Journal of Computer Science & Systems Biology - Open Access

Research Article
OPEN ACCESS Freely available online
doi:10.4172/jcsb.1000065

JCSB/Vol.3 Issue 5

# Modularity and Distribution of Sulfur Metabolism Genes in Bacterial Populations: Search and Design

**Andrew Kuznetsov**

*University of Freiburg, Germany; Institute of Biology of the Southern Seas, Ukraine*

## Abstract

Biological Engineering involves global DNA sampling and modular design from genetic parts. A new approach reflected by natural history is based on the recognition of interchangeable DNA fragments that move around the world due to horizontal gene transfer. According to the large scale, metagenomics provide opportunities to sequence whole genomes within environmental populations. Annotated gene sequences, protein structures, and metabolic data can be used to design small biosystems from interchangeable genetic parts, the same as from functional modules.

To illustrate this, the 21 genes for sulfur metabolism were inferred from the genome of bacterium *Vesicomyosocius okutanii* HA, and the distribution of two gene clusters (dissimilatory sulfite reductase – *dsr* and sulfur-oxidation – *sox*) within environmental samples was investigated. The correlation between the *dsr* and *sox* clusters for the experimental set of 41 stations was R = 0.86 which demonstrates the complementarity of *dsr* and *sox* metabolic pathways in environmental populations. Hypothetical functions were assigned using comparisons with known proteins. The 18 reads from symbionts of gutless worm *Olavius algarvensis* showed a high identity to large AprA protein from *V.okutanii*. In addition, comparative 3D modeling of hypothetical DsrB protein revealed sulfite reductase ferredoxin-like half domain, sulfite reductase 4Fe-4S domain, and a repressor of phase-1 flagellin.

The simplistic reconstruction of sulfur metabolism from parts and examples of hierarchical modularity in nature are given. The origin of modularity is considered in the context of minimal cell and horizontal gene transfer. The role of ancient sulfur metabolism in modularization is discussed under the umbrella of iron-sulfur world theory (Wächtershäuser, 1988), deep-hot biosphere model (Gold, 1992), and radiolysis hypothesis (Garzón and Garzón, 2001). The reverse engineering approach based on natural genetic modules is proposed for understanding early life.

**Keywords:** Constructive biology; Sulfur metabolism; Metagenomics; Gene geography; Horizontal gene transfer; Minimal genome; Origin of life; Compositional evolution

## Introduction

Instead of small gradual changes like point mutations, the mechanisms of compositional evolution combine interdependent genetic modules that have evolved previously in parallel (Watson, 2006). Examples of compositional mechanisms in nature include hybridization (Rieseberg et al., 2003), horizontal gene transfer (HGT) (Doolittle, 1980; Jain et al., 2002), and symbiotic encapsulation (Margulis, 1970; Merezhkovsky, 1909), as exhibited in the history of major evolutionary transitions (Maynard Smith and Szathmary, 1995). Both gradual and compositional mechanisms of biological evolution are mediated by natural selection (Darwin, 1859). On the other hand, progress in the development of complex man made systems, like software products, has demonstrated a benefit of modular design that sometimes leads to an ignorance of evolution. In this light, the search for natural building blocks in DNA records and attempts to design artificial biological systems from those complementary modules might help us appreciate the beginning of life and the dynamics of evolutionary progress up to the modern life forms. As Richard Watson wrote: 'the existence of modularity in nature is now becoming testable' (Watson, 2006). To prove this statement, metagenomics deals with the global gene set where each genome is studied as part of a biological community and DNA sequences are analyzed from the viewpoint of ecology.

Let us consider sulfur metabolism in a symbiotic community as an inspiration for a 'protocell'. This approach will help us to recognize modules in complex systems. For instance, hemoglobin in the giant tube worm *Riftia pachyptila* binds hydrogen sulfide ($H_2S$) which allows the worm to reduce toxicity and transmit the sulfide into symbiotic bacteria which supply the host with nutrition (Flores et al., 2005; Chabasse et al., 2006). In general, reduced inorganic sulfur compounds can be used as electron donors. Oxidation of sulfide to sulfate takes place in the bacterial cytoplasm by using dissimilatory sulfite reductase (DsrAB, EC 1.8.1.2) operating in reverse, adenosine-5'-phosphosulfate reductase (APS reductase, EC 1.8.99.2) and ATP sulfurylase (EC 2.7.7.4) (Pott and Dahl, 1998). Genes involved in oxidative sulfur metabolism and $CO_2$ fixation via the Calvin-Benson-Bassham cycle were characterized in some metagenomics projects (Beller et al., 2006; Scott et al., 2006; Kuwahara et al., 2007). These genes are mainly localized in clusters. The large *dsr* cluster encodes the dissimilatory sulfide oxidation and provides the assembly of electron transport complex III (Dahl et al., 2005; Pires et al., 2006). The *sox* genes cluster is used to oxidize thiosulfate taking place in the periplasmic space (Friedrich et al., 2001). Some genes encoding sulfide quinone oxidoreductase, rhodenase, and ATP sulfurylase are under peculiar transcriptional control (Beller et al., 2006). Sulfate, a final product of the sulfur oxidation under anoxic conditions, is

**Corresponding author:** Andrew Kuznetsov, University of Freiburg, Germany; Institute of Biology of the Southern Seas, Ukraine, E-mail: andrei_kouznetsov@hotmail.com

**Received** December 02, 2009; **Accepted** November 26, 2010; **Published** November 28, 2010

**Citation:** Kuznetsov A (2010) Modularity and Distribution of Sulfur Metabolism Genes in Bacterial Populations: Search and Design. J Comput Sci Syst Biol 3: 091-106. doi:10.4172/jcsb.1000065

pumped out of the symbiotic cell by a sulfate permease. For example, see Figures 2 and 3A.

The aim of this paper is to map proteins involved in sulfur metabolism within various biosystems. Molecular reconstructions are tested for signs of modularity. The article poses a problem of origin of modules in biological systems and particularly addresses the compatibility of modules due to the lateral gene transfer. The methodology combines the analysis of environmental databases and the design of metabolic pathways. The scheme includes a transport of hydrogen sulfide, accumulation of polysulfide, sulfide, and thiosulfate oxidation into sulfate, and finally a sulfate export. Produced electrons will be used to ATP generation via a respiratory electron-transport chain. This 'energetic module' will be installed on an imaginable 'minimal cell' with a set of housekeeping genes (Gabaldón et al., 2007; Gibson et al., 2008; Gil et al., 2004; Glass et al., 2006; Koonin, 2003). The gene assembling includes *sox* and *dsr* clusters from the sulfur oxidizing bacterium *Candidatus Vesicomyosocius okutanii* HA with a small completely sequenced genome (~1 Mb), which is a symbiont of a deep-sea clam *Calyptogena okutanii* (Kuwahara et al., 2007). The artificial system assumes a particular advantage for symbiotic bacteria and their host. Some ideas for this model were retrieved from (Dahl et al., 2005; Pires et al., 2006), even though the symbiotic relationships might be more complex (Arndt et al., 2001). Preliminary data from global DNA sampling and their analysis, as well as common ideas concerning biological design and extension of life based on natural interchangeable modules have been considered.

## Materials and Methods

Databases, such as the National Center for Biotechnology Information – NCBI (Benson et al., 1998), Expert Protein Analysis System – ExPASy (Appel et al., 1994; Gasteiger et al., 2003), Kyoto Encyclopedia of Genes and Genomes – KEGG (Kanehisa and Goto, 2000) and Integrated Genomics – ERGO (Overbeek et al., 2003), were used to find proteins and gene assemblies in order to infer sulfide oxidation metabolic networks that may be small enough for optimization, synthesis, and experimental evaluation. Data obtained were compared with environmental samples available from Community Cyberinfrastructure for Advanced Marine Microbial Ecology Research and Analysis – CAMERA (Rusch et al., 2007). To be clear, CAMERA is a metagenomic data set with tools for querying the databases and for visualization. For instance, the 'Sorcerer II' Global Ocean Sampling (GOS) expedition puts together microbial communities every 200 miles around the globe. In addition, CAMERA v1.3.2 contains a vertical profile and deep-water samples of marine microbial communities collected in Hawaii and in the eastern Mediterranean, 150 genomes of ocean microbes, symbionts of Mediterranean gutless oligochaete *Olavius algarvensis* and hydrothermal vent polychaete *Alvinella pompejana*, as well as other collections including samples from the acid mine drainage biofilm and rich farm soil (Seshadri et al., 2007). These sequence data were scanned for genes coding sulfur metabolism.

To reduce the complexity of a system and to model dynamics of biochemical networks, the JDesigner v2.1 simulator was used (see link to manual in references). The networks were described by first order differential equations (ODE) with arbitrary parameters allowing a qualitative interpretation. CAMERA Fragment Recruitment Viewer enabled the comparison of complete bacterial genomes with environmental databases (Rusch et al., 2007). A family of BLAST programs was applied for quick searches in large databases (Altschul et al., 1997). The significance of hits was estimated by a bit score and by an expectation value. Bit score is a measure of similarity between the hit and the query that is derived from a raw score by normalization, the E-value revealing how many times an expected result could occur by chance. The higher the score, the better the alignment; the lower the E-value, the more significant the score. Strong values for a bit score near 400 and the expectation below $1e^{-100}$ were preferred during alignments and superior querying. E-value threshold, larger than 1.0, was biologically relevant against environmental DNA datasets in preliminary scans. T-COFFEE v5.68-7.71 servers were used for the automatic multiple sequence alignments and an annotation (Notredame et al., 2000); the improved visualization was with BOXSHADE v3.21 utility. Empirical evaluation of the quality of multiple sequence alignment was prepared within regular parameters; residues with a yellow to red background are correctly aligned more than 80%. Protein domain identification was carried out within InterProScan (Mulder et al., 2007) and PFAM (Bateman et al., 2000; Bateman et al., 2002) databases. The comparative protein structure analysis was done with default parameters by three server machines which complemented each other – ESyPred3D (Lambert et al., 2002), 3D-JIGSAW v2.0 (Bates et al., 2001) and Geno3D v2.0 (Combet et al., 2002). Protein visualization was made by an SPDB viewer v4.0 (Guex and Peitsch, 1997). Links to the online servers are mentioned in the references. Some techniques such as the normalization and correlation analysis are described in results.

| Product Name | Start | End | Strand | Length | Gene ID | Locus |
|---|---|---|---|---|---|---|
| ATP sulfurylase | 98093 | 99301 | + | 402 | 5172354 | *sat* |
| adenylylsulfate reductase membrane anchor | 99516 | 100385 | + | 289 | 5171802 | *aprM* |
| adenylylsulfate reductase β-subunit | 100417 | 100896 | + | 159 | 5172128 | *aprB* |
| adenylylsulfate reductase | 100896 | 102779 | + | 627 | 5172122 | *aprA* |
| sulfur oxidation protein SoxB | 172596 | 174485 | + | 629 | 5172731 | *soxB* |
| sulfur oxidation protein SoxA | 770792 | 771607 | - | 271 | 5172420 | *soxA* |
| sulfur oxidation protein SoxZ | 771635 | 771937 | - | 100 | 5171932 | *soxZ* |
| sulfur oxidation protein SoxY | 771971 | 772414 | - | 147 | 5172000 | *soxY* |
| sulfur oxidation protein SoxX | 772425 | 772772 | - | 115 | 5171986 | *soxX* |
| intracellular sulfur oxidation protein DsrR | 817196 | 817537 | - | 113 | 5172350 | *dsrR* |
| intracellular sulfur oxidation protein DsrP | 818938 | 820140 | - | 400 | 5172402 | *dsrP* |
| intracellular sulfur oxidation protein DsrO | 820166 | 820897 | - | 243 | 5172419 | *dsrO* |
| intracellular sulfur oxidation protein DsrJ | 820894 | 821277 | - | 127 | 5172414 | *dsrJ* |
| putative glutamate synthase (NADPH) small subunit | 821307 | 823271 | - | 654 | 5172337 | *dsrL* |
| intracellular sulfur oxidation protein DsrK | 823327 | 824892 | - | 521 | 5172342 | *dsrK* |
| intracellular sulfur oxidation protein DsrM | 824894 | 825667 | - | 257 | 5172339 | *dsrM* |
| intracellular sulfur oxidation protein DsrC | 825744 | 826067 | - | 107 | 5172365 | *dsrC* |
| intracellular sulfur oxidation protein DsrB | 827224 | 828297 | - | 357 | 5172320 | *dsrB* |
| intracellular sulfur oxidation protein DsrA | 828373 | 829674 | - | 433 | 5172315 | *dsrA* |
| rhodanese family protein | 950273 | 950752 | - | 159 | 5171848 | COSY_0913 |
| sulfide-quinone reductase | 995954 | 997240 | + | 428 | 5172159 | *sqr* |
| **SUM** | 19656 | | | 6538 | 21 | |

**Table 1: Genes coding sulfur metabolism in the bacterium *Vesicomyosocius okutanii* HA.**

Journal of Computer Science & Systems Biology - Open Access

Research Article
OPEN ACCESS Freely available online
doi:10.4172/jcsb.1000065

JCSB/Vol.3 Issue 5

The resume of data available from CAMERA v1.3.2 used in experiments is as follows. The Hawaii Ocean Time Series (HOT) station collected genomes down to 4000 m (DeLong et al., 2006). The other metagenomics study was in the Ionian abyssal plain, a deep flat basin between Sicily and Greece in the Eastern Mediterranean, whose deep waters are free from an intrusion of cool polar waters that feed the bottom of the World Ocean. The next sample is the deep-sea hydrothermal vent polychaete worm, *Alvinella pompejana*, which maintains diverse microbial symbionts. The microbial community was isolated from the dorsal integument of *A.pompejana* collected from the East Pacific Rise at a depth of 2500 m. Another worm, Mediterranean gutless oligochaete *Olavius algarvensis*, lives with four microbial symbionts having different metabolic pathways. The symbiotic bacterial consortium demonstrates mutualistic relationships with their host and with each other (Woyke et al., 2006). The purpose of the Acid Mine Drainage biofilm sequencing project was to investigate the diversity of metabolic pathways in the microbial communities at Iron Mountain, California, i.e. sulfur oxidation, nitrogen fixation, and iron oxidation. The sample obtained from the underground was localized within a pyrite ($FeS_2$) ore body. This biogeochemical system provided a coupling between microbial iron oxidation and acidification due to pyrite dissolution (Lo et al., 2007; Tringe et al., 2005). Waseca County is well-known for rich, black soil that produces record crops every year. Surface soil (0-10 cm) was collected from a farm in Waseca County, Minnesota. Microscopic analyses revealed the presence of various prokaryotic organisms in the sample. PCR-amplified 16S rRNA sequences from the sample confirmed numerous bacterial and archaeal lineages (Tringe et al., 2005). Additional information about samples is available from the CAMERA web page.

