

Research Article

A Database of Plastid Protein Families from Red Algae and Apicomplexa and Expression Regulation of the *moeB* Gene

Oleg A. Zverkov, Alexandr V. Seliverstov, and Vassily A. Lyubetsky

*Institute for Information Transmission Problems of the Russian Academy of Sciences (Kharkevich Institute),
Bolshoy Karetny Pereulok 19, Moscow 127994, Russia*

Correspondence should be addressed to Oleg A. Zverkov; zverkov@iitp.ru

Received 18 June 2014; Revised 29 August 2014; Accepted 13 September 2014

Academic Editor: William H. Piel

Copyright © 2015 Oleg A. Zverkov et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

We report the database of plastid protein families from red algae, secondary and tertiary rhodophyte-derived plastids, and Apicomplexa constructed with the novel method to infer orthology. The families contain proteins with maximal sequence similarity and minimal paralogous content. The database contains 6509 protein entries, 513 families and 278 nonsingletons (from which 230 are paralog-free, and among the remaining 48, 46 contain at maximum two proteins per species, and 2 contain at maximum three proteins per species). The method is compared with other approaches. Expression regulation of the *moeB* gene is studied using this database and the model of RNA polymerase competition. An analogous database obtained for green algae and their symbiotic descendants, and applications based on it are published earlier.

1. Introduction

The concept of orthology and construction of orthology databases are important areas of bioinformatic research. However, the orthology relationship is not yet decisively formalized and some of its important features may depend on taxonomic context of the data and properties of particular organelles. Mathematically, identification of orthologs corresponds to building clusters in a graph with its vertices assigned gene or protein sequences. The majority of clustering methods utilize various strategies to weight the graph edges with subsequent construction of “highly connected components,” that is, clusters resulting from a certain clustering procedure.

The edge weight reflects similarity of amino acid sequences generated in various pairwise alignment procedures, intron content and positioning, protein domain architecture, gene synteny, and so forth. Usually the weights are computed with global alignment using the Needleman-Wunsch algorithm, or local alignment using BLAST. Various clustering approaches were proposed, from specifically organized partitioning of the spanning tree of the initial graph (the originally proposed algorithm ClusterZSL, refer to [1]) to time estimation of random walk on a graph (the OrthoMCL algorithm).

In the latter algorithm based on Markov clustering, walk within a cluster is long, and jumps between clusters are rare [2]. Due to heuristic nature of these processes, comparison of the algorithms cannot be formalized, especially in the absence of standard benchmarking data. The description of OrthoMCL implicitly states that its convergence is difficult to discuss even in hypothesis.

The algorithm ClusterZSL essentially differs from commonly employed methods, including OrthoMCL, by not using the mutual-best-hit criterion. For a pair of genomes, a gene may produce none or many best hits; the latter is especially the case when considering suboptimal hits that may in fact represent true orthologs. In contrast with other methods, ClusterZSL also minimizes the amount of paralogs in each cluster that in general seems a reasonable property. ClusterZSL can consider gene positioning in DNA and orthologous context of the gene neighborhood. A version of this algorithm that uses gene synteny was applied to various chordate animals and will be described in a separate publication.

The algorithm ClusterZSL and its computer program implementation possess the computational complexity of maximum n^2 accurate to a coefficient. The OrthoMCL uses matrix multiplication, the operation with the minimal complexity

TABLE 1: Orthologs of *moeB* in plastids of rhodophyte algae as inferred with ClusterZSL and their genomic neighborhoods.

