Evolution and Systematics of Plastids of Rhodophytic Branch

V.A. Lyubetsky, R.A. Gershgorin, L.I. Rubanov, A.V. Seliverstov, and O.A. Zverkov

Institute for Information Transmission Problems of the Russian Academy of Sciences (Kharkevich Institute), Bolshoy Karetny per. 19, build.1, Moscow 127051 Russia, lyubetsk@iitp.ru

The genomes of plastids, semiautonomous organelles originating from cyanobacteria, have been studied in algae of the phylum Rhodophyta as well as in the species with plastids of secondary or tertiary origin from those of Rhodophyta. These include species of the superphyla Alveolata and Heterokonta (classes Bacillariophyceae, Bolidophyceae, Chrysophyceae, Dictyochophyceae, Eustigmatophyceae, Phaeophyceae, Xanthophyceae, and Raphidophyceae) as well as phyla Cryptophyta and Haptophyta. Their growth temperature ranges from -1.8°C in *Phaeocystis antarctica* (Smith *et al*. 1999) to 56°C in algae of the class Cyanidiophyceae living in hot springs. The unicellular alga *Triparma laevis* related to diatoms lives at 0°–10°C and cannot be found at temperatures above 15°C (Ichinomiya, Kuwata 2015). According to Claquin *et al.* (2008), the optimal growth temperatures for *Lepidodinium chlorophorum* is about 22°C; *Emiliania huxleyi*, 23°C; *Thalassiosira* pseudonana, 25°C; and *Pseudo-nitzschia fraudulenta*, 21°C. Global warming can substantially change the range of all algae. Similar changes have been observed in the past (Li et al. 2016).

Intergenic lengths of three types were calculated in algal plastids: between convergent genes, between divergent genes, and between tandem genes. If neighboring genes overlap, the distance between them is set equal to zero. It is not unusual for convergent and tandem genes to overlap. The data obtained suggested that the median distances between convergent and tandem genes are roughly similar for many species; the median distance between divergent genes is about three times as much; statistically, the distance between genes decreases with the optimal growth temperature. The shortest median distance between divergent genes was observed in *Cyanidioschyzon merolae* living in hot springs as well as in apicomplexan parasites in homeotherms. All species with the median distance between divergent genes below 140 b.p. live at high temperature.

The revealed relationship between the genome structure and optimal growth temperature makes it possible to evaluate the capacity of species with plastids to adapt to low temperature environment. Specifically, we propose that the adaptation is possible if the intergenic distance is relatively high. This can be applied in breeding the varieties resistant to very high or very low temperatures, in particular, those containing xenoplastids. We have found a specific organization of plastids in *Leucocytozoon caulleryi* and *Plasmodium* spp., which confirms the critical role of apicoplasts in infecting homeothermic cells, namely, short intergenic distances and elongated N-termini of proteins targeted to the plastid (Seliverstov *et al.* 2015).

A phylogenetic tree was built based on the proteins encoded in plastids of most considered species as well as those representing the outgroup (*Cyanophora paradoxa*, *Chlorokybus atmophyticus*, *Mesostigma viride*). Another tree was generated using highly conserved elements (HCEs) identified in the complete plastid genomes of all considered species. These trees are in a good agreement.

Previously, the order Parmales was assigned to golden algae (class Chrysophyceae) (Booth, Marchant 1987). According to the current NCBI taxonomy, it belongs to the class Synurophyceae within Stramenopiles. Ichinomiya *et al.* (2016) argued that the alga *Triparma laevis* belongs to the class Bolidophyceae and is close to diatoms, which is well supported by our protein and HCE trees. Thus, there are no grounds to believe that *Triparma laevis* is a close relative of *Ochromonas* sp. CCMP1393 (Chrysophyceae); on the contrary, it represents another class related to diatoms (Bacillariophyceae). Although previously *Trachydiscus minutus* was assigned to yellow-green algae (Xanthophyceae), Přibyl *et al.* (2012) assign to *T. minutus* the class Eustigmatophyceae. This is confirmed by our protein tree: *T. minutus* forms a clade with *Nannochloropsis* spp., while *Ochromonas sp. CCMP1393* is a related branch. On the HCE tree, *T. minutus* forms a clade with *Ochromonas sp. CCMP1393* (Chrysophyceae) but is separate from *Nannochloropsis* spp. On the HCE tree, *Cryptomonas paramecium* is distant from other cryptophyte algae, which can be attributed to a significant reduction of the plastid genome related to the loss of photosynthesis. Minor divergence of the HCE tree from the protein tree is observed in *Aureococcus anophagefferens* and *Aureoumbra lagunensis*, which form no clade, as well as in *Pavlova lutheri*, a branch close to Prymnesiophyceae, which forms no clade with other haptophyte algae. Otherwise,