## Results

### Comparison of four thiotrophic bacterial genomes with CAMERA global databases

Initially, I needed to have a suitable thiotrophic reference to scan global databases for sulfur genes. The CAMERA Fragment Requirement Viewer was used to compare complete genomes of two thiobacteria and two chemoautotrophic endosymbionts with subsequent databases: GOS – All Metagenomic Sequence Reads (N), Metagenome of Marine NaCl-Saturated Brine and Microbial Community Genomics at the HOT/ALOHA (Figure 1). Freeliving microorganisms such as *Thiomicrospira crunogena* XCL-2 and *Thiobacillus denitrificans* ATCC 25259 demonstrated multiple homologies with different stations (depicted in color). Spikes with more than 90% identity corresponded to 5S, 16S, and 23S ribosomal RNA, various tRNAs, and housekeeping genes that are common in many bacteria. *T.crunogena* with a genome of 2 427 674 nt in length presented a maximal number of matching sequences (63 134) which is difficult to study. A large set of *T.denitrificans* genes showed about 60-85% similarity to the microbiota at Sargasso Station 11 (Figure 1B, cyan). In contrast, symbiotic bacteria *Candidatus Vesicomyosocius okutanii* HA and *Candidatus Ruthia magnifica* Cm represented less complex patterns of homology allowing analysis and reasonable interpretations (Figure 1C,D). Bacterium *V.okutanii* was chosen as a reference for future research.

### Gene collection

Efficient browsing of NCBI, KEGG, and ERGO databases inferred metabolic pathways for a 'minimal' sulfur metabolism within bacterium *Vesicomyosocius okutanii* HA. This symbiotic microorganism has a small single chromosome (1 022 154 nt) and is annotated (Kuwahara et al., 2007). As a result of the search, the 21 genes comprising 19 656 nt from *V.okutanii* were assigned to the hydrogen sulfide oxidation into sulfate (Table 1). The bacterium *V.okutanii* demonstrated a very compact organization of *sox* and

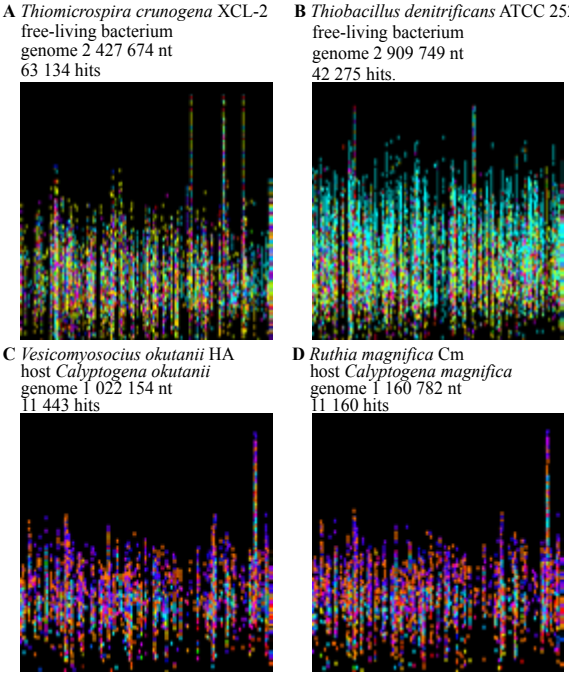



**Figure 1: Fragment requirement plots for four thiobacteria in CAMERA Viewer.** Abscissa is the genome length, ordinate is the 50 to 100% homology; comparisons with the databases, such as All Metagenomic Sequence Reads (N) – Global Ocean Sampling Expedition, Metagenome of Marine NaCl-Saturated Brine, and Microbial Community Genomics at the HOT/ALOHA; the visualization of stations is done by inverted colors.

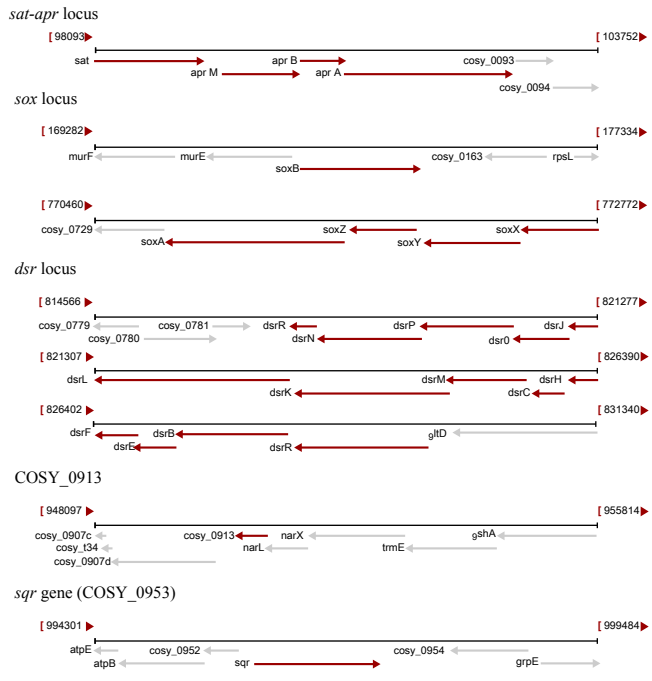
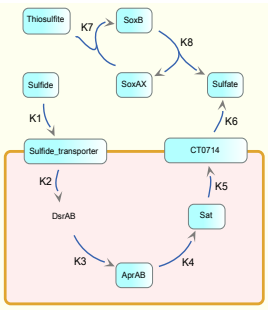



**Figure 2: Fragments of *V.okutanii* HA genetic map with genes coding sulfur metabolism.**

A




B

| | | |
|---|---|---|
| Input | | Sulfide = 10<br>Thiosulfite = 10 |
| Output | | Sulfate = 0 |
| Participants | in | Sulfide transporter = 0<br>DsrAB = 0<br>AprAB = 0<br>Sat = 0<br>CT0714 = 0 |
| | out | SoxAX = 1<br>SoxB = 1 |
| Parameters | | $k_{1,3\text{-}8} = 0.1$<br>$k_2 = 0.1; 0.01$ |

**Figure 3: A simplified model of the bacterial sulfur metabolism. A** – network, **B** – initial conditions and parameters of the ODE model.

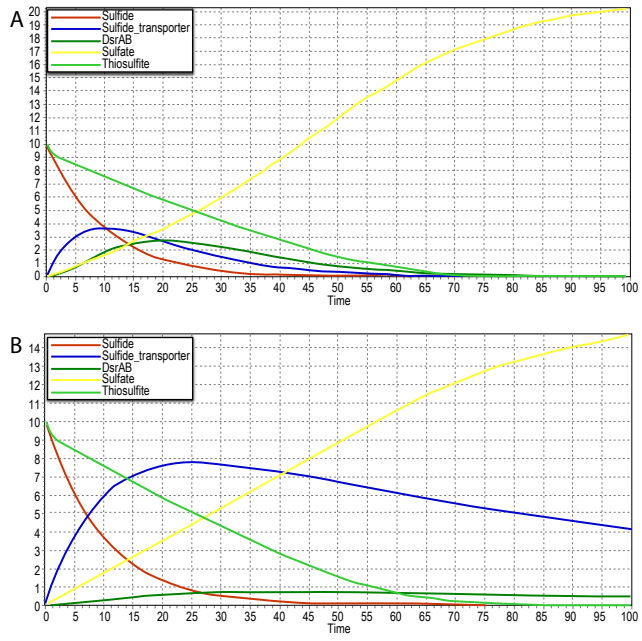



**Figure 4: Role of hydrogen sulfide transport in sulfate production. A** – strong transport, $k_2 = 0.1$; **B** – weak transport, $k_2 = 0.01$; the sulfate output is yellow.

*dsr* genes clusters. The architecture of *sox* and *dsr* clusters, as well as *sqr* and *apr* genes is represented in Figure 2. To complete the metabolic system, some additional genes from *Chlorobium tepidum*, *Paracoccus pantotropus,* and *Riftia pachyptila* are suggested below. Genes and their products are described according to the metabolic fluxes (see appendix for details):

(i) *sox* sulfur-oxidizing multienzyme system (*V.okutanii soxXYZA* and *soxB* genes),
(ii) sulfide-quinone oxidoreductase (*V.okutanii sqr* gene, EC 1.8.5.-),
(iii) rhodenase (*V.okutanii* COSY_0913 gene, EC 2.8.1.1),
(iv) dissimilatory-siroheme-sulfite-reductase complex including the electron-transport complex III (*V.okutanii dsr* genes),
(v) APS reductase (*V.okutanii aprABM* genes) and ATP sulfurylase (*V.okutanii sat* gene),
(vi) putative sulfate transporter (*C.tepidum* CT0714 gene).

It should be mentioned that *Vesicomyosocius okutanii* HA lost *soxC* and *soxD* genes from the genome which may have increased the electron flow in *Paracoccus pantotropus* GB17. These missing genes could have been accepted from *P.pantotropus.* The haemoglobin from the deep-sea tube worm *Riftia pachyptila* is also an attractive subject for engineering because of the significant role of hydrogen sulfide transport through its body to symbiotic cells. Why would an artificial symbiotic bacterium not produce a secreted form of vestimentiferan hemoglobin in order to support itself in the host?

## Modeling of hydrogen sulfide oxidation within JDesigner

Even the small genome of bacterium *V.okutanii* demonstrates a great complexity of sulfur metabolism; chemical reactions run stepwise and parallel and involve many different enzymes. The JDesigner simulator was used to explore connections between main nodes in the biochemical network and to model metabolic fluxes via the ordinary differential equations. The nodes represented enzymes; the arcs denoted streams of substrates and their products (Figure 3A).

Now let us look at a simplified sulfur oxidation system where hydrogen sulfide and thiosulfite are inputs and the sulfate is the output. Enzymes SoxAX and SoxB convert thiosulfite into sulfate in the periplasm. A hypothetical sulfide transporter pumps $HS^-$ ions into cells. Enzymes DsrAB, AprAB and Sat transform the sulfide into sulfate in the cytoplasm. CT0714 permease releases the final product from the cell. The parameters were chosen as shown in Figure 3B. Initial values of sulfide and thiosulfite were high (=10) opposite the low range of sulfate (=0). The same low rations were assigned to cytoplasmic components, such as sulfide transporter, DsrAB, AprAB, Sat, and CT0714. Periplasmic proteins SoxAX and SoxB have intermediate initial values (=1). All coefficients beyond $k_2$ are equal to 0.1. The coefficient $k_2$ is a feature of a flux between the sulfide transporter and DsrAB complex. Accordingly, the parameter $k_2$ with values 0.1 and 0.01 describe the intensive and slow transport of hydrogen sulfide into the system.

The time course simulation revealed the role of sulfide transport in the sulfate production. The system reached a plateau for a new high-output stationary state at $k_2 = 0.1$ very fast (Figure 4A). In contrast, the simulation of the weak transport at $k_2 = 0.01$ showed a slow growth of output sulfate ratio, i.e. about 14 units versus 20 units at the end of the simulation (Figure 4B). The results support the assumption that the import of nutrition may be a limiting factor for the symbiotic metabolism. Multiple inputs would have led to a more reliable survival of organisms.