Class	Species	Locus	Protein MoeB	Genomic context
Bangiophyceae	<i>Porphyra purpurea</i>	NC_000925	NP_053945.1	(<i>trnW</i>)- <i>ORF75-moeB</i>
Bangiophyceae	<i>Porphyridium purpureum</i>	NC_023133	YP_008965710.1	(<i>ORF144</i>)- <i>ycf38-moeB</i>
Bangiophyceae	<i>Pyropia haitanensis</i>	NC_021189	YP_007947865.1	(<i>trnW</i>)- <i>ORF75-moeB</i>
Bangiophyceae	<i>Pyropia perforata</i>	NC_024050	YP_009027619.1	(<i>trnW</i>)- <i>ORF75-moeB</i>
Bangiophyceae	<i>Pyropia yezoensis</i>	NC_007932	YP_537017.1	(<i>trnW</i>)- <i>moeB</i>
Bangiophyceae	<i>Cyanidioschyzon merolae</i>	NC_004799	NP_849016.1	(<i>trnW</i>)- <i>moeB</i>
Bangiophyceae	<i>Cyanidium caldarium</i>	NC_001840	NP_045115.1	(<i>trnW</i>)- <i>moeB</i>
Florideophyceae	<i>Calliarthron tuberculosis</i>	NC_021075	YP_007878185.1	(<i>trnW</i>)- <i>moeB</i>
Florideophyceae	<i>Chondrus crispus</i>	NC_020795	YP_007627343.1	(<i>trnW</i>)- <i>moeB</i>
Florideophyceae	<i>Gracilaria salicornia</i>	NC_023785	YP_009019560.1	(<i>trnW</i>)- <i>moeB</i>
Florideophyceae	<i>Gracilaria tenuistipitata</i>	NC_006137	YP_063552.1	(<i>trnW</i>)- <i>moeB</i>
Florideophyceae	<i>Grateloupia taiwanensis</i>	NC_021618	YP_008144807.1	(<i>trnW</i>)- <i>moeB</i>

The *moeB* orthologs are also denoted by *moeB*, irrespective of corresponding original annotations. Genes on the opposite strand to *moeB* are given in brackets.

n^ω , where the exponent ω is a parameter. For the Gauss algorithm $\omega = 3$ and for the Strassen algorithm $\omega = \log_2 7 \approx 2.81$ [3]. An asymptotically faster algorithm is known, which, however, takes advantage only with matrices of very high order and is practically of little use [4]; also refer to [5, 6]. Further concerns with the OrthoMCL algorithm are the estimation of the number of iterations (including matrix multiplications) and proof of convergence. The convergence requirement is obviously met with ClusterZSL. The running time of OrthoMCL appears to be much longer than that of ClusterZSL, at least with our testing data. Due to high scalability, performance of ClusterZSL does not depend on the amount of CPUs, which is a valuable practical property; the authors are unaware of attempts to assess the scalability of OrthoMCL.

Compare ClusterZSL with the algorithm used in the Ensembl database. Both start from the spanning tree. On later stages, the Ensembl algorithm relies in many respects on multiple alignments of leaf proteins, the task exponential in computational complexity if the alignment is optimized [7]. For alignment construction, the algorithm integrates the *M*-Coffee algorithm [8] or Mafft for larger data [9]. Both mentioned alignment procedures are heuristic and do not guarantee global minimization of the used functional. The ClusterZSL algorithm does not utilize multiple alignment.

Worth mentioning is another clustering method to establish orthology that was previously used by the authors. When the size of the clusters is known, for example, in studies of multicomponent systems where the length of the orthologous series is known for one component, the most dense cluster of the known size is constructed using the algorithm described in [10, 11]. We do not compare with phylogenetic methods here; for instance, refer to [12]. Note that the phylogenetic position of a species or protein belonging to any species is not always known.

The problem of the transcription factor regulon definition is of great interest. In red algae, the only plastid-encoded transcription factors are Ycf27, Ycf28, Ycf29, and RbcR (Ycf30). Of little information on them, the RbcR binding sites are known to vary even among close species [13], which hampers

their detection. We will consider this problem on the example of the factor Ycf28, which, as it turned out, regulates the expression of the gene *moeB*.

In this study, the gene *moeB*, which is itself an important object of research, is tackled in a case study of gene expression regulation using ClusterZSL. This gene encodes an E1-like family enzyme involved in molybdopterin and thiamine biosynthesis. This family includes proteins that catalyze the adenylation by ATP of the carboxyl group of the C-terminal glycine in sulfur carrier proteins, for example, Moad or ThiS. Bacterial proteins with domains characteristic for this family are described in [14]. The *moeB* gene is present in plastids of all sequenced Rhodophyta; refer to Table 1. Its ortholog in *Porphyra purpurea* and *Pyropia* spp. is *ORF382*, in *Cyanidium caldarium* *chlN*. In *P. perforata* the neighboring genes *moeB* and *ORF382* encode the N- and C-termini of the MoeB protein.

