 58,108 |
| L.Sum.O.500m | Yes | Yes | No | No | Low GTA[d] | 42,118 |
| L.Sum.O.2000m | Yes | No | No | No | Low GTA[d] | 68,516 |
| L.Spr.O.2000m | No | No | Yes | No | Low sporadic | 55,332 |
| M.Fall.O.1000m | No | No | No | Yes | Low GTA[d] | 225,833 |
| Photic | | | | | | |
| L.Sum.O.10m[b] | No | No | No | No | None | 165,256 |
| L.Spr.C.10m[b] | No | No | No | No | None | 107,244 |
| L.Spr.I.10m[b] | No | No | No | No | None | 92,415 |
| L.Win.O.10m[b] | No | No | No | No | None | 192,685 |
| M.Fall.C.10m[b] | No | No | No | No | None | 303,519 |
| M.Fall.I.10m[b] | No | No | No | No | None | 321,754 |
| M.Fall.O.10m[b] | No | No | No | No | None | 203,238 |
| M.Fall.O.105m[b] | No | No | No | No | None | 156,509 |
| SFC.Spr.C.5m[b] | No | No | No | No | None | 487,339 |
| SFD.Spr.C.5m[b] | No | No | No | No | None | 645,463 |
| SFS.Spr.C.5m[b] | No | No | No | No | None | 504,826 |
| STC.Spr.C.5m[b] | No | No | No | No | None | 821,404 |
| GD.Spr.C.8m | Yes | Yes | No | Yes | Low GTA[d] | 116,855 |
| GF.Spr.C.9m | Yes | Yes | No | Yes | Low GTA[d] | 82,739 |
| L.Spr.O.10m | Yes | Yes | No | Yes | Low GTA[d] | 75,036 |
| M.Fall.I.42m | Yes | Yes | No | Yes | Low GTA[d] | 31,528 |

[a]Samples were evaluated for gene transfer agents (GTAs) and sporadic contamination (see Results, section on 'Ruling out bacterial contamination') based on 16S analysis.
[b]Only samples indicated as having no contamination were used in the analysis.
[c,d]High GTAs are described as viromes with greater than 100 hits to a single bacterial species, low GTAs have less than 50 hits as determined by analysis (see Additional file 2: Figure S2).
[b]Sample metadata are further described by Hurwitz and Sullivan [21].