## Worldwide distribution of *sox* and *dsr* genes clusters

Enzymes of sulfur metabolic network coded by *sox* cluster (5 genes) and *dsr* cluster (14 genes) from *V.okutanii* were selected for a future search in seven metagenomic databases; (1) 'Sorcerer II' Global Ocean Sampling (GOS) collection, (2) bacterial consortium from the gutless worm *Olavius algarvensis*, (3) symbionts of polychaete *Alvinella pompejana*, (4) probes from the Deep Mediterranean, Ionian 3 Km station, (5) data from Hawaii ocean station ALOHA, 130 m, (6) farm soil microbial community, Waseca County, and (7) samples from the Richmond acid mine. To compare polypeptide sequences with DNA targets, the TBLASTN program was used. Queries were formulated for each protein. CAMERA services looked for stations with the maximal number of hits and best matching sequences. Identities to *sox* and *dsr* strings were found at 41 stations (Figure 5A). The strong positive pair correlation R = 0.86 between *sox* and *dsr* genes distribution in these samples was revealed. The absolute number of hits was higher for *dsr* cluster than for *sox* cluster due to the prevailing number of genes in the *dsr* cluster (14 vs. 5). This inconvenience was eliminated by normalization – the number of hits was divided in each sample by

Research Article
OPEN ACCESS Freely available online
doi:10.4172/jcsb.1000065
JCSB/Vol.3 Issue 5

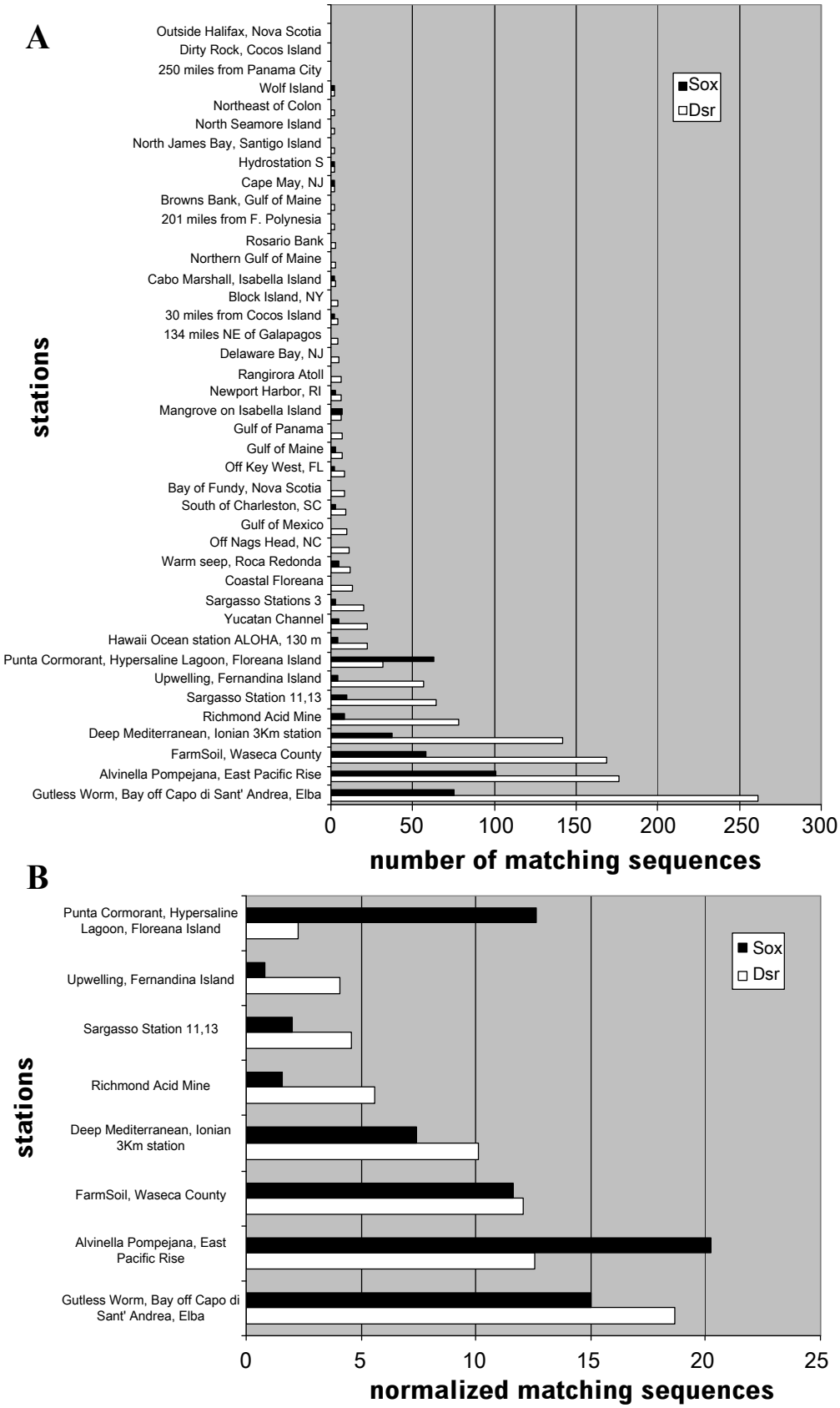



**Figure 5: Distribution of *sox* and *dsr* genes by stations**, **A** – all positives, **B** – best stations

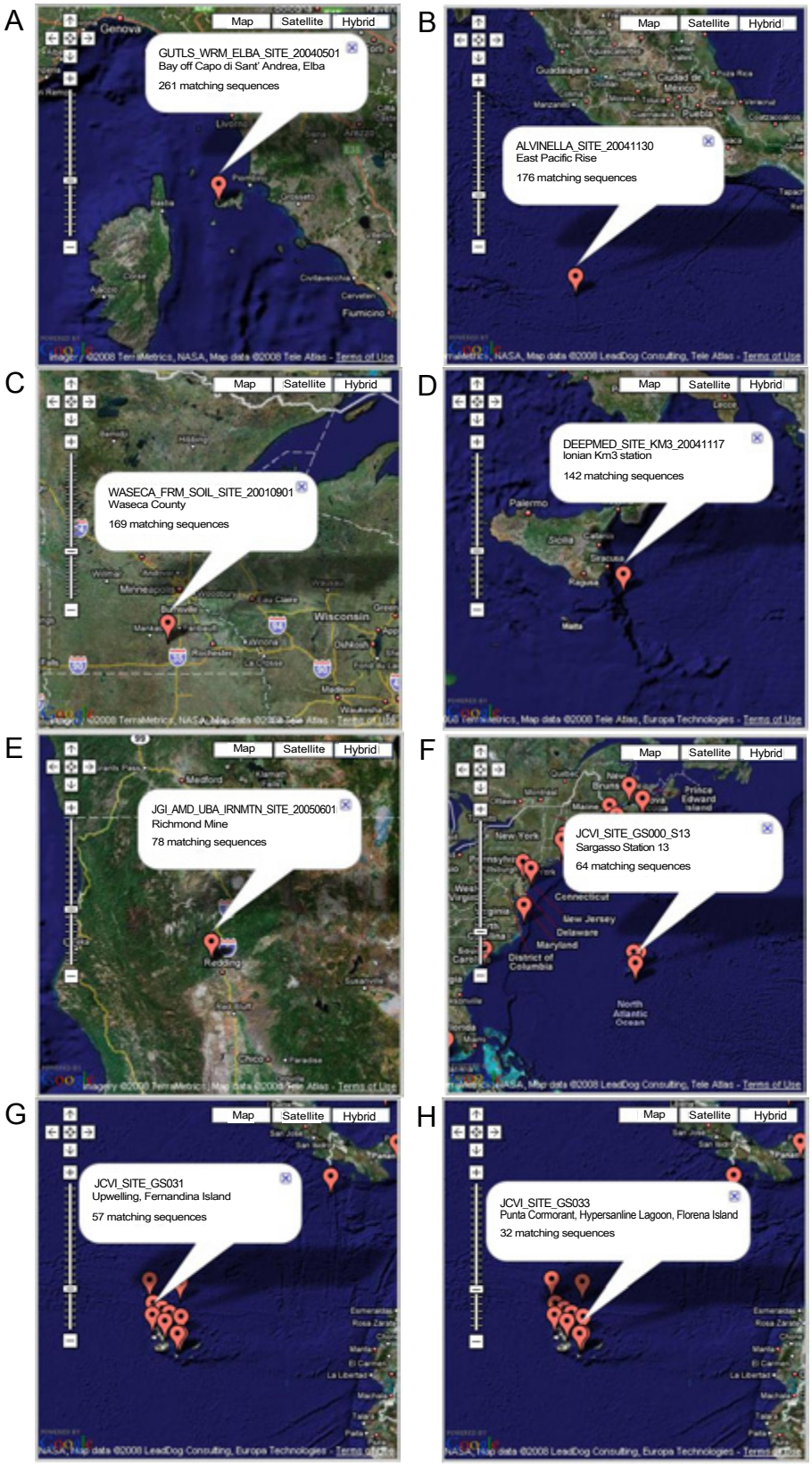



**Figure 6: Location of experimental stations with a plenty of Dsr matching sequences. A** – Gutless Worm, Bay of Capo di Sant' Andrea, Elba; **B** – Alvinella Pompejana, East Pacific Rise; **C** – Farm Soil, Waseca County; **D** – Deep Mediterranean, Ionian 3 Km station; **E** – Richmond Acid Mine; **F** – Sargasso Station 13; **G** – Upwelling, Fernandina Island, Galapagos; **H** – Punta Cormorant, Hypersaline Lagoon, Floreana Island, Galapagos.

the number of genes in a cluster (Figure 5B). The normalization did not change the pattern of distribution on the whole. The number of hits in the examined data decreased exponentially, station by station, and only a small fraction of samples presented a sufficient content of genes for sulfur metabolism.

Then the eight best stations from the initial 41 stations were chosen for a more detailed analysis: (1) Gutless Worm, Bay of Capo di Sant' Andrea, Elba, (2) Alvinella Pompejana, East Pacific Rise, (3) Farm Soil, Waseca County, (4) Deep Mediterranean, Ionian 3 Km station, (5) Richmond Acid Mine, (6) Sargasso See Station 11, 13, (7) Upwelling, Fernandina Galapagos Island, (8) Punta Cormorant Hypersaline Lagoon, Floreana Island, and Galapagos Archipelago. This final collection included microorganisms from different niches such as host environments (cases 1 and 2 with a maximal number of hits), rich farm soil, deep-see anoxic water, acid mine drainage, epipelagial zone, upwelling zone, and hypersaline lagoon (Figure 6). Microbes from these diverse environments involve similar genes in sulfur metabolism. Nevertheless, the distribution of *sox* and *dsr* genes was not the same for all samples. The normalization revealed disproportions in *sox* and *dsr* gene allocation for 8 preeminent stations (Figure 5B). Bacterial symbionts of Mediterranean gutless worm *Olavius algarvensis* from the Bay of Capo di Sant' Andrea and microbiota from the deep eastern Mediterranean demonstrated a privilege for *dsr* genes in samples. In contrast, symbionts from the hydrothermal vent polychaete *Alvinella pompejana* utilize mostly *sox* genes. Microbes from the farm soil in Waseca County presented both *dsr* and *sox* metabolic pathways. The content of *dsr* and *sox* genes decreased sufficiently at the next stations. The *dsr* genes dominated in microbiota from Richmond acid mine, Sargasso See station 11, 13 and the upwelling at Fernandina Island. Amazingly, the microbes living in the Punta Cormorant hypersaline lagoon show an increased role of *sox* genes for sulfur metabolism in this extreme environment.

These facts show an obvious difference between *sox* and *dsr* genes clusters. Ecological distributions of those genes appeared depending on environmental conditions. The geographical location of genomes revealed an importance of the periplasmic *sox* component in the oxidative sulfur metabolism for particular environments like the hypersaline lagoon.

## Investigation of individual genes and proteins involved in sulfur metabolism

In addition to Sox and Dsr proteins, I compared the S-quinone, AprA and Sat proteins from bacterium *V.okutanii* with environmental datasets by CAMERA BLAST tools. The Gutless Worm and GOS (P) databases were used in the experiments. Values of identities with the top score were chosen to represent results. Two heterologous proteins, such as CT0714 from *C.tepidum* and Tcr_0602 (soxZ) from *T.crunogena*, were considered as a negative control in the correlation analysis Table 2. I observed a high positive linear relationship between gutless worm *O.algarvensis* bacterial symbiotic consortium and GOS (P) microbial databases with the correlation coefficient R = 0.85 for Dsr proteins and R = 0.83 for Sox proteins. The correlation was smaller (R = 0.67) in the third group – S-quinone, AprA, and Sat proteins. The control group – CT0714, Tcr_0602 – illustrated a negative correlation. Dsr proteins in the line from DsrA to DsrL came across with more than 55% identity whereas the proteins DsrB, DsrE, and DsrC demonstrated higher than 70% identity for both databases. The rest of the group, polypeptides from DsrJ to DsrR, had low identities. The DsrR polypeptide displayed minimal identities 35% and 20% for both databases. Only SoxX protein from the *sox* genes cluster represented about 70% identity for two target datasets. Surprisingly, AprA polypeptide from the third protein group showed maximal identities 85% and 75% for the gutless worm and GOS (P), respectively. Control proteins CT0714 and Tcr_0602 presented relatively small identities.