traditionally close plastids compose clades on both trees. The location of plastids of different dinoflagellate species on the tree confirms their independent origin. Our trees suggest that diatoms were the donor of plastids in *Durinskia baltica* and *Kryptoperidinium foliaceum*; and haptophytes were the donor of plastids in *Karlodinium veneficum*. On both trees, *Pavlova lutheri* occurs in the clade including *Karlodinium veneficum*, which is separate from other haptophyte algae of the class Prymnesiophyceae. On the other hand, *Karlodinium veneficum* neighbors the clade composed of apicomplexan parasites, *Chromera velia*, and *Vitrella brassicaformis*. The two latter species are close relatives of Apicomplexa, and together with dinoflagellates and ciliophorans compose the superphylum Alveolata, which agrees with other published data.

The class Cyanidiophyceae (which belongs to the subdivision Cyanidophytina of the phylum Rhodophyta) forms a separate clade on the protein tree but is a part of the common Rhodophyta clade on the HCE tree. This indicates a close relationship between plastids of all Rhodophyta. As cyanidiophyceans adapted to living at high temperature, their proteins rapidly evolved, which explains why the tree of proteins deviates from the tree of species.

A good agreement of small subtrees of proteins and HCEs confirms the applicability of the HCE approach to determine the phylogenetic position of species or identify at recent plastid donor. Significantly, HCE identification requires no genome annotation. For instance, HCEs are less prone to temperature-induced modifications than proteins in Cyanidiophyceae, which makes HCEs indispensable in the studies of thermophiles.

To expand the studies in (Gershgorin *et al.* 2015, Lyubetsky *et al.* 2016), the scenarios of chromosome rearrangements were deduced in rhodophytic plastids. In particular, the scenarios demonstrate the similarity of chromosome structures in sporozoan apicoplasts and rhodophytic plastids, which agrees with our hypothesis of the common origin of expression regulation in genes from these species, including the common regulatory pattern of translation initiation in the genes coding for DNA-dependent RNA polymerase beta chain and the protein SufB involved in iron-sulfur cluster formation. The similarity of chromosome structures is observed in rhodophytic and cryptophytic plastids. On the other hand, our results indicate an early and independent segregation of diatom and haptophyte plastids.

Plastid genomes were retrieved from GenBank. Highly conserved elements were found using the original algorithm based on the identification of dense subgraphs, which was described in (Rubanov et al. 2016) and tested in (Gershgorin et al. 2017).

The research was supported by the Russian Science Foundation, project no. 14-50-00150. The calculations were performed at the Joint Supercomputer Center of the Russian Academy of Sciences.

1. W.O. Smith Jr., D.M. Nelson, S. Mathot (1999) *J Plankton Res*, **21**(8): 1519–1536.

2. M. Ichinomiya, A. Kuwata (2015) *Aquat Microb Ecol*, **75**: 207–223.

3. P. Claquin, I. Probert, S. Lefebvre, B. Veron (2008) *Aquat Microb Ecol*, **51**:1–11.

4. G. Li, H. Dong, W. Hou, S. Wang, H. Jiang, J. Yang, G. Wu (2016) *Scientific Reports*, **6**:19769.

5. A.V. Seliverstov, O.A. Zverkov, S.N. Istomina, S.A. Pirogov, P.S. Kitsis (2015) *BioMed Research International*, **2015**:452958.

6. B.C.Booth, H.J. Marchant (1987) *J Phycol*, **23**:245–260.

7. M. Ichinomiya, A. Lopes dos Santos, P. Gourvil, S. Yoshikawa, M. Kamiya, K. Ohki, S. Audic, C. de Vargas, M.-H. Noël, D. Vaulot, A. Kuwata (2016) *The ISME Journal*, **10**:2419–2434.

8. P. Přibyl, M. Eliáš, V. Cepák, J. Lukavský, P. Kaštánek (2012) *J Phycol*, **48**(1):231–242.

9. R.A. Gershgorin, K.Yu. Gorbunov, A.V. Seliverstov, V.A. Lyubetsky (2015) *Proceedings of the 39th IITP RAS Interdisciplinary Conference & School "Information Technology and Systems 2015" (ITaS'15)*, Sochi, Russia, Sep 7–11 2015, Moscow: IITP, 2015, p. 105–120.

10. V.A. Lyubetsky, R.A. Gershgorin, A.V. Seliverstov, K.Yu. Gorbunov (2016) *BMC Bioinformatics*, **17**:40.

11. L.I. Rubanov, A.V. Seliverstov, O.A. Zverkov, V.A. Lyubetsky (2016) *BMC Bioinformatics*, **17**:385.

12. R.A. Gershgorin, K.Yu. Gorbunov, O.A. Zverkov, L.I. Rubanov, A.V. Seliverstov, V.A. Lyubetsky (2017) *Life*, **7**:9.