
Abstract

The Division of Dinophyta is one of the most diverse eukaryotic groups. Usually, members of the dinoflagellates are photoautotrophic or heterotrophic biflagellated unicellular organisms. Together with the Apicomplexa and Ciliophora they represent the Alveolata, which are one part of the SAR supergroup. In order to determine the phylogenetic relationships of dinoflagellates, single genes (or parts of them) of the nuclear-encoded rRNA operon are commonly used. Phylogenetic analyses have mostly employed either the SSU rRNA gene or the first approximately 800-1400 bases from the 5'-terminus of the LSU rDNA. Very few studies were based on both rRNA genes or even on the complete rRNA operon. The published phylogenetic trees usually show a low resolution and vary in their results, which are difficult to reproduce. It thus seems to be difficult to clarify the relationship between the different groups of dinoflagellates and their basal topologies, using only the rRNA genes. What are the reasons for this? Does it make a difference, which region of the nrRNA operon is used for phylogenetic analysis? To answer these questions a congruent taxon sampling was used, which included only taxa with sequences of the entire rRNA operon. Based on this alignment, the influence of different alignment methods, the numbers and choice of the selected characters, as well as the phylogenetic methods themselves were investigated for the phylogenetic reconstruction. The resulting trees thereby show topological differences between the analyses of the different partitions of the rRNA operon.

The results suggest that the complete rRNA operon or at least a combination of SSU and partial LSU rDNA enhance resolution in molecular phylogenetic analyses of dinoflagellates. The analysis also revealed that the alignment of dinoflagellate sequences contains a systematic error. Both, the distinctive regions of the rRNA genes, as well as the individual groups of dinoflagellates show different base frequencies. The variations of the base frequencies primarily affect the pyrimidin bases cytosine and thymine/uracil. This twofold pyrimidine bias can not be compensated by the usual analytical methods and leads to unresolved phylogenies of the dinoflagellates. Different LBA-artefacts also seem to have a negative impact on different topologies of their phylogeny. Such analytical methods like NJ(LogDet), which take non-homogeneous base frequencies for different taxa into consideration, are less affected by this twofold pyrimidine bias. Accordingly, it would be useful for the resolution of dinoflagellate phylogenies to use such methods for the analysis. Furthermore, different analytical methods can be adjusted (e. g. the jumping evolution model, currently available only for analysis of amino acid sequences) to compensate these errors.

Zusammenfassung

Die Abteilung der Dinophyta umfasst eine der diversesten eukaryotischen Gruppen von vorwiegend heterotrophen oder photoautotrophen einzelligen Flagellaten. Sie bilden zusammen mit den Apikomplexa und Ciliophora die Alveolata und sind somit ein Bestandteil der SAR Supergruppe. Zur Bestimmung von phylogenetischen Verwandtschaften innerhalb der Dinoflagellaten werden zumeist einzelne Gene des kernkodierten rRNA-Operons analysiert. Für diese phylogenetischen Analysen werden entweder die SSU rDNA oder die ersten etwa 800-1400 Basen des 5'-Endes der LSU rDNA benutzt. Hierbei wird nur für wenige der publizierten Phylogenien eine Kombination der beiden Gene verwendet. Noch seltener wird das gesamte rRNA Operon analysiert. Die veröffentlichten Stammbaumrekonstruktionen weisen zumeist eine geringe Auflösung auf, variieren dabei in ihren Ergebnissen und sind selten reproduzierbar. Es erweist sich demnach als schwierig die Verwandtschaftsverhältnisse der einzelnen Dinoflagellatengruppen untereinander, sowie die Topologie an ihrer Basis, mittels der rRNA-Gene aufzuklären. Um nachzuvollziehen ob diese Unterschiede daher rühren, dass verschiedene Bereiche des rRNA-Operons für die Analysen herangezogen wurden, oder ob andere Umstände dies hervorrufen, wurden verschiedene Untersuchungen durchgeführt. Zu Beginn wurde eine kongruente Taxonauswahl so aufgebaut, sodass die Sequenzen aller enthaltenen Taxa das gesamte rRNA-Operon abdeckten. Auf der Basis dieses Alignments wurde der Einfluss unterschiedlicher Alignierungsmethoden, die Art der Merkmalsauswahl, sowie die Anwendung verschiedener Analyseverfahren auf die Stammbaumrekonstruktion untersucht. Die resultierenden Bäume zeigten topologische Unterschiede zwischen den Analysen der verschiedenen Partitionen des rRNA Operons. Generell verbesserte sich bei den Analysen die phylogenetische Auflösung für die Dinoflagellaten, sobald beide Gene für diese verwendet wurden. Die Auswertung ergab weiterhin, dass sich in den Alignments der Sequenzen der Dinoflagellaten ein systematischer Fehler verbirgt. Dabei weisen sowohl die verschiedenen Bereiche der rRNA Gene, als auch die einzelnen Gruppen der Dinoflagellaten, unterschiedliche Basenfrequenzen auf. Die Schwankungen der Basenfrequenzen betreffen dabei vor allem die Pyrimidinbasen Cytosin und Thymin/Uracil. Dieser zweifache Pyrimidin-Fehler kann von den gängigen Analyseverfahren nicht kompensiert werden und resultiert so in unaufgelösten Phylogenien der Dinoflagellaten. Auch verschiedene LBA-Artefakte scheinen die Dinoflagellatenphylogenien an verschiedenen Stellen negativ zu beeinflussen. Analysemethoden wie zum Beispiel NJ(LogDet), die nicht von einer homogenen Basenfrequenz für die verschiedenen Taxa ausgehen, sind von diesem zweifachen Pyrimidin-Fehler weniger betroffen. Demnach wäre es für die Auflösung der Dinoflagellatenphylogenien dienlich, auch solche Analysemethoden zu verwenden. Weiterführend könnten Analyseansätze (z. B. das Jumping-Evolutionsmodell) angepasst werden, die momentan nur für die Auswertung von Aminosäuresequenzen zur Verfügung stehen, um diese Fehler auszugleichen.