Data revealed a high similarity between *V.okutanii* and queried

| Protein | Identities, % | | Protein | Identities, % | |
|---|---|---|---|---|---|
| | Gutless Worm | GOS (P) | | Gutless Worm | GOS (P) |
| DsrA | 56 | 61 | **SoxX** | 69 | 73 |
| **DsrB** | 72 | 75 | SoxY | 44 | 62 |
| **DsrE** | 72 | 72 | SoxZ | 41 | 44 |
| DsrF | 61 | 67 | SoxA | 43 | 42 |
| DsrH | 42 | 46 | SoxB | 49 | 56 |
| **DsrC** | 73 | 76 | *correlation* | *0.83* | |
| DsrM | 47 | 54 | S-quinone | 54 | 59 |
| DsrK | 64 | 64 | **AprA** | 85 | 75 |
| DsrL | 57 | 77 | Sat | 55 | 72 |
| DsrJ | 42 | 45 | *correlation* | *0.67* | |
| DsrO | 52 | 50 | | | |
| DsrP | 52 | 49 | CT0714 | 35 | 57 |
| DsrN | 43 | 59 | Tcr_0602 | 45 | 53 |
| DsrR | 35 | 20 | | | |
| *correlation* | *0.85* | | *correlation* | *-1* | |

red – high identity, blue – low identity

**Table 2: Identities in the CAMERA BLAST search for individual proteins from bacterium *V.okutanii* to the reads from Gutless Worm and GOS (P) databases.**

| Genome | Length | GC Content | Coding | Molecule | Genes | Proteins |
|---|---|---|---|---|---|---|
| *Porcine circovirus* | 1 768 | 48% | 93% | circular ssDNA | 2 | 2 |
| *Buchnera aphidicola* | 416 380 | 20% | 85% | circular dsDNA | 397 | 357 |
| *Nanoarchaeum equitans* | 490 885 | 31% | 92% | circular dsDNA | 582 | 536 |
| *Mycoplasma genitalium* | 580 076 | 31% | 90% | circular dsDNA | 525 | 477 |
| *Mesoplasma florum* | 793 224 | 27% | 92% | circular dsDNA | 717 | 682 |
| *Vesicomyosocius okutanii* | 1 022 154 | 31% | 85% | circular dsDNA | 978 | 937 |
| *Aquifex aeolicus* | 1 551 335 | 43% | 92% | circular dsDNA | 1580 | 1529 |
| *Acanthamoeba polyphaga* | 1 181 404 | 27% | 86% | linear dsDNA | 1258 | 911 |
| *Thiomicrospira denitrificans* | 2 201 561 | 34% | 92% | circular dsDNA | 2163 | 2096 |
| *Thiomicrospira crunogena* | 2 427 734 | 43% | 89% | circular dsDNA | 2259 | 2196 |
| *Thiobacillus denitrificans* | 2 909 809 | 66% | 92% | circular dsDNA | 2879 | 2827 |
| *Escherichia coli* | 4 639 675 | 50% | 85% | circular dsDNA | 4466 | 4133 |

**Table 3: Comparison of viral and bacterial genomes that demonstrates an intersection between their sizes.**

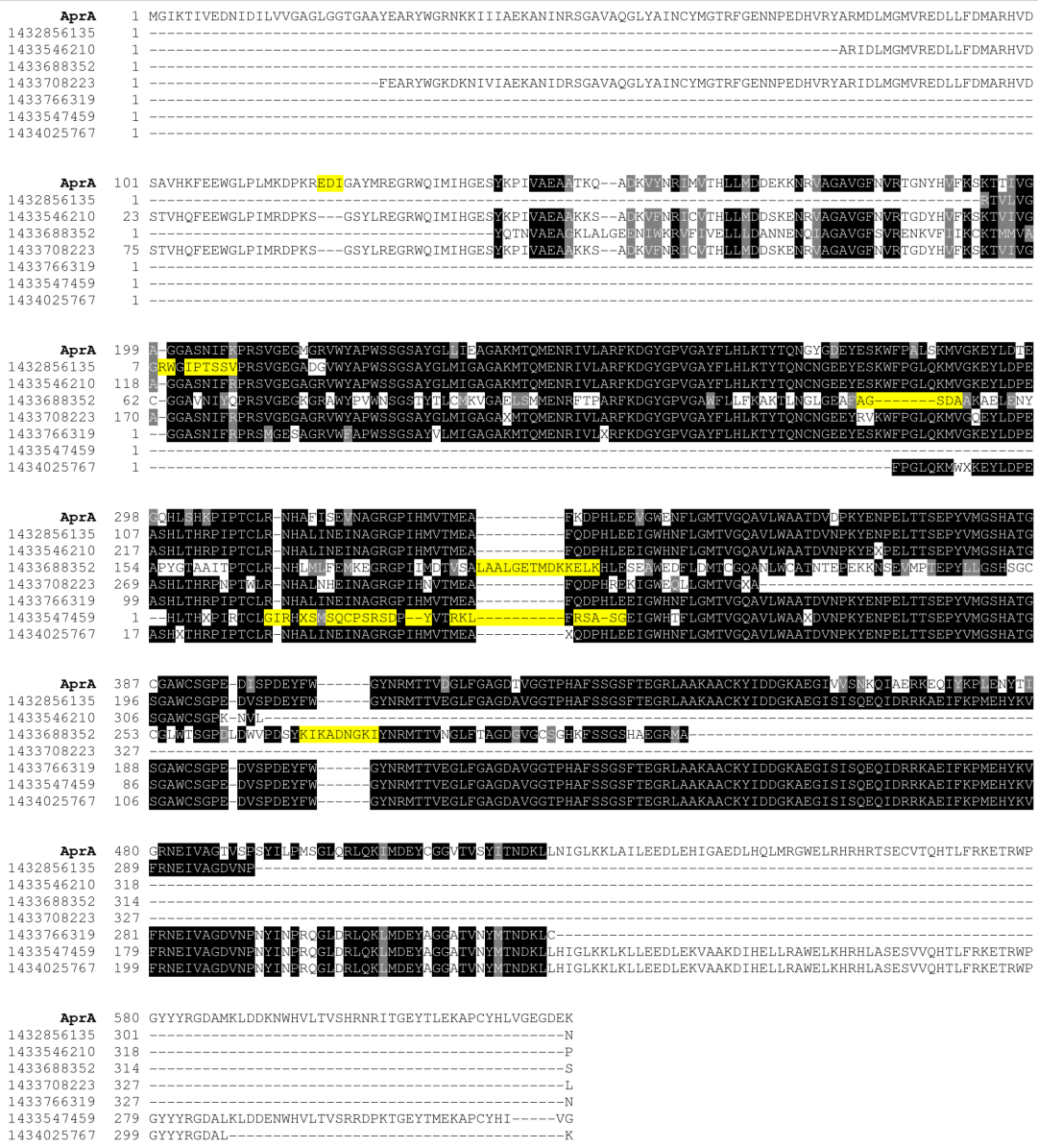


**Figure 7: Alignment of the protein reads from microorganisms living in the gutless worm *O.algarvensis* to the AprA polypeptide from bacterium *V.okutanii***
AprA – protein from *V.okutanii*, Gene ID: 148244258; numbers – selected reads from the Gutless Worm database; yellow – sufficient insertions, deletions and exchanges

databases for the AprA protein, as well as other cytoplasmic polypeptides, such as DsrB, DsrE and DsrC whereas DsrR chaperon-like protein exhibited low identity and a potential diversity in ecology.

## Discovery of aprA-gene family in thiotrophic symbionts of gutless worm *O.algarvensis*

One should remember that the thioautotrophic bacterium *Vesicomyosocius okutanii* is an intracellular symbiont harboring in the gill epithelial cells of the deep-sea clam, *Calyptogena okutanii* (Kuwahara et al., 2007). In my experiments, the adenylylsulfate reductase AprA query sequence from *V.okutanii* bacterial genome demonstrated a high identity to the readings from the symbiotic microbial community of gutless marine oligochaete *Olavius algarvensis* that inhabits the shallow sub littoral sediments on the northwest coast of Elba Island, Italy (Figure 6A). Endosymbiotic sulfate-reducing bacteria, living within this worm produce sulfide that can serve as an energy source for complementary sulfide-oxidizing symbionts (Dubilier et al., 2001). I found 18 subject DNA fragments with an E-value less than $1.35e^{-70}$, a bit score higher than 267.7, and an alignment length of 185 to 336 aa in the Gutless Worm database. The best search result is the sequence GUTLS WRM ELBA READ 1433546210 with an 85% identity. An additional search in the Gene Bank using the BLASTP program revealed a significant degree of identity (92%) between the read 1433546210 and the adenylylsulfate reductase (Gene ID: 3673410) from the β-proteobacterium *Thiobacillus denitrificans* ATCC 25259. A narrow subtree of AprA protein family with a close distance from AprA-like protein (read 1433546210) to homologous members of β- and γ-proteobacteria (data not shown) can indicate that the read 1433546210 belongs to a proteobacterium living in the gutless worm *O.algarvensis*.

Each of the 18 hypothetical AprA protein sequences was compared with each other and with the query – AprA protein from bacterium

Journal of Computer Science & Systems Biology - Open Access

Research Article
OPEN ACCESS Freely available online
doi:10.4172/jcsb.1000065

JCSB/Vol.3 Issue 5

*V.okutanii*. The T-COFFEE automatic server found a perfect matching between short experimental reads (about 300 aa) and the large AprA polypeptide (627 aa) covering nearly all of the AprA sequence. After the manual elimination of similar fragments, I finished with the seven final sequences demonstrating numerous exchanges, insertions, and deletions (Figure 7). The read 1433708223 (327 aa long) was similar to N-part of AprA protein; and the read 1433547459 (321 aa) matched C-terminus with some variations. The most irregular read 1433688352 (314 aa) demonstrated multiple exchanges, one deletion and 2 insertions with flanking rearrangements. In addition, read 1432856135 (301 aa) showed an insertion/exchange with possible neighboring compensations in its own N-terminus. The read 1433547459 (321 aa) is very interesting because it has a long region including 28 amino residues with 2 deletions, 1

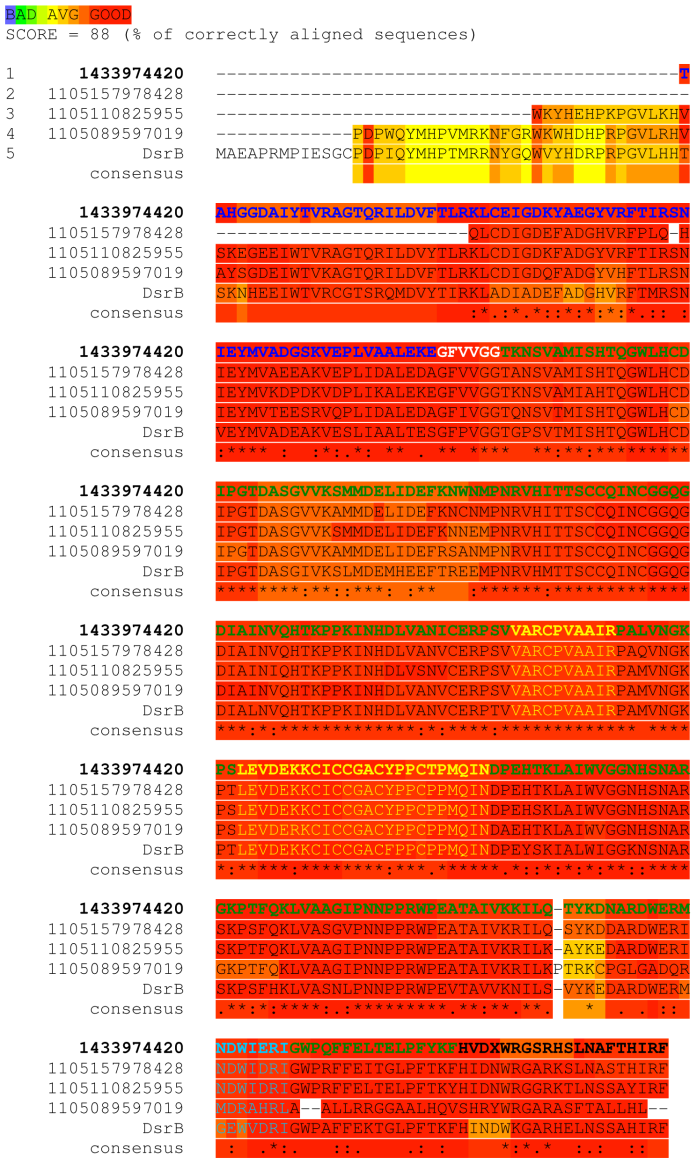



**Figure 8: Multiple alignments of DsrB-like reads from different GOS collections 1** – GUTLS WRM ELBA READ 1433974420, **2** – Sargasso Station 13 JCVI PEP 1105157978428, **3** – Upwelling Fernandina Island JCVI PEP 1105110825955, **4** – Punta Cormorant Hypersalin JCVI PEP 1105089597019, **5** – DsrB protein from *V.okutanii* (Gene ID: 148244930); color characters are predicted domains, see text and legends for Figure 9, 10; background is the quality of alignment.


**Figure 9: Domains within hypothetical DsrB protein from a gutless worm's symbiont.** All the 313 amino residues of the sequence GUTLS_WRM_ELBA_READ_1433974420 are presented. The sulfite reductase ferredoxin-like half domain, position 1 – 67, is blue. The sulfite reductase 4Fe-4S domain, 74 – 293, is green. The 4Fe-4S binding site, 165 – 174, and 4Fe-4S binding site, 184 – 207, are yellow insertions. The repressor of phase-1 flagellin, 270 – 277, is red. Predicted protein structures were retrieved from the following servers: ESyPred3D, position 11 – 149 (bold cursive), 3D-JIGSAW, 153 – 206 (bold) and GENO3D, 207 – 299 (cursive).

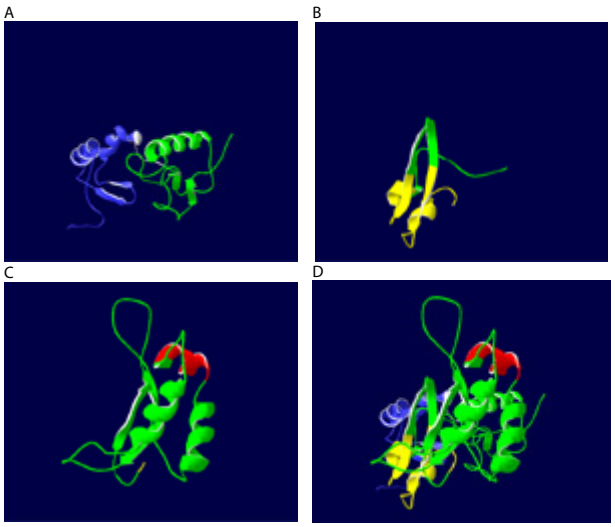



**Figure 10: Protein structures predicted from GUTLS WRM ELBA READ 1433974420. A** – by ESyPred3D server, **B** – by 3D-JIGSAW server, **C** – by Geno3D server, **D** – manual assembling of DsrB protein from the A, B and C fragments within SPDBV program; sulfite reductase ferredoxin-like half domain (blue), large sulfite reductase 4Fe-4S domain (green), 4Fe-4S binding sites (yellow), repressor of phase-1 flagellin (red).

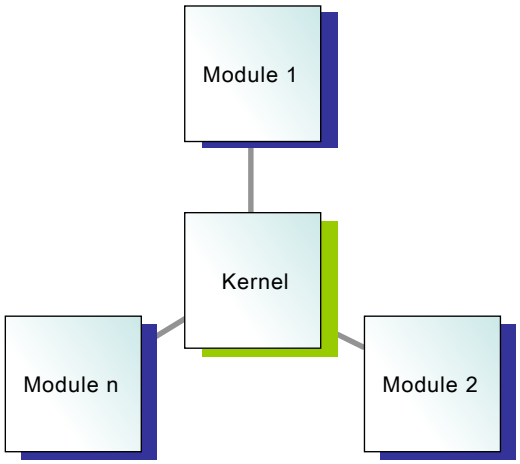



**Figure 11: Modules are channeling a system toward unique energy sources and a particular environment.**

insertion, and numerous amino acid exchanges. Moreover, AprA protein from bacterium *V.okutanii* has Glu-Asp-Ile insertion without visible compensatory rearrangements in comparison with the reads 1433546210 and 1433708223 which may say something about the early history or the neutrality of this event. As well, single amino acid variations were found in all experimental reads.

In addition, the 7 reads (1432856135, 1433546210, 1433547459, 1433688352, 1433708223, 1433766319 and 1434025767) were used as queries to search the NCBI Gene Bank. Many homologies with Apr proteins (adenylylsulfate reductase, adenosine-5-phosphosulfate reductase) from different species were found; β-, γ-, and Δ-proteobacteria, green sulfur bacteria, and uncultured bacteria, for instance, endosymbionts of bivalves as the intracellular sulfur-oxidizing endosymbiont *Ruthia magnifica* Cm from the clam *Calyptogena magnifica*, or symbionts of mussels *Bathymodiolus brevior*, *Bathymodiolus azoricus* and *Bathymodiolus thermophilus*, sulfur-oxidizing bacterial symbionts from gutless marine oligochaeta *Inanidrilus makropetalos* and from the giant tube worm *Riftia pachyptila*. The last data confirmed a vast diversity of AprA-like proteins discovered in bacterial symbionts of the gutless worm *O.algarvensis*. The heterogeneity might be a result of the large and dissimilar microbial community utilizing different APS reductases. This diversity provoked intriguing questions about their origin, complementary functions, and the speed of microevolution.

## Multiple alignments of AprA-, DsrA-, DsrB- and SoxX-protein families

AprA, DsrA, DsrB, and SoxX proteins are objects for future research because of their importance in sulfur metabolism and their high degree of identity in previous experiments (Table 2, bold characters). Top matching reads for the entire proteins were selected from the best samples, such as *O.algarvensis* symbionts, *A.pompejana* symbionts, Waseca farm soil, deep Mediterranean, Richmond acid mine, Sargasso station, upwelling at Fernandina Island, and hypersaline lagoon on Floreana Island. These reads were used for comparison within each protein family by T-COFFEE multiple sequence alignment program. The highest homology (71%) was found within AprA-protein family. The DsrA and DsrB families demonstrated 61% and 66% identity respectively. The SoxX-protein family had a low matching score (48%). However, these results were affected by the different quality of databases; many hypothetical proteins were represented by their fragments. A good example is shown in Figure 8, namely the multiple alignment for reduced DsrB-protein family (identity 88%). The high identity of the core alignment to the annotated DsrB protein from *V.okutanii* supposed a biological significance of this conserved sequence.

## Identification of domains within the hypothetical DsrB protein

The DNA sequence GUTLS WRM ELBA READ 1433974420, with positions 35-973 nt coding a hypothetical DsrB protein 313 aa in length (Figure 8, upper bold string), was chosen for an annotation. This polypeptide sequence was submitted to InterProScan and PFAM databases. The InterProScan server identified the query as a sulfite reductase, dissimilatory type β-subunit encoded by *dsrB*-type gene. The polypeptide comprises a ferredoxin-like hemoprotein β-component in the N-terminus typical for sulfite/nitrite reductases and a 4Fe-4S region in the C-terminus also usual for sulfite and nitrite reductases. More detailed information about the protein domains in the query sequence was retrieved from the PFAM server (see annotated data in Figure 9). In addition to sulfite/nitrite reductase ferredoxin-like half domain (blue), I uncovered two 4Fe-4S binding sites (yellow) and a region for the repressor of phase-1 flagellin (red) within the large sulfite reductase 4Fe-4S domain (green). The recapitulated active center of the hypothetical DsrB sulfite reductase explains the conservatism of the area obtained in multiple alignments of fragments from different sources (Figure 8, yellow characters). An unexpected insertion, the repressor for flagellin synthesis, with a very conservative Arg residue and less conservative Asp and Ile residues, was also found in other members of the DsrB family (Figures 8, 9, aqua characters). This finding may be evidence of an autoregulation that couples energy metabolism and bacterial movement.

## 3D reconstruction of the protein encoded in GUTLS WRM ELBA READ 1433974420

Several comparative modeling servers, e.g. ESyPred3D, 3D-JIGSAW, and Geno3D were used to predict 3D structure of the polypeptide encoded in the read 1433974420. Each individual server

Research Article
OPEN ACCESS Freely available online
doi:10.4172/jcsb.1000065
JCSB/Vol.3 Issue 5

retrieved only a part of the submitted sequence. The job of ESyPred3D server is depicted in Figure 9 as a bold cursive, the message from 3D-JIGSAW is shown in bold characters, and Geno3D result is represented as a cursive string. These server machines build three-dimensional models for proteins based on homologues of known polypeptide structures. The first 3D model was built by ESyPred3D program using the 2akj chain A with 2.8 Angstroms resolution. This template shared 14.2% identity with the query sequence. The 2akj protein is a spinach ferredoxin nitrate oxidoreductase (EC 1.7.7.1) from chloroplasts whose enzyme catalyzes a limiting step in siroheme biosynthesis (Swamy et al., 2005). The other server, the Geno3D offered a large set of potential templates with a top hit for the sulfite reductase, dissimilatory-type β subunit from *Halorhodospira halophila* DSM 244/SL1 (Gene ID: 4710578, score 323, expectation $4e^{-87}$, identities 73%). Finally, using the protein fragment retrieved by ESyPred3D server, I was able to reconstruct the C-part of the sulfite reductase ferredoxin-like half domain (blue) and the N-terminus of sulfite reductase 4Fe-4S domain (green). See Figures 9 and 10 for details. The 3D-JIGSAW server predicted two internal 4Fe-4S binding sites (yellow) within the large sulfite reductase domain (green). Geno3D server built the C-terminus of the sulfite reductase 4Fe-4S domain. The modeling started with the last Asn residue at the second 4Fe-4S binding site and finished after a repressor of phase-1 flagellin (red). Summarizing the jobs of all three servers, it was possible to reconstruct a 3-dimensional structure of the whole polypeptide encoded in the read 1433974420 except for 10 aa in the N-terminus and 14 aa in the C-terminus.

I identified two parallel α-helixes on the surface of the protein core and two antiparallel β-sheets that possibly interact with the siroheme molecule inside the sulfite reductase ferredoxin-like complex (Figure 10A, blue). Remarkably, the two loops of 4Fe-4S binding site (Figure 10B, yellow) are close enough to permit the possible interaction with an iron-sulfur cluster. The repressor of phase-1 flagellin represents a small α-helix outside the core protein (Figure 10C, red). The assemblage of those three polypeptide chains into a complete protein molecule by SPDBV viewer was not easy because of an intricacy of the sulfite reductase 4Fe-4S domain (green) and a chaotic character of polypeptide terminus within each of 3D models (Figure 10D). However, the result demonstrated the possibility of reconstructing the protein structure from an environmental DNA read. The 3-dimensional model obtained is in close proximity to the intact DsrB-like protein sequence from the symbiont of amazing gutless worm *Olavius algarvensis* found in the Bay of Capo di Sant' Andrea on the Elba Island in Italy.

## Discussion

### A minimal gene set

Starting with a minimal cell, it is difficult to define a minimal set of genes that support life (Koonin, 2000) because the complexity of life is entirely dependent on the specific environment. The minimalist approach leads to an evident controversy concerning the definition of life. For instance, the virus *Porcine circovirus* comprises 1 768 nt and consists of only 2 genes (Table 3). The minimal known symbiotic bacterium *Buchnera aphidicola* is much bigger (Perez-Brocal et al., 2006) comprising 416 380 nt and consisting of 397 genes which might confirm a difference between viruses and bacteria, as well as a distinction between nonliving and living objects. However, a huge virus like *Acanthamoeba polyphaga* comprises 1 181 404 nt and consists of 1 258 genes which is more than in some bacteria, such as *Nanoarchaeum equitans*, *Mycoplasma genitalium*, *Mesoplasma florum*, *Vesicomyosocius okutanii,* and *Aquifex aeolicus.* In this case, early ideas about the reconstruction of an ancestor genome by comparing complete modern bacterial genomes (Mushegian and Koonin, 1996) may not be enough to make an artificial cell. An alternative could be a more realistic community metabolism in a special niche.

Resent generalizations say that biological systems are modular; and life plays with interdependent modules (Watson, 2006; Woese, 2002). Evolution acts on a combination of individual genes, gene clusters, and functional genetic systems including genetic and protein networks, as well as cellular pathways that are called here 'modules' (Bork et al., 1998; Marcotte et al., 1999; Overbeek et al., 1999). Metagenomic data and their annotation and analysis have revealed that the full record of proteins is far from being complete (Yooseph et al., 2007). Proteins may have appeared very early parallel to RNA and lipids worlds (Guimarães, 1994; Segré et al., 2001; Takahashi and Mihara, 2004; Weissmann, 2005). Perhaps, even the functional modules evolved very early and contributed to cellularization much earlier than previously imagined (discussed below) (Lindahl, 2004). My state-of-the-art research assumes a relatively complex kernel, a gene set comprising basic 'living' functions, such as core replication, repair, transcription, translation, several central metabolic pathways, possibly cytoskeleton and membrane forming, as well as cell division (Gabaldón et al., 2007; Gibson et al., 2008; Gil et al., 2004; Glass et al., 2006; Koonin, 2003). This approach includes the establishment of modules, and their combination and hierarchical design. For instance, sulfur metabolism might be organized within a single module. Modules perform special functions related to the particular environment, e.g. the hydrogen sulfide conversion (Figure 11).

### Synthetic symbiont

An autonomous agent which exploits a reach environment and complements certain functions in the host organism, e.g. sulfur energetic metabolism was considered as a synthetic symbiont. It assumed that the host provided a nutrition transport ($H_2S$) and removed wastes ($SO_4^{2-}$). In agreement with the 'modular logics', the 21 genes from the genome of the symbiotic bacterium *Vesicomyosocius okutanii* HA were inferred (Table 1, Figure 2) and seven important enzyme activities were considered in the simplified model of food transport and processing (Figure 3). The simulation of hydrogen sulfide oxidation showed that the import of nutrition may be a limiting factor for symbiotic metabolism (Figure 4). Following dynamic modeling, the effective transport system and multiple diets might have provided a more reliable survival of the symbiotic system. That is why the sulfide- and sulfate-transporters have been proposed for use in Engineering. In some cases, a contact of host tissues with hydrogen sulfide can be reduced by the sulfide binding to hemoglobin in the blood (Arp and Childress, 1983). The artificial symbiotic bacterium would produce a secreted $H_2S$ transmitter similar to the hemoglobin from the giant tube worm *R.pachyptila* to protect the host organism from toxic sulfide.

The set of coupled genes (Table 1, Figure 2) and the metabolic model Figure 3 were similar to the sulfur metabolism in *Chlorobium tepidum* TLS. Bacterium *C.tepidum* can utilize thiosulfate and sulfide as a source of sulfur. The genes involved in sulfur oxidation are organized in clusters. For example, one cluster includes most of the *sox* genes to oxidize thiosulfate. Three flavoprotein reductases may oxidize sulfide to form polysulfides as a reserve. Further oxidation to sulfate takes place in the cytoplasm by using dissimilatory siroheme sulfite reductase (DsrABC), adenosine-5'-phosphosulfate reductase

(APS reductase), and ATP sulfurylase. The sulfate exporter CT0714 and a thiosulfate importer are identified in the genome of *C.tepidum* (Eisen et al., 2002).

## Russell-Martin-Koonin scenario of origin of life. Role of sulfur on early life

A parsimonious evolutionary scenario considers the origin of genomes and cells within inorganic compartments (Koonin and Martin, 2005). A population of retrovirus-like elements was imagined as a universal common ancestor of modern life. They existed within small geological cavities. RNA molecules were subject to variations. Natural selection acted on compartment contents. The primitive genetic elements evolved massively and parallel to mechanisms involved in replication, recombination, transmission between compartments and colonization (Koonin, 2003; Koonin and Martin, 2005; Woese, 1998). In my opinion, circulation of coding molecules between compartments would be analogous to contemporary horizontal gene transfer. A dramatic bottle-neck effect happened when highly parallel chemical systems on the basis of polypeptides, RNA, and lipids were critically interconnected and naturally encoded in the DNA sequences. The details of these events remain obscure. All those interconnected roots converged into hypothetical singularity as a result of self-organization and gave origin to the upper part of the genealogical tree of life.

Sulfur was one of the important elements for the prebiotic chemistry of early Earth and occurred in a variety of redox states, ranging from -2 in sulfide to +6 in sulfate. Wächtershäuser (1988) expressed the idea that the earliest living 'organisms' were mineralbased systems converting $CO_2$ and $H_2S$ into organic compounds on the surface of pyrite crystals ($FeS_2$). This two-dimensional metabolism first began at hydrothermal vents. Russell and Martin (2004) proposed a similar hypothesis that primordial metabolism arose within small pyrite cavities, and chimneys formed on the ocean floor through which $CO_2$ and $H_2$ bubbled up. This 'hydrothermal reactor' model assumed the emergence of life in the sulfide system by pyrite catalysis and primitive mechanism of $CO_2$ fixation by geochemical pathways. Products of the original reactions catalyzed the following ones; the prehistoric metabolic network arose and included new participants. As an argument, iron-sulfide is found in the active center of many enzymes, such as adenosine-5'-phosphosulfate reductase (APS reductase, EC 1.8.99.2). The modern acid mine drainage microbial community at Richmond Mine Iron Mountain in California demonstrates 8 *sox* and 78 *dsr* matching DNA sequences (Figures 5A, 6E). Recently, DNA from a 2.8-km deep gold mine in South Africa was assembled into a genome of bacterium *Candidatus Desulforudis audaxviator*. This sulfate-reducing chemoautotrophic thermophile utilizes radiolitically generated sources of energy and nutrients by using machinery shared with archaea. Horizontally acquired genes, such as *aprA*, *dsrE* and *sat* have been revealed (Chivian et al., 2008).

## Distribution of oxidative sulfur metabolism

Knowledge about chemosynthetic symbionts is still fragmentary because of the difficulty in cultivating symbionts and their hosts in a laboratory. Rather than isolating pure cultures, the high-throughput sequencing was applied to environmental samples to obtain information about the individual genomes. New methods have been applied to study bacteria involved in the biochemistry of the global sulfur cycle (Eisen, 2007; Meyer and Kuever, 2008; Tripp et al., 2008). In the present research, 4 complete bacterial genomes were compared with 3 global databases and visualized by CAMERA Fragment Requirement Viewer (Figure 1). Small symbiotic bacterium *V.okutanii* HA was chosen as a preferred reference from other thiotrophic microorganisms, such as *Ruthia magnifica* Cm, *Thiomicrospira crunogena* XCL-2, and *Thiobacillus denitrificans* ATCC 25259. After that, reads from 41 positive stations around the world were carefully investigated by CAMERA BLAST tools for an identity to *sox* and *dsr* genes from bacterium *V.okutanii*. Instead of evenly distributed *sox* and *dsr* genes by stations, numerous *sox* and *dsr* sequences were found mostly in the specific niches, such as the host's environment; farm soil, deep-see anoxic water, acid mine drainage, epipelagial zones, upwelling zones and hypersaline lagoon (Figures 5, 6). Although, *sox* and *dsr* genes demonstrated high correlation in experimental set ($R = 0.86$), samples with disproportions of *sox* and *dsr* genes were also found. One possible reason for this difference might be the heterogeneity of populations as a result of bacterial specialization and compartmentalization of reactions. Data obtained exhibited the complementarity of *sox* and *dsr* genes in various niches depending on the specific environment.

Current investigation has revealed a homology of AprA protein, as well as DsrB, DsrE, and DsrC proteins from *V.okutanii* with the corresponding reads in CAMERA databases (Table 2). In contrast, DsrR polypeptide represented a low identity that may be a consequence of variability in chaperone genes. The 18 subject reads from gutless worm *O.algarvensis* bacterial symbionts demonstrated a high similarity to AprA query polypeptide from *V.okutanii*. The protein family that was discovered, adenosine-5-phosphosulfate reductases, showed an immense diversity following different lineages. I supposed that these unknown symbionts belonged to β- and γ-proteobacteria. Other authors described chemoautotrophic sulfur-oxidizing β-proteobacteria and γ-proteobacterial sulfate reducing symbionts in *O.algarvensis* (Dubilier et al., 2001; Woyke et al., 2006).

My results may not reflect phylogeny because the BLAST local sequence alignment removes the most divergent regions from the sequences, thus obscuring the relationships between sequences. Due to the low quality of databases (random and partial DNA reads), there were some difficulties during the multiple sequence alignments with the T-COFFEE program. It could be that many hypothetical proteins were represented only by their fragments. Nevertheless, I was able to reveal different patterns of variability for AprA and DsrB proteins. Numerous insertions and deletions in AprA polypeptides were found, as well as prevailed amino acid exchanges in DsrB proteins (compare Figure 7 and 8).

Kuever and co-workers (Meyer et al., 2007; Meyer and Kuever, 2007; Meyer and Kuever 2008) supposed that *apr* genes are distributed between photo- and chemotrophs including invertebrate symbionts with anaerobic or facultative anaerobic lifestyles. The sulfur-oxidizing prokaryotes (SOP) turn into two phylogenetic lineages. The proteins of Apr II lineage, e.g. from *Chlorobiaceae*, are similar to the enzymes from the sulfate-reducing prokaryotes (SRP). This clustering is in contradiction to the dissimilatory sulfite reductase phylogeny (Dsr) and specifies the possible horizontal *aprA* gene transfer from SRP to SOB. Another paper (Klein et al., 2001) illustrated that DsrAB sequences generated a tree incompatible with the corresponding 16S rRNA phylogeny. Evidence for *dsr* genes horizontal transfer was found within β-proteobacteria, affecting *Desulfobacula toluolica*. The dsrAB genes in *Archaeoglobus* species are also the result of an ancient horizontal gene transfer from a bacterial donor. Maybe, the horizontal gene transfer played a significant role in the distribution of sulfur metabolism between populations (Boucher et al., 2003). Perhaps, sulfur-oxidation (*sox*) and dissimilatory sulfite reductase (*dsr*) clusters, which represent different functional submodules for the energetic sulfur metabolism, were developed independently and accepted many times by various inhabitants of our planet.

Journal of Computer Science & Systems Biology - Open Access

Research Article
OPEN ACCESS Freely available online
doi:10.4172/jcsb.1000065
JCSB/Vol.3 Issue 5

## Horizontal gene transfer

HGT is widespread (Andersson, 2005; Boucher et al., 2003). Communal planetary genes persist in bacterial genomes and have a tendency to cluster. Groups of travelling genes are involved in the basic molecular processes and are engaged in adaptation. It could be these genes have existed since the transition of nonliving to living matter and provided a mechanism for acquiring new functions during the exploration of various environments. The horizontal gene transfer could be a very old mechanism supporting a compatibility of modules (Brown, 2003). The ancient virus world should also be mentioned because of the essential role of viruses in evolution. About $10^{28}$ bp of DNA are transduced by phages every year in the World Ocean (Paul et al., 2002). Viruses can contribute to global horizontal gene transfer by moving DNA between species and biomes (Sano et al., 2004). Microbial biodiversity may change due to migration of viruses and cross-species genetic exchanges. In a recent metagenomics project, the photosystem I gene cluster was identified in genomes of marine cyanophages and is thought to be a function in photosynthesis of their hosts (Sharon et al., 2009).

Additionally, the species of lower metazoans can be an attractive target for gene transfer, e.g a massive HGT in bdelloid rotifers was well documented recently (Gladyshev et al., 2008). The horizontal gene transfer mediated by sperm was revealed in the bivalve mollusk *Mytilus galloprovincialis* Lam. (Guerra et al., 2005; Kuznetsov et al., 2001). HGT involves particular molecular mechanisms both for prokaryotes (Smith et al., 1981) and eukaryotes (Pittoggi et al., 2006; Sciamanna et al., 2003; Zani et al., 1995). Expansion of protein, non-protein-coding RNA, and microRNA families, as well as repeated elements have been identified in the genome of the platypus *Ornithorhynchus anatinus* that could also be interpreted as a result of massive HGT in early vertebrate evolution, although researchers have considered only the 'traditional' vertical process (Warren et al., 2008). In addition, a transmission of symbiotic bacteria was described – horizontal and vertical transfer of symbionts. For example, juvenile tube worms *Riftia pachyptila* have to acquire the bacteria from local surroundings in each generation because adult worms are completely dependent on symbiotic bacteria (Nussbaumer et al., 2006). In contrast, *Vesicomyidae* clams transmit their thiotrophic endosymbionts vertically between generations via the egg cytoplasm (Stewart et al., 2008). Gene transfer from intracellular bacteria to multicellular eukaryotes is well established for 4 insect and 4 nematode species that range from more than 1 Mb to less than 500 bp insertions. A near complete genome of *Wolbachia pipientis* was found in the chromosome of the widespread tropical fruit fly *Drosophila ananassae* (Hotopp et al., 2007).

Boaden (1975) proposed the origin of multicellular organisms from a sulfide system (rocky roots thiobios hypothesis). He assumed fauna with anaerobic metabolism in all stages of individual development in the sulfide zone. Boaden further suggested so as to the first metazoa were anaerobes living in reduced conditions and that some animals in the sulfide zone might retain the biochemical pathways of the earliest metazoan. The iron content of deep-sea sediments shows that deep ocean was anoxic at 580 million years ago. However, last evolution scenarios considered the deep-ocean oxygenation shortly before the appearance of the Ediacara biota at about 575 Ma which is earliest eukaryotic fossils of up to meter scale soft bodied organisms and colonies (Canfield et al., 2007). As we have seen in the evolution scene, genetic information attempts to come into eukaryotes within different participants (Katz, 2002; Kidwell, 1993). Sulfur metabolism is dispersed around the world (Figure 5A, 6).

## Reconstruction of DsrB protein

The read 1433974420 from GUTLS WRM ELBA dataset exhibited a conserved protein core (Figure 8) that helped me detect domain structures within the polypeptide, such as (1) sulfite reductase ferredoxin-like half domain, (2) large sulfite reductase 4Fe-4S domain including two 4Fe-4S binding sites, and (3) a repressor for flagellin synthesis (Figure 9). Moreover, the comparative structural analysis by ESyPred3D, 3D-JIGSAW, and Geno3D servers allowed the 3D reconstruction of an almost complete hypothetical sulfite reductase, DsrB protein (Figure 10). Just compare the discovered domains side by side with known proteins.

As has been considered by other scientists, sulfite reductase is a main enzyme for both dissimilation of oxidized anions and biosynthetic assimilation of sulfur (Hansen, 1994). Sulfite reductase is a multisubunit enzyme composed of dimers of either $\alpha/\beta$ or $\alpha/\beta/\gamma$ subunits, each containing a siroheme and iron-sulfur cluster prosthetic center (Pierik et al., 1992). For instance, the well-known *Escherichia coli* sulfite reductase (SiR) is a complex enzyme composed of two proteins; a flavoprotein alpha-component (SiR-FP) plus a hemoprotein betacomponent (SiR-HP), demonstrating $8\alpha$-$4\beta$ quaternary structure (Zeghouf et al., 2000). SiR-FP contains both FAD and FMN, while SiR-HP includes 4Fe-4S cluster coupled to a siroheme through a cysteine bridge. SiR-HP has a two-fold symmetry which causes a unique threedomain $\alpha/\beta$ fold that controls assembly and enzymatic activity (Crane et al., 1995).

4Fe-4S domain is a superfamily whose members bind to iron-sulfur clusters. Structure of the domain is antiparallel β-sheets. Members include bacterial-type ferredoxins and various reductases and dehydrogenases. Ferredoxins are iron-sulfur proteins mediating electron transfer in a variety of metabolic processes. They are comprehensive from bacteria to mammals and fall into several groups (Eck and Dayhoff, 1966; George et al., 1985). Most ferredoxins contain at least one conserved domain including 4 Cys residues that bind to the 4Fe-4S cluster. During the evolution of bacterial ferredoxins, gene duplications, transpositions, and fusion occurred resulting in the appearance of proteins with multiple ironsulfur centers, e.g. bi-cluster 2[4Fe-4S] and polyferredoxins, iron-sulfur subunits of bacterial succinate dehydrogenase/fumarate reductase, pyruvate-flavodoxin oxidoreductase, formate hydrogenlyase and dehydrogenase complexes, NADH:ubiquinone reductase, and others. In some bacterial ferredoxins, one of the duplicated domains lost some conserved Cys residues. These domains lacked their iron-sulfur binding property. 3D structures are known both for mono- and bi-cluster 4Fe-4S ferredoxins (Duee et al., 1994; Fukuyama et al., 1989).

Flagellin subunits assemble into filaments of bacterial flagella. I speculated that the insertion of a repressor for flagellin synthesis into the 4Fe-4S domain might be an example of unusual regulation coupling the energy metabolism and bacterial movement. It may also be interesting evidence of the hierarchical modularity, where the flagellin-inhibiting helix is a small protein building block of low hierarchy, which is able to interconnect two larger functional modules, such as mobility and sulfur metabolism on a higher hierarchical state. DNA-binding repressor of phase-1 flagellin encoded by *fljA* gene was identified in the genomes of *Salmonella abony* and *Salmonella typhimurium* (McClelland et al., 2001). The peptide sequence Phe-Asp-Trp-Val-Ser-Arg-Ile at the position 166-172 from *Salmonella* proteins is similar to the sequence Asn-Asp-Trp-Ile-Glu-Arg-Ile from the discovered polypeptide (Figures 9, 10, red). The virulence regulation by nutrient limitation is known for *Legionella*

*pneumophila* (Bachman and Swanson, 2001). This fact suggests that the symbiont associated with *O.algarvensis*, the worm lacking a mouth, gut, and nephridia might encompass a free-living stage and virulent features.

## Conclusion, the complexity of biological systems

Unicellular life emerged about 3.5-3.8 billion years ago and another 2.5 billion years passed before multicellular organisms made their appearance. Nevertheless, it took less than one billion years for architecturally remarkable species to evolve (Fenchel, 2002; Wacey, 2009). The speed and complexity of evolution has increased visibly. It seems that gradual evolution might not explain the major evolutionary transitions indicated by fossil records (e.g. *Opabinia regalis*, Burgess Shale fauna, Cambrian). On the other hand, HGT can alter the speed of evolution and design by a propagation of 'constructive mutations' across geographical populations (Beiko and Charlebois, 2007; Hao et al., 2004; Jain et al., 2003; Meyer and Kuever, 2007). Functional combination of genes by HGT could check the fitness of interaction between genes and other components in the whole system (Watson, 2006). If life stores genetic information in a modular form, then the modular nature of genetic information makes it possible to accept useful units from other species to swap the global gene set (Beiko et al., 2005; Simonson et al., 2005). It supposes an idea regarding the dramatic level of complexity that geochemical, biochemical, and genetic systems should attain before they come alive as a result of modularization. The lithotrophic biosphere extends many kilometres into the deep sea and Earth crust (Gold, 1992). Our restricted knowledge of this kind of life stems from sampling and drilling. Deep sea and subterranean habitats may point out that lithotrophy contributes much more to the biomass of Earth than thought (Falkowski et al., 2008). It can even give insight into possible life on dwarf planets and satellites with a thermal silicate core and oceans beneath an ice surface, such as Pluto, Enceladus, Callisto, Ganymede, and Europa (Garzón and Garzón, 2001; Gold, 1992; Vance et al., 2007).

Nature has always been a source of metaphors and inspiration. In recent years, it has provoked many successful techniques. When nature is complex, some problems can be solved through design by analogy and the development of novel systems. The results of this research show that anaerobic sulfur metabolism of bacteria-invertebrate symbioses within the giant tube worm *Riftia pachyptila*, the deep-sea clam *Calyptogena okutanii*, hydrothermal vent polychaete *Alvinella pompejana* and the gutless oligochaete *Olavius algarvensis*, offers a great potential for future exploration of anoxic zones. Data recovered from official databases by NCBI and CAMERA browsers, as well as an understanding of biological complexity will help target sulfur metabolism, the chemosynthetic origin of life, and the synthetic forms of life (Kuznetsov, 2009; Kuznetsov et al., 2007). My assumption is that biological complexity developed during a 'scalable acceleration' during which the more intricate and interconnected biological parts were successively involved in the course of compositional evolution by the natural processes of replication, shuffling, selection, and transfer.

### Acknowledgments

The author is grateful to Nelli Sergeeva and Maksim Gulin for initially inspiring this work, to Mikhail Kats, Jan Kuever, Peer Staehler, and Hubert Bernauer for their encouragement and suggestions, and to Elena Georgieva and Ursula Link. My deep thanks to Josef Maier, Sven Panke, Victor de Lorenzo and the anonymous referees for their criticism and recommendations. Special acknowledgments go to Steven Benner who advised me on the allegory of the metabolic enzyme that moves around by lateral transfer, and to Vladik Avetisov who pointed out the limits of the gradualistic view on evolution. Many thanks to Bert Schnell who put life into the manuscript. All annotated data were obtained from NCBI, CAMERA and PFAM public databases. This work is partly supported by the European Science Foundation.

### References

1. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, et al. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25: 3389-3402.
2. Andersson JO (2005) Lateral gene transfer in eukaryotes. Cell Mol Life Sci 62: 1182-1197.
3. Appel RD, Bairoch A, Hochstrasser DF (1994) A new generation of information retrieval tools for biologists: the example of the ExPASy WWW server. Trends Biochem Sci 19: 258-260.
4. Arndt C, Gaill F, Felbeck H (2001) Anaerobic sulfur metabolism in thiotrophic symbioses. J Exp Biol 204: 741-750.
5. Arp AJ, Childress JJ (1983) Sulfide Binding by the Blood of the Hydrothermal Vent Tube Worm Riftia pachyptila. Science 219: 295-297.
6. Bachman MA, Swanson MS (2001) RpoS co-operates with other factors to induce Legionella pneumophila virulence in the stationary phase. Mol Microbiol 40: 1201-1214.
7. Bateman A, Birney E, Cerruti L, Durbin R, Etwiller L, et al. (2002) The Pfam protein families database. Nucl Acids Res 30: 276-280.
8. Bateman A, Birney E, Durbin R, Eddy SR, Sonnhammer ELL, et al. (2000) The Pfam protein families database. Nucleic Acids Res 28: 263-266.
9. Bates PA, Kelley LA, MacCallum RM, Sternberg MJE (2001) Enhancement of protein modeling by human intervention in applying the automatic programs 3D-JIGSAW and 3D-PSSM. Proteins 45: 39-46.
10. Beiko RG, Charlebois RL (2007) A simulation test bed for hypotheses of genome evolution. Bioinformatics 23: 825- 831.
11. Beiko RG, Harlow TJ, Ragan MA (2005) Highways of gene sharing in prokaryotes. Proc Natl Acad Sci U S A 102: 14332-14337.
12. Beller HR, Chain PS, Letain TE, Chakicherla A, Larimer FW (2006) The genome sequence of the obligately chemolithoautotrophic, facultatively anaerobic bacterium Thiobacillus denitrificans. J Bacteriol 188: 1473-1488.
13. Benson DA, Boguski MS, Lipman DJ, Ostell J, Francis Ouellette BF (1998) GenBank. Nucleic Acids Res 25: 1-7.
14. Boaden PJS (1975) Anaerobiosis, meiofauna, and early metazoan evolution. Zool Scr 4: 21-24.
15. Bork P, Dandekar T, Diaz-Lazcoz Y, Eisenhaber F, Huynen M, et al. (1998) Predicting function: from genes to genomes and back. J Mol Biol 283: 707-725.
16. Boucher Y, Douady CJ, Papke RT, Walsh DA, Boudreau ME, et al. (2003) Lateral gene transfer and the origins of prokaryotic groups. Annu Rev Genet 37: 283-328.
17. BOXSHADE, http://www.ch.embnet.org/software/BOX_form.html
18. Brown JR (2003) Ancient horizontal gene transfer. Nat Rev Genet 4: 121-132.
19. CAMERA, http://camera.calit2.net/
20. Canfield DE, Poulton SW, Narbonne GM (2007) Late-Neoproterozoic deep-ocean oxygenation and the rise of animal life. Science 315: 92-95.
21. Chabasse C, Bailly X, Sanchez S, Rousselot M, Zal F (2006) Gene structure and molecular phylogeny of the linker chains from the giant annelid hexagonal bilayer hemoglobins. J Mol Evol 63: 365-374.
22. Chivian D, Brodie EL, Alm EJ, Culley DE, Dehal PS, et al. (2008) Environmental genomics reveals a single-species ecosystem deep within Earth. Science 322: 275-278.
23. Combet C, Jambon M, Deléage G, Geourjon C (2002) Geno3D: automatic comparative molecular modelling of protein. Bioinformatics. 18: 213-214.
24. Crane BR, Siegel LM, Getzoff ED (1995) Sulfite reductase structure at 1.6 A: evolution and catalysis for reduction of inorganic anions. Science 270: 59-67.
25. Dahl C, Engels S, Pott-Sperling AS, Schulte A, Sander J, et al. (2005) Novel genes of the dsr gene cluster and evidence for close interaction of Dsr proteins during sulfur oxidation in the phototrophic sulfur bacterium Allochromatium vinosum. J Bacteriol 187: 1392-1404.

Research Article
OPEN ACCESS Freely available online
doi:10.4172/jcsb.1000065
JCSB/Vol.3 Issue 5

26. Darwin Ch (1859) On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life, John Murray, London, 204-208.

27. DeLong EF, Preston CM, Mincer T, Rich V, Hallam SJ, et al. (2006) Community genomics among stratified microbial assemblages in the ocean's interior. Science 311: 496-503.

28. Doolittle WF (1980) Revolutionary concepts in evolutionary cell biology. Trends Biochem Sci 5: 146-149.

29. Dubilier N, Mülders C, Ferdelman T, de Beer D, Pernthaler A, et al. (2001) Endosymbiotic sulfate-reducing and sulfide-oxidizing bacteria in an oligochaete worm. Nature 411: 298-302.

30. Duee ED, Fanchon E, Vicat J, Sieker LC, Meyer J, et al. (1994) Refined crystal structure of the 2[4Fe-4S] ferredoxin from Clostridium acidurici at 1.84 A resolution. J Mol Biol 243: 683-695.

31. Eck RV, Dayhoff MO (1966) Evolution of the Structure of Ferredoxin Based on Living Relics of Primitive Amino Acid Sequences. Science 152: 363-366.

32. Eisen JA, Nelson KE, Paulsen IT, Heidelberg JF, Wu M, et al. (2002) The complete genome sequence of Chlorobium tepidum TLS, a photosynthetic, anaerobic, green-sulfur bacterium. Proc Natl Acad Sci U S A. 99: 9509-9514.

33. Eisen JA (2007) Environmental shotgun sequencing: Its potential and challenges for studying the hidden world of microbes. PLoS Biol 5: e82.

34. ERGO, http://ergo.integratedgenomics.com/ERGO/

35. ESyPred3D, http://www.fundp.ac.be/sciences/biologie/urbm/bioinfo/esypred/

36. ExPASy, http://www.expasy.org/

37. Falkowski PG, Fenchel T, Delong EF (2008) The microbial engines that drive Earth's biogeochemical cycles. Science 320: 1034-1039.

38. Fenchel T (2002) The Origin and Early Evolution of Life, Oxford University Press, New York.

39. Flores JF, Fisher CR, Carney SL, Green BN, Freytag JK, et al. (2005) Sulfide binding is mediated by zinc ions discovered in the crystal structure of a hydrothermal vent tubeworm hemoglobin. Proc Natl Acad Sci U S A 102: 2713-2718.

40. Friedrich CG, Rother D, Bardischewsky F, Quentmeier A, Fischer J (2001) Oxidation of reduced inorganic sulfur compounds by bacteria: Emergence of a common mechanism? Appl Environ Microbiol 67: 2873-2882.

41. Fukuyama K, Matsubara H, Tsukihara T, Katsube Y (1989) Structure of [4Fe-4S] ferredoxin from Bacillus thermoproteolyticus refined at 2.3 A resolution. Structural comparisons of bacterial ferredoxins. J Mol Biol 210: 383- 398.

42. Gabaldón T, Peretó J, Montero F, Gil R, Latorre A, et al. (2007) Structural analyses of a hypothetical minimal metabolism. Philos Trans R Soc Lond B Biol Sci 362: 1751-1762.

43. Garzón L, Garzón ML (2001) Radioactivity as a significant energy source in prebiotic synthesis. Orig Life Evol Biosph 31: 3-13.

44. Gasteiger E, Gattiker A, Hoogland C, Ivanyi I, Appel RD, et al. (2003) ExPASy: The proteomics server for in-depth protein knowledge and analysis. Nucleic Acids Res 31: 3784-3788.

45. Geno3D, http://geno3d-pbil.ibcp.fr

46. George DG, Hunt LT, Yeh LS, Barker WC (1985) New perspectives on bacterial ferredoxin evolution. J Mol Evol 22: 20-31.

47. Gibson DG, Benders GA, Andrews-Pfannkoch C, Denisova EA, Baden-Tillson H, et al. (2008) Complete chemical synthesis, assembly and cloning of a Mycoplasma genitalium genome. Science 319: 1215-1220.

48. Gil R, Silva FJ, Peretó J, Moya A (2004) Determination of the core of a minimal bacterial gene set. Microbiol Mol Biol Rev 2004 68: 518-537.

49. Gladyshev EA, Meselson M, Arkhipova IR (2008) Massive horizontal gene transfer in bdelloid rotifers. Science 320: 1210-1213.

50. Glass JI, Assad-Garcia N, Alperovich N, Yooseph S, Lewis MR, et al. (2006) Essential genes of a minimal bacterium. Proc Natl Acad Sci USA 103: 425-430.

51. Gold T (1992) The deep, hot biosphere. Proc Natl Acad Sci U S A 89: 6045-6049.

52. Guerra R, Carballada R, Esponda P (2005) Transfection of spermatozoa in bivalve molluscs using naked DNA. Cell Biol Int 29: 159-164.

53. Guex N, Peitsch MC (1997) SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. Electrophoresis 18: 2714-2723.

54. Guimarães RC (1994) Linguistics of biomolecules and the protein-first hypothesis for the origins of cells. Journal of Biol Physics 20:193-99.

55. Hansen TA (1994) Metabolism of sulfate-reducing prokaryotes. Antonie Van Leeuwenhoek 66: 165-185.

56. Hao W, Golding GB (2004) Patterns of bacterial gene movement. Mol Biol Evol 21: 1294-1307.

57. Hotopp JC, Clark ME, Oliveira DC, Foster JM, Fischer P, et al. (2007) Widespread lateral gene transfer from intracellular bacteria to multicellular eukaryotes. Science 317: 1753-1756.

58. InterProScan, http://www.ebi.ac.uk/InterProScan/

59. Jain R, Rivera MC, Moore JE, Lake JA (2003) Horizontal gene transfer accelerates genome innovation and evolution. Mol Biol Evol 20: 1598-1602.

60. Jain R, Rivera MC, Moore JE, Lake JA (2002) Horizontal gene transfer in microbial genome evolution. Theor Popul Biol 61: 489-495.

61. JDesigner, http://www.sys-bio.org/software/jdesigner.htm

62. JDesigner, the manual, http://public.kgi.edu/~hsauro/sysbio/papers/JDBooklet.pdf

63. Kanehisa M, Goto S (2000) KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res 28: 27-30.

64. Katz LA (2002) Lateral gene transfers and the evolution of eukaryotes: theories and data. Int J Syst Evol Microbiol 52: 1893-1900.

65. KEGG, http://www.genome.jp/kegg/

66. Kidwell MG (1993) Lateral transfer in natural populations of eukaryotes. Annu Rev Genet 27: 235-256.

67. Klein M, Friedrich M, Roger AJ, Hugenholtz P, Fishbain S, et al. (2001) Multiple lateral transfers of dissimilatory sulfite reductase genes between major lineages of sulfate-reducing prokaryotes. J Bacteriol 183: 6028-6035.

68. Koonin EV, Martin W (2005) On the origin of genomes and cells within inorganic compartments. Trends Genet 21: 647-654.

69. Koonin EV (2003) Comparative genomics, minimal gene-sets and the last universal common ancestor. Nat Rev Microbiol 1: 127-136.

70. Koonin EV (2000) How many genes can make a cell: the minimal-gene-set concept. Annu Rev Genomics Hum Genet 1: 99-116.

71. Kuwahara H, Yoshida T, Takaki Y, Shimamura S, Nishi S, et al. (2007) Reduced genome of the thioautotrophic intracellular symbiont in a deep-sea clam, Calyptogena okutanii. Curr Biol 17: 881-886.

72. Kuznetsov A, Sergeeva N, Cholodov V, Erokhin V (2007) Perspectives on artificial ecosystems in the Black Sea anoxic zone. ECSB 2007-The European Conference on Synthetic Biology: design, programming and optimization of biological systems, Sant Feliu de Guixols, Spain, 29: 50-52.

73. Kuznetsov A (2009) Synthetic Biology as a proof of Systems Biology. Chapter V in the Handbook of Research on Systems Biology Applications in Medicine. Ed. Andriani Daskalaki. IGI Global. 97-115.

74. Kuznetsov AV, Pirkova AV, Dvorianchikov GA, Panfertsev EA, Gavriushkin AV, et al. (2001) [Study of the transfer of foreign genes into mussel Mytilus galloprovincialis Lam. eggs by spermatozoa]. Ontogenez 32: 309-318.

75. Lambert C, Leonard N, De Bolle X, Depiereux E (2002) ESyPred3D: Prediction of proteins 3D structures. Bioinformatics 18: 1250-1256.

76. Lindahl PA (2004) Stepwise evolution of nonliving to living chemical systems. Orig Life Evol Biosph 34: 371-389.

77. Lo I, Denef VJ, Verberkmoes NC, Shah MB, Goltsman D, et al. (2007) Strain-resolved community proteomics reveals recombining genomes of acidophilic bacteria. Nature 446: 537-541.

78. Marcotte EM, Pellegrini M, Ng HL, Rice DW, Yeates TO, et al. (1999) Detecting protein function and protein-protein interactions from genome sequences. Science 285: 751-753.

79. Margulis L (1970) Origin of Eukaryotic Cells. New Haven: Yale University Press.

80. Maynard Smith J, Szathmary E (1995) The Major Transitions in Evolution. New York: Oxford University Press Inc.

81. McClelland M, Sanderson KE, Spieth J, Clifton SW, Latreille P, et al. (2001) Complete genome sequence of Salmonella enterica serovar Typhimurium LT2. Nature 413: 852-856.

82. Merezhkovsky KS (1909) The Theory of Two Plasms as the Basis of Symbiogenesis, a New Study or the Origins of Organisms. Kazan.

83. Meyer B, Imhoff JF, Kuever J (2007) Molecular analysis of the distribution and phylogeny of the soxB gene among sulfuroxidizing bacteria - evolution of the Sox sulfur oxidation enzyme system. Environ Microbiol 9: 2957-2977.

84. Meyer B, Kuever J (2008) Phylogenetic Diversity and Spatial Distribution of the Microbial Community Associated with the Caribbean Deep-water Sponge Polymastia cf. corticata by 16S rRNA, *aprA,* and *amoA* Gene Analysis. Microb Ecol 56: 306-321.

85. Meyer B, Kuever J (2007) Phylogeny of the alpha and beta subunits of the dissimilatory adenosine-5'-phosphosulfate (APS) reductase from sulfate-reducing prokaryotes - origin and evolution of the dissimilatory sulfate-reduction pathway. Microbiology 153: 2026-2044.

86. Mulder NJ, Apweiler R, Attwood TK, Bairoch A, Bateman A, et al. (2007) New developments in the InterPro database. Nucleic Acids Res 35: D224-228.

87. Mushegian AR, Koonin EV (1996) A minimal gene set for cellular life derived by comparison of complete bacterial genomes. Proc Natl Acad Sci USA 93: 10268-10273.

88. NCBI, http://www.ncbi.nlm.nih.gov/

89. Notredame C, Higgins DG, Heringa J (2000) T-Coffee: A novel method for fast and accurate multiple sequence alignment. J Mol Biol 302: 205-217.

90. Nussbaumer AD, Fisher CR, Bright M (2006) Horizontal endosymbiont transmission in hydrothermal vent tubeworms. Nature 441: 345-348.

91. Overbeek R, Fonstein M, D'Souza M, Pusch GD, Maltsev N (1999) The use of gene clusters to infer functional coupling. Proc Natl Acad Sci USA 96: 2896-2901.

92. Overbeek R, Larsen N, Walunas T, D'Souza M, Pusch G, et al. (2003) The ERGO genome analysis and discovery system. Nucleic Acids Res 31: 164-171.

93. Paul JH, Sullivan MB, Segall AM, Rohwer F (2002) Marine phage genomics. Comp Biochem Physiol B Biochem Mol Biol 133: 463-476.

94. Perez-Brocal V, Gil R, Ramos S, Lamelas A, Postigo M, et al. (2006) A small microbial genome: the end of a long symbiotic relationship? Science 314: 312-313.

95. PFAM, http://pfam.janelia.org/

96. Pierik AJ, Duyvis MG, van Helvoort JM, Wolbert RB, Hagen WR (1992) The third subunit of desulfoviridin-type dissimilatory sulfite reductases. Eur J Biochem 205: 111-115.

97. Pires RH, Venceslau SS, Morais F, Teixeira M, Xavier AV, et al. (2006) Characterization of the Desulfovibrio desulfuricans ATCC 27774 DsrMKJOP complex--a membrane-bound redox complex involved in the sulfate respiratory pathway. Biochemistry 45: 249-262.

98. Pittoggi C, Beraldi R, Sciamanna I, Barberi L, Giordano R, et al. (2006) Generation of biologically active retro-genes upon interaction of mouse spermatozoa with exogenous DNA. Mol Reprod Dev 73:1239-1246.

99. Pott AS, Dahl C (1998) Sirohaem sulfite reductase and other proteins encoded by genes at the dsr locus of Chromatium vinosum are involved in the oxidation of intracellular sulfur. Microbiology 144:1881-1894.

100. Rieseberg LH, Raymond O, Rosenthal DM, Lai Z, Livingstone K, et al. (2003) Major ecological transitions in wild sunflowers facilitated by hybridization. Science 301: 1211-1216.

101. Rusch DB, Halpern AL, Sutton G, Heidelberg KB, Williamson S, et al. (2007) The Sorcerer II Global Ocean Sampling expedition: northwest Atlantic through eastern tropical Pacific. PLoS Biol 5: e77.

102. Russell MJ, Martin W (2004) The rocky roots of the acetyl-CoA pathway. Trends Biochem Sci 29: 358-363.

103. Sano E, Carlson S, Wegley L, Rohwer F (2004) Movement of viruses between biomes. Appl Environ Microbiol 70: 5842-5846.

104. Sauve V, Bruno S, Berks BC, Hemmings AM (2007) The SoxYZ complex carries sulfur cycle intermediates on a peptide swinging arm. J Biol Chem 282: 23194-23204.

105. Sciamanna I, Barberi L, Martire A, Pittoggi C, Beraldi R, et al. (2003) Sperm endogenous reverse transcriptase as mediator of new genetic information. Biochem Biophys Res Commun 312: 1039-1046.

106. Scott KM, Sievert SM, Abril FN, Ball LA, Barrett CJ, et al. (2006) The genome of deep-sea vent chemolithoautotroph *Thiomicrospira crunogena* XCL-2. PLoS Biol 4: e383.

107. Segré D, Ben-Eli D, Deamer DW, Lancet D (2001) The lipid world. Orig Life Evol Biosph 31: 119-145.

108. Seshadri R, Kravitz SA, Smarr L, Gilna P, Frazier M (2007) CAMERA: a community resource for metagenomics. PLoS Biol 5: e75.

109. Sharon I, Alperovitch A, Rohwer F, Haynes M, Glaser F, et al. (2009) Photosystem I gene cassettes are present in marine virus genomes. Nature 461: 258-262.

110. Simonson AB, Servin JA, Skophammer RG, Herbold CW, Rivera MC, et al. (2005) Decoding the genomic tree of life. Proc Natl Acad Sci USA 1: 6608-6613.

111. Smith HO, Danner DB, Deich RA (1981) Genetic transformation. Annu Rev Biochem 50: 41-68.

112. SPDBV, http://www.expasy.ch/spdbv/

113. Stewart FJ, Young CR, Cavanaugh CM (2008) Lateral symbiont acquisition in a maternally transmitted chemosynthetic clam endosymbiosis. Mol Biol Evol 25: 673-687.

114. Swamy U, Wang M, Tripathy JN, Kim SK, Hirasawa M, et al. (2005) Structure of spinach nitrite reductase: implications for multi-electron reactions by the iron-sulfur:siroheme cofactor. Biochemistry 44: 16054-16063.

115. Takahashi Y, Mihara H (2004) Construction of a chemically and conformationally self-replicating system of amyloid-like fibrils. Bioorg Med Chem 12: 693-699.

116. Tringe SG, von Mering C, Kobayashi A, Salamov AA, Chen K, et al. (2005) Comparative metagenomics of microbial communities. Science 308: 554-557.

117. Tripp HJ, Kitner JB, Schwalbach MS, Dacey JW, Wilhelm LJ, et al. (2008) SAR11 marine bacteria require exogenous reduced sulfur for growth. Nature 452:741-744.

118. Vance S, Harnmeijer J, Kimura J, Hussmann H, Demartin B, et al. (2007) Hydrothermal systems in small ocean planets. Astrobiology 7: 987-1005.

119. Wacey D (2009) Early Life on Earth: A Practical Guide. Springer 31.

120. Wächtershäuser G (1988) Before enzymes and templates: theory of surface metabolism. Microbiol Rev 52: 452-84.

121. Warren WC, Hillier LW, Marshall Graves JA, Birney E, Ponting CP, et al. (2008) Genome analysis of the platypus reveals unique signatures of evolution. Nature 453: 175-183.

122. Watson RA (2006) Compositional Evolution: The Impact of Sex, Symbiosis, and Modularity on the Gradualist Framework of Evolution. Vienna Ser Theor Biol: A Bradford Book.

123. Weissmann C (2005) Birth of a prion: spontaneous generation revisited. Cell 122: 165-8.

124. Woese CR (2002) On the evolution of cells. Proc Natl Acad Sci USA 99: 8742-8747.

125. Woese CR (1998) The universal ancestor. Proc Natl Acad Sci USA 95: 6854-6859.

126. Woyke T, Teeling H, Ivanova NN, Huntemann M, Richter M, et al. (2006) Symbiosis insights through metagenomic analysis of a microbial consortium. Nature 443: 950-955.

127. Yooseph S, Sutton G, Rusch DB, Halpern AL, Williamson SJ, et al. (2007) The Sorcerer II Global Ocean Sampling expedition:expanding the universe of protein families. PLoS Biol 5: e16.

128. Zani M, Lavitrano M, French D, Lulli V, Maione B, et al. (1995) The mechanism of binding of exogenous DNA to sperm cells: factors controlling the DNA uptake. Exp Cell Res 217: 57-64.

129. Zeghouf M, Fontecave M, Coves J (2000) A simplifed functional version of the Escherichia coli sulfite reductase. J Biol Chem 275: 37651-3756.

130. 3D-JIGSAW, http://www.bmm.icnet.uk/servers/3djigsaw/