As evident from Table 1, the neighbor of *moeB* on the opposite strand is *trnW* that encodes the tryptophanyl-tRNA. In *Porphyra purpurea*, *Pyropia haitanensis*, and *Pyropia perforata* the genes *trnW* and *moeB* are separated by the short coding frame *ORF75*. The only exception is *Porphyridium purpureum*, where the neighborhood of *moeB* lacks a reliably highly transcribed gene on the opposite strand; refer to [15–23].

In this study we describe a database ClusterZSL of orthologous plastid proteins in red algae, secondary and tertiary rhodophyte-derived plastids, and Apicomplexa (the RedLine at May 2014 from the GenBank; also refer to <http://lab6.iitp.ru/ppc/redline50/>), constructed with the same algorithm ClusterZSL.

We use it in a case study of transcription regulation of the *moeB* gene. An analogous database obtained for green algae and their symbiotic descendants (the green line) and its applications are published in [1, 24–26].

Some recent papers ([27] et al.) glance upon plastid proteins the database CpBase, <http://chloroplast.ocean.washington.edu/>. It represents 35 plastomes from RedLine in comparison with 50 plastomes represented in the database ClusterZSL. The authors are not aware of the description of

TABLE 2: Orthologs of Ycf28 in plastids of Rhodophyta as inferred with ClusterZSL.

Class	Species	Locus	Protein Ycf28	Bit score	E-value
Bangiophyceae	<i>Porphyra purpurea</i>	NC.000925	NP_053952.1	50.9	9.5e - 14
Bangiophyceae	<i>Porphyridium purpureum</i>	NC.023133	YP_008965713.1	48.6	5.0e - 13
Bangiophyceae	<i>Pyropia haitanensis</i>	NC.021189	YP_007947872.1	52.9	2.2e - 14
Bangiophyceae	<i>Pyropia perforata</i>	NC.024050	YP_009027626.1	53.3	1.7e - 14
Bangiophyceae	<i>Pyropia yezoensis</i>	NC.007932	YP_537023.1	55.2	4.4e - 15
Bangiophyceae	<i>Cyanidioschyzon merolae</i>	NC.004799	NP_849012.1	29.5	4.5e - 07
Bangiophyceae	<i>Cyanidium caldarium</i>	NC.001840	NP_045121.1	55.8	2.8e - 15
Florideophyceae	<i>Calliarthron tuberculosum</i>	NC.021075	YP_007878179.1	43.9	1.5e - 11
Florideophyceae	<i>Chondrus crispus</i>	NC.020795	YP_007627337.1	31.7	9.1e - 08
Florideophyceae	<i>Gracilaria salicornia</i>	NC.023785	YP_009019566.1	29.0	6.5e - 07
Florideophyceae	<i>Gracilaria tenuistipitata</i>	NC.006137	YP_063558.1	32.6	4.8e - 08
Florideophyceae	<i>Grateloupia taiwanensis</i>	NC.021618	YP_008144797.1	33.6	2.4e - 08

The last two columns contain estimates for the Pfam Crp-like helix-turn-helix domain (PF13545).

the method, which the CpBase has been constructed with, as well as the details related to it.

2. Materials and Methods

All plastid proteins are available in GenBank [28]. Orthology was established with the ClusterZSL algorithm described in [1] and applied previously in [24–26]. The algorithm parameters were set to $H = 0.6$, $L = 0$. Gene annotations were verified with the Pfam [29] and Prosite [30] databases.

Promoters were predicted using an algorithm described in [24, 31, 32]. For different σ -subunits of bacterial type RNA polymerases it utilizes data on mutation profiles of the *psbA* promoter in *Sinapis alba* [33] and other experimentally studied promoters [34].

In searches for motifs in the 5'-leader regions of *moeB* we used the original algorithm published in [35, 36] and the WEB service MEME [37], although the motifs were not detected.

The notion of the phylogenetic distribution (profile) is defined in [26]: for a given gene/protein g , it is a function on a given set S of species that equals (for all s from S) +1 if g is present in s , and -1 otherwise.

In Section 3 we essentially exploit the originally proposed model of RNA polymerase competition [38, 39]. The model describes the following situation. In DNA locus transcription many RNA polymerases involved simultaneously bind with the promoters of their type and elongate along their chains, possibly towards each other. This leads to the interaction of RNA polymerases, both between each other and with various protein and structural factors on DNA and RNA. As a result, the transcription levels of the genes significantly change, right up to inability to initiate the transcription of the divergent located gene (below in this role *moeB*), when an actively transcribed gene (resp., *trnW*) plays against it, provided the intergenic region is not organized in a special way.

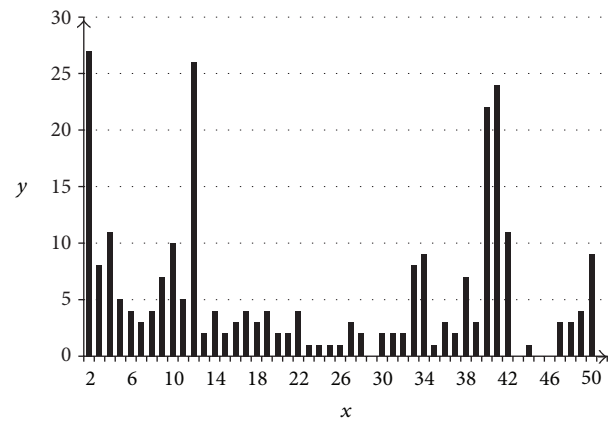


FIGURE 1: Distribution of cluster (y -axis, the ordinate) versus species (x -axis, the abscissa) numbers.

3. Results

We report the database ClusterZSL (<http://lab6.iitp.ru/ppc/>) of plastid protein families from red algae, secondary and tertiary rhodophyte-derived plastids, and Apicomplexa (the RedLine). The families contain proteins with maximal sequence similarity and minimal paralogous content and are built using the ClusterZSL algorithm. The database contains 6005 protein entries, 513 families, and 278 nonsingletons (from which 230 are paralog-free, and among the remaining 48, 46 contain at maximum two proteins per species and 2 at maximum three proteins per species). The comparison of the obtained protein families with the biological annotations indicates their good conformity.

Tables 1 and 2 describe two clusters of the database. Figure 1 presents a diagram of species content in inferred clusters.

Standard bacterial type promoters were not detected in the 5'-leader regions of *moeB*. However, the {A, T}-rich

regions found upstream *moeB* may represent functioning –10 promoter boxes. Based on modeling RNA polymerases competition we suggest that the promoters of *moeB* are located in between *moeB* and *trnW* (refer to the Conclusions) and differ distinctly from the common template.

The presented database allows comparing a cluster of a gene (e.g., *moeB*) with all other clusters. Phylogenetic distributions of *moeB* and Ycf28 coincide; for example, there is a unique transcription factor, which is encoded in a plastid if and only if *moeB* is encoded in it; it is Ycf28. That indicates that the best hit against *moeB* is Ycf28, a transcription factor.

The lack of detected –35 box for *moeB* naturally suggests that Ycf28 is an activator. Based on the same modeling, we surmise that the Ycf28 binding sites are located in between genes *moeB* and *trnW*. The only exception might be *Porphyridium purpureum*. The Ycf28 binding motif itself was not identified, probably due to the variability of binding sites.

Note that the 5'-UTRs of *moeB* are usually short and allow for very limited secondary RNA folding [40]. No conserved structures potentially regulating translation initiation were found that also suggests presence of transcription regulation.

4. Conclusions

The Ycf28 proteins are present in plastids of all Rhodophyta; refer to Table 2. In *Cyanidioschyzon merolae* and *Porphyridium purpureum* this protein is notably shorter.

In the presented database, phylogenetic distributions of *moeB* and transcription factor Ycf28 coincide. This observation leads to the suggestion that Ycf28 is a transcription regulation factor for *moeB*. The factor Ycf28 is a close homolog of the cyanobacterial transcription factor NtcA involved in regulation of nitrogen metabolism [41, 42]. Among cyanobacterial genes under the NtcA regulation only two have homologs in plastids. These are the genes of the factor itself and the regulatory protein GlnB from the family PII [43]. However, GlnB is rarely found in plastids, and the corresponding 5'-UTRs lack the conserved motif typically binding NtcA in cyanobacteria [41, 42]. This may suggest that the plastid-encoded Ycf28 and cyanobacterial NtcA are involved in different regulations.

In most species, presence of the actively transcribed tRNA gene *trnW* on the opposite strand precludes *moeB* transcription from a promoter located upstream that of *trnW* due to inevitable strong RNA polymerase competition. An important role of such competition in expression of closely located antidiirected genes is substantiated in modelling and various experiments on gene expression. Such evidence includes data on bacterial type RNA polymerases σ -subunit knockout in plastids of *Arabidopsis thaliana* and data for mitochondrial RNA polymerases of the phage type [38, 39]. Therefore, the *moeB* promoter is likely to be located in between genes *moeB* and *trnW* and requires transcription initiation due to absence of an evident –35 box. Considering polymerase competition at these genes, the transcription factor binding site is likely to occur in the same region between the genes. Indeed, a binding site within an intensively transcribed region is unlikely effective due to interference of the factor with RNA polymerases.

Notably, short conserved motifs adjoining {A, T}-rich regions at their 3'-end are commonly found upstream *moeB*. This may be related to a low GC-content in plastids of most species. However, the predicted location of the binding site makes the putative mechanism of expression regulation specific to *moeB*.

Conflict of Interests

The authors declare that they have no conflict of interests.

Acknowledgments

The authors are deeply grateful to the editor of the paper for his valuable comments. Also the authors would like to thank L. Rusin for valuable discussions and help with preparing the paper. Research was partly funded by the Russian Foundation for Basic Research (Grant 13-04-40196-H).

References

- [1] V. A. Lyubetsky, A. V. Seliverstov, and O. A. Zverkov, "Elaboration of the homologous plastid-encoded protein families that separate paralogs in magnoliophytes," *Mathematical Biology and Bioinformatics*, vol. 8, no. 1, pp. 225–233, 2013 (Russian).
- [2] S. van Dongen and C. Abreu-Goodger, "Using MCL to extract clusters from networks," *Methods in Molecular Biology*, vol. 804, pp. 281–295, 2012.
- [3] V. Strassen, "Gaussian elimination is not optimal," *Numerische Mathematik*, vol. 13, pp. 354–356, 1969.
- [4] D. Coppersmith and S. Winograd, "Matrix multiplication via arithmetic progressions," *Journal of Symbolic Computation*, vol. 9, no. 3, pp. 251–280, 1990.
- [5] F. Le Gall, "Powers of tensors and fast matrix multiplication," in *Proceedings of the 39th International Symposium on Symbolic and Algebraic Computation (ISSAC '14)*, pp. 296–303, July 2014.
- [6] A. V. Smirnov, "The bilinear complexity and practical algorithms for matrix multiplication," *Computational Mathematics and Mathematical Physics*, vol. 53, no. 12, pp. 1781–1795, 2013.
- [7] A. J. Vilella, J. Severin, A. Ureta-Vidal, L. Heng, R. Durbin, and E. Birney, "EnsemblCompara GeneTrees: complete, duplication-aware phylogenetic trees in vertebrates," *Genome Research*, vol. 19, no. 2, pp. 327–335, 2009.
- [8] I. M. Wallace, O. O'Sullivan, D. G. Higgins, and C. Notredame, "M-Coffee: Combining multiple sequence alignment methods with T-Coffee," *Nucleic Acids Research*, vol. 34, no. 6, pp. 1692–1699, 2006.
- [9] K. Katoh and D. M. Standley, "MAFFT multiple sequence alignment software version 7: improvements in performance and usability," *Molecular Biology and Evolution*, vol. 30, no. 4, pp. 772–780, 2013.
- [10] A. E. Galashov and A. V. Kel'manov, "A 2-approximate algorithm to solve one problem of the family of disjoint vector subsets," *Automation and Remote Control*, vol. 75, no. 4, pp. 595–606, 2014.
- [11] A. V. Kel'manov and S. M. Romanchenko, "FPTAS for solving a problem of search for a vector subset," *Diskretnyi Analiz i Issledovanie Operatsii*, vol. 21, no. 3, pp. 41–52, 2014 (Russian).

- [12] C. M. Zmasek and S. R. Eddy, "RIO: analyzing proteomes by automated phylogenomics using resampled inference of orthologs," *BMC Bioinformatics*, vol. 3, article 14, 2002.
- [13] A. Minoda, A. P. M. Weber, K. Tanaka, and S.-Y. Miyagishima, "Nucleus-independent control of the rubisco operon by the plastid-encoded transcription factor Ycf30 in the red alga *Cyanidioschyzon merolae*," *Plant Physiology*, vol. 154, no. 3, pp. 1532–1540, 2010.
- [14] M. S. Cortese, A. B. Caplan, and R. L. Crawford, "Structural, functional, and evolutionary analysis of *moeZ*, a gene encoding an enzyme required for the synthesis of the *Pseudomonas* metabolite, pyridine-2,6-bis(thiocarboxylic acid)," *BMC Evolutionary Biology*, vol. 2, no. 1, article 8, 2002.
- [15] J. Collén, B. Porcel, W. Carré et al., "Genome structure and metabolic features in the red seaweed *Chondrus crispus* shed light on evolution of the Archaeplastida," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 110, no. 13, pp. 5247–5252, 2013.
- [16] M. S. Depriest, D. Bhattacharya, and J. M. López-Bautista, "The plastid genome of the red macroalga *Grateloupia taiwanensis* (Halymeniaceae)," *PLoS ONE*, vol. 8, no. 7, Article ID e68246, 2013.
- [17] G. Glöckner, A. Rosenthal, and K. Valentin, "The structure and gene repertoire of an ancient red algal plastid genome," *Journal of Molecular Evolution*, vol. 51, no. 4, pp. 382–390, 2000.
- [18] J. C. Hagopian, M. Reis, J. P. Kitajima, D. Bhattacharya, and M. C. De Oliveira, "Comparative analysis of the complete plastid genome sequence of the red alga *Gracilaria tenuistipitata* var. *liui* provides insights into the evolution of rhodoplasts and their relationship to other plastids," *Journal of Molecular Evolution*, vol. 59, no. 4, pp. 464–477, 2004.
- [19] J. Janouškovec, S.-L. Liu, P. T. Martone et al., "Evolution of red algal plastid genomes: ancient architectures, introns, horizontal gene transfer, and taxonomic utility of plastid markers," *PLoS ONE*, vol. 8, no. 3, Article ID e59001, 2013.
- [20] N. Ohta, M. Matsuzaki, O. Misumi et al., "Complete sequence and analysis of the plastid genome of the unicellular red alga *Cyanidioschyzon merolae*," *DNA Research*, vol. 10, no. 2, pp. 67–77, 2003.
- [21] M. E. Reith and J. Munholland, "Complete nucleotide sequence of the *Porphyra purpurea* chloroplast genome," *Plant Molecular Biology Reporter*, vol. 13, no. 4, pp. 333–335, 1995.
- [22] L. Wang, Y. Mao, F. Kong et al., "Complete sequence and analysis of plastid genomes of two economically important red algae: *Pyropia haitanensis* and *Pyropia yezoensis*," *PLoS ONE*, vol. 8, no. 5, Article ID e65902, 2013.
- [23] M. A. Campbell, G. Presting, M. S. Bennett, and A. R. Sherwood, "Highly conserved organellar genomes in the Gracilariaceae as inferred using new data from the Hawaiian invasive alga *Gracilaria salicornia* (Rhodophyta)," *Phycologia*, vol. 53, no. 2, pp. 109–116, 2014.
- [24] V. A. Lyubetsky, A. V. Seliverstov, and O. A. Zverkov, "Transcription regulation of plastid genes involved in sulfate transport in viridiplantae," *BioMed Research International*, vol. 2013, Article ID 413450, 6 pages, 2013.
- [25] O. A. Zverkov, L. Y. Rusin, A. V. Seliverstov, and V. A. Lyubetsky, "Study of direct repeats in micro evolution of plant mitochondria and plastids based on protein clustering," *Moscow University Biological Sciences Bulletin*, vol. 68, no. 2, pp. 58–62, 2013.
- [26] O. A. Zverkov, A. V. Seliverstov, and V. A. Lyubetsky, "Plastid-encoded protein families specific for narrow taxonomic groups of algae and protozoa," *Molecular Biology*, vol. 46, no. 5, pp. 717–726, 2012.
- [27] S. R. Starkenburg, K. J. Kwon, R. K. Jha et al., "A pangenomic analysis of the Nannochloropsis organellar genomes reveals novel genetic variations in key metabolic genes," *BMC Genomics*, vol. 15, no. 1, article 212, 2014.
- [28] D. A. Benson, M. Cavanaugh, K. Clark et al., "GenBank," *Nucleic Acids Research*, vol. 41, no. 1, pp. D36–D42, 2013.
- [29] M. Punta, P. C. Coggill, R. Y. Eberhardt et al., "The Pfam protein families database," *Nucleic Acids Research*, vol. 40, no. 1, pp. D290–D301, 2012.
- [30] C. J. A. Sigrist, E. de Castro, L. Cerutti et al., "New and continuing developments at PROSITE," *Nucleic Acids Research*, vol. 41, no. 1, pp. D344–D347, 2013.
- [31] V. A. Lyubetsky, L. I. Rubanov, and A. V. Seliverstov, "Lack of conservation of bacterial type promoters in plastids of Streptophyta," *Biology Direct*, vol. 5, article 34, 2010.
- [32] A. V. Seliverstov, E. A. Lysenko, and V. A. Lyubetsky, "Rapid evolution of promoters for the plastome gene *ndhF* in flowering plants," *Russian Journal of Plant Physiology*, vol. 56, no. 6, pp. 838–845, 2009.
- [33] A. Homann and G. Link, "DNA-binding and transcription characteristics of three cloned sigma factors from mustard (*Sinapis alba* L.) suggest overlapping and distinct roles in plastid gene expression," *European Journal of Biochemistry*, vol. 270, no. 6, pp. 1288–1300, 2003.
- [34] E. A. Lysenko, "Plant sigma factors and their role in plastid transcription," *Plant Cell Reports*, vol. 26, no. 7, pp. 845–859, 2007.
- [35] V. A. Lyubetsky and A. V. Seliverstov, "Some algorithms related to finite groups," *Information Processes*, vol. 3, no. 1, pp. 39–46, 2003 (Russian).
- [36] V. A. Lyubetsky and A. V. Seliverstov, "Note on cliques and alignments," *Information Processes*, vol. 4, no. 3, pp. 241–246, 2004.
- [37] T. L. Bailey, M. Boden, F. A. Buske et al., "MEME SUITE: tools for motif discovery and searching," *Nucleic Acids Research*, vol. 37, no. 2, pp. W202–W208, 2009.
- [38] V. A. Lyubetsky, O. A. Zverkov, L. I. Rubanov, and A. V. Seliverstov, "Modeling RNA polymerase competition: the effect of σ -subunit knockout and heat shock on gene transcription level," *Biology Direct*, vol. 6, article 3, 2011.
- [39] V. A. Lyubetsky, O. A. Zverkov, S. A. Pirogov, L. I. Rubanov, and A. V. Seliverstov, "Modeling RNA polymerase interaction in mitochondria of chordates," *Biology Direct*, vol. 7, article 26, 2012.
- [40] A. A. Vladimirov, "Non-crossing matchings," *Problems of Information Transmission*, vol. 49, no. 1, pp. 54–57, 2013.
- [41] M. I. Muro-Pastor and F. J. Florencio, "Regulation of ammonium assimilation in cyanobacteria," *Plant Physiology and Biochemistry*, vol. 41, no. 6–7, pp. 595–603, 2003.
- [42] K. V. Lopatovskaya, A. V. Seliverstov, and V. A. Lyubetsky, "NtcA and NtcB regulons in cyanobacteria and rhodophyta chloroplasts," *Molecular Biology*, vol. 45, no. 3, pp. 522–526, 2011.
- [43] K. Forchhammer, "PII signal transducers: novel functional and structural insights," *Trends in Microbiology*, vol. 16, no. 2, pp. 65–72, 